PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
FACULTAD DE FÍSICA
INSTITUTO DE ASTROFÍSICA

REDSHIFT-SPACE DISTORTIONS WITH SPLIT DENSITIES

by

Enrique Paillas Villavicencio

Thesis submitted to the Faculty of Physics
of Pontificia Universidad Católica de Chile, as one of the requirements
to qualify for the academic Ph.D. degree in Astrophysics.

Supervisor	:	Dr. Nelson Padilla
Co-supervisors	:	Dr. Ariel Sánchez
		Dr. Yan-Chuan Cai
Correctors	:	Dr. Felipe Barrientos
		Dr. Ezequiel Treister
		Dr. Jeremy Tinker

June, 2021

Santiago, Chile

To you, 2,000 years from now.

Personal acknowledgements

I deeply appreciate the support and knowledge that my supervisors, Nelson Padilla, Ariel Sánchez and Yan-Chuan Cai, have given me during these years. This work would not have been possible without their mentoring. I would also like to express my gratitude to my family and friends, whose love and encouragement have kept me going throughout this journey.

During the course of my graduate studies, I have been supported by CONICYT-PCHA/Doctorado Nacional (2017-21170093). I also acknowledge support from CONICYT project Basal AFB-170002 and Fondecyt 1191813.

Abstract[1]

Accurate modelling of redshift-space distortions (RSD) is challenging in the non-linear regime for two-point statistics e.g. the two-point correlation function (2PCF). We take a different perspective to split the galaxy density field according to the local density, and cross-correlate those densities with the entire galaxy field. We demonstrate that combining a series of cross-correlation functions (CCFs) offers improvements over the 2PCF as follows: 1. The distribution of peculiar velocities in each split density is nearly Gaussian. This allows the Gaussian streaming model for RSD to perform accurately for a wide range of scales. 2. The probability distribution function of the density field at small scales is non-Gaussian, but the CCFs of split densities capture the non-Gaussianity, leading to improved cosmological constraints over the 2PCF. We can obtain unbiased constraints on the growth parameter $f\sigma_{12}$ at the per-cent level, and Alcock-Paczynski (AP) parameters at the sub-per-cent level with the minimal scale of $15\,h^{-1}$Mpc. This is a $\sim$30 per cent and $\sim$6 times improvement over the 2PCF, respectively. 3. Baryon acoustic oscillations (BAO) are contained in all CCFs of split densities. Including BAO scales helps to break the degeneracy between the line-of-sight and transverse AP parameters, allowing independent constraints on them. We test our methodology on N-body simulations and apply it to the BOSS DR12 galaxy samples, obtaining constraints for the growth rate of structure at different redshifts.

[1]This thesis uses verbatim text of a manuscript by the author published in the Monthly Notices of the Royal Astronomical Society journal (Paillas et al., 2021).

Resumen

El modelado de las distorsiones en el espacio de redshift (RSD) es un desafío en el régimen no lineal para estadísticas de dos puntos, e.g. la función de correlación de dos puntos (2PCF). Tomamos una perspectiva diferente y dividimos el campo de densidad de galaxias de acuerdo a la densidad local y correlacionamos las densidades con todo el campo de galaxias. Demostramos que la combinación de una serie de funciones de correlación cruzada (CCF) ofrece mejoras sobre la 2PCF de la siguiente manera: 1. La distribución de velocidades peculiares en cada densidad dividida es casi Gaussiana. Esto permite que el modelo de transmisión gaussiana para RSD funcione con precisión para una amplia gama de escalas. 2. La distribución de probabilidad del campo de densidad a escalas pequeñas no es gaussiana, pero las CCF de densidades divididas capturan la no gaussianidad, lo que conduce a restricciones cosmológicas mejoradas sobre la 2PCF. Podemos obtener restricciones insesgadas en el parámetro de crecimiento $f\sigma_{12}$ a un nivel porcentual, y los parámetros de Alcock-Paczynski (AP) a un nivel sub-porcentual usando una escala mínima de $15\,h^{-1}$Mpc. Esto es una mejora de ~ 30 por ciento y ~ 6 veces sobre la 2PCF, respectivamente. 3. Las oscilaciones acústicas de bariones (BAO) están contenidas en todos los CCF de densidades divididas. La inclusión de escalas BAO ayuda a romper la degeneración entre los parámetros AP transversales y a lo largo de la línea de vision. Probamos nuestra metodología en simulaciones de N-cuerpos y la aplicamos a las muestras de galaxias BOSS CMASS y LOWZ, obteniendo restricciones para la tasa de crecimiento de la estructura en diferentes épocas.

Contents

Chapter 1

Introduction

In our standard cosmological picture, the Λ cold dark matter (ΛCDM) model, our present Universe seems to have evolved from a very homogeneous initial state, where small-scale quantum fluctuations from the inflationary epoch gave rise to density perturbations that grew over time due to gravitational instabilities. The rate of growth of these overdensities has always been at contest with the expansion rate of the Universe (Linder & Polarski, 2019). As such, precise measurements of this growth rate of structure at different redshifts constitute valuable tests for ΛCDM, as well as for alternative cosmological models, such as modifications to the theory of gravity (Copeland et al., 2006; Linder & Cahn, 2007; Jennings et al., 2011; Joyce et al., 2015; Koyama, 2016; Paillas et al., 2019).

It is often said that we are living in the era of precision cosmology, where large-scale galaxy surveys have granted us information to achieve almost percent-level constraints on the parameters that describe the global properties of our Universe. A powerful method for constraining the growth rate of cosmic structure is through the analysis of redshift-space distortions (RSD, Jackson, 1972; Kaiser, 1987). This effect arises because we estimate distances to extra-galactic sources through their redshift. In doing so, we assume that the recession velocity of these objects is only sourced by the cosmological expansion. In practice, galaxies also exhibit peculiar velocities that are induced by gravitational interactions with neighbouring structures. These peculiar velocities perturb the cosmological redshifts, which produces distortions in the clustering pattern of galaxies in *redshift space*[1]. Since peculiar velocities are induced by gravity, an accurate modelling of this effect grants us the possibility to estimate the rate at which structure is assembling (Kaiser, 1987).

As an example, Fig. 1.1 illustrates the effect that RSD would produce when observing a distant overdense region in the Universe, such as a massive galaxy cluster. The left-hand side panel shows the positions of the cluster galaxies in real space, where the arrows represent their peculiar motions: galaxies in the outskirts are attracted by the cluster's gravitational pull, and are therefore falling towards the centre of the overdensity in a more or less coherent way. The galaxies in the centre, however, have already been through many interactions with nearby galaxies in this crowded environment, and therefore exhibit roughly random motions. This is how the cluster would look like if we knew precisely what are the distances to each galaxy in this group. Note that in this example, the cluster is close to spherical in real space. The easiest way to estimate distances to each galaxy is through their redshift. The observer, who is looking from a very distant location from the bottom of the image, measures a galaxy redshift by taking a spectrum of its light, and assumes that it is sourced by the expansion of the Universe. Assuming a flat ΛCDM Universe at the matter-dominated era, the comoving[2] distance to the galaxy can be calculated as

$$d_C(z) = \frac{c}{H_0} \int_0^z \frac{dz'}{\sqrt{\Omega_m (1+z')^3 + \Omega_\Lambda}} \,, \tag{1.1}$$

where c is the speed of light, H_0 is the Hubble constant, and Ω_m and Ω_Λ are the present-day matter density and dark energy densities, respectively. This is all fine in

[1]Throughout this thesis, I will make the distinction between real space, which is the true, unaltered distribution of galaxies in a cosmological volume, and redshift-space, which is a distorted version of the galaxy distribution we see when mapping redshifts to distances.

[2]*Comoving* distances, d_C, do not change as the Universe expands, as opposed to *physical* distances, d_P, which scale as $d_P = d_C a$, where a is the scale factor of the Universe at a certain epoch.

theory. In practice, however, the galaxy peculiar motions along the observer's line of sight (LOS) will also contribute to the redshift. On large scales, the coherent motion of galaxies falling towards the overdensity will make the cluster appear squashed along the LOS in redshift space (shown in the right-hand side panel), an effect that was first described by Kaiser (1987). On small scales, the random galaxy motions will make the cluster look elongated along the LOS, which is commonly known as the Fingers of God effect (FoG, Jackson, 1972). Even though this cartoon exaggerates the physical scales and the degree of distortion that is produced in redshift space, it serves to illustrate some of the recurring effects that we will find throughout this thesis. We can imagine a similar situation around underdense regions in the Universe, such as cosmic voids. On large scales, galaxies often would recede away from the centre in a coherent way as the void expands and evacuates its matter towards the void walls. The net effect is that an expanding void that is spherical in real space, would appear elongated on large scales along the line of sight in redshift space (Cai et al., 2016).

RSD has been widely used in observations to measure the growth rate of structure and test theories of gravity with galaxy redshift surveys such as BOSS (Alam et al., 2017), eBOSS (eBOSS Collaboration et al., 2020), VIPERS (Pezzotta et al., 2017), GAMA (Blake et al., 2013), the 6 degree Field Galaxy Survey 6dFGS (Beutler et al., 2012), the Subaru FMOS galaxy redshift survey (Okumura et al., 2016) and the WiggleZ Dark Energy Survey (Blake et al., 2011). These measurements have achieved a 5-10 percent level of accuracy on the estimation of the growth rate of structure at various redshifts. The next generation of spectroscopic redshift surveys such the Dark Energy Spectroscopic Instrument (DESI, Levi et al., 2019) and *Euclid* (Laureijs et al., 2011) promise even better parameter constraints, reaching the percent-level of accuracy.

The extraction of cosmic growth rate information from observations requires robust theoretical frameworks that can model RSD across different scales. Great efforts

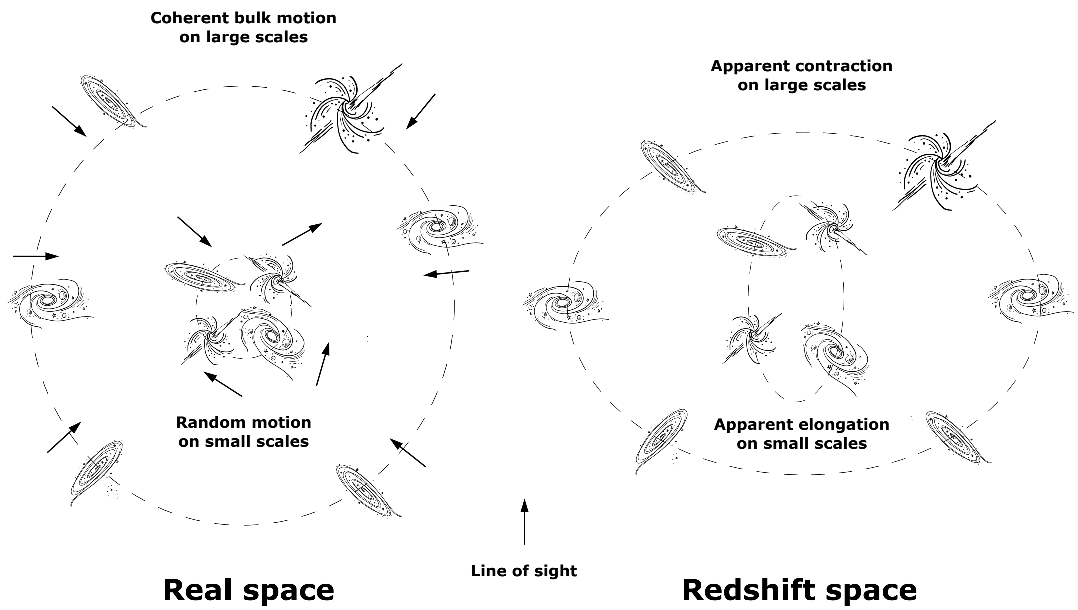

Figure 1.1: An illustration of how do redshift-space distortions affect the observed geometry of an overdensity, such as a galaxy cluster. The left-hand side panel shows the true distribution of galaxies in real space, where arrows represent the galaxy peculiar motions. On large scales, galaxies fall coherently towards the potential well, while on small scales their motions are fairly random. Due to the perturbation of peculiar velocities to galaxy redshifts, the cluster appears squashed along the line of sight on large scales, and elongated on small scales.

have been made to improve the modelling of RSD in terms of measurements of the two-point correlation function (2PCF, e.g. Sheth et al., 2001a; Reid & White, 2011; Bianchi et al., 2016) or the power spectrum (e.g. Scoccimarro, 2004; Seljak & McDonald, 2011; Chen et al., 2020). The galaxy 2PCF, often denoted as $\xi(r)$, is a measure of the degree of clustering of a distribution of galaxies in a cosmological volume. It is defined as the excess probability of finding a pair of galaxies separated by a given distance r, with respect to a random distribution. The power spectrum can be thought of as the Fourier space analog of the 2PCF. These two-point statistics characterising the variance of the density field are able to capture all information if the field is Gaussian, which is the case in the so-called 'linear regime', either at early times in the Universe or at very large scales, where non-linear gravitational evolution has not yet taken place. In the non-linear regime, however, the density field becomes non-Gaussian, so the variance of the field becomes an incomplete statistic. In a non-Gaussian field, the 2PCF will not be able to extract all the information. This is a major limitation for two-point statistics (Einasto et al., 2020).

Among the different theoretical frameworks, the streaming model (Peebles, 1980; Fisher, 1995) has been widely used for modelling RSD in configuration space. This model provides a mapping between the real and redshift-space correlation functions, which, as we will see later, allows us to put constraints on the growth rate of structure. A crucial ingredient for the model to work accurately is the distribution of pairwise[3] velocities of the galaxies that are being analysed. This distribution is known to be highly non-Gaussian at the quasi-linear and non-linear scales, making it challenging to model (e.g. Scoccimarro, 2004). Extensive effort has been made to model the pairwise velocity distribution (Kuruvilla & Porciani, 2018; Cuesta-Lazaro et al., 2020). Reid & White (2011) have shown that, under the assumption of a Gaussian pairwise velocity distribution, taking the true real-space 2PCF, and the mean streaming velocity and velocity dispersion profiles from simulations, the predictions of the streaming model agree with direct redshift-space measurements at a few percent level at a scale of $s \sim 25\,h^{-1}$Mpc, and fail at smaller scales. Kuruvilla & Porciani (2018) have shown that dropping the Gaussian assumption, with the full distribution function, the model works perfectly for the two-dimensional redshift-space correlation function. Alternatively, Tinker (2007) showed that the skewed PDF of the pairwise velocity arises when combining halo pairs from environments with different densities. By splitting a sample of haloes into different quantiles according to

[3]In this context, the term pairwise indicates that a quantity is measured relative to another another object. The galaxy pairwise velocity, for instance, refers to the relative peculiar velocity vector for a given pair of galaxies.

their local number density, they verified that the pairwise velocities at fixed density are approximately Gaussian, and this helps to improve the accuracy of the streaming model.

Alternative approaches for overcoming the limitation of two-point measurements include higher-order statistics, such as the three-point correlation function and the bispectrum (e.g. Sefusatti et al., 2006; Gil-Marín et al., 2012; Slepian & Eisenstein, 2015, 2017), non-linear transformation to re-Gaussianise the density field (Neyrinck et al. 2009, 2011; Neyrinck 2011; Wang et al. 2011), density split statistics for weak lensing analysis (Gruen et al., 2016, 2018; Friedrich et al., 2018), counts-in-cells statistics (Szapudi & Pan, 2004; Klypin et al., 2018; Jamieson & Loverde, 2020; Mandal & Nadkarni-Ghosh, 2020), and using the concept of separate Universe to model density-dependent two-point statistics (Wagner et al. 2015; Chiang et al. 2015), which, in essence, corresponds to a three-point quantity. These approaches usually provide complementary cosmological constraints by accessing information from the non-Gaussian field.

More recently, there has been increasing attention in using RSD around cosmic voids, taking the advantage of the milder density contrasts around them, which may be better described by linear dynamics (Hamaus et al., 2014; Pollina et al., 2017). Paz et al. (2013) employed the Gaussian streaming model for extracting velocity profiles around voids; Hamaus et al. (2015) applied the Gaussian streaming model for voids to constrain growth parameters, with the assumption that the density and streaming velocity are linearly coupled. Cai et al. (2016) wrote down how the redshift-space correlation function is mapped to its real-space version when only streaming velocities are accounted for. When expanding it to the linear order, it works well for small density perturbations, but by definition does not work when δ is relatively large, i.e. near void centres. Nadathur et al. (2019b) derived expressions for RSD around voids and convolved them with a Gaussian velocity distribution, showing that it helps to improve the performance of the model, and stressed that the

Gaussian streaming model provides a poor fit to their data. These models, though somewhat disputed at the details, have been applied to observational data, and in some cases, have led to significant improvement for the constraints of AP and growth parameters (e.g. Hamaus et al., 2017; Achitouv, 2019; Hawken et al., 2020; Correa et al., 2020). Nevertheless, like other beyond-two-point statistics, RSD around voids are an interesting development, but there is no obvious reason to include only voids in the analysis.

The research presented in this thesis builds upon the idea of separate universes, counts-in-cells and density split statistics for weak lensing. I will present a framework to model RSD around different density environments. The method consists in splitting random positions of the spherically smoothed galaxy density field in different quantiles, and computing CCFs between these positions and the entire galaxy field in redshift space. This is in essence density split in three dimensions (DS). The CCF between regions in each quantile with the galaxy field corresponds to the stacked galaxy number density around environments of different depths. DS can also be seen as a generalisation of the void-galaxy CCF, as it naturally includes voids, clusters and intermediate-density regions in a general framework to exploit their combined constraining power on cosmology.

With the aforementioned setup, the main question I wish to address, which constitutes the central idea of this thesis, is : *does the combination of a series of CCFs contain a different amount of cosmological information than the conventional 2PCF?* A closely related question is: does DS make the modelling more accurate for RSD than the standard 2PCF? This work shows that with DS, the distribution of the velocities in each density quantile are well-fit by a Gaussian function. This allows the Gaussian streaming model to perform accurately at almost all scales. Perhaps more importantly, the CCFs of split 3D density naturally captures the non-Gaussian distribution of the density, and this leads to improved cosmological constraints over two-point statistics.

I test this methodology on cosmological N-body simulations, showing that it enables us to recover unbiased constraints for the growth rate of cosmic structure when simultaneously modelling dynamical and geometrical distortions in the data. I also show how to apply this framework to observational galaxy samples by using the BOSS CMASS and LOWZ catalogues from SDSS DR12, providing estimates for the growth rate of structure at $z = 0.38$ and $z = 0.61$.

The vast majority of the cosmological measurements and analyses presented in this thesis were carried out using software that I designed and implemented from scratch for this project. All of these codes and routines are made available at my GITHUB repository[4], in the hope that it can serve someone else in the future, and also following the spirit of transparent and reproducible science.

[4]https://github.com/epaillas

Chapter 2

Redshift-space distortions with split densities

To compare the power of constraining cosmology between the 2PCF and DS using RSD, we need to have adequate dynamical distortion models for both the 2PCF and DS. I will employ the Gaussian streaming model (Fisher, 1995) for both of them. In addition, we also need to account for the geometrical distortions, the Alcock-Paczynski effect (Alcock & Paczynski, 1979). I will start by introducing these two distortion effects. Then, I will describe the density split algorithm, which is the main basis of the cosmological analysis of this thesis.

2.1 Dynamical distortions

To the linear order in velocity, the observed redshift of a distant galaxy is the sum of two components, the cosmological redshift and the redshift due to its peculiar velocity sourced by gravity. The observed redshift-space distance of a galaxy is then

$$\mathbf{s} = \mathbf{r} + \frac{v_\parallel}{aH}\hat{\mathbf{z}}, \tag{2.1}$$

where $\mathbf{r}$ and $\mathbf{s}$ are real- and redshift-space distance vectors, $v_\parallel$ is the peculiar velocity along the line-of-sight direction $\hat{\mathbf{z}}$; a is the scale factor of the Universe, H is the Hubble parameter at a. With mass conservation, we have

$$[1 + \xi^s(\mathbf{s})]\,\mathrm{d}^3\mathbf{s} = [1 + \xi(\mathbf{r})]\,\mathrm{d}^3\mathbf{r}\ , \tag{2.2}$$

where $\xi(\mathbf{r})$ and $\xi^s(\mathbf{s})$ denote the real and redshift-space correlation functions, which measure the excess probability of finding a galaxy pair separated by a given scale. The streaming model (Peebles, 1980) provides a mapping between the real-space correlation function to the redshift-space anisotropic correlation function:

$$1 + \xi^s(s_\perp, s_\parallel) = \int [1 + \xi(r)]\,\mathcal{P}(v_\parallel, \mathbf{r})\mathrm{d}v_\parallel, \tag{2.3}$$

where $r^2 = r_\parallel^2 + r_\perp^2$ is the real-space separation and $v_\parallel = aH(s_\parallel - r_\parallel)$ is the pairwise line-of-sight velocity, which has a probability distribution $\mathcal{P}(v_\parallel, \mathbf{r})$. Note that for a given r, the PDF for the line-of-sight velocities depends on the subtended angle from the line of sight θ, as it has the contribution from both the radial and tangential components. Assuming a Gaussian form for $\mathcal{P}(v_\parallel, \mathbf{r})$, and that the density field is also Gaussian, neglecting higher order terms, the mapping becomes:

$$\begin{aligned} 1 + \xi^s(s_\perp, s_\parallel) = