Forecasts for WEAVE-QSO: 3D clustering and connectivity of critical points with Lyman-α tomography

K. Kraljic [1]★ C. Laigle,[2] C. Pichon,[2,3] S. Peirani,[2,4] S. Codis,[5] J. Shim [6] C. Cadiou [7]
D. Pogosyan,[8] S. Arnouts,[1] M. Pieri,[1] V. Iršič,[9,10] S. S. Morrison [1,11] J. Oñorbe [12]
I. Pérez-Ràfols [13] and G. Dalton[14,15]

[1]*Aix Marseille Université, CNRS, CNES, LAM, F-13621 Marseille, France*
[2]*CNRS and Sorbonne Université, UMR 7095, Institut d'Astrophysique de Paris, 98 bis, Boulevard Arago, F-75014 Paris, France*
[3]*IPhT, DRF-INP, UMR 3680, CEA, L'Orme des Merisiers, Bât 774, F-91191 Gif-sur-Yvette, France*
[4]*Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Bd de l'Observatoire, CS 34229, F-06304 Nice Cedex 4, France*
[5]*AIM, CEA, CNRS, Université Paris-Saclay, Université Paris Diderot, Sorbonne Paris Cité, F-91191 Gif-sur-Yvette, France*
[6]*Korea Institute for Advanced Study, 85 Hoegiro, Dongdaemun-gu, Seoul 02455, Republic of Korea*
[7]*Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK*
[8]*Department of Physics, University of Alberta, 412 Avadh Bhatia Physics Laboratory, Edmonton, Alberta T6G 2J1, Canada*
[9]*Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[10]*Cavendish Laboratory, University of Cambridge, 19 J. J. Thomson Ave., Cambridge CB3 0HE, UK*
[11]*Department of Astronomy, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA*
[12]*Facultad de Física, Universidad de Sevilla, Avda. Reina Mercedes s/n. Campus de Reina Mercedes, E-41012 Sevilla, Spain*
[13]*Sorbonne Université, Université Paris Diderot, CNRS/IN2P3, LPNHE, 4 Place Jussieu, F-75252 Paris, France*
[14]*Department of Physics, University of Oxford, Parks Rd, Oxford OX1 3RH, UK*
[15]*TFC RALSpace, UKRI, Harwell Campus, Swindon OX1 10QX, UK*

ABSTRACT

The upcoming WEAVE-QSO survey will target a high density of quasars over a large area, enabling the reconstruction of the 3D density field through Lyman-α (Ly-α) tomography over unprecedented volumes smoothed on intermediate cosmological scales (≈ 16 Mpc h^{-1}). We produce mocks of the Ly-α forest using Ly-α Mass Association Scheme, and reconstruct the 3D density field between sightlines through Wiener filtering in a configuration compatible with the future WEAVE-QSO observations. The fidelity of the reconstruction is assessed by measuring one- and two-point statistics from the distribution of critical points in the cosmic web. In addition, initial Lagrangian statistics are predicted from the first principles, and measurements of the connectivity of the cosmic web are performed. The reconstruction captures well the expected features in the auto- and cross-correlations of the critical points. This remains true after a realistic noise is added to the synthetic spectra, even though sparsity of sightlines introduces systematics, especially in the cross-correlations of points with mixed signature. Specifically, the most striking clustering features involving filaments and walls could be measured with up to 4σ of significance with a WEAVE-QSO-like survey. Moreover, the connectivity of each peak identified in the reconstructed field is globally consistent with its counterpart in the original field, indicating that the reconstruction preserves the geometry of the density field not only statistically, but also locally. Hence, the critical points' relative positions within the tomographic reconstruction could be used as standard rulers for dark energy by WEAVE-QSO and similar surveys.

Key words: surveys – galaxies: evolution – galaxies: formation – cosmology: large-scale structure of Universe.

1 INTRODUCTION

The geometry and cosmic evolution of large-scale structure are our best probes to make sense of the accelerated expansion of the Universe. At $z > 2$, the Lyman-α (Ly-α) forest absorption towards bright background sources is observable from ground-based optical instruments and can be used at intermediate (~ 1 Mpc h^{-1}) to large (~ 200 Mpc h^{-1}) scales as a tracer of the underlying density field. The prospect of using tomography of the Ly-α forest for reconstructing

the cosmic web (Bond, Kofman & Pogosyan 1996) has a long history (see e.g. Pichon et al. 2001; D'Odorico et al. 2006; Caucci et al. 2008; Gallerani, Kitaura & Ferrara 2011; Kitaura, Gallerani & Ferrara 2012; Cisewski et al. 2014; Ozbek, Croft & Khandai 2016; Japelj et al. 2019; Horowitz et al. 2019, 2021b) and is now within reach from current (e.g. CLAMATO: Lee & White 2016; Krolewski et al. 2018; Lee et al. 2018, eBOSS-Stripe 82: Ahumada et al. 2020; Ravoux et al. 2020, LATIS: Newman et al. 2020) and upcoming quasar or star-forming galaxy surveys (e.g. WEAVE-QSO: Pieri et al. 2016, Pieri et al. in preparation, PFS: Takada et al. 2014 or DESI: DESI Collaboration 2016). Such reconstruction represents an unparalleled opportunity, as it gives us access to many large and intermediate

★ E-mail: katarina.kraljic@lam.fr

scales (Bernardeau et al. 2002). Its success relies on the orders-of-magnitude better sensitivity of detection of neutral hydrogen in absorption (when compared to emission), along Gpc-long lines of sight (Petitjean, Mueket & Kates 1995; Rauch 1998). Hence, Ly-α tomography provides means to characterize the expansion-driven geometry of the cosmic web in the lead up to the epoch of dark energy domination.

Depending on the design of the surveys (sampling of background sources, availability of quasar and/or galaxy spectra, spectral resolution, and signal-to-noise ratio, SNR), different scales and volumes will be accessible, making the tomographic reconstruction either more suitable for studies focused on co-evolution of galaxies and the intergalactic-medium (if filaments can be reconstructed at the Mpc-scale) or for cosmological analysis (if large volumes are available). Using idealized mocks, the pioneering work of Caucci et al. (2008) demonstrated that the topology of the cosmic web traced either by Minkowski functionals, such as the genus (Hamilton, Gott & Weinberg 1986), or the skeleton (Sousbie, Pichon & Kawahara 2011) could be well recovered with this method. In the same vein, Horowitz et al. (2019) showed that cosmic web structure classification from eigenvalues and eigenvectors of the pseudo-deformation tensor could be accurately performed (see also Lee & White 2016; Krolewski et al. 2017, for the first application of pseudo-deformation tensor analysis on tomographic maps), while Horowitz et al. (2021b) focused on proto-cluster identification (see also Stark et al. 2015a, for the first proto-cluster detection, and e.g. Cisewski et al. 2014; Ozbek et al. 2016; Japelj et al. 2019, for complementary mock-based analyses of the quality of the reconstruction). In particular, several reconstruction methods have been presented in the literature. Wiener filtering is the classical approach, but different procedures have also been proposed, either involving a sophistication of the standard Wiener Filter (e.g. Li, Horowitz & Cai 2021), such as e.g. a forward modelling approach (e.g. Porqueres et al. 2020; Horowitz et al. 2019, 2021b).

Encouraged by these theoretical pursuits, three-dimensional reconstruction of the density field from the Ly-α forest has already been successfully performed in observational surveys, notably with the CLAMATO program (see Horowitz et al. 2021a, for the latest release) and eBOSS-Stripe 82 (Ravoux et al. 2020). The Ly-α forest has proven to be a powerful tracer of the density field, particularly sensitive to intermediate densities: Therefore, tomographic reconstruction should allow us to characterize the geometry of the weakly overdense and underdense regions of the Universe, i.e. specifically the walls and filaments of the cosmic web (see Fig. 1 for an illustration).

The clustering properties of maxima of 3D density fields were predicted for Gaussian random field by Regos & Szalay (1995) and revisited more recently by Baldauf et al. (2021). Such predictions provide insight on their dependence over cosmological parameters. More recently, Shim et al. (2021) systematically investigated the statistical properties of all critical points (i.e. the loci of zero gradient) of the cosmic field of ΛCDM simulations, and in particular the number counts and clustering properties of wall-like and filament-like saddle points.[1] As they trace the relative position of walls and filaments (beyond the more standard peaks and voids), these saddle points help characterize the global geometry and evolution of the cosmic web (Cadiou et al. 2020). They define the underlying topology, which makes them robust to most systematics (e.g. biasing). In particular, wall-saddle clustering is a convenient measurement to probe the typical size of voids, which otherwise can be characterized by more
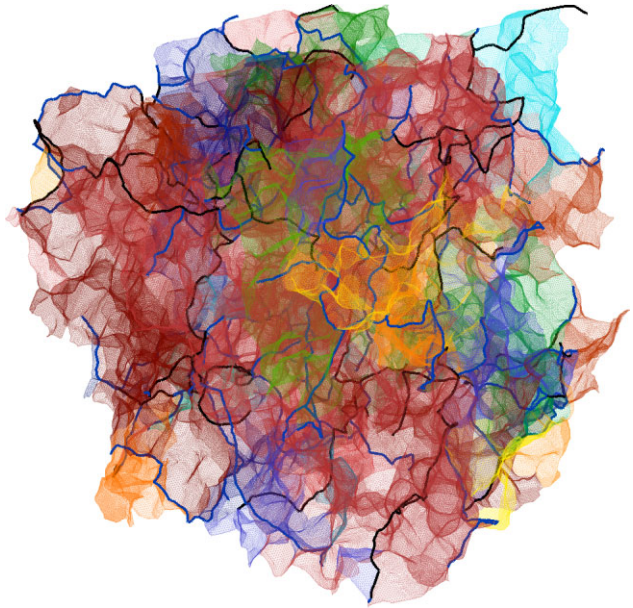
Figure 1. The walls (colour coded randomly) and filaments (dark colour for all filaments, light colour for filaments of higher persistence) extracted from the DM density field of one of the mocks smoothed at a scale of 16 Mpc h^{-1} (see Section 3.5 for more details on the cosmic web identification). The purpose of the reconstruction performed in this study is to recover as accurately as possible the geometry of this cosmic web, since it defines the metric in which we can constrain dark energy. In order to assess this accuracy, we focus on the number counts and clustering properties of the critical points associated with peaks, voids, filaments and walls of the cosmic web. We also compute the connectivity of its nodes.

classical approaches (as done by e.g. Stark et al. 2015b, using a measure of underdensities within tomographic maps of Ly-α forest flux). On the other hand, the cross-correlation of peaks and filament-type saddles is sensitive to the typical length of filaments. These sets of points probe less biased regions than galaxy surveys (Desjacques, Jeong & Schmidt 2018), hence their dynamics are better captured by perturbation theory (Gay, Pichon & Pogosyan 2012). Shim et al. (2021) showed that the statistical properties of the set of critical points such as the size of the exclusion zones are weakly dependent on redshift, hence could in principle be used as standard rulers (Lazkoz, Nesseris & Perivolaropoulos 2008; Appleby et al. 2021) to constrain alternative cosmology models (e.g. Bamba et al. 2012).

The WEAVE-QSO survey, as part of the wider WHT Extended Aperture Velocity Explorer (WEAVE, Dalton et al. 2012) survey, is potentially well suited to deriving cosmologically meaningful statistics with critical points. Its large volume will make it possible to probe the large-scale structure over several thousands of square degrees allowing us to characterize the geometry of weakly overdense and underdense regions of the Universe, while its high density will allow to reach scales of ~ 16 Mpc h^{-1}. Could the Ly-α tomography reconstruction performed on surveys such as WEAVE-QSO be precise enough to measure the clustering of critical points and distinguish between different cosmological models?

In this paper, we investigate to what extent the cosmic web is correctly recovered with this technique by focusing on the clustering of critical points of the density field and their connectivity. We model the WEAVE-QSO survey and constrain the impact of its specificity on our ability to extract cosmological information from the clustering of critical points. In particular, we will investigate what sets of critical pairs are least impacted by systematics and

---

[1]Recall that a saddle point is a point where the gradient is null, but that is neither a minimum nor a maximum.

uncertainties inherited from the reconstruction. We will also measure the individual connectivity of peaks identified in the reconstructed field and compare it to their counterparts in the original field, in order to verify that the geometry of the reconstructed field is robust not only statistically, but also locally.

This paper is organized as follows: In Section 2, we describe the method used to produce the mocks and the corresponding estimators. In Section 3, we assess the quality of the reconstruction using one- and two-point statistics of critical points, and their connectivity. We discuss our results in Section 4 and conclude in Section 5.

Finally, Appendix A investigates the expected accuracy on the reconstruction for future surveys (different configurations of sampling and noise), Appendix C provides an illustration of the distribution of critical points in reconstructed fields, Appendix B sketches the steps involved in the computation of the Lagrangian two-point functions for the dark matter ΛCDM model smoothed over the relevant scales, Appendix D provides a measurements of summary statistics, while Appendix E looks into the impact of the choice of rarity and smoothing length.

2 MOCK DATA AND ESTIMATION METHODS

Let us first describe briefly our mocks, reconstruction method, critical point extraction, and clustering estimators.

2.1 Modelling the Ly-α forest

We rely on five ΛCDM simulation snapshots of $(1 \, \mathrm{Gpc} \, h)^{-3}$ at $z = 2.5$, run with *WMAP7* cosmological parameters (Komatsu et al. 2011): $\Omega_m = 0.272$, $\Omega_\lambda = 0.728$, $\sigma_8 = 0.81$, $n_s = 0.967$. Each *N*-body simulation follows the time evolution of 2048^3 dark matter (DM) particles using GADGET2 (Springel 2005). Fig. 1 shows the cosmic web of one of these simulations.

An estimator for the Ly-α forest flux from the DM has been synthesized using the hydrodynamical HORIZON-AGN simulation (Dubois et al. 2014), with LyMAS2, an improved version of the Ly-α Mass Association Scheme (LyMAS; Peirani et al. 2014, 2022). In brief, the optical depth of Ly-α absorption is calculated along about one million sightlines extracted from HORIZON-AGN, based on the neutral hydrogen (H I) density, whose evolution and distribution in the simulation is impacted by metal-dependent cooling, photoionization and heating from the UV background, feedback and metal enrichment. We work in the distant observer limit, assuming the sightlines are all parallel to one side of the box. LyMAS2 assigns a Ly-α flux in redshift space to each pixel in the *N*-body simulation assuming that the 3d flux correlations are mainly driven by the correlations of the underlying DM (over)density and velocity dispersion fields.

To produce each large Ly-α forest mock catalogue, we have then derived first the DM density and the velocity dispersion fields (on regular grids of 4096^3) from each gadget simulation at $z = 2.5$. To save computational time throughout this work, the initial grids of each simulated volume are resampled on a Cartesian grid of 512^3 (we kept 1 pixel of 8), and therefore the resulting resolution of the Ly-α spectra is $1.95 \, \mathrm{Mpc} \, h^{-1}$ per pixel, which corresponds to a resolution of about 2.6 Å per pixel in the Ly-α forest at $z = 2.3$, or equivalently ~R1560. We note that our simulated spectra are at the BOSS resolution and are less resolved than the spectra that will be obtained with WEAVE (R5000), but this difference should not

impair our analysis, since this latter relies on smoothed fields.[2] More details on the LyMAS2 implementation and Ly-α mock production are given in Peirani et al. (2022). As a reminder, Ly-α flux is equal to $\exp(-\tau)$, where τ is the Ly-α optical depth, which, at first order, scales like the H I density. In the following, the H I field is estimated as simply being $-\log(\mathrm{flux})$.

Note that redshift-space distortions are included in the DM and flux (and consequently H I) simulated fields but not in the Gaussian random field (GRF) predictions: this should be kept in mind when interpreting our results, e.g. in Fig. 8. We checked though that when smoothing the fields at the scales considered in this work (12–16 $\mathrm{Mpc} \, h^{-1}$), redshift-space distortions do not alter our results at the level of the expected accuracy, i.e. their impact remains within uncertainties of the reconstruction. Note that $\sigma_{DM} \sim 0.1$ when smoothed over 16 $\mathrm{Mpc} \, h^{-1}$ at redshift 2.5, so we are probing the regime accessible to perturbation theory for critical points (Gay et al. 2012).

### 2.2 Mimicking the distribution of sightlines

2.2.1 Quasar counts and separation between sightlines

The number of observed quasar spectra defines the overall achievable transverse resolution of the reconstructed map. Fig. 2 presents the expected separation between sightlines as a function of redshift for different magnitude cuts, where m_r is the magnitude in the *r* filter passband. To compute it, we use the Palanque-Delabrouille et al. (2016a,b) counts (their table 7, model PLE), which provides the number of quasars per 100 deg² per bins of redshift ($\Delta(z) = 1$) and per bins of magnitudes in the *r* band ($\Delta(m) = 0.5$). The top panel of Fig. 2 displays these counts as a function of magnitude cuts, after having interpolated them on a finer redshift grid ($\Delta(z) = 0.2$). Throughout this work we use only the portion of the rest-frame spectra between the Ly-α (1216 Å) and the Ly-β (1026 Å) wavelength, to avoid contamination of the Ly-α forest by Ly-β absorption lines. In practice, we should also exclude from the analysis the fraction of the sightlines in the direct vicinity of the quasar (over $\sim 30 \, \mathrm{Mpc} \, h^{-1}$) to mitigate the proximity effect (where the H I distribution is affected by the ionizing UV flux from the quasar). Accounting for this effect and excluding the corresponding fraction of sightlines would only lead to a slight increase of the mean separation between sightlines from 14.6 to 15.6 $\mathrm{Mpc} \, h^{-1}$. For example, at $z \sim 2.3$, lines of sight are therefore usable along at most ~165 $\mathrm{Mpc} \, h^{-1}$. This allows us to estimate the mean separation between sightlines, which is presented on the bottom panel of Fig. 2. We also show, for comparison, the mean separation between sightlines reached in eBOSS-Stripe 82 (Ravoux et al. 2020).[3]

We emphasize that we choose the mean separation between sightlines for the correlation length used in the reconstruction (Section 3). Ravoux et al. (2020) made a different choice on eBOSS-Stripe 82, using instead the mean distance to the closest sightline, which returns a smaller value (about 10 $\mathrm{Mpc} \, h^{-1}$ instead of $\sim 20 \, \mathrm{Mpc} \, h^{-1}$ at $z = 2.5$). While this choice allows them to reach smaller scales in regions where sightlines are well clustered, the quality of their reconstruction is degraded in other places where sightlines are sparser (which they subsequently masked in the reconstructed map). Our

---

[2]Given that what primarily limits the resolution of the reconstructed map is the transverse separation between sightlines, the spectra are not exploited at the WEAVE fiducial spectral resolution. Indeed, several spectral resolution elements contribute to a given volume element in the reconstructed density field.

[3]For eBOSS-Stripe 82, L_T has been estimated using the pixel map made publicly available at https://zenodo.org/record/3737781#.YUJbnn069h.
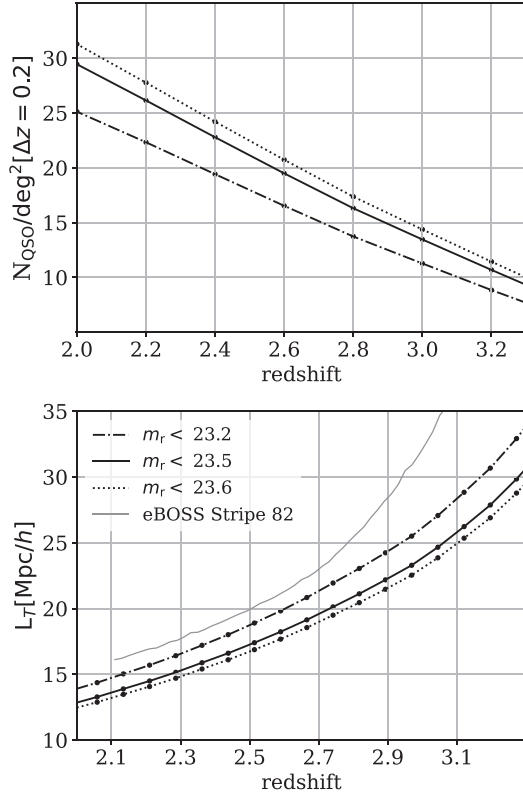
Figure 2. Top panel: the number of quasars per deg^2 per bin of $\Delta z = 0.2$ as a function of the limiting magnitude. These counts follow from Palanque-Delabrouille et al. (2016a,b). The configuration investigated in this study (corresponding to the HIGHDENS footprint, $m_r < 23.5$) is represented by the solid black line. Bottom panel: mean separation between sightlines as a function of redshift, for the same limiting magnitude thresholds as in the top panel, assuming that the full length between the Ly-α and Ly-β wavelengths is usable. In grey, we have overplotted for comparison the mean separation L_T in eBOSS-Stripe 82.

choice of taking the mean separation between sightlines is more conservative and allows us to obtain a more homogeneous quality of the reconstruction everywhere, which is required to derive robust statistics on the clustering of critical points.

2.2.2 WEAVE-QSO survey specifications

The WEAVE-QSO survey is expected to begin in 2022 and will observe in various modes and configurations towards a variety of survey goals (Jin et al, in preparation; Pieri et al in preparation). Two samples with particularly high density are of interest for the present work. We stress that the precise details of the survey plan may evolve over the coming years.

2.2.2.1 The 'WIDE' footprint Over the 6000 deg^2 of the WIDE footprint, a near-complete sample of quasars at $2.5 < z < 3$ with $m_r < 23.5$ will be observed by WEAVE-QSO with a spectral resolution R5000, corresponding to $\sim$48 sightlines per deg^2. This equates to a comoving volume of $\sim$13.6 (Gpc$\,h$)$^{-3}$ (assuming the same cosmology as our simulation). The target selection for this footprint is provided by the J-PAS survey (Benitez et al. 2014).

2.2.2.2 The 'HIGHDENS' footprint Over the 418 deg^2 of the HIGHDENS footprint, all quasars with $m_r < 23.5$ and $z > 2.15$ will
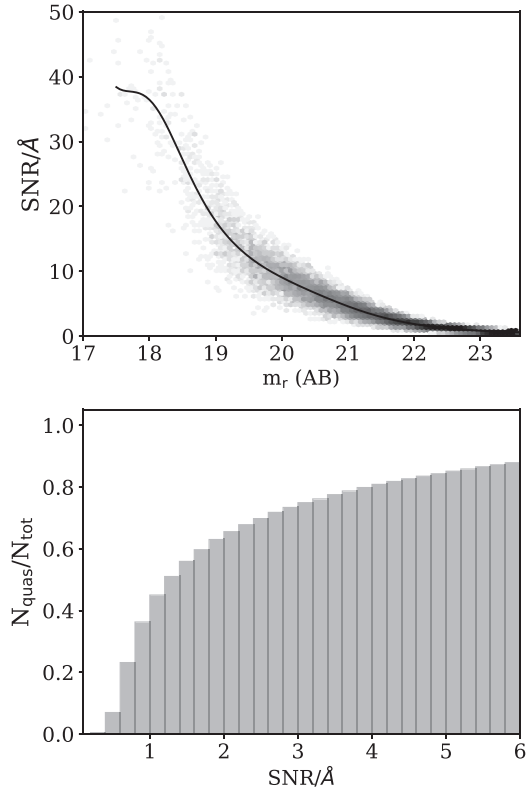
Figure 3. Top panel: SNR/Å (defined in the continuum of the Ly-α forest) as a function of magnitude in the r band as adopted in this study. Bottom panel: corresponding cumulative count of quasars as a function of SNR/Å. These distributions correspond to reasonable forecasts for the WEAVE-QSO survey (Morrison 2019; Jin et al. in preparation). We expect about 50 per cent of quasars to have an SNR/Å larger than 1.4.

be observed with WEAVE at R5000, corresponding to $\sim$111 sightlines per deg^2. This configuration therefore reaches higher density compared to the WIDE footprint through its extension to $2.15 < z < 2.5$ (see Fig. 2). The volume covered by this area over this redshift range is however only about 0.73 (Gpc$\,h$)$^{-3}$. This footprint is also targeted by J-PAS and is placed within the WIDE footprint above. Its boundary is defined by the HETDEX main 'spring' field (Gebhardt et al. 2021).

In this work, we adopt a simulated distribution of quasars that matches the redshift and the density expected on the HIGHDENS footprint, but we note that the WIDE footprint provides an equivalent distribution (in terms of limiting magnitude) over its higher redshift range.

2.2.2.3 SNR distribution The top panel of Fig. 3 presents the SNR distribution as a function of magnitude that is expected to be compatible with the WEAVE-QSO main sample (Jin et al. in preparation). These values correspond to the SNR/Å of the continuum in the Ly-α forest. The cumulative number of quasars as a function of SNR/Å is shown in the bottom panel. We expect about 50 per cent quasars to have an SNR/Å larger than 1.4.

2.2.3 Adopted configurations

We describe below the different sets of spectra used in our study to mimic WEAVE-QSO observations ($\mathcal{R}_{WQ}$) and to test the effect of the different sources of noise: sparsity of the sightline distribution,

irregularity of their spatial distribution, noise on spectra ($\mathcal{R}_{I,r}$, $\mathcal{R}_{I,U}$, and $\mathcal{R}_{\text{noise}}$).

2.2.3.1 Fiducial realization, $\mathcal{R}_{\text{WQ}}$ Without attempting to perfectly reproduce the WEAVE-QSO selection function in one of the two above-mentioned footprints, we simply define a density of sightlines that is compatible with the specifications in the HIGHDENS footprint. At $z \sim 2.3$ and for $m_r < 23.5$, we populate each of our five $(1 \, \text{Gpc} \, h)^{-3}$ simulated boxes with about ~ 30 quasars per $(100 \, \text{Mpc} \, h^{-1})^3$, their positions being drawn from a uniform distribution and quasars being assigned randomly a magnitude so that we reproduce the counts presented in the top panel of Fig. 2. The flux along the sightlines is perturbed with white Gaussian noise, with a magnitude-dependent SNR, the resulting distribution of which is shown in Fig. 3. Note that because quasar positions are scattered randomly, clustering will inevitably occur ('Poisson clumping'), however the projected small-scale clustering of our quasar distribution will be reduced with respect to what is measured in the real Universe. In particular, the fact that bright quasars (hence with high SNR) are expected to be more clustered (e.g. Shen et al. 2009) could degrade a bit the homogeneity of the quality of our reconstructed map in real data (boosting the accuracy in patches where quasars are well clustered, and conversely degrading it elsewhere) and lead to results slightly worse than those obtain for $\mathcal{R}_{\text{WQ}}$ at the same quasar density.

The mean sightline separation, estimated as the square root of the area of the field divided by the number of sightlines in thin redshift slices, is about 15 Mpc h^{-1}. We note though that over $2.15 < z < 2.5$, i.e over the redshift range covered by the HIGHDENS footprint, this separation is expected to vary between 14 and 17 Mpc h^{-1}.

We did not implement continuum errors on the reconstruction in this work. Full WEAVE-QSO data mocks are expected including quasar spectral templates and subsequent continuum fitting in order to perform a more error assessment. We note, however, that neighbouring sightlines are not expected to share the same continuum fitting errors in the least biased continuum fitting methods available (e.g. Bautista et al. 2017). Therefore the probing of structures by many sightlines will reduce continuum errors through averaging over these sightlines, and will therefore be subdominant. Also, although this could induce fake correlations in the resulting density field, we do not expect it to bias the clustering of the rarest critical points, because continuum fitting inaccuracy will not generate sharp features in the spectra.

Finally, we note that our simulation set covers a volume of more than six times the expected volume of the HIGHDENS footprint. This should be factored in when discussing the statistical significance of our results in the context of the WEAVE-QSO survey. At the same time, the WIDE footprint is about 2.7 times larger than our simulated set, but covering a higher redshift range ($2.5 < z < 3$), which would lead to a larger separation between sightlines (as can be inferred from Fig. 2).We discuss in Section 4.2 how the statistical significance of our results would change depending on the footprint.

2.2.3.2 Realization without noise on spectra, $\mathcal{R}_{I,r}$ For testing the impact of noise on spectra, we also produce a realization identical to $\mathcal{R}_{\text{WQ}}$, but without perturbing the Ly-α flux to mimic noise on spectra. Comparing $\mathcal{R}_{I,r}$ to $\mathcal{R}_{\text{WQ}}$ allows to test the impact of noise in the Ly-α forest on the quality of the reconstruction.

2.2.3.3 Realization with regular distribution of sightlines, $\mathcal{R}_{I,U}$ For convergence study, we produce a realization with no noise on spectra and the same density of sightlines but with quasars regularly distributed across the box (so that we have exactly one sightline every

$\sim 15 \text{Mpc} \, h^{-1}$). Comparing $\mathcal{R}_{I,U}$ to $\mathcal{R}_{I,r}$ allows to test the impact of shot noise, i.e non-regularity in the distribution of sightlines, on the quality of the reconstruction.

2.2.3.4 Realization with noise only, $\mathcal{R}_{\text{noise}}$ In order to test the significance of our measurement, we also perform the tomographic reconstruction on 10 sets of sightlines (covering 1 $(\text{Gpc} \, h)^{-3}$ each, as for the other realizations) containing only white Gaussian noise (with an rms equal to the rms of the large-scale structure fluctuations in the simulated Ly-α forest, but without correlations along and between sightlines), with the same parametrization as for the other reconstructed sets. In the following, we will quantify the deviation from the signal produced by this 'noise-only' realization.

2.3 Inversion of the Ly-α forest

2.3.1 Reconstruction method

The three-dimensional distribution of the Ly-α flux contrast δ is reconstructed by interpolating between the lines of sight using Wiener filtering in comoving space (see Pichon et al. 2001; Caucci et al. 2008; Lee et al. 2018).[4]

Let $\mathbf{D}$ be the one-dimensional array representing the data set (all sightlines placed end to end), and $\mathbf{M}$ is the three-dimensional array of the field estimated from the data. Maximizing the penalized likelihood of the data given an assumed (zero mean Gaussian) prior for the flux contrast field yields

$$\mathbf{M} = \mathbf{C}_{\delta^3 d_\delta}(\mathbf{C}_{\delta\delta} + \mathbf{N})^{-1}\mathbf{D}, \tag{1}$$

where $\mathbf{C}_{\delta^3 d_\delta}$ is the mixed parameter-data covariance matrix, and $\mathbf{C}_{\delta\delta}$ is the data covariance matrix. We assume that the noise is uncorrelated, therefore the noise covariance matrix can be expressed as $\mathbf{N} = n^2\mathbf{I}$. In addition, for simplicity (and as commonly assumed in the literature), we assume a normal covariance matrix prior:

$$\mathbf{C}_{\delta\delta}(x_1, x_2, \mathbf{x}_{1\text{T}}, \mathbf{x}_{2\text{T}}) = \sigma_{\delta\delta}^2 \, e^{-\frac{|x_1 - x_2|^2}{2L_x^2}} e^{-\frac{|\mathbf{x}_{1\text{T}} - \mathbf{x}_{2\text{T}}|^2}{2L_\text{T}^2}}, \tag{2}$$

where $(x_i, \mathbf{x}_{i\text{T}})$ are the coordinates of points along and perpendicular to the line of sight. The mixed parameter-data covariance matrix $\mathbf{C}_{\delta^3 d_\delta}$ is taken of the same form. In principle, one could get directly a better estimate of the data-data covariance matrix from the simulation. Ozbek et al. (2016) investigated how the reconstruction depends on the form used for the covariance matrix, and concluded that, at the scales they considered (~ 30 Mpc h^{-1}), the precise form of the covariance matrix has little impact. However, at the scales of few Mpc h^{-1} probed in our study, the covariance matrix of the underlying flux–density contrast field is expected to deviate more strongly from this normal form. Because generating covariance matrices is a computationally intensive process, investigating how the normal approximation impacts the quality of the reconstruction is beyond the scope of the present work.

The reconstruction depends on the normalization, $n^2/\sigma_{\delta\delta}^2$, involving the ratio of the noise matrix amplitude (used for stabilizing the reconstruction) and the data-data prior covariance amplitude, and also on the correlation lengths L_x and L_T, along and perpendicular to the line of sight, respectively.

---

[4]See also, e.g. Li et al. (2021) for an alternative reconstruction method.
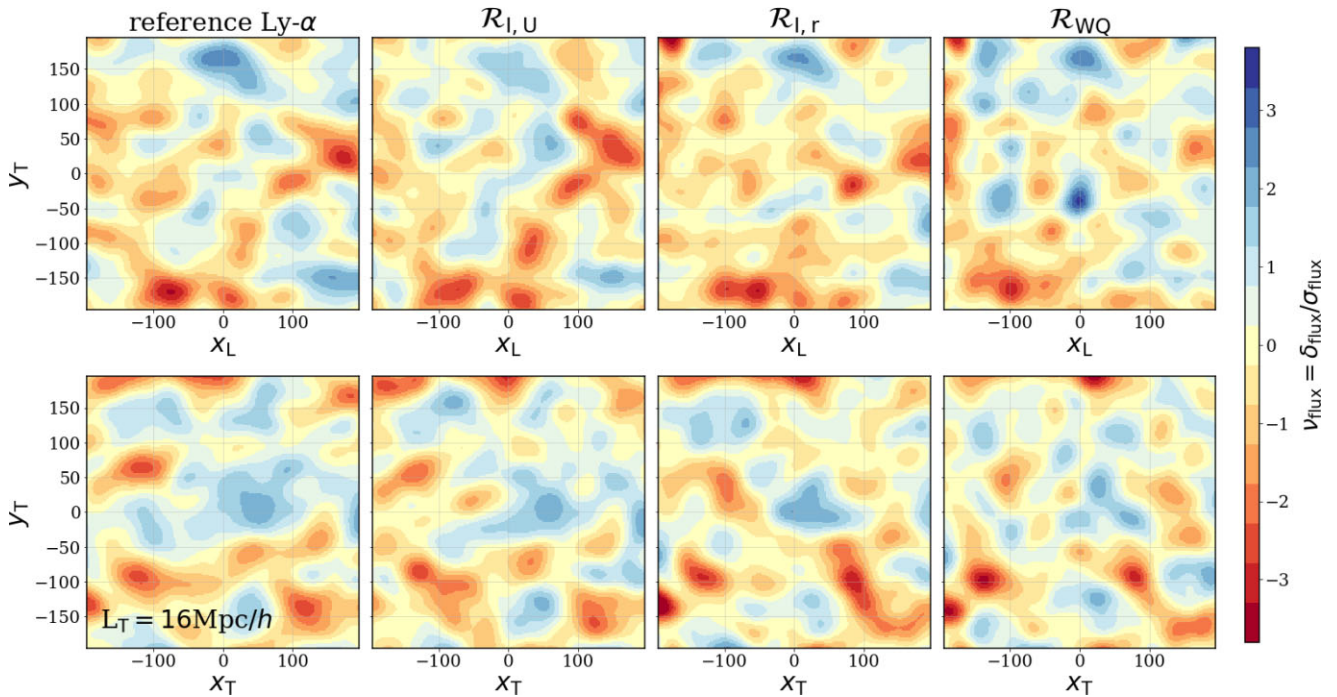
Figure 4. A visualization of 32 Mpc h^{-1} thick slices (parallel to the line of sight) in the flux contrast (in units of rms fluctuations) of the reference Ly-α field (left-hand panel, smoothed at 16 Mpc h^{-1}) and the reconstructed fields ($L_T = 16 \sim$ Mpc h^{-1}) when increasing sparsity/noise in the input data set: a sparse but regular distribution of sightlines without noise on spectra ($\mathcal{R}_{I,U}$, middle left-hand panel), a random distribution of sightlines without noise added to spectra ($\mathcal{R}_{I,r}$, middle right-hand panel), and a random distribution of sightlines with a realistic SNR distribution ($\mathcal{R}_{WQ}$, right-hand panel). As sparsity and noise increase, structures tend to be more disconnected (which will create more critical points, in particular wall-like and filament-like saddle points). The bottom panel shows transverse slices (i.e. perpendicular to the line of sight) through the same fields. The smoothing scale is identical in the top and bottom rows.

2.3.2 Specific settings towards WEAVE-QSO

For the $\mathcal{R}_{WQ}$ mocks, n^2 is determined from the noise on each sightline, so that the contribution of noisy sightlines to the reconstructed map is filtered. However, we set a cap to SNR = 16 to avoid the reconstruction being dominated by a few sightlines with very high SNR. The variance $\sigma_{\delta\delta}^2 = 0.06$ was directly estimated from the variance of the Ly-α flux on the noiseless simulated spectra. For the data covariance matrix (encoding the correlation in the input simulation, the resolution of which is ~ 2 Mpc h^{-1}, see Section 2.1), we use a correlation length of 2 Mpc h^{-1} in the three directions, while for the parameter-data covariance matrix we adopt $L_x = 2$ Mpc h^{-1} (which corresponds to our spectral resolution) and $L_T = 16$ Mpc h^{-1}. This value for the transverse correlation length L_T is slightly larger than the mean separation between sightlines, and is chosen because we cannot hope to reconstruct structures at a smaller scale than roughly the mean distance between sightlines. Appendix A explores how the reconstruction degrades when decreasing L_T, while Appendix E shows its impact on the clustering of critical points.

In order to obtain an isotropic field, which is more convenient to use in our analysis to investigate the clustering of critical points, the reconstructed three-dimensional map is subsequently smoothed with an anisotropic Gaussian kernel of standard deviation 2 Mpc h^{-1} in the transverse direction and 16 Mpc h^{-1} in the longitudinal direction, which ensures a globally isotropic reconstruction at a scale of $\sqrt{16^2 + 2^2} \sim 16.1$ Mpc h^{-1}.

Because of the noise on spectra, some pixels on the input data set can exhibit (non-physical) flux values larger than 1 or smaller than 0. Before performing the reconstruction, we cap these values to 1 and 0 respectively. Matrix inversion is performed through lower-upper (LU) decomposition, and to save time, the reconstruction is

performed in parallel on overlapping boxes with a larger buffer regions of width $2.5 \times L_T$ (we checked that decreasing the width of the buffer region leads to spurious critical points).

After having performed the reconstruction on all set of simulated spectra ($\mathcal{R}_{I,U}$, $\mathcal{R}_{I,r}$, $\mathcal{R}_{WQ}$, and $\mathcal{R}_{noise}$), the flux contrast in the reconstructed map is converted into a pseudo H I density, using the following transformation: $f: \delta \mapsto -\log((\delta + 1) \times \langle F \rangle)$, where $\langle F \rangle = 0.795$ is the mean Ly-α flux in the simulation. In practice, given that the fluctuations in the flux contrast are of small amplitude (because the field is smoothed at such large scales), $f(\delta) \sim -\delta + \log(\langle F \rangle)$. The same transformation is applied to the simulated Ly-α reference field after smoothing with an isotropic Gaussian kernel of standard deviation 16 Mpc h^{-1}.

Throughout this paper, these smoothed reconstructed maps converted into the pseudo H I density are also compared with the original DM and HI smoothed with an isotropic Gaussian kernel of standard deviation 16 Mpc h^{-1}. We note that the only difference between the Ly-α reference field and the H I reference field is that for the former smoothing has been applied on the flux before converting it into pseudo H I density, while for the latter the order of the transformations is reversed. Note finally that in what follows all statistics are computed in units of the root mean square (rms) fluctuations of the field.

2.3.3 Visualization of the reconstruction

Fig. 4 shows the reconstruction in the various configurations in units of the rms fluctuations of each field. Projection in slices of thickness 32 Mpc h^{-1} (twice the smoothing scale) parallel and perpendicular to the line of sight are plotted in the top and bottom
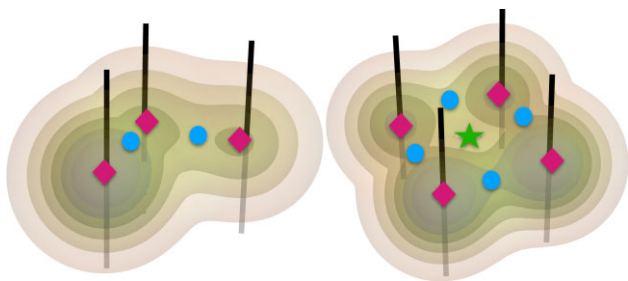
Figure 5. The impact of adding noise along the line of sight on the topology of the reconstructed field hence the number of critical points. For example, in the left-hand panel, positive noise along three (vertical) LOS (separated by more than the smoothing scale) are sufficient to generate two spurious filament-like saddle points (blue circle), on top of the three generated maxima (magenta diamond). In the right-hand panel, positive noise along four LOS generates an extra spurious wall-like (green star) saddle point in between the line of sights. Illustratively, uncorrelated noise could lead to such configurations, respectively, 1/4 and 1/8th of the time.

Table 1. Mean number of peaks ($\mathcal{P}$), filaments ($\mathcal{F}$), walls ($\mathcal{W}$), and voids ($\mathcal{V}$) for the DM density, the H I density, the Ly-α reference field, and the three types of reconstruction used in this work at smoothing scale 16 Mpc h^{-1}.

	$\mathcal{P}$	$\mathcal{F}$	$\mathcal{W}$	$\mathcal{V}$
DM (density)	2045 ± 38	6264 ± 62	6170 ± 58	1958 ± 17
	1908 ± 39	5273 ± 59	5236 ± 37	1874 ± 17
H I (density)	2161 ± 27	6599 ± 72	6736 ± 68	2080 ± 26
	2013 ± 25	5587 ± 63	5544 ± 46	1976 ± 23
Ly-α	2132 ± 27	6627 ± 64	6644 ± 59	2156 ± 24
	2002 ± 27	5604 ± 56	5633 ± 48	2031 ± 26
$\mathcal{R}_{I,U}$	2582 ± 24	7678 ± 61	7531 ± 82	2409 ± 38
	2376 ± 25	6565 ± 51	6410 ± 69	2242 ± 33
$\mathcal{R}_{I,r}$	2697 ± 34	7858 ± 115	7742 ± 86	2578 ± 41
	2463 ± 22	6717 ± 30	6606 ± 50	2375 ± 37
$\mathcal{R}_{WQ}$	3003 ± 34	8883 ± 143	8856 ± 192	2964 ± 28
	2718 ± 25	7627 ± 45	7603 ± 73	2714 ± 28
$\mathcal{R}_{noise}$	3202 ± 49	9576 ± 150	9583 ± 146	3191 ± 39
	2894 ± 41	8196 ± 105	8200 ± 97	2911 ± 36

Notes. The second line reports the number of critical points after the borders removal. The errors represent the standard deviations.

panels respectively, for the Ly-α reference field (extreme left-hand panel), the reconstructed field with a regular distribution of sightlines $\mathcal{R}_{I,U}$ (middle left-hand panel), the reconstructed field with a random distribution of sightlines $\mathcal{R}_{I,r}$ (middle right-hand panel), and the reconstructed field with a realistic noise on spectra $\mathcal{R}_{WQ}$ (extreme right-hand panel). Adding noise on spectra and increasing sparsity creates more structures, which will be reflected in change in the topology of the excursion set and therefore in the number and clustering of critical points.

Fig. 5 describes qualitatively how noise along the line of sight may induce the appearance of extra critical points in their vicinity. Assuming that the noise dominates the large-scale structure density within that region, the reconstruction of spurious overdensities will bridge the field in between sightlines. Such bridges will contain extra filament-like and possibly wall-like saddle points (shown using the same colour coding as the previous figures). This effect would mostly impact sightline separations larger than the smoothing scale.

2.4 Critical point statistics estimators

The critical points of a field are the points where the gradient vanishes (Milnor 1963). They are classified by the sign of the eigenvalues of the matrix of second derivatives of the field: the signature of a critical point is the number of negative eigenvalues, from three for peaks to zero for voids (similarly to e.g. T-web classification; Hahn et al. 2007). For their identification, we implement a local quadratic estimator based on a second-order Taylor-expansion of the density field (see Gay et al. 2012, appendix G).

In the following, the critical points are extracted from the *smoothed density* fields. In the main text, the adopted smoothing scale is 16 Mpc h^{-1}. For DM and H I, we first smooth the density fields with an isotropic 3D Gaussian kernel (of standard deviation 16 Mpc h^{-1}) before extracting the critical points. For both the Ly-α reference field (smoothed at 16 Mpc h^{-1}) and the reconstructed fields $\mathcal{R}_{I,U}$, $\mathcal{R}_{I,r}$, $\mathcal{R}_{WQ}$, and $\mathcal{R}_{noise}$, we applied first the transformation to pseudo H I density (described at the end of Section 2.3.2) before extracting the critical points.

Following Shim et al. (2021), we rely on the so-called Davis–Peebles estimator (Davis & Peebles 1983) for the (cross)-correlations, ξ_{ij}, which is given by

$$1 + \xi_{ij}(r) = \frac{\langle C_i C_j \rangle}{\sqrt{\langle C_i R_j \rangle \langle C_j R_i \rangle}} \sqrt{\frac{N_{R_i} N_{R_j}}{N_{C_i} N_{C_j}}}. \quad (3)$$

Here C_i stands for a catalogue of critical points $i \in \{\mathcal{P}, \mathcal{F}, \mathcal{W}, \mathcal{V}\}$, while R_i is a catalogue with randomly distributed points following a uniform probability distribution in the same volume. The expectation, $\langle XY \rangle$, measures the number count of pairs of critical points X and Y separated by r. The sample size, N_{R_i}, of the random catalogue is a factor of 100 larger than the corresponding size of the simulated set N_{C_i}.

To avoid edge effects, we choose to discard from the statistics the critical points that are closer than ≈ 9.8 Mpc h^{-1} (five grid cells) from the side of the box.[5]

3 STATISTICS OF CRITICAL POINTS

Let us now quantify our ability to recover the statistics of critical points from WEAVE-QSO-like survey. Since we follow closely the choices (estimator, rarity) made in Shim et al. (2021), which strongly impacts the shape of the correlation functions, we will mostly focus our discussion on the relative difference between the input and recovered relations, given that the origin of the features have been addressed in that paper.

3.1 Critical points' total number counts

Table 1 shows the numbers of critical points of each type (peaks, filaments, walls and voids) in the flux contrast of the Ly-α and three reconstructed fields. For completeness, we also provide the corresponding numbers for the DM and H I density fields, as well as for the noise-only field $\mathcal{R}_{noise}$ that will serve as a reference for the quantification of the reconstructed signal. For the DM density

---

[5]Note that the simulation boxes are periodic by construction, however, this periodicity is not fully preserved by our implementation of the reconstruction. As confirmed by the visual inspection of the distribution of critical points near the borders, the choice of removing 5 grid cells is quite conservative. However, we checked that choosing a larger value, e.g. corresponding to L_T, does not affect our subsequent results.

Table 2. Fraction of the mean number of filaments over peaks ($\mathcal{F}/\mathcal{P}$), filaments over walls ($\mathcal{F}/\mathcal{W}$), and voids over peaks ($\mathcal{V}/\mathcal{P}$) for the Ly-α reference field, the three types of reconstruction used in this work, and noise-only field at smoothing scale 16 Mpc h^{-1}.

	$\mathcal{F}/\mathcal{P}$	$\mathcal{F}/\mathcal{W}$	$\mathcal{V}/\mathcal{P}$	$\mathcal{W}/\mathcal{V}$
Ly-α	3.11 ± 0.02	0.99 ± 0.01	1.01 ± 0.01	3.08 ± 0.02
$\mathcal{R}_{I,U}$	2.97 ± 0.02	1.02 ± 0.01	0.93 ± 0.01	3.08 ± 0.03
$\mathcal{R}_{I,r}$	2.91 ± 0.03	1.02 ± 0.01	0.96 ± 0.01	3.0 ± 0.03
$\mathcal{R}_{WQ}$	2.96 ± 0.03	1.0 ± 0.01	0.98 ± 0.01	2.98 ± 0.03
$\mathcal{R}_{\mathrm{noise}}$	2.99 ± 0.02	0.99 ± 0.01	0.99 ± 0.01	3.0 ± 0.02

field, the number of peaks is higher than the number of voids and the number of filaments is higher than the number of walls. This deviates, as expected, from GRF predictions, for which, due to the symmetry, the number of extrema (peaks and voids) is predicted to be equal, and similarly for saddle points (filaments and walls), as is the case of the noise-only field $\mathcal{R}_{\mathrm{noise}}$. In addition, for GRF, the ratio of the number of filaments to the number of peaks (or walls to voids) is predicted to be exactly $(29\sqrt{15} + 18\sqrt{10})/(29\sqrt{15} - 18\sqrt{10}) \approx 3.05$, see Appendix B.

It is also expected (see Gay et al. 2012), that at first non-Gaussian perturbative order, the total number of extrema (voids and peaks) and the total number of saddles (filaments and walls) is preserved, and therefore also their ratio. Given the smoothing scales involved, we should not be far from this regime. Indeed, for the DM density field, filaments and walls are found to be about three times (3.1 and the ratio is ~ 2.8 after the borders removal) more abundant than peaks and voids. Similarly, for the H I density field derived from the DM density, the ratio between the number of saddles (filaments together with walls) and the number of extrema (peaks together with voids) is close to 3 (3.1 and ~ 2.8 after the borders removal), and the number of identified peaks is larger than the number of voids; see Table 2. This ratio is half the mean connectivity of the cosmic web (Codis, Pogosyan & Pichon 2018; see Section 3.5). As such its robustness is not unexpected.

In contrast, filament-type saddles are found to be less numerous than walls. This could be caused by the effect of reduced periodicity of the H I field as suggested by the fact that the trend reverses after the removal of the critical points near the borders. In practice, we remove five pixels (corresponding to ≈ 9.8 Mpc h^{-1}) from each side of the box along each direction to ensure consistency w.r.t the reconstructed fields without periodicity.

For the Ly-α reference field, the number of voids is higher than the number of peaks and the number of walls is higher than the number of filaments even after boundary trimming. However, the ratio of saddles over extrema is preserved (~ 3.1 and 2.8 after the boundary removal; see Table 2).

For all types of reconstruction, the number of peaks is larger than the number of voids and the number of filaments is larger than the number of walls and the ratio of saddles over extrema is also close to 3. However, the number of critical points in any reconstructed field is higher than the number of critical points of the original field. This fraction is lowest for $\mathcal{R}_{I,U}$ (1.15 at $L_T = 16$ Mpc h^{-1}) and highest for $\mathcal{R}_{WQ}$ (1.35 at $L_T = 16$ Mpc h^{-1}). This is expected, since $\mathcal{R}_{I,U}$ corresponds to the less noisy reconstruction (regular distribution of sightlines and no noise on spectra). Overall, all the different sources of noise (sparsity of the sightline distribution, irregularity of their spatial distribution, noise on spectra) result in an increase of the number of critical points. For all types of reconstruction, this fraction is slightly higher for peaks than for voids.

Nonetheless, as expected for sufficiently large volumes (Shim et al. 2021), the ratio between the number of peaks and walls over filaments and voids remains close to one for all fields and reconstructions.

3.2 One-point function of critical points

Let us now study the distribution of the critical points number counts as a function of rarity in both the original and reconstructed fields. The rarity of the critical point is defined as $\nu \equiv \delta/\sigma$, with the overdensity contrast of the smoothed density field $\delta \equiv \rho/\bar{\rho} - 1$ and σ the rms fluctuation of the field $\sigma^2 \equiv <\delta^2>$. Our purpose in choosing rarity is to sample populations that represent the same abundance for a given type of critical points. This allows us to limit the number of configurations we investigate. The lack of overlap in rarity values induces exclusion zones in correlation functions (see Section 3.3 and appendix B of Shim et al. 2021, for a more extended discussion). For illustration, Fig. C1 shows the projection of ν, in slices of thickness 32 Mpc h^{-1} (twice the smoothing scale) for the Ly-α reference field (extreme left-hand panel) and the three reconstructed field $\mathcal{R}_{I,U}$ (middle left-hand panel), $\mathcal{R}_{I,r}$ (middle right-hand panel), and $\mathcal{R}_{WQ}$ (extreme right-hand panel) with overplotted the identified critical points (in red for peaks, in blue for filaments, in green for walls and in orange for voids). The visual inspection of these maps confirms that more critical points are identified in all reconstructed fields compared to the reference Ly-α flux. However, in spite of projection effects, it can be seen, that several critical points are quite robustly identified across all fields. A quantitative measure of this effect is provided in the remainder of this paper.

To quantify the effect of bias, Fig. 6 shows the relative number counts of critical points (top panel) and their difference (bottom panel) as function of rarity for the DM and H I density fields (left-hand panel) and for DM density and Ly-α reference fields (right-hand panel). For each type of critical point, the number counts are normalized by the total number of critical points. Let us also stress that the rms fluctuation of each field used to define ν is computed independently for each of them. In particular, this means the linear bias is factored in (see e.g. Bagla, Khandai & Datta 2010, for the H I linear bias). The bottom panel of the figure thus allows us to probe the non-linear bias of different critical points. The maximum amplitude of the rarity distribution is higher for voids (walls) than for peaks (filaments) with the effect being stronger for DM density field. The distributions clearly show a positively-skewed rarity driven by gravitational clustering (as expected; see Gay et al. 2012). At low rarity, more voids are identified in the H I density fields compared to the DM density field, while at rarity corresponding to the maximum of the rarity distributions the trend is reversed. Similar behaviour is found for rarity distribution of walls, such that there are more walls at low rarity in the H I density field compared to DM density field, while at intermediate rarity the number of walls is higher in the DM field. On the other hand, the trends are reversed for filaments and peaks. At their respective intermediate rarity (near the maxima of rarity distributions), more filaments (peaks) are identified in the H I density field than in the DM density field, while at highest rarity the number of filaments (peaks) is higher for the DM. Interestingly, the differences of the relative number counts (bottom panel) for peaks and voids are antisymmetric with respect to $\nu = 0$ (similarly for filaments and walls), which is likely to reflect the fact that these critical points are oppositely biased tracers. Qualitatively, similar trends are found when comparing Ly-α fluxes with the DM density field, with enhanced relative differences highlighting the observational bias associated with measuring the fluxes. This is reflected by the
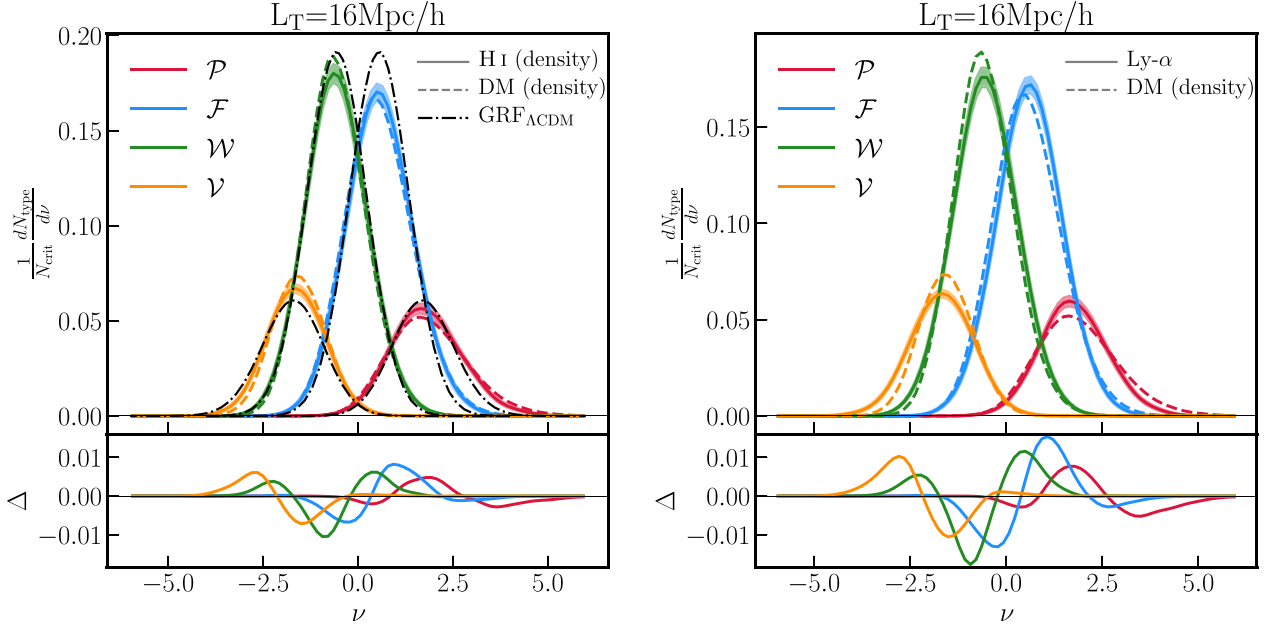
Figure 6. Effect of bias. Top panel: relative number counts of critical points ($\mathcal{P}$: peaks, $\mathcal{F}$: filaments, $\mathcal{W}$: walls, and $\mathcal{V}$: voids) as functions of their rarity ν for the H I density and Ly-α (solid lines, left- and right-hand panels, respectively), compared to DM density fields (dashed lines) and Gaussian random field with ΛCDM power spectrum (dash–dotted lines) at the smoothing scale of 16 Mpc h^{-1}. The shaded area corresponds to the standard deviation across five mocks, shown for clarity only for the H I field. Bottom panel: difference of the relative number counts of critical points in the H I and DM density (left-hand panel), and Ly-α and DM density (right-hand panel).

differences in the tails of the PDFs, capturing rare events such as the flux saturation of the densest peaks.

Let us now focus on the comparison between the Ly-α reference field and the three types of reconstruction. Fig. 7 shows the number counts of critical points as a function of rarity for the Ly-α flux (solid lines) together with the three configurations adopted for the reconstruction, namely $\mathcal{R}_{I,U}$ (dashed lines), $\mathcal{R}_{I,r}$ (dotted lines), and $\mathcal{R}_{WQ}$ (dash–dotted lines). As already noted in Section 3.1, the number of critical points in any reconstructed field is larger than in the Ly-α reference field, with smallest differences for $\mathcal{R}_{I,U}$ and largest for $\mathcal{R}_{WQ}$. The differences show a dependence on the rarity of a critical point. For filaments and walls they are confined to regions of intermediate rarity (in the vicinity of the maxima of the number counts distributions), while for peaks and voids the differences are more uniformly distributed over a much larger range of rarities. While the largest differences between the Ly-α and reconstructed fields are measured for filaments and walls, cumulative differences (for all rarities) are larger for peaks and voids (see Section 3.1).
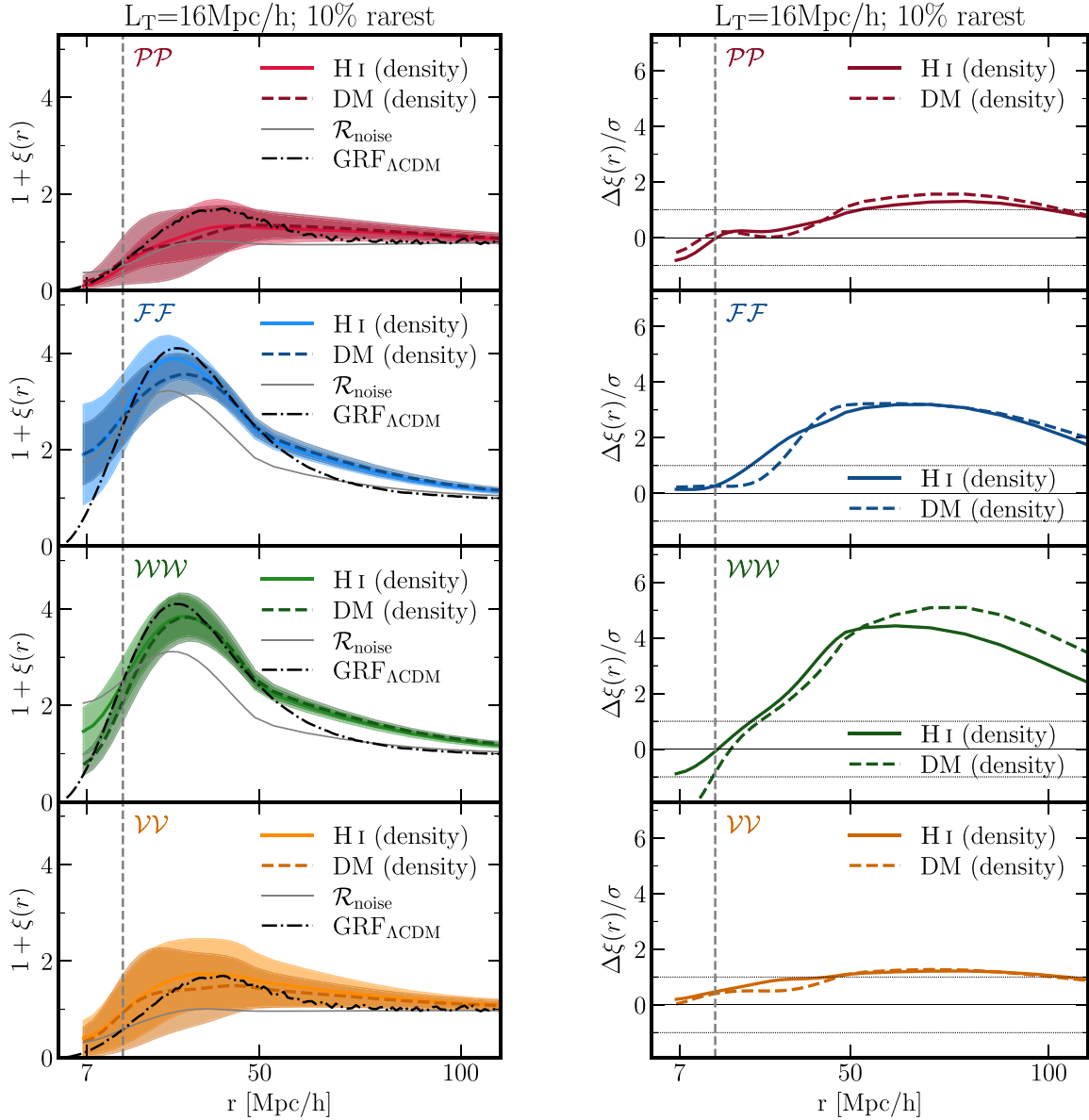
3.3 Autocorrelation of critical points

Let us now move to the two-point statistics, starting with auto-correlation functions. Throughout the paper, we show results for the 10 per cent rarest critical points selected as follows. For peaks and filaments (respectively, voids and walls), the critical points are extracted above (respectively, below) the rarity threshold $\nu_{\mathrm{type,\,c}}$ yielding a given abundance, i.e. $N_{\mathrm{type}}(\nu \geq \nu_{\mathrm{type,\,c}})/N_{\mathrm{type}}$ (respectively, $N_{\mathrm{type}}(\nu \leq \nu_{\mathrm{type,\,c}})/N_{\mathrm{type}}$).

The choice of the 10 per cent rarity is a compromise between very rare outstanding events and therefore noisy measurements on the one hand, and less rare events with less enhanced characteristic features on the other hand (see also Shim et al. 2021, for a discussion
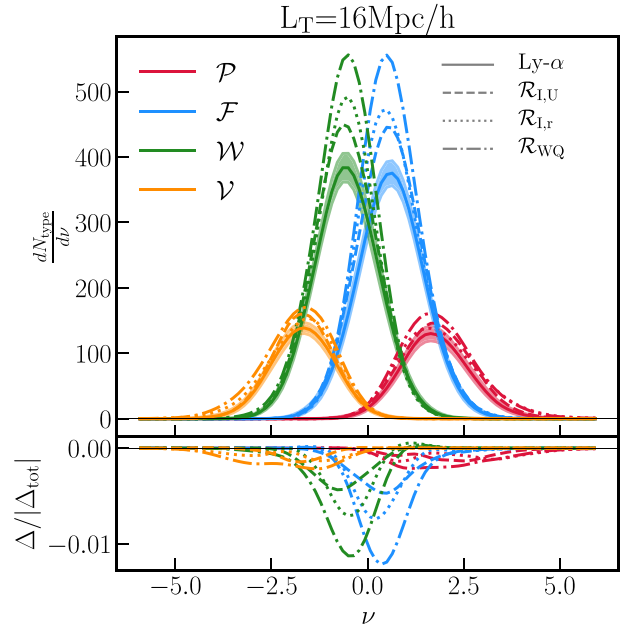
Figure 7. Top panel: absolute number counts of critical points ($\mathcal{P}$: peaks, $\mathcal{F}$: filaments, $\mathcal{W}$: walls, and $\mathcal{V}$: voids) as functions of their rarity ν for the Ly-α (solid lines) and reconstructed fields $\mathcal{R}_{I,U}$ (dashed lines), $\mathcal{R}_{I,r}$ (dotted lines), and $\mathcal{R}_{WQ}$ (dash–dotted lines). Bottom panel: difference of the number counts of critical points of a given type normalized by the absolute value of the difference of the total number of critical points in the original and reconstructed fields. The shaded area corresponds to the error on the mean across the five mocks, shown for clarity only for the Ly-α reference field.
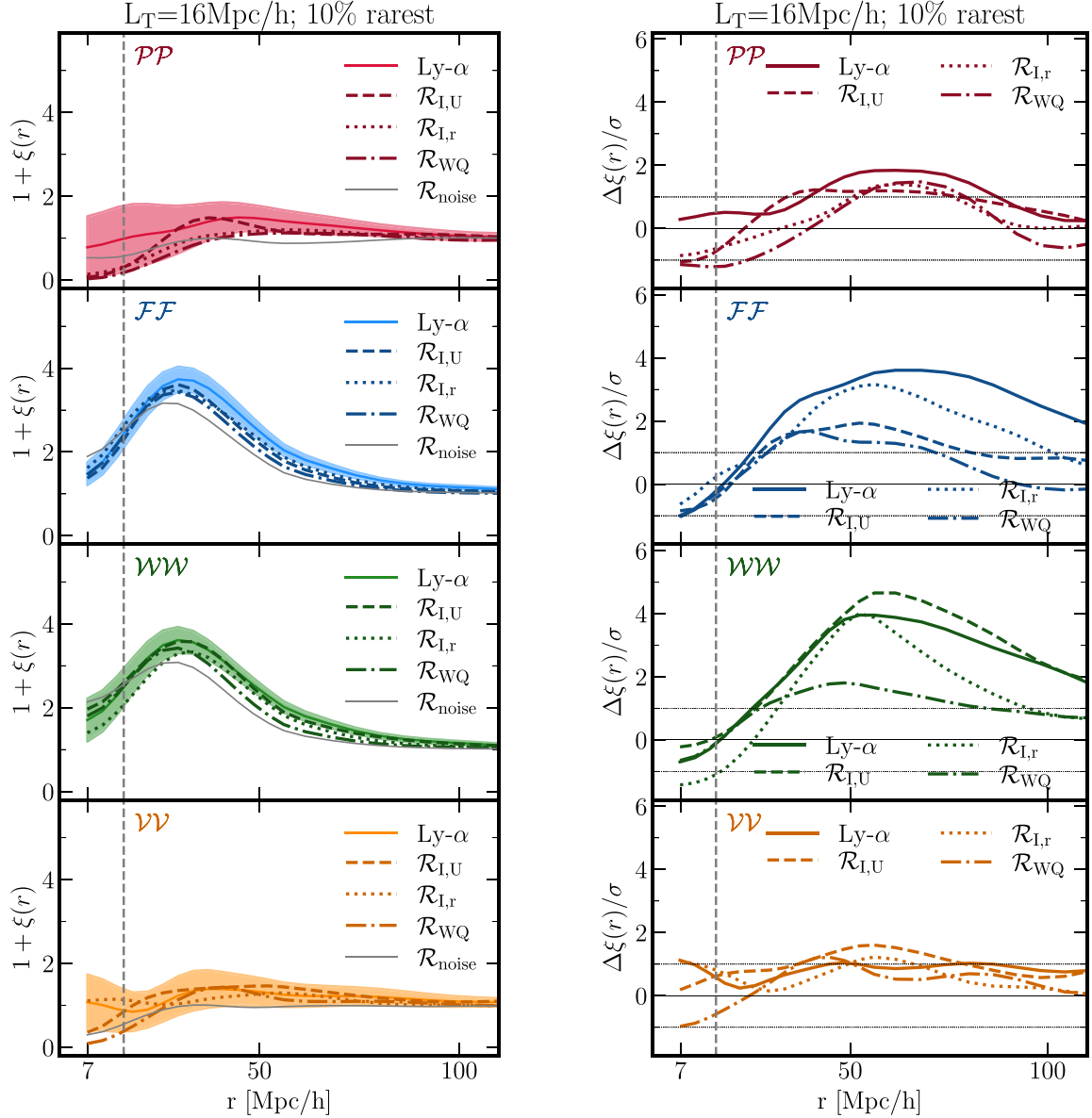
Figure 8. Left-hand panel: autocorrelations of critical points with 10 per cent abundance. $\mathcal{PP}$ (peak-peak), $\mathcal{FF}$ (filament-filament), $\mathcal{WW}$ (wall-wall), and $\mathcal{VV}$ (void-void) correlations are shown for H I density field (coloured solid lines), DM field (coloured dashed lines), and field containing noise-only (grey solid line). Predicted autocorrelations of critical points for ΛCDM spectrum are shown for comparison (black dash–dotted lines). Smoothing of the fields is 16 Mpc h^{-1}. The shaded area corresponds to the error on the mean across five mocks. Right-hand panel: differences of autocorrelations of critical points with respect to the noise in the units of σ of a given field (i) and noise ($\mathcal{R}_{\text{noise}}$), i.e. $\sigma = \sqrt{\sigma_i^2 + \sigma_{\mathcal{N}}^2}$, that we use to assess the significance of the measured signal.

regarding this choice). The relative impact of this specific choice of rarity is addressed in Appendix E.

Fig. 8 (left-hand panels) shows the autocorrelations of the 10 per cent rarest critical points in the H I density (colour solid lines) and DM density (colour dashed lines) fields. The autocorrelations for noise-only field ($\mathcal{R}_{\text{noise}}$; thin grey solid line) that are used to quantify the significance are shown together with the autocorrelations in the Gaussian random field with ΛCDM power spectrum, for a guidance only (black dash–dotted lines). As expected, given the smoothing scale, there is a very good agreement between the H I and DM density fields. The autocorrelations of all critical points follow qualitatively similar trends. At small separations, they are negative ($\xi(r) < 0$), representing a region of anticlustering or

exclusion. They then increase, and reach a positive maximum at $\approx 2L_{\text{T}}$ for filaments and walls and ≈ 2.5–$3L_{\text{T}}$ for peaks and voids (see Table D1), before decreasing towards zero in the regime of large separations. Filaments and walls show enhanced clustering at small separations, the maximum of their autocorrelation function occurs earlier, and their exclusion region is narrower compared to peaks and voids. These differences between the autocorrelations of saddles and extrema potentially manifest different density and curvature conditions around extrema- and saddle-points (Shim et al. 2021).

To quantify the significance of the signal contained in these autocorrelations, we compute the difference between the density (H I or DM) and noise-only fields autocorrelation, in units of the standard deviation ($\sigma = \sqrt{\sigma_i^2 + \sigma_{\mathcal{N}}^2}$, with σ_i and $\sigma_{\mathcal{N}}$ being the
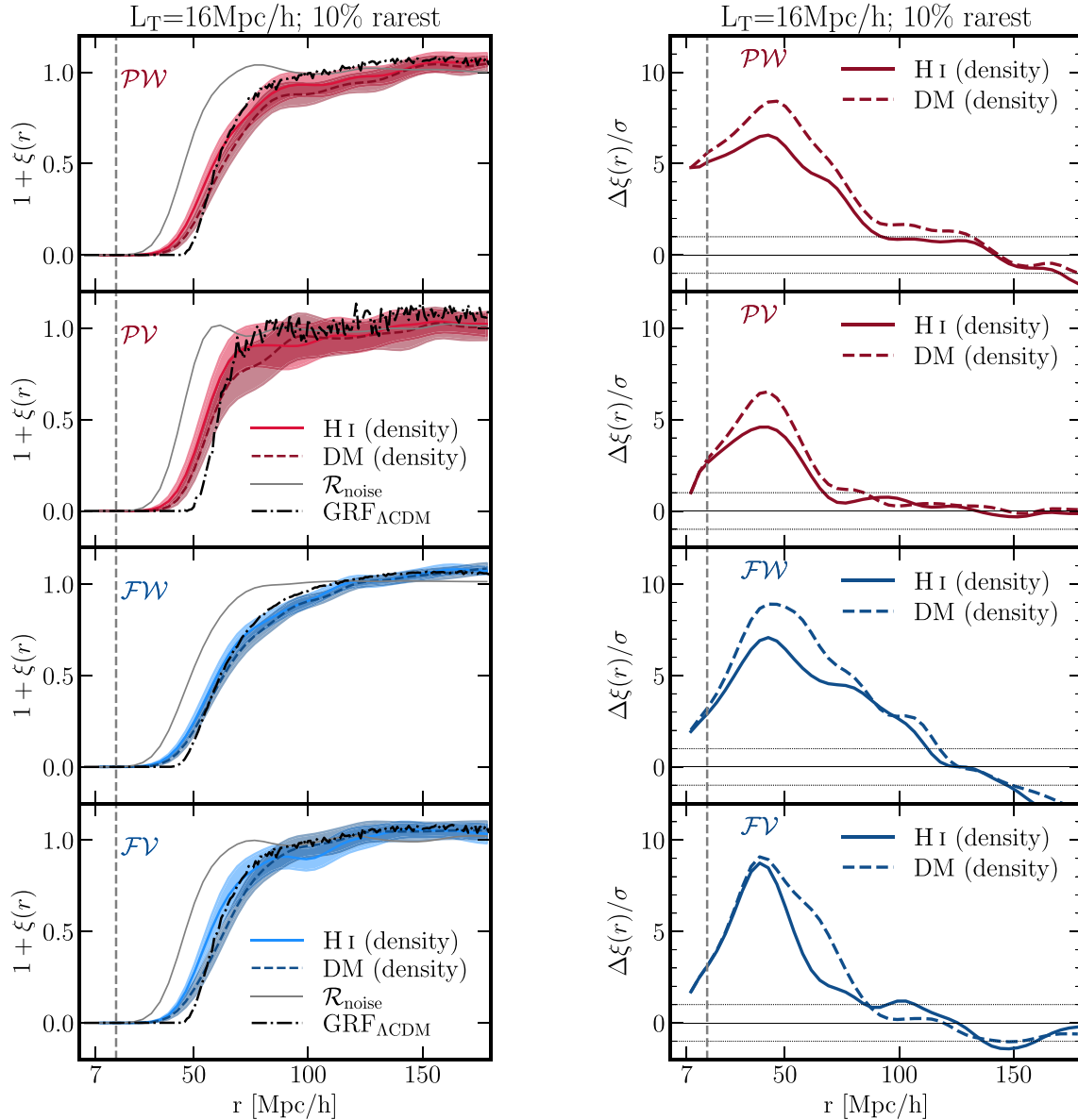
Figure 9. Autocorrelations of critical points with 10 per cent abundance for the Ly-α flux and reconstructed fields (left-hand panel), and their relative difference with respect to the noise (right-hand panel). $\mathcal{PP}$ (peak-peak), $\mathcal{FF}$ (filament-filament), $\mathcal{WW}$ (wall-wall), and $\mathcal{VV}$ (void-void) correlations are shown for Ly-α field (solid lines) and reconstructed fields $\mathcal{R}_{I,r}$ (dashed lines), $\mathcal{R}_{I,U}$ (dotted lines), and $\mathcal{R}_{WQ}$ (dash–dotted lines). The fields is smoothed over 16 Mpc h^{-1}. The shaded area corresponds to the error on the mean across five mocks. For the sake of clarity, only errors for Ly-α are shown, those of $\mathcal{R}_{I,U}$, $\mathcal{R}_{I,r}$, and $\mathcal{R}_{WQ}$ are comparable. As expected, the $\mathcal{FF}$ and $\mathcal{WW}$ correlations are best recovered.

standard deviations of the fields i, H I or DM density, and noise-only, respectively) of the corresponding density field. As shown on Fig. 8 (right-hand panels) these differences are highest for filaments and walls, at $\approx$ 3-4σ, while for peaks and voids, the significance of the measured signal is at the level of $\approx 1\sigma$.

Fig. 9 shows the autocorrelations of the critical points with abundance of 10 per cent in the Ly-α reference field, together with the three reconstructions (left-hand panels) at $L_T = 16$ Mpc h^{-1}. The corresponding differences between the autocorrelations of a given field and noise-only field are shown on the right-hand panels. These autocorrelations show qualitative behaviour similar to the H I and DM density correlations, with an anticlustering at small separations, a maximum at $\approx 2L_T$ for filaments and walls and at ≈ 2.5–$3L_T$ for

peak and voids (see Table D1), and a decrease towards zero at large separations. The quality of the reconstruction is typically the best for the regular distribution of lines of sight ($\mathcal{R}_{I,U}$), it decreases for the reconstruction with the random distribution ($\mathcal{R}_{I,r}$) and degrades further when the noise on the spectra is added ($\mathcal{R}_{WQ}$). The autocorrelations of filaments and walls are better recovered than those of peaks and voids. To quantify the quality of the signal contained in the Ly-α flux, we once again compute the differences of autocorrelations with respect to the noise-only field, shown on the right-hand side of Fig. 9. Remarkably, for filaments and walls, the most striking features in the autocorrelation functions can be measured with up to 5σ of significance for $\mathcal{R}_{I,U}$ and up to 2σ for $\mathcal{R}_{WQ}$. Similar level of significance of the measured features in the autocorrelations is found for the critical

Figure 10. Cross-correlations of critical points with 10 per cent abundance (left-hand panel) and their relative difference with respect to the noise in the units of the total sigma (multiplied by -1 for convenience, right-hand panel). $\mathcal{PW}$ (peak-wall), $\mathcal{PV}$ (peak-void), $\mathcal{FW}$ (filament-wall), and $\mathcal{FV}$ (filament-void) correlations. All these correlations ' main features are well detected, since they involve $\mathcal{F}$ and $\mathcal{W}$. As expected, the GRF correlations, which do not include peculiar velocities, have wider exclusion zones.

points with 20 per cent rarity (see Fig. E1). This is in line with the conclusions that saddle point statistics are more advantageous to use for extracting cosmological information (Gay et al. 2012; Shim et al. 2021) because the cosmic evolution of saddle-points is less non-linear than that of extrema-points (see fig. 10 of Gay et al. 2012). Conversely, the constraining power for peaks and voids is limited at best to 1σ.
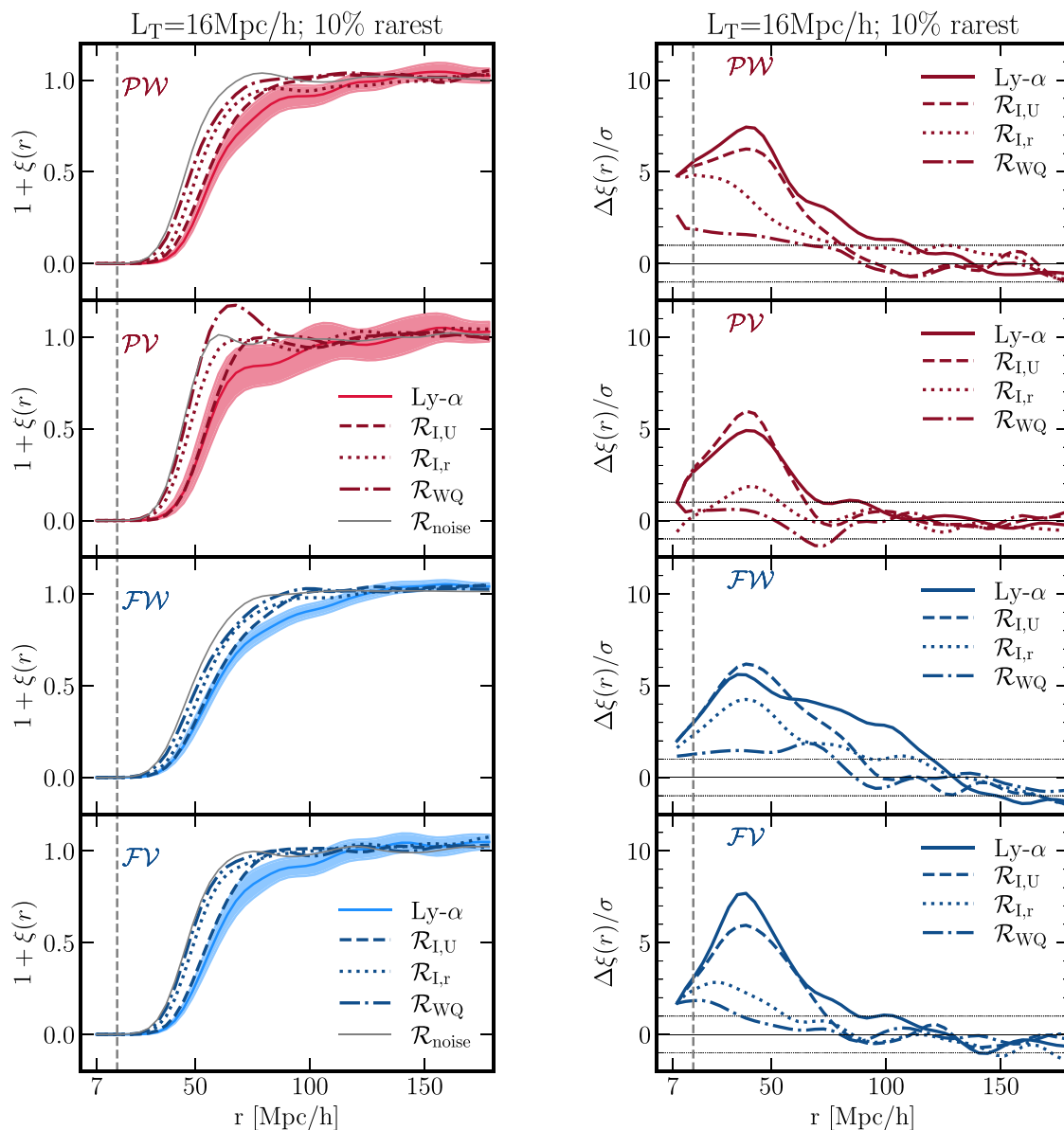
3.4 Cross-correlation of critical points

Let us start by considering the cross-correlation functions of over-dense and underdense critical points, i.e. cross-correlations $\mathcal{PW}$, $\mathcal{PV}$, $\mathcal{FW}$, and $\mathcal{FV}$. The cross-correlations of critical points of

the same overdensity sign, i.e. $\mathcal{PF}$ and $\mathcal{WV}$ are addressed in Section 3.4.2.

3.4.1 Overdense and underdense critical points

Fig. 10 shows such cross-correlation functions with 10 per cent abundance for the H I and DM density fields (coloured solid and dashed lines, respectively). As in the case of autocorrelations, there is a very good agreement between the H I and DM density fields and the cross-correlations of overdense and underdense critical points follow qualitatively similar trends. However, the cross-correlations of overdense and underdense critical points are very different from the autocorrelations. At small separations there is an exclusion zone where $\xi(r)$ is close to -1, but then these cross-correlations
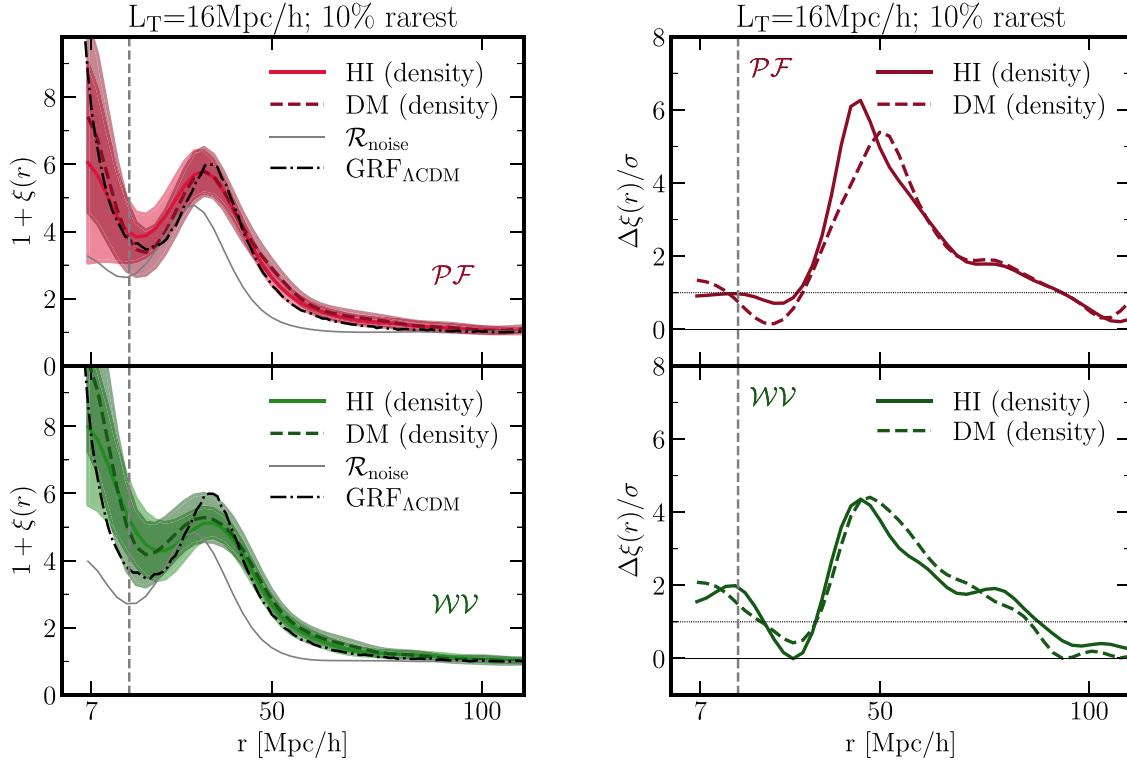
Figure 11. Left-hand panel: cross-correlations of critical points with 10 per cent abundance. Left-hand panel: $\mathcal{PW}$ (peak-wall), $\mathcal{PV}$ (peak-void), $\mathcal{FW}$ (filament-wall), and $\mathcal{FV}$ (filament-void) correlations for the original field (solid) and three reconstructions, $\mathcal{R}_{I,U}$ (dotted), $\mathcal{R}_{I,r}$ (dotted), and $\mathcal{R}_{WQ}$ (dash–dotted). Right-hand panel: difference of cross-correlations between the noise and other fields.

monotonically increase, reaching zero at large separations. The size of the exclusion zone for H I and DM density field is significantly larger ($>4\sigma$) compared to the the noise-only field for all cross-correlations of under and overdense critical points (see Table D2). Note that the cross-correlations are negative at all separations, meaning that the overdense and underdense critical points are always anticorrelated. The appearance of the exclusion zone and anticlustering are consequences of the fact that these critical point pairs are oppositely biased tracers of the underlying DM density field. This is due to both curvature and density continuity constraints that force the positively and negatively biased critical points to strongly separate, as mentioned in Shim et al. (2021).

The significance of the outstanding features contained in these cross-correlations is again quantified. As shown on the right-hand side of Fig. 10, the exclusion zone is very well constrained, with a significance level of up to $\approx 8\sigma$.

Moving on to the comparison between the Ly-α and the reconstructed fields, Fig. 11 shows their cross-correlations (left-hand panel) together with the normalized differences with respect to the noise-only field (right-hand panel). For all fields, all four types of cross-correlation functions exhibit the same features, (i) an exclusion zone at small separations, (ii) a monotonic increase towards zero at large separations, and (iii) anticorrelation ($\xi(r) < 0$) at all scales. The quality of the reconstruction is again highest for the regular distribution of the sightlines ($\mathcal{R}_{I,U}$), decreases for the random distribution of the sightlines ($\mathcal{R}_{I,r}$) and is lowest with added noise on the spectra ($\mathcal{R}_{WQ}$). The quality of the reconstruction also shows a variation with the type of the cross-correlation, in particular for the realistic configuration $\mathcal{R}_{WQ}$. While for $\mathcal{R}_{I,U}$, the exclusion zone is constrained up to $\approx 6\sigma$ for all cross-correlations, for $\mathcal{R}_{WQ}$ it is only at the level of $\approx 1\sigma$ for peak-wall ($\mathcal{PW}$), filament-wall ($\mathcal{FW}$), and filament-void ($\mathcal{FV}$), and it decreases well below 1σ

Figure 12. Cross-correlations of critical points with 10 per cent abundance (left-hand panel) at the smoothing scale $L_T = 16$ Mpc h^{-1} for H I (solid coloured lines) and DM density (dashed coloured lines) fields and their differences relative to the noise in units of sigma (right-hand panel). $\mathcal{PF}$ (peak-filament) and $\mathcal{WV}$ (wall-void) correlations are shown on the top and bottom panels, respectively. The vertical dashed grey line indicates the smoothing scale.

for peak-void ($\mathcal{PV}$) correlations (see Table D2 for the size of the exclusion zone for all fields). Contrarily to autocorrelations, the significance of the cross-correlations of overdense and underdense critical points increases with decreased rarity, in particular for $\mathcal{FW}$, where the level of significance reaches 4σ at 20 per cent abundance (see Fig. E2).

3.4.2 Same overdensity sign critical points

The two remaining cross-correlation functions are between the critical points of the same overdensity sign, i.e. peak-filament ($\mathcal{PF}$) and wall-void ($\mathcal{WV}$) correlations.

Fig. 12 shows these cross-correlations for H I and DM density fields (coloured solid and dashed lines, respectively) and 10 per cent abundance, together with the noise-only (thin grey line) and GRF with ΛCDM power spectrum (black dash–dotted line) for a comparison. The cross-correlations of the critical points of same overdensity sign have a fundamentally different behaviour compared to the cross-correlations of overdense and underdense critical points (Fig. 10). They diverge at zero separation without exhibiting any exclusion zone, nor anticlustering at small separations; they have a local maximum at intermediate separations and finally, they approach zero at large separations. As discussed in Shim et al. (2021), the divergence at zero separation is expected for two critical points with signatures difference of one and with overlapping ranges of density values. This behaviour is connected to the merging rate of these critical points at $r \to 0$ when the field is smoothed on increasing scales (Cadiou et al. 2020). The position of the local maximum r_{max} is an expected geometrical feature of the cosmic web. This enhanced probability to find high peaks near high filament-type saddle points in the peak-

filament cross-correlations (which defines a statistically preferred distance between peaks- and filament-type saddles) is a measure of the typical length of filaments between two peaks, estimated as twice r_{max}. For H I and DM density fields, r_{max} is $\approx 2L_T$ (≈ 33.8 and ≈ 31.5 Mpc h^{-1}, respectively), therefore the typical length of filaments in both field is ≈ 64 Mpc h^{-1} (see Table D1).

In the cosmic web framework, the local maximum in wall-void cross-correlations can also be interpreted as the typical radius of voids, corresponding to ≈ 29.6 and ≈ 33.8 Mpc h^{-1} ($\approx 2L_T$) for H I and DM density field, respectively (see also Table D1).

Finally, Fig. 13 shows the cross-correlation functions of the critical points with the same overdensity sign and with 10 per cent abundance for Ly-α reference field, and is compared to three reconstruction configurations, along with the noise-only field (left-hand panels). The right-hand panels show the relative differences in units of standard deviation with respect to the noise. All salient features of these cross-correlations are recovered by all reconstructions considered in this work. For the peak-filament cross-correlation, the position of the local maximum (r_{max}) is best captured by the $\mathcal{R}_{I,U}$ reconstruction. However, for all fields, it is at $\approx 2L_T$ as in the case of H I and DM density fields. For the height of the local maximum (h_{max}), it is $\mathcal{R}_{I,r}$ that is closest to the reference field Ly-α (see also Table D1). For the wall-void cross-correlation, the r_{max} is equally captured by all types of reconstruction, again at $\approx 2L_T$, while for the h_{max}, $\mathcal{R}_{I,U}$ and $\mathcal{R}_{I,r}$ show a better match compared to $\mathcal{R}_{WQ}$; however, the measured values are comparable, within the error bars.

The existence of a local maximum is constrained with a significance up to 4σ for $\mathcal{R}_{I,U}$ and up to 1.5σ–2σ for $\mathcal{R}_{WQ}$ for both peak-filament and wall-void cross-correlations, in contrast with the significance of the cross-correlations of overdense and underdense critical points.
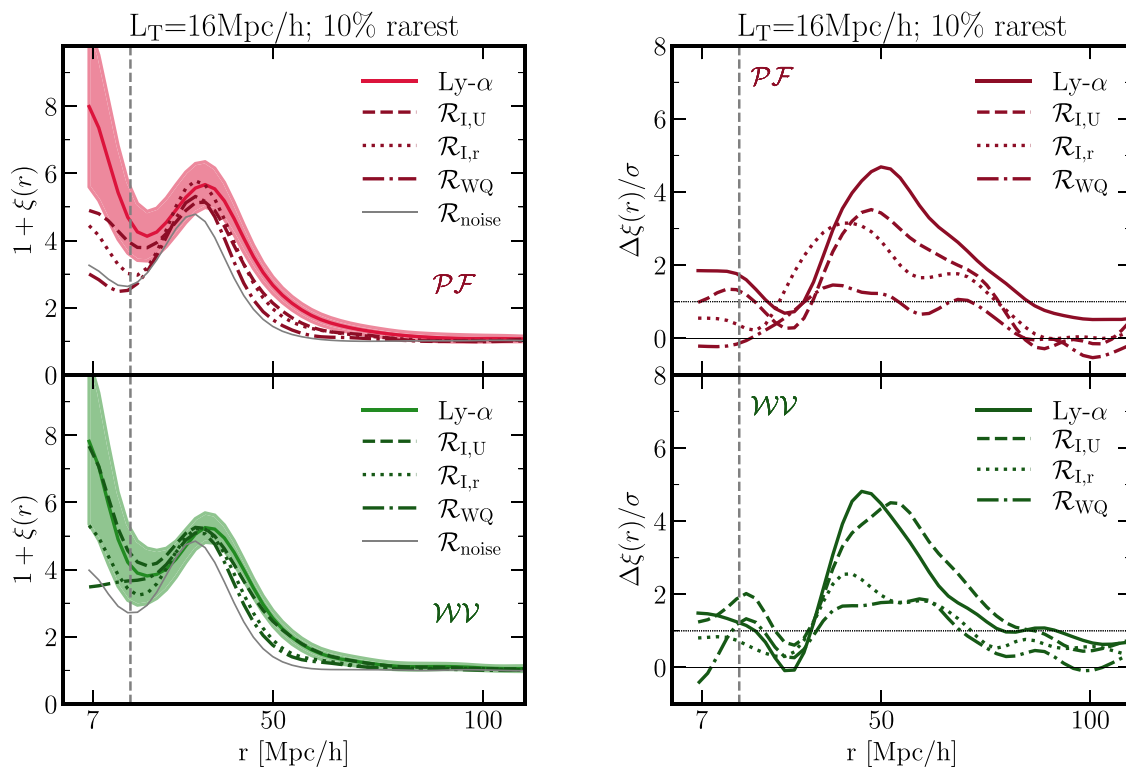
Figure 13. Left-hand panel: cross-correlations $\mathcal{PF}$ (peak-filament) and $\mathcal{WV}$ (wall-void) with 10 per cent abundance at the smoothing scale $L_T = 16$ Mpc h^{-1} for the Ly-α, the reconstructed fields, and noise-only field (left-hand panel). Right-hand panel: the differences of the cross-correlation functions with respect to the noise in the units of total sigma ($\sqrt{\sigma_i^2 + \sigma_{\mathcal{N}}^2}$, with i corresponding to σ of Ly-α, $\mathcal{R}_{I,U}$, $\mathcal{R}_{I,r}$, and $\mathcal{R}_{WQ}$). The vertical dashed grey line indicates the smoothing scale.

3.5 Cosmic connectivity of critical points

Let us finally revisit our clustering results from the slightly different angle of topology (see Caucci et al. 2008, for a first investigation with tomographic reconstruction). Indeed, the relative positions of saddles and peaks impact the filamentary structure emerging from peaks, while the geometry of tunnels of given iso-contours is set by the positions of wall- and filament-saddles. Morse theory (Milnor 1963) establishes a close relationship between the distribution of critical points of the field on the one hand, and the topology of its excursion sets (the iso-contours of the field) on the other hand. The number of connected components within the excursion is one such quantity, and it is controlled by the connectivity of the field, defined as the number of ridges branching out of a given peak towards a given saddle point (Codis et al. 2018).[6] Let us therefore measure this connectivity, in all the fields analysed in this study, using the ridge tracer algorithm DISPERSE (Sousbie et al. 2011)[7] and assess our ability to recover it via a WEAVE-QSO like survey.

Fig. 14 shows the PDF of the connectivity (top panel), the dependence of the median connectivity on the rarity ν (middle panel) for DM and H I density fields, Ly-α flux and the three reconstructed fields at $L_T = 16$ Mpc h^{-1}. While the median connectivity is 5 for all fields, except for Ly-α for which it is 6, the mean connectivity is close

to 6 for all fields. This is completely expected from Table 2 since it should match twice the ratio of the number of saddles to peaks (Pichon et al. 2010). Overall, there is a good agreement for both the PDF of the connectivity and the dependence of the connectivity on ν across all considered fields. As expected, the mean connectivity increases for peaks of higher rarity (Codis et al. 2018), a feature that is recovered via the mock survey. This property seems robust, which is expected since connectivity reflects the underlying topology, hence does not discriminate well small features changes across the various fields.

Going one step further, we measured the cross-match of the connectivity based on closest peak identification. In practice, for each peak in the reference field, we associate the closest peak in the matched field, without imposing any other condition on the match. We then compare the connectivities of these matched peaks. The result is presented in the bottom panel of Fig. 14.[8] It reflects our ability to reconstruct the precise geometry of the field not only statistically, but also locally: the less noisy the reconstruction, the better the agreement. Conversely, the noise-only field falls back to the expected mean connectivity of 6.1 (see Section 3.1). Overall, there is a good agreement between the recovered connectivity and the original one on a peak-to-peak basis. This is not unexpected given the consistency required by Morse theory. It does highlight that not only the critical points' relative distances are well preserved on average (as shown above from their clustering), but more generally their relative positions within the cosmic web (which controls the

[6] For instance, two distinct sets of iso-contours connect when reaching the height of the saddle point in between.

[7] For our purpose, DISPERSE was run on the regular grid of density or flux contrast, depending on the field used, with a persistence threshold of 0.08 so as to obtain a total number of critical points comparable with the numbers given in Table 1.

[8] Note that the matching in the reverse order gives almost identical results.
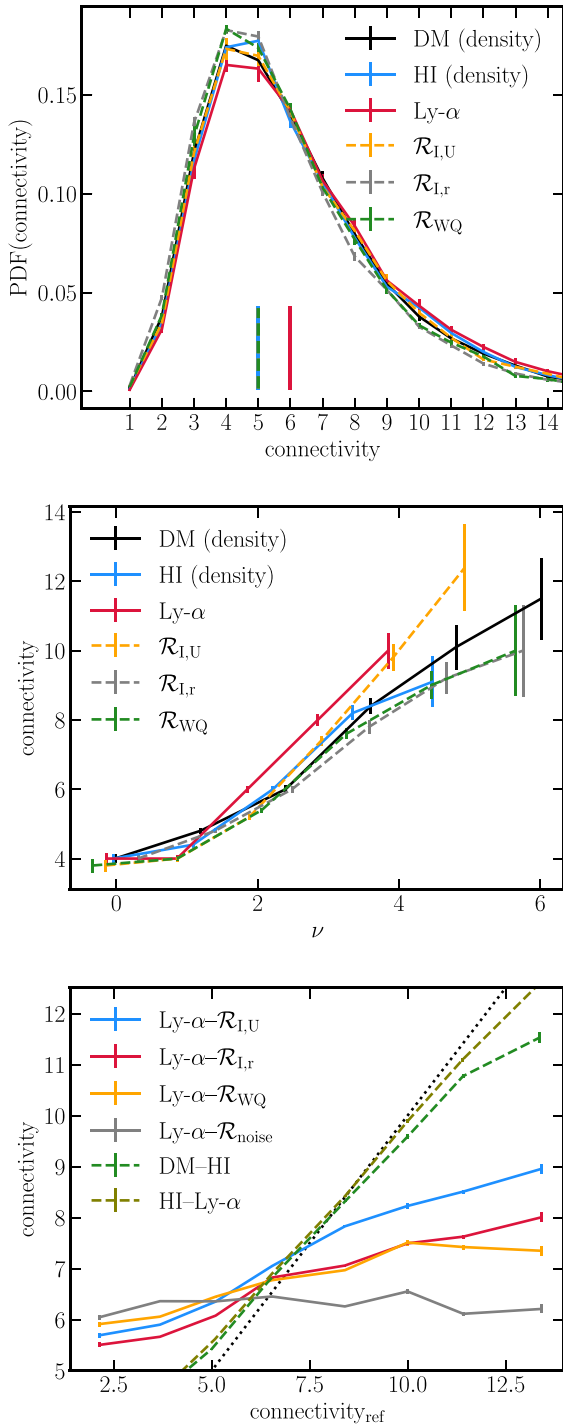
Figure 14. Top panel: PDF of the connectivity for DM and H I density fields, Ly-α flux, and the three reconstructed fields. The vertical lines represent the medians of the connectivity across five mocks. This median is 5 for all fields but Ly-α for which it is 6. The mean values are close to 6 for all fields. Middle panel: connectivity as a function of rarity ν for the fields considered in this work. The mean connectivity of the peaks increases with rarity for all fields. Bottom panel: the connectivity of matched peaks across the range of fields considered. For each match, the reference field is the first in the label. The closer the field to the reference one, the better the match. Note that both peak mismatch and change in the local geometry impact the connectivity. The black dotted line is the diagonal. The error bars are computed as the mean error across five mocks.

topology of the excursion set) is also recovered at various degrees, depending on the level of noise and the distribution of sightlines.

Measuring the connectivity of density peaks on a cluster-by-cluster basis could prove useful beyond cosmology, e.g. in the context of understanding the formation of galaxy groups and clusters (Darragh Ford et al. 2019; Kraljic et al. 2020; Lokken et al. 2021), using tomographic surveys probing smaller scales compared to those exploited in this work, such as e.g. CLAMATO (Lee & White 2016), LATIS (Newman et al. 2020) or upcoming PFS survey (Takada et al. 2014). The size of the exclusion zone on these scales could be used to constrain the geometry of the warm-hot inter-galactic medium bubbles. Beyond the connectivity, one could quantify the orientation and strength of filaments around reconstructed peaks, as was investigated in 2D through stacking of lensing maps (Codis et al. 2017; Gouin et al. 2017). Note finally that tomography would also allow us to compute the dual connectivity of voids, which may prove more robust than that of peaks.

4 DISCUSSION

Let us discuss globally our main findings in terms of summary statistics, error budgets, and upcoming surveys.

4.1 Summary statistics

We start by computing physically motivated summary statistics capturing the relative evolution of the outstanding features in the two-point function of the critical points as a function of SNR and sampling strategy. Specifically, for the autocorrelations and cross-correlations of critical points of the same overdensity sign we use the radius of their maximum $r_{max, ij}$ and the corresponding height $h_{max, ij}$, while for the cross-correlations of overdense and underdense critical points we use the size of the exclusion zone $r_{ex, ij}$.

We recall that $r_{max, ij}$ is defined as the separation at which the autocorrelation $\mathcal{C}_i\mathcal{C}_i$ or cross-correlation $\mathcal{C}_i\mathcal{C}_j$ peak (see Table D1), $h_{max, ij}$ is the height of the maximum of autocorrelation $1 + \mathcal{C}_i\mathcal{C}_i$ or cross-correlation $1 + \mathcal{C}_i\mathcal{C}_j$ (see Table D1), and $r_{ex, ij}$ corresponds to the radius at which the cross-correlation $\mathcal{C}_i\mathcal{C}_j$ correlation departs from -1. In practice, due to the noise on the measurement, we allow for a departure from this value by 0.01 (see Table D2).

From these numbers we extract ratios for the three reconstructed fields and the noise-only field with respect to the reference field Ly-α, and study how these ratios vary with the parameters of the reconstruction method. These ratios are shown in Tables D3 and D4. All of these ratios confirm the overall conclusion based on the detailed analysis of the two-point correlation functions, that is, globally, the quality of the reconstruction is the highest for the regular distribution of sightlines ($\mathcal{R}_{I,U}$), it degrades for their random distribution ($\mathcal{R}_{I,r}$) and is further reduced when the realistic noise on the spectra is added ($\mathcal{R}_{WQ}$). Such a trend is best captured by the ratios of $r_{ex, ij}$ (Table D2). The ratios of $r_{max, ij}$ and $h_{max, ij}$ (Table D3) do not allow us to well discriminate between the reconstructions and typically do not capture all the details contained in two-point statistics.

Note that even though our choice of summary statistics is physically motivated, it is not ideal to highlight the differences between the reconstruction at the (relatively high) level of noise in the WEAVE-QSO mocks and the reference Ly-α flux, with respect to to the noise-only field, and therefore to constrain cosmology. In particular, the slopes and widths of the maximum auto- and cross-correlations might be more powerful for discriminating between different underlying power spectra. Beyond summary statistics derived from single auto-

and cross-correlations, one could also consider measuring the ratio between complementary auto and cross-correlations, e.g. $\mathcal{FF}/\mathcal{WW}$, $\mathcal{PV}/\mathcal{FW}$, etc., as a way to get rid of some systematics inherited from the noise introduced in the reconstruction.

4.2 How to improve the error budgets in future surveys?

4.2.1 Decreasing rarity

Part of the discrepancy between the original and reconstructed field comes from the non-regular sampling of the density field, which, in turn, depends on the spatial distribution of background sources. When sightlines are randomly distributed, in some regions of the reconstructed volume they will be less clustered than the correlation length set at the reconstruction stage (taken here as the mean inter-sightline distance), which will inevitably degrade the count and correlation of critical points. This can be seen for example on Fig. 11, where the discrepancy is the largest between the random distribution of sightlines (dotted lines) and the regular distribution of sightlines (dashed lines). Adding extra-noise on the spectra (dash–dotted line) has on overall a smaller impact. To minimize the consequence of this random sampling (without changing the survey design), one could increase the correlation length in the reconstruction and carry out the study at a larger scale. Obviously, when considering a fixed area on the sky, increasing the smoothing scale would also increase the statistical uncertainties since the number of volume elements would decrease. To mitigate this effect, we could choose to include in the statistics less rare critical points (e.g. taking all the 20 per cent rarest rather than only the 10 per cent rarest). Appendix E shows indeed that the signal is more significant while decreasing rarity.

4.2.2 Combining galaxies and quasars background sources

Another straightforward but costly way to mitigate the impact of shot noise due to the random sightline distribution could be to increase the number of sightlines. In the HIGHDENS footprint, we assumed that all quasars brighter than 23.5 in the r band will be observed. Given that the galaxy number counts dominate over the quasar number counts for $r > 22.5$, one could complement the survey with bright star-forming galaxies in order to efficiently increase the number of sightlines in the reconstruction. This is the strategy adopted in other surveys (e.g. in PFS: Takada et al. 2014), but at the price of observing a smaller area.

4.2.3 Increasing the volume of the survey

Recall that in this study we explored a configuration compatible with the HIGHDENS footprint in terms of sightline density, but, as mentioned above, our simulation set covers ∼6.8 times the expected volume in this footprint. As a consequence, one should expect the size of the errorbars to be about $\sqrt{6.8} \simeq 2.6$ times larger when restricting ourselves to a volume comparable to the HIGHDENS footprint.

On the other hand, to improve the overall significance of the detection, one could consider carrying out the analysis on a larger volume by exploiting the WIDE footprint, which will cover a much larger area (6000 deg²). However, the limited redshift window ($2.5 < z < 3$) will be higher than the one studied here. For the same magnitude limit in the background quasar distribution, this will translate into a lower sightline density. More precisely, in the WIDE footprint, the reconstruction could be performed at a scale $L_\mathrm{T} \simeq 19 \, \mathrm{Mpc} \, h^{-1}$ over

13.6 (Gpc h)$^{-3}$, resulting into ∼2.0 × 10⁶ volume elements, while in the HIGDENS footprint, the reconstruction could be performed at a scale of $L_\mathrm{T} \simeq 16 \, \mathrm{Mpc} \, h^{-1}$ over 0.7 (Gpc h)$^{-3}$, resulting into ∼1.8 × 10⁵ volume elements. Therefore, the size of the errorbars is expected to be divided by $\sqrt{11.1} \simeq 3.3$ when using the WIDE instead of the HIGDENS footprint. In other words, the errorbars in the WIDE footprint are expected to be $3.3/2.6 \simeq 1.3$ times smaller than those displayed in this paper, turning a 4σ detection into a 5σ one.

It remains to be seen though whether the scientific gain (from the point of view of constraining cosmology through the statistics of critical points) is higher in the HIGDENS footprint (high density of quasars, small volume, lower redshift) or in the WIDE one (lower density of quasars, much larger volume, higher redshift). This question can be fully addressed only after having explored which scale/rarity is the most effective for cosmology (Shim et al. in preparation).

4.3 Prospects

From a cosmological perspective, given that the radii of exclusion zone and maximum correlation negligibly evolve with time (Shim et al. 2021), one can make use of these particular scales as standard rulers to measure the expansion of the Universe because they are analytically predictable from first principles and nearly redshift-independent. Because the correspondence between redshift and distance of an object depends on the underlying cosmology, the characteristic clustering scales of critical points will remain constant and match the theoretical prediction only when the correct cosmological parameters are adopted. Requesting such a match yields an estimator for the corresponding parameters. It then becomes crucial to assess the ability to recover these characteristic clustering scales from observations. Our forecasts (Figs 9, 11, and 13 for 10 per cent rarity, and the corresponding Figs E1, E2, and E3 for 20 per cent rarity) show that the cross-correlations $\mathcal{PF}$ and $\mathcal{WV}$ are constrained at the 2σ level, the autocorrelations $\mathcal{FF}$, $\mathcal{WW}$ at up to ∼2σ level, the cross-correlations $\mathcal{PW}$, $\mathcal{PV}$, $\mathcal{FV}$ up to 3σ, 1σ, 3σ level, respectively, and $\mathcal{FW}$ up to the ∼4σ level, which represents the strongest significance for the WEAVE-QSO-like configuration. Thus, to provide tighter constraints on cosmological parameters it is more advantageous to use the characteristic features of the two-point functions involving filaments and/or walls (e.g. the exclusion zone in the $\mathcal{FW}$, $\mathcal{PW}$, $\mathcal{FV}$ cross-correlations, or the radius at a local maximum in the $\mathcal{FF}$, $\mathcal{WW}$ autocorrelations and in the $\mathcal{PF}$, $\mathcal{WV}$ cross-correlations) as standard rulers because they can be measured with a higher significance. Using these features as cosmic rulers complements current approaches that rely on the BAO scale to measure cosmological parameters. Interestingly, these scales associated with the two-point correlation functions are smaller than BAO's and thus probe different part of the power spectrum, with more modes available within a given survey geometry. Shim et al. (in preparation) investigate the cosmology dependence of the clustering of critical points, exploring alternative cosmology models. Eventually, connecting their results and ours will allow us to make efficient cosmic forecasts from Ly-α tomography, relying on both one (Gay et al. 2010; Codis et al. 2013) and two-point (Shim et al. 2021) statistics predictions.

While the present paper was focused on the technical specification of WEAVE-QSO, other upcoming tomographic surveys such as PFS or DESI could help to improve error budgets. It would be of interest to calibrate the best compromise one should make in terms of surface area, depth, tracers (Lyman-break galaxies versus

QSOs) and expected SNR. It would clearly be an asset to complement spectroscopic surveys with photometric redshift ones, as they could be integrated into the reconstruction (even though non trivially, see Horowitz et al. 2021b, for a recent effort to use multiple tracers to reconstruct density field). It would of course also be of interest to quantify the clustering of critical points on intensity maps in two dimensions, and our ability to extract such points from the corresponding surveys.

Moving beyond critical points this could be further completed by investigating the cosmic evolution of critical lines, e.g. connecting saddles together (Pogosyan et al. 2009), through the statistics of their (differential) length, as has been attempted for galactic catalogues (Sousbie et al. 2008) and in ELT mocks (Japelj et al. 2019).

5 CONCLUSIONS

Mocks were used to asses our ability to recover the connectivity and clustering properties of critical points of the reconstructed large-scale structure from Ly-α tomography in the context of a realistic quasar survey configuration (WEAVE-QSO). The mocks were produced with the LyMAS (Peirani et al. 2014, 2022).

Our main findings are the following:

(i) *General.* As expected, the quality of reconstruction decreases with randomness in the distribution of lines of sight and with the inclusion of noise on the spectra. Conversely, the measured signal increases with decreasing rarity of the critical points and with increasing smoothing scale, but at the expense of less marked features.

(ii) *Critical points' number counts.* The total number of the critical points is larger in the reconstructed field compared to the original (reference) field by about 15 per cent for $\mathcal{R}_{I,U}$, 19 per cent for $\mathcal{R}_{I,r}$, and 35 per cent for $\mathcal{R}_{WQ}$. This fraction is slightly higher for peaks than for voids. However, as expected, reconstructed filaments and walls are about three times more abundant than peaks and voids, while the ratio between the number of peaks and walls over filaments and voids is close to 1 for all the reconstructions.

(iii) *Autocorrelations of critical points.* The reconstruction captures the main expected features of the autocorrelation functions: exclusion zones at small separations, maxima at ≈ 2–$3L_T$ and convergence towards zero at large separations, in particular for saddles (even for high rarity).

(iv) *Cross-correlations of overdense and underdense critical points.* The large exclusion zone at small separations and monotonic increase towards zero at large separations are well recovered. The amplitude of these cross-correlations is however systematically higher compared to the original Ly-α field for all explored reconstruction configurations.

(v) *Cross-correlations of the same overdensity sign critical points.* Again, the reconstruction recovers the correlation's main features: divergence at zero separation, lack of negative correlations, and exclusion zone at small scales, presence of a local maximum at similar separation, when compared to the autocorrelations.

(vi) *Resilience of saddles.* Fortunately, the (cross-) correlations involving the least non-linear critical points (walls, filaments), which display the least amount of variation with redshift are also those that are best reconstructed from WEAVE-like Ly-α tomography. This validates a posteriori using the clustering of saddle points as a novel cosmic probe. The significance of autocorrelations reaches 2σ (1σ) for walls (filaments). It is up to 4σ for the cross-correlations of filaments and walls of 20 per cent abundance and up to 2σ for the cross-correlations peak-filament and wall-void.

(vii) *Connectivity.* The topology of the recovered field, as traced by its connectivity, is in good agreement with the initial one both statistically and in the vicinity of given peaks. This is consistent with the persistence of the clustering properties of critical points with respect to tomographic reconstruction.

Our conclusions highlight that the main features of the two-point correlation functions of critical points can be recovered with a good degree of confidence in a WEAVE-QSO-like tomographic surveys (Takada et al. 2014; DESI Collaboration 2016; Pieri et al. 2016). As they show little evolution with redshift (Shim et al. 2021), their clustering should provide useful complementary estimators for dark energy experiments.

DATA AVAILABILITY

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

REFERENCES

Ahumada R. et al., 2020, ApJS, 249, 3
Appleby S., Park C., Hong S. E., Hwang H. S., Kim J., Tonegawa M., 2021, ApJ, 907, 75
Bagla J. S., Khandai N., Datta K. K., 2010, MNRAS, 407, 567
Baldauf T., Codis S., Desjacques V., Pichon C., 2021, Phys. Rev. D, 103, 083530
Bamba K., Capozziello S., Nojiri S., Odintsov S. D., 2012, Ap&SS, 342, 155
Bautista J. E. et al., 2017, A&A, 603, A12
Benitez N. et al., 2014, preprint (arXiv:1403.5237)
Bernardeau F., Colombi S., Gaztañaga E., Scoccimarro R., 2002, Phys. Rep., 367, 1
Bond J. R., Kofman L., Pogosyan D., 1996, Nature, 380, 603
Cadiou C., Pichon C., Codis S., Musso M., Pogosyan D., Dubois Y., Cardoso J. F., Prunet S., 2020, MNRAS, 496, 4787

Caucci S., Colombi S., Pichon C., Rollinde E., Petitjean P., Sousbie T., 2008, MNRAS, 386, 211

Cisewski J., Croft R. A. C., Freeman P. E., Genovese C. R., Khandai N., Ozbek M., Wasserman L., 2014, MNRAS, 440, 2599

Codis S., Pichon C., Pogosyan D., Bernardeau F., Matsubara T., 2013, MNRAS, 435, 531

Codis S., Gavazzi R., Pichon C., Gouin C., 2017, A&A, 605, A80

Codis S., Pogosyan D., Pichon C., 2018, MNRAS, 479, 973

D'Odorico V. et al., 2006, MNRAS, 372, 1333

Dalton G. et al., 2012, in McLean I. S., Ramsay S. K., Takami H., eds, Proc. SPIE Conf. Ser. Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV. SPIE, Bellingham, p. 84460P

Darragh Ford E. et al., 2019, MNRAS, 489, 5695

Davis M., Peebles P. J. E., 1983, ApJ, 267, 465

DESI Collaboration, 2016, preprint (arXiv:1611.00036)

Desjacques V., Jeong D., Schmidt F., 2018, Phys. Rep., 733, 1

Dubois Y. et al., 2014, MNRAS, 444, 1453

Gallerani S., Kitaura F. S., Ferrara A., 2011, MNRAS, 413, L6

Gay C., Pichon C., Le Borgne D., Teyssier R., Sousbie T., Devriendt J., 2010, MNRAS, 404, 1801

Gay C., Pichon C., Pogosyan D., 2012, Phys. Rev. D, 85, 023011

Gebhardt K. et al., 2021, ApJ, 923, 217

Gouin C., Gavazzi R., Codis S., Pichon C., Peirani S., Dubois Y., 2017, A&A, 605, A27

Hahn O., Porciani C., Carollo C. M., Dekel A., 2007, MNRAS, 375, 489

Hamilton A. J. S., Gott J., Richard I., Weinberg D., 1986, ApJ, 309, 1

Horowitz B., Lee K.-G., White M., Krolewski A., Ata M., 2019, ApJ, 887, 61

Horowitz B. et al., 2021a, preprint (arXiv:2109.09660)

Horowitz B., Zhang B., Lee K.-G., Kooistra R., 2021b, ApJ, 906, 110

Japelj J. et al., 2019, A&A, 632, A94

Kitaura F.-S., Gallerani S., Ferrara A., 2012, MNRAS, 420, 61

Komatsu E. et al., 2011, ApJS, 192, 18

Kraljic K. et al., 2020, MNRAS, 491, 4294

Krolewski A., Lee K.-G., Lukić Z., White M., 2017, ApJ, 837, 31

Krolewski A. et al., 2018, ApJ, 861, 60

Lazkoz R., Nesseris S., Perivolaropoulos L., 2008, J. Cosmol. Astropart. Phys., 2008, 012

Lee K.-G., White M., 2016, ApJ, 831, 181

Lee K.-G. et al., 2018, ApJS, 237, 31

Li Z., Horowitz B., Cai Z., 2021, ApJ, 916, 20

Lokken M. et al., 2021, preprint (arXiv:2107.05523)

Milnor J., 1963, Morse Theory. Princeton Univ. Press, Princeton, NJ

Morrison S. S., 2019, PhD Thesis

Newman A. B. et al., 2020, ApJ, 891, 147

Ozbek M., Croft R. A. C., Khandai N., 2016, MNRAS, 456, 3610

Palanque-Delabrouille N. et al., 2016a, A&A, 587, A41

Palanque-Delabrouille N. et al., 2016b, A&A, 589, C2

Peirani S., Weinberg D. H., Colombi S., Blaizot J., Dubois Y., Pichon C., 2014, ApJ, 784, 11

Peirani S. et al., 2022, MNRAS, in press

Petitjean P., Mueket J. P., Kates R. E., 1995, A&A, 295, L9

Pichon C., Vergely J. L., Rollinde E., Colombi S., Petitjean P., 2001, MNRAS, 326, 597

Pichon C., Gay C., Pogosyan D., Prunet S., Sousbie T., Colombi S., Slyz A., Devriendt J., 2010, in Alimi J.-M., Fuözfa A., eds, AIP Conf. Ser. Vol. 1241, Invisible Universe. Am. Inst. Phys., New York, p. 1108

Pieri M. M. et al., 2016, in Reylé C., Richard J., Cambrésy L., Deleuil M., Pécontal E., Tresse L., Vauglin I., eds, SF2A-2016: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics, p. 259

Pogosyan D., Pichon C., Gay C., Prunet S., Cardoso J. F., Sousbie T., Colombi S., 2009, MNRAS, 396, 635

Porqueres N., Hahn O., Jasche J., Lavaux G., 2020, A&A, 642, A139

Rauch M., 1998, ARA&A, 36, 267

Ravoux C. et al., 2020, J. Cosmol. Astropart. Phys., 2020, 010

Regos E., Szalay A. S., 1995, MNRAS, 272, 447

Shen Y. et al., 2009, ApJ, 697, 1656

Shim J., Codis S., Pichon C., Pogosyan D., Cadiou C., 2021, MNRAS, 502, 3885

Sousbie T., Pichon C., Courtois H., Colombi S., Novikov D., 2008, ApJ, 672, L1

Sousbie T., Pichon C., Kawahara H., 2011, MNRAS, 414, 384

Springel V., 2005, MNRAS, 364, 1105

Stark C. W., White M., Lee K.-G., Hennawi J. F., 2015a, MNRAS, 453, 311

Stark C. W., Font-Ribera A., White M., Lee K.-G., 2015b, MNRAS, 453, 4311

Takada M. et al., 2014, PASJ, 66, R1

APPENDIX A: QUALITY VERSUS SNR AND L_T

Fig. A1 quantifies the pixel-to-pixel correlation between the 3D maps of the original and the reconstructed flux contrast in units of the rms fluctuations, for the different configuration studied in this work (as presented in Section 2.2.3). On each panel is displayed the Pearson correlation coefficient, that we can use as a single metric to assess the overall agreement between both fields.

Let us now use this metric to quantify the degradation of the correlation between the original and the reconstructed fields. We performed the reconstruction with a similar density of sightlines but different levels of Gaussian white noise added on the Ly-α forest of each sightline prior to the reconstruction. The corresponding Pearson
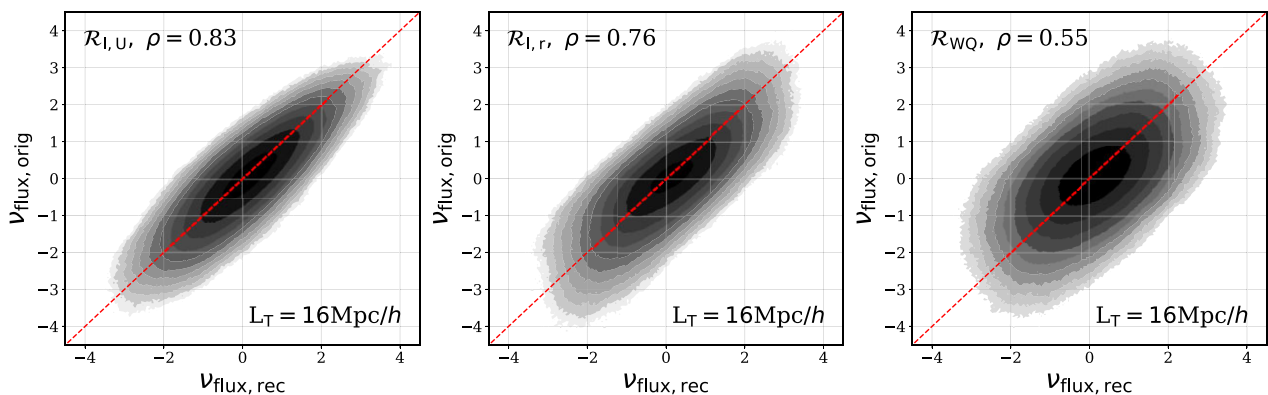
Figure A1. The original flux density contrast (in units of rms fluctuations of the field) plotted against the reconstructed flux density contrast for the different noise configurations envisaged in this study, i.e. regular distribution of sightlines ($\mathcal{R}_{I,U}$), random distribution of sightlines ($\mathcal{R}_{I,r}$), and fiducial configuration ($\mathcal{R}_{WQ}$). Also indicated is the Pearson correlation coefficient ρ between the two fields.
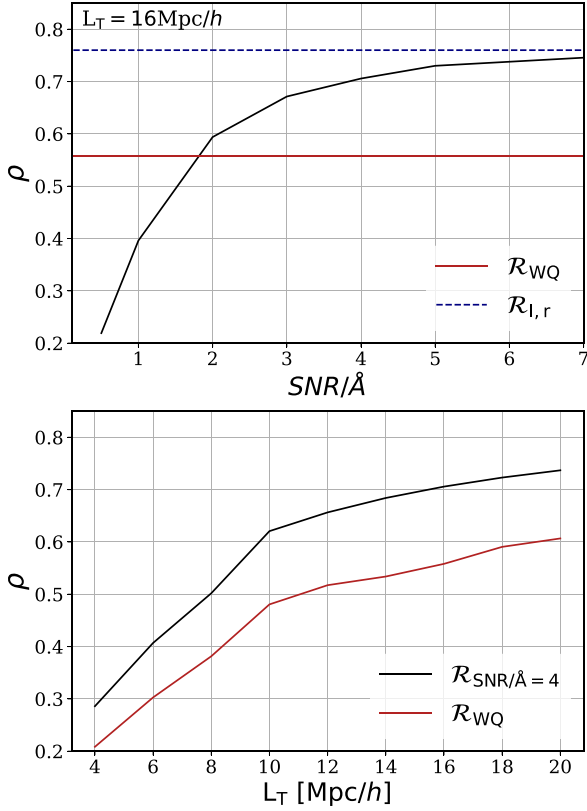
Figure A2. Top panel: Pearson correlation coefficient between the Ly-α reference field and the reconstructed fields (with $L_T = 16\,\mathrm{Mpc}\,h^{-1}$) for different realizations of the SNR on the sightlines. Also indicated as straight lines are the correlation coefficients for $\mathcal{R}_{WQ}$ (red) and $\mathcal{R}_{I,r}$ (dashed blue). Bottom panel: correlation coefficient between the original and the reconstructed fields when different correlation scales L_T have been adopted when performing the reconstruction, adopting a constant SNR/$\mathring{A} = 4$ (black) or a realistic SNR distribution (red).

correlation coefficient is displayed on the top panel of Fig. A2. On overall, increasing the SNR/$\mathring{A}$ brings the reconstructed and original fields in better agreement. However, one also note that the correlation coefficient reaches a plateau at SNR/$\mathring{A} > 4$. This suggests that, at high SNR/$\mathring{A}$ (on spectra), the noise budget starts to be dominated by shot noise due to the finite sampling and clustering of sightlines.

On the bottom panel of Fig. A2, the impact of choosing different correlation lengths L_T when performing the reconstruction (see Section 3) is quantified. The reconstruction are performed with the same density of sightlines (as for $\mathcal{R}_{WQ}$) and with a realistic SNR distribution (similar to $\mathcal{R}_{WQ}$) on the one hand, and with a constant SNR/$\mathring{A} = 4$ on the other hand. As a reminder, $L_T = 16\,\mathrm{Mpc}\,h^{-1}$ was chosen as it was reflecting the mean separation between sightlines. Unsurprisingly, one notes that the agreement improves when a larger

L_T is used in the reconstruction (several sightlines will contribute to the same volume element in the reconstructed map, which will enhance the signal over the noise). Interestingly, the agreement between the original and reconstructed fields is relatively well preserved even when decreasing L_T below its fiducial value, down to $\sim 10\,\mathrm{Mpc}\,h^{-1}$, from which it dropped brutally. One should note, however, that, when decreasing L_T, the quality of the reconstruction will be spatially less and less homogeneous, due to the non-regular distribution of the sightlines. The Pearson correlation coefficient, which is a global metric, is not very sensitive to this degradation, but this effect might dramatically impacts the count and correlation of the critical points.

APPENDIX B: PREDICTIONS IN THE LINEAR REGIME

Throughout the main text, we have compared measurements for the critical points counts and cross-correlations with Gaussian random fields predictions. To compute those predictions, we rely on the formalism extensively described in Shim et al. (2021) (we refer the reader to Appendix A of this paper for more details) with a few modifications. We first compute the covariance matrix of a Gaussian random field and its first and second derivatives at two spatial positions separated by a distance r and characterized by a power spectrum given by the linear power spectrum used in the simulation smoothed with a Gaussian kernel of length $L_T = 16$ Mpc h^{-1}. From this 20×20 covariance matrix, we compute the corresponding joint probability distribution function (PDF) of the field and its first and second derivatives and then evaluate through an MCMC integration scheme the probability of finding two critical points with specified signatures separated by r in order to get their cross-correlation functions. Each point of the function is evaluated by generating 10 millions random numbers satisfying the joint PDF with a zero gradient constraint and is kept only if the critical points conditions are fulfilled (signatures and density thresholds). The result is displayed with dash–dotted black lines on the figures of the main text.

APPENDIX C: CRITICAL POINTS VISUALIZATION

Fig. C1 shows for illustration purposes only a qualitative comparison of a randomly selected region from the full volume in the reference Ly-α field (left-hand panel), and the three reconstructed fields, $\mathcal{R}_{I,U}$ (middle left-hand panel), $\mathcal{R}_{I,r}$ (middle right-hand panel), and $\mathcal{R}_{WQ}$ (right-hand panel) at smoothing scale $L_T = 16$ Mpc h^{-1}. For each reconstruction, as more structures are created, the number of critical points is correspondingly higher compared to the Ly-α reference field. However, in spite of the projection effects (the projection is performed on 32 Mpc h^{-1} thick slice), several critical points are robustly identified.
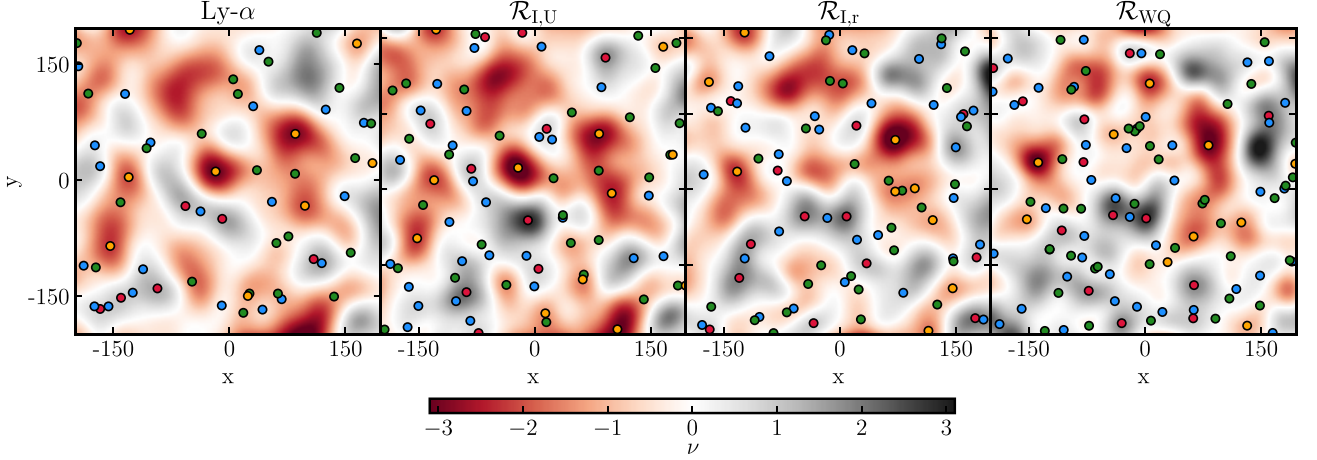
Figure C1. A visualization of a 32 Mpc h^{-1} thick and 400 Mpc h^{-1} wide (in both directions) slices (randomly selected from the full box) of rarity with overplotted critical points colour-coded by their type (peaks in red, filaments in blue, walls in green and voids in orange) for the reference Ly-α field (left-hand panel), and the three reconstructed fields, $\mathcal{R}_{I,U}$ (middle left-hand panel), $\mathcal{R}_{I,r}$ (middle right-hand panel), and $\mathcal{R}_{WQ}$ (right-hand panel) at $L_T = 16$ Mpc h^{-1}.

APPENDIX D: SUMMARY STATISTICS

This Appendix provides a summary of measured positions and heights of the maxima of autocorrelations and cross-correlations $\mathcal{PF}$, $\mathcal{WV}$, sizes of the exclusion zones of cross-correlations $\mathcal{PW}$, $\mathcal{PV}$, $\mathcal{FW}$, $\mathcal{FV}$ and summary statistics presented in the main text.

Table D1 summarizes the position ($r_{\max, ij}$) and height ($h_{\max, ij}$) of the maximum of autocorrelations and cross-correlations $\mathcal{PF}$, $\mathcal{WV}$ for all the fields used in this work at the smoothing scale $L_T = 16$ Mpc h^{-1}.

Table D2 shows the size of the exclusion zone ($r_{ex, ij}$) of cross-correlations $\mathcal{PW}$, $\mathcal{PV}$, $\mathcal{FW}$, $\mathcal{FV}$ for all the fields used in this work at the smoothing scale $L_T = 16$ Mpc h^{-1}.

Tables D3 and D4 report the ratios of $r_{\max, ij}$, $h_{\max, ij}$, and $r_{ex, ij}$ with respect to the Ly-α reference field for the reconstructed and noise-only fields at the smoothing scale $L_T = 16$ Mpc h^{-1}.

Table D2. Size of the exclusion zone $r_{ex, ij}$ (in [Mpc h^{-1}]) of cross-correlations $\mathcal{PW}$, $\mathcal{PV}$, $\mathcal{FW}$, $\mathcal{FV}$ for 10 PER CENT rarity, for all the fields used in this work at the smoothing scale $L_T = 16$ Mpc h^{-1}.

	$\mathcal{PW}$	$\mathcal{PV}$	$\mathcal{FW}$	$\mathcal{FV}$
H I	44.23 ± 0.82	47.24 ± 1.26	38.22 ± 1.26	43.48 ± 0.76
DM	45.74 ± 0.67	49.49 ± 0.67	42.73 ± 1.06	44.98 ± 1.34
Ly-α	44.95 ± 1.52	49.51 ± 1.27	38.09 ± 0.68	43.42 ± 0.83
$\mathcal{R}_{I,U}$	40.38 ± 0.68	46.47 ± 0.83	36.57 ± 0.68	41.14 ± 1.08
$\mathcal{R}_{I,r}$	41.14 ± 1.08	41.90 ± 1.27	36.57 ± 0.68	41.14 ± 1.08
$\mathcal{R}_{WQ}$	36.58 ± 0.68	43.42 ± 1.74	32.77 ± 0.68	38.86 ± 0.83
$\mathcal{R}_{noise}$	36.19 ± 0.55	40.00 ± 0.55	32.77 ± 0.48	36.58 ± 0.48

Note. The errors are the standard deviations of the mean across all mocks.

Table D1. Position ($r_{\max, ij}$ [Mpc h^{-1}]) and height ($h_{\max, ij}$) of the maximum of autocorrelations ($\mathcal{PP}$, $\mathcal{FF}$, $\mathcal{WW}$, $\mathcal{VV}$) and cross-correlations $\mathcal{PF}$, $\mathcal{WV}$ for 10 per cent rarity, for all the fields used in this work at the smoothing scale $L_T = 16$ Mpc h^{-1}.

		$\mathcal{PP}$	$\mathcal{FF}$	$\mathcal{WW}$	$\mathcal{VV}$	$\mathcal{PF}$	$\mathcal{WV}$
$r_{\max, ij}$	H I	30.82 ± 3.45	25.13 ± 0.67	28.92 ± 1.44	36.50 ± 3.93	33.78 ± 1.72	29.64 ± 1.99
	DM	37.92 ± 2.07	30.34 ± 1.82	27.97 ± 1.82	35.55 ± 3.18	31.48 ± 1.72	33.78 ± 1.46
	Ly-α	39.77 ± 7.42	27.33 ± 1.35	24.31 ± 1.35	32.64 ± 1.6	33.78 ± 1.72	35.63 ± 1.51
	$\mathcal{R}_{I,U}$	35.66 ± 3.14	29.61 ± 2.14	30.36 ± 3.10	34.91 ± 4.86	32.86 ± 1.05	31.94 ± 0.41
	$\mathcal{R}_{I,r}$	40.96 ± 3.03	25.82 ± 1.07	28.09 ± 1.73	51.13 ± 3.96	29.64 ± 1.20	31.48 ± 1.13
	$\mathcal{R}_{WQ}$	51.83 ± 7.09	27.34 ± 0.83	26.58 ± 1.66	43.24 ± 3.65	31.48 ± 0.65	34.24 ± 0.41
	$\mathcal{R}_{noise}$	37.38 ± 6.44	26.20 ± 0.36	29.61 ± 1.51	40.37 ± 2.88	31.09 ± 0.63	29.87 ± 0.47
$h_{\max, ij}$	H I	3.85 ± 0.45	5.66 ± 0.43	5.76 ± 0.42	5.15 ± 0.44	7.77 ± 0.67	7.38 ± 0.31
	DM	3.64 ± 0.49	5.11 ± 0.45	6.34 ± 0.19	4.83 ± 0.64	8.47 ± 0.53	7.67 ± 0.25
	Ly-α	3.35 ± 0.38	4.90 ± 0.12	4.76 ± 0.09	3.34 ± 0.19	7.39 ± 0.61	6.68 ± 0.31
	$\mathcal{R}_{I,U}$	2.86 ± 0.31	4.64 ± 0.30	4.31 ± 0.27	2.59 ± 0.17	7.03 ± 0.38	6.82 ± 0.22
	$\mathcal{R}_{I,r}$	2.12 ± 0.21	4.49 ± 0.15	4.40 ± 0.09	2.39 ± 0.21	7.29 ± 0.19	7.16 ± 0.36
	$\mathcal{R}_{WQ}$	1.75 ± 0.07	4.34 ± 0.13	4.53 ± 0.23	2.54 ± 0.23	7.22 ± 0.12	6.64 ± 0.49
	$\mathcal{R}_{noise}$	2.09 ± 0.22	4.17 ± 0.12	3.88 ± 0.13	1.96 ± 0.10	6.43 ± 0.18	6.27 ± 0.19

Note. The errors are the standard deviations of the mean across all mocks.

Table D3. Summary statistics.[1]

		$\mathcal{PP}$	$\mathcal{FF}$	$\mathcal{WW}$	$\mathcal{VV}$	$\mathcal{PF}$	$\mathcal{WV}$
$r_{\max,\,ij}$	$\mathcal{R}_{I,U}$	0.89 ± 0.19	1.08 ± 0.09	1.25 ± 0.13	1.07 ± 0.15	0.97 ± 0.06	0.89 ± 0.04
	$\mathcal{R}_{I,r}$	1.03 ± 0.20	0.95 ± 0.06	1.15 ± 0.09	1.57 ± 0.11	0.88 ± 0.06	0.88 ± 0.05
	$\mathcal{R}_{\mathrm{WQ}}$	1.30 ± 0.26	1.0 ± 0.06	1.09 ± 0.09	1.21 ± 0.11	0.93 ± 0.05	0.96 ± 0.04
	$\mathcal{R}_{\mathrm{noise}}$	0.94 ± 0.25	0.96 ± 0.05	1.22 ± 0.08	1.24 ± 0.10	0.92 ± 0.05	0.84 ± 0.04
$h_{\max,\,ij}$	$\mathcal{R}_{I,U}$	0.85 ± 0.14	0.95 ± 0.07	0.91 ± 0.06	0.78 ± 0.08	0.95 ± 0.10	1.02 ± 0.06
	$\mathcal{R}_{I,r}$	0.63 ± 0.12	0.92 ± 0.04	0.92 ± 0.03	0.72 ± 0.09	0.99 ± 0.09	1.07 ± 0.07
	$\mathcal{R}_{\mathrm{WQ}}$	0.52 ± 0.09	0.89 ± 0.04	0.95 ± 0.05	0.76 ± 0.09	0.98 ± 0.08	0.99 ± 0.09
	$\mathcal{R}_{\mathrm{noise}}$	0.62 ± 0.12	0.85 ± 0.03	0.82 ± 0.03	0.59 ± 0.06	0.87 ± 0.08	0.94 ± 0.05

[1]*Note*. Ratios of $r_{\max,\,ij}$ and $h_{\max,\,ij}$ with respect to the Ly-α reference field for the reconstructed and noise-only fields at the smoothing scale $L_T = 16\ \mathrm{Mpc}\,h^{-1}$ and for 10 PER CENT rarity.

Table D4. Summary statistics.

	$\mathcal{PW}$	$\mathcal{PV}$	$\mathcal{FW}$	$\mathcal{FV}$
$\mathcal{R}_{I,U}$	0.89 ± 0.04	0.94 ± 0.03	0.96 ± 0.03	0.95 ± 0.03
$\mathcal{R}_{I,r}$	0.92 ± 0.04	0.85 ± 0.04	0.96 ± 0.03	0.95 ± 0.03
$\mathcal{R}_{\mathrm{WQ}}$	0.81 ± 0.03	0.88 ± 0.04	0.86 ± 0.03	0.89 ± 0.03
$\mathcal{R}_{\mathrm{noise}}$	0.81 ± 0.03	0.81 ± 0.03	0.86 ± 0.02	0.84 ± 0.02

Note. Ratios of $r_{\mathrm{ex},\,ij}$ with respect to the Ly-α reference field for the reconstructed and noise-only fields at the smoothing scale $L_T = 16\ \mathrm{Mpc}\,h^{-1}$ and for 10 PER CENT rarity.

APPENDIX E: IMPACT OF RARITY AND SMOOTHING

Let us explore the impact of rarity and smoothing on the two-point correlation functions.

Starting with rarity, we complement the 10 per cent rarity results presented in the main text with rarities of 5 and 20 per cent. Fig. E1 shows the differences of the autocorrelations of the Ly-α reference field and the three reconstructions, $\mathcal{R}_{I,U}$, $\mathcal{R}_{I,r}$, and $\mathcal{R}_{\mathrm{WQ}}$, with respect to the noise-only field $\mathcal{R}_{\mathrm{noise}}$ for the critical points with abundance of 5 (left-hand panels) and 20 per cent (right-hand panels) at $L_T = 16\ \mathrm{Mpc}\,h^{-1}$. As was the case for 10 per cent rarity (see Fig. 9), the most striking features in the autocorrelation functions are measured for filaments and walls. The increased significance with decreased rarity is notable for the $\mathcal{WW}$ autocorrelations, where the significance increases from about 2.5σ (1σ) for $\mathcal{R}_{I,U}$ ($\mathcal{R}_{\mathrm{WQ}}$) at 5 per cent abundance to up to 5σ (2σ) at 20 per cent abundance. For the $\mathcal{FF}$ autocorrelations, this significance is mush less striking. The significance of the outstanding features in the autocorrelations of the critical points at 10 per cent abundance is comparable to that of 20 per cent abundance.

Fig. E2 compares the differences of the cross-correlations of Ly-α reference field and the three reconstructions, $\mathcal{R}_{I,U}$, $\mathcal{R}_{I,r}$, and $\mathcal{R}_{\mathrm{WQ}}$, with respect to the noise-only field $\mathcal{R}_{\mathrm{noise}}$ for the overdense and underdense critical points with abundance of 5 per cent (left-hand panels) and those of 20 per cent (right-hand panels) at $L_T = 16\ \mathrm{Mpc}\,h^{-1}$. In contrast to the overall mild increase of the significance of the outstanding features contained in the autocorrelations, their significance increases strikingly for all cross-correlations of under and overdense critical points, i.e. $\mathcal{PW}$, $\mathcal{PV}$, $\mathcal{FW}$, and $\mathcal{FV}$.

While for $\mathcal{PW}$ and $\mathcal{PV}$ the significance increases by about the factor of 2 between 5 and 20 per cent abundance (up to 8σ at 20 per cent for $\mathcal{R}_{I,U}$), for $\mathcal{FW}$, this factor is even higher (between factor of 2.5 for $\mathcal{R}_{I,U}$ and factor of 4 for $\mathcal{R}_{\mathrm{WQ}}$). The most striking increase of the significance of outstanding features with decreasing rarity is seen for $\mathcal{FV}$, where for $\mathcal{R}_{I,U}$, the significance increases from 3σ to 9σ, and for $\mathcal{R}_{\mathrm{WQ}}$, it is from 0σ to up to 3σ.

Similarly, but to a lesser extent, the significance increases with decreasing rarity for the cross-correlations of the critical points with the same overdensity sign, i.e. $\mathcal{PF}$ and $\mathcal{WV}$, as was shown in Fig. E3.

Let us now examine the impact of smoothing on the two-point correlation functions. We will focus on the comparison between the H I and DM density fields. Fig. E4 shows the autocorrelations of the 10 per cent rarest critical points in the original (H I density; colour solid lines) and DM density (colour dashed lines) fields, as in Fig. 8, but smoothed at the scale $L_T = 12\ \mathrm{Mpc}\,h^{-1}$. The autocorrelations of all critical points follow qualitatively similar trends, as in the case of 16 Mpc h^{-1} smoothing scale, with a good agreement between the H I and DM density fields. As expected, the maximum of the autocorrelations is again reached at $\approx 2L_T$ for filaments and walls and ≈ 2.5–$3L_T$ for peaks and voids (see Table E2). The positions and heights of the maxima are much better constrained at $L_T = 12\ \mathrm{Mpc}\,h^{-1}$ smoothing, compared to 16 Mpc h^{-1}.

Fig. E5 shows the cross-correlation function of under and over-dense critical points with 10 per cent abundance for the H I and DM density fields (coloured solid and dashed lines, respectively), following Fig. 10, but smoothed at the scale $L_T = 12\ \mathrm{Mpc}\,h^{-1}$. As for the autocorrelations, there is a better agreement between the H I and DM density fields compared to the smoothing $L_T = 16\ \mathrm{Mpc}\,h^{-1}$. The size of the exclusion zone is identical for the H I and DM density fields and it is smaller compared to the fields smoothed at $L_T = 16\ \mathrm{Mpc}\,h^{-1}$ (see Table E1).

Fig. E6 shows the cross-correlation function of the critical points of the same overdensity sign for H I and DM density fields (coloured solid and dashed lines, respectively) and 10 per cent abundance, following Fig. 12, but at the smoothing scale of 12 Mpc h^{-1}. The agreement between the two fields with decreased smoothing scale is again confirmed. For both H I and DM density fields, the position of maxima $r_{\max,\,i}$ is ≈ 26.9 Mpc h^{-1} ($\approx 2L_T$). The height of the maxima $h_{\max,\,i}$ is also in a good agreement between the two fields (see Table E2).
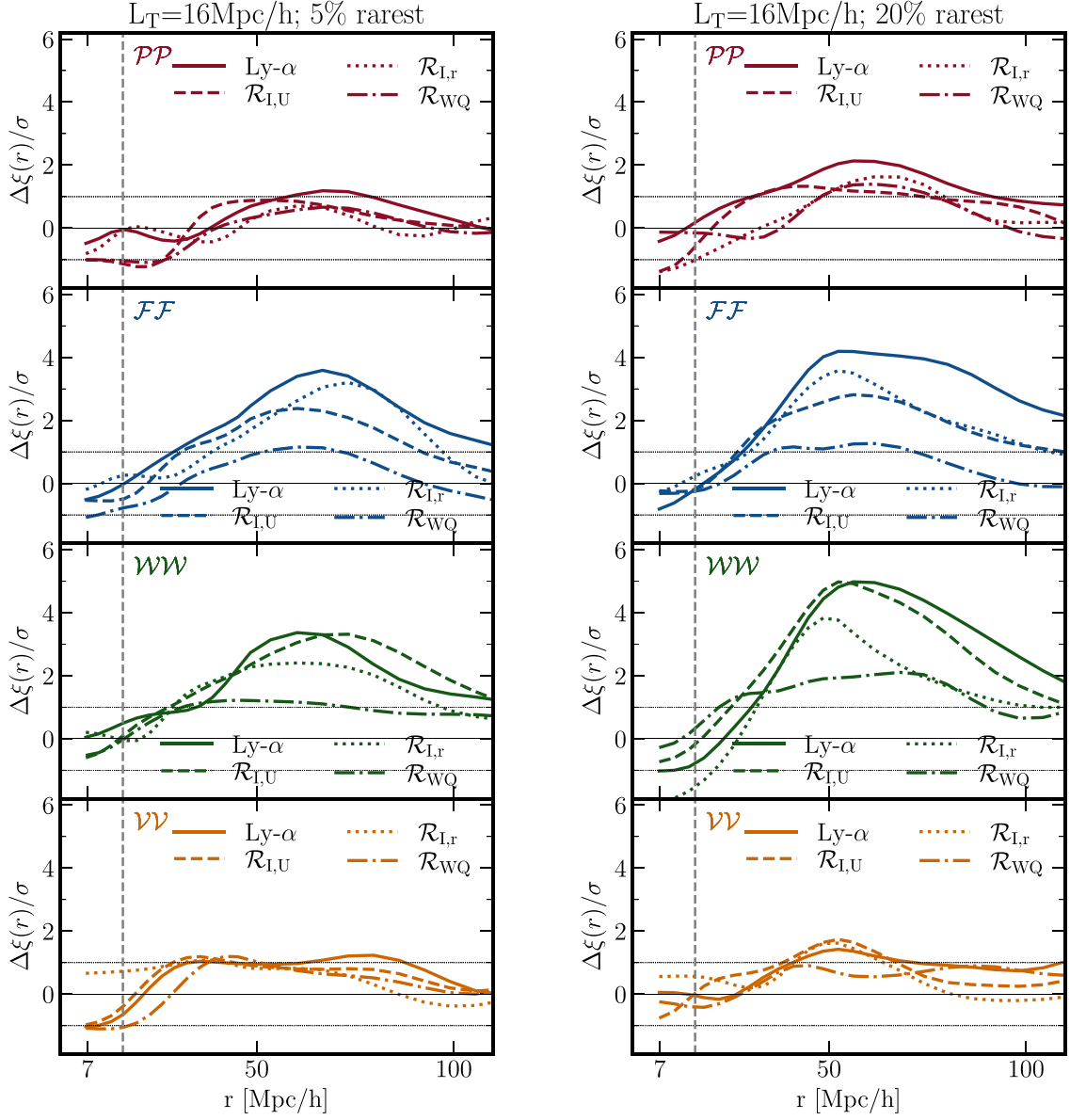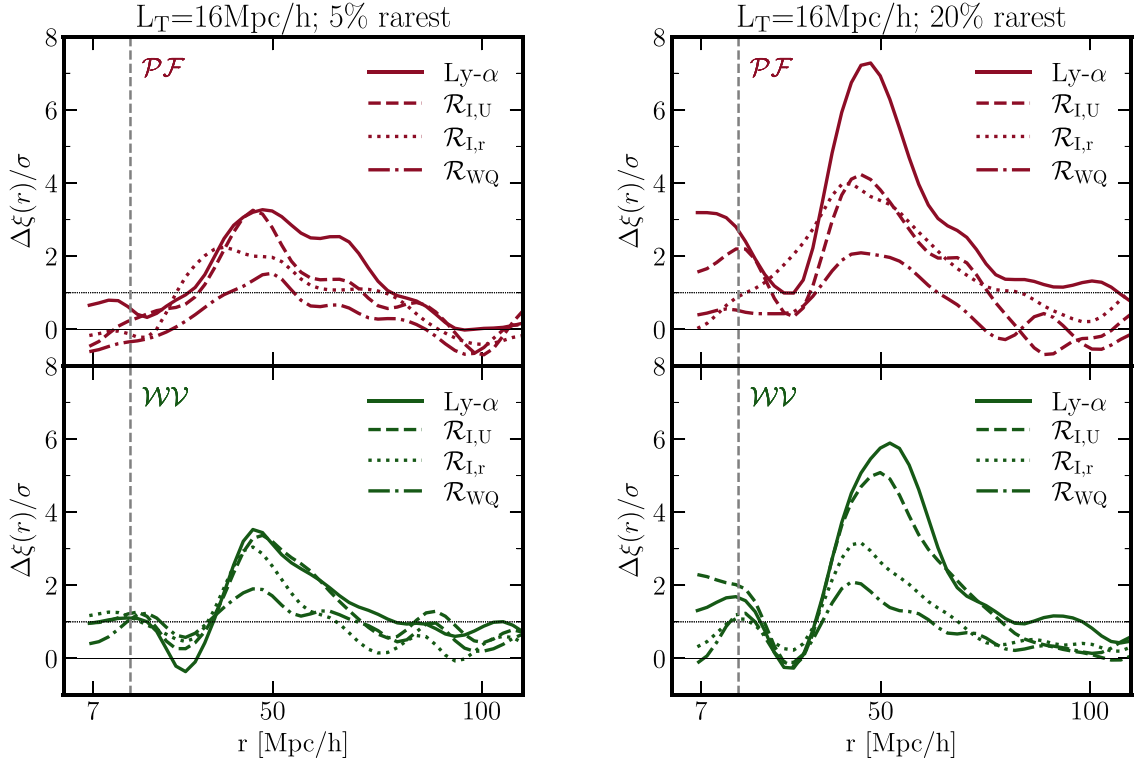
Figure E1. Differences of autocorrelations of critical points of Ly-α and three reconstructed fields with respect to the noise for 5 (left-hand panel) and 20 per cent rarity (right-hand panel) at the smoothing scale $L_T = 16$ Mpc h^{-1}. Vertical dashed grey lines indicate the smoothing scale. While with decreasing rarity the significance of the differences does not change for $\mathcal{VV}$ autocorrelations and it only slightly increases for $\mathcal{PP}$ and $\mathcal{FF}$, it is enhanced by a factor of about 1.5–2 for $\mathcal{WW}$. The differences obtained for 20 per cent rarity (right-hand panel) are comparable to those of 10 per cent (see Fig. 9).
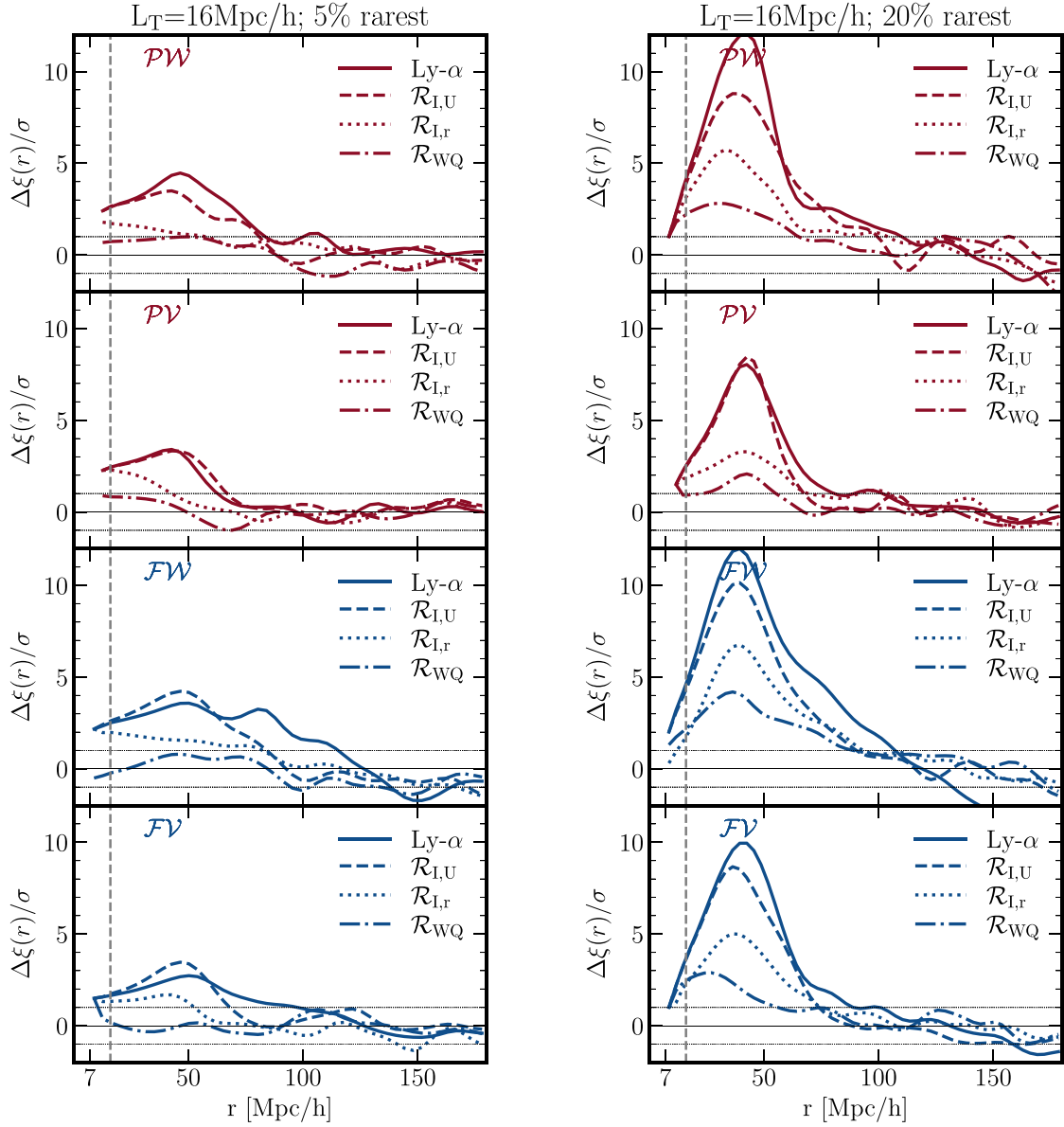
Figure E2. Differences of cross-correlations of overdense and underdense critical points of Ly-α and three reconstructed fields with respect to the noise with 5 (left-hand panel) and 20 per cent rarity (right-hand panel) at the smoothing scale $L_T = 16\,\mathrm{Mpc}\,h^{-1}$. The vertical dashed grey line indicates the smoothing scale. For all cross-correlations, the significance of their differences between all fields and the noise only field increases by about a factor of 2 with decreased rarity (from 5 to 20 per cent).

Figure E3. Differences of cross-correlations of critical points with the same sign of overdensity of Ly-α and three reconstructed fields with respect to the noise with 5 (left-hand panel) and 20 per cent rarity (right-hand panel) at the smoothing scale $L_T = 16$ Mpc h^{-1}. The vertical dashed grey line indicates the smoothing scale. As in the case of under and overdense critical points, for all cross-correlations, there is about a factor of up to 2 increase in the significance of differences with decreased rarity.
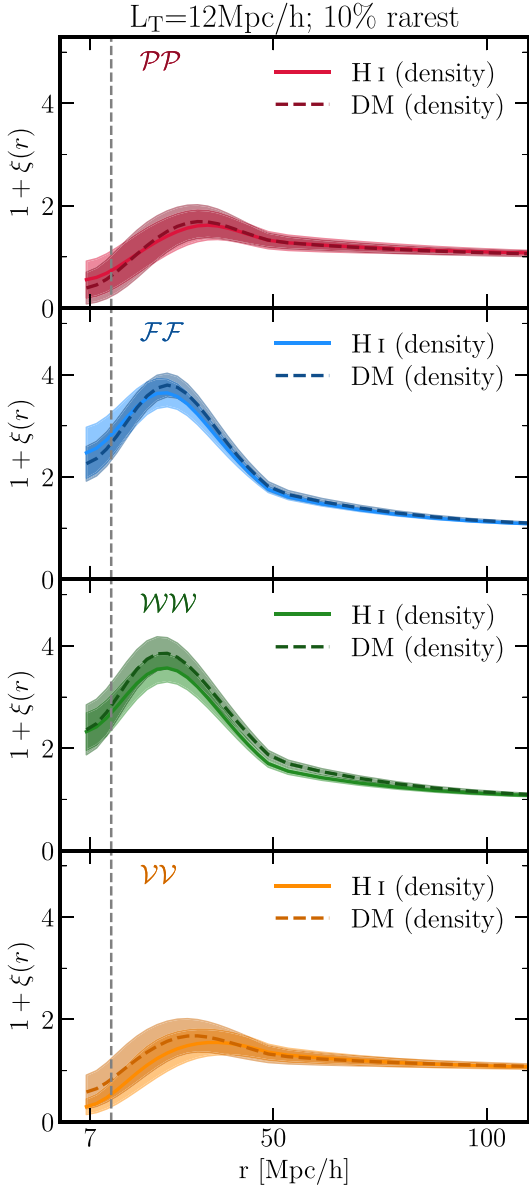
Figure E4. Autocorrelations of critical points with 10 per cent abundance. $\mathcal{PP}$ (peak-peak), $\mathcal{FF}$ (filament-filament), $\mathcal{WW}$ (wall-wall), and $\mathcal{VV}$ (void-void) correlations (from the top to bottom panels) are shown for H I density field (coloured solid lines) and DM field (coloured dashed lines) at the smoothing scale of 12 Mpc h^{-1}. The shaded area corresponds to the error on the mean across five mocks. The vertical dashed grey line indicates the smoothing scale.
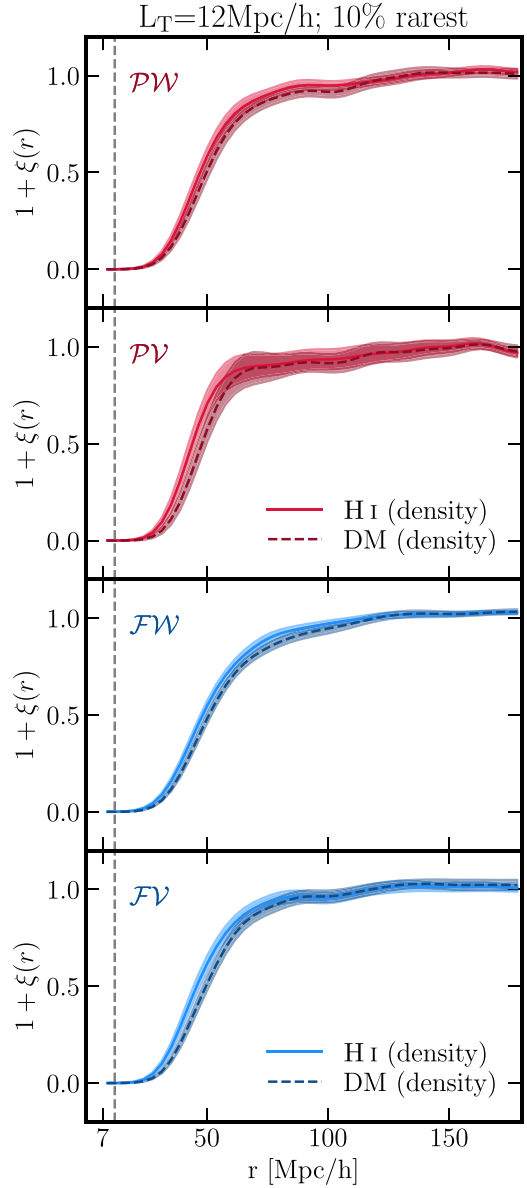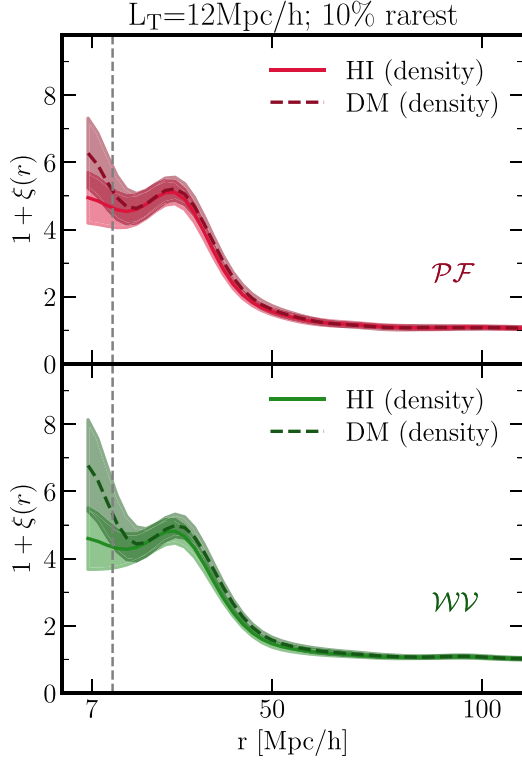
Figure E5. Cross-correlations of critical points with 10 per cent abundance. $\mathcal{PW}$ (peak-wall), $\mathcal{PV}$ (peak-void), $\mathcal{FW}$ (filament-wall), and $\mathcal{FV}$ (filament-void) correlations (from the top to bottom panels) are shown for the H I density field (coloured solid lines) and DM field (coloured dashed lines) at the smoothing scale of 12 Mpc h^{-1}. The shaded area corresponds to the error on the mean across five mocks. The vertical dashed grey line indicates the smoothing scale.

Table E1. Size of the exclusion zone $r_{\mathrm{ex},\,ij}$ (in [Mpc h^{-1}]) of cross-correlations $\mathcal{PW}$, $\mathcal{PV}$, $\mathcal{FW}$, $\mathcal{FV}$ for 10 PER CENT rarity, for H I and DM density fields at the smoothing scale $L_T = 12$ Mpc h^{-1}.

	$\mathcal{PW}$	$\mathcal{PV}$	$\mathcal{FW}$	$\mathcal{FV}$
H I	30.69 ± 1.26	33.70 ± 0.82	28.44 ± 0.67	31.45 ± 1.06
DM	32.95 ± 0.82	35.96 ± 1.25	31.45 ± 0.0	33.70 ± 0.82

Note. The errors are the standard deviations of the mean across all mocks.

Figure E6. Cross-correlations of critical points with 10 per cent abundance at the smoothing scale $L_T = 12\,\mathrm{Mpc}\,h^{-1}$ for H I (solid coloured lines) and DM (dashed coloured lines) density fields. The $\mathcal{PF}$ (peak-filament) and $\mathcal{WV}$ (wall-void) correlations are shown on the top and bottom panels, respectively. The vertical dashed grey line indicates the smoothing scale, while vertical colored lines mark the position of the maximum of each curve.

Table E2. Position ($r_{\mathrm{max},\,ij}$ [Mpc h^{-1}]) and height ($h_{\mathrm{max},\,ij}$) of the maximum of autocorrelations and cross-correlations $\mathcal{PF}$, $\mathcal{WV}$ for 10 PER CENT rarity, for H I and DM density fields at the smoothing scale $L_T = 12\,\mathrm{Mpc}\,h^{-1}$.

		$\mathcal{PP}$	$\mathcal{FF}$	$\mathcal{WW}$	$\mathcal{VV}$	$\mathcal{PF}$	$\mathcal{WV}$
$r_{\mathrm{max},\,ij}$	H I	34.13 ± 0.42	25.61 ± 1.24	24.66 ± 0.79	27.50 ± 3.85	26.88 ± 0.65	25.49 ± 1.05
	DM	30.34 ± 3.10	23.24 ± 1.03	20.39 ± 0.67	29.87 ± 2.42	25.95 ± 0.82	29.18 ± 0.65
$h_{\mathrm{max},\,ij}$	H I	2.86 ± 0.21	4.93 ± 0.28	4.94 ± 0.24	3.15 ± 0.38	5.10 ± 0.34	4.81 ± 0.35
	DM	3.15 ± 0.18	4.83 ± 0.11	4.96 ± 0.26	3.03 ± 0.23	6.34 ± 0.19	6.15 ± 0.29

Note. The errors are the standard deviations of the mean across all mocks.

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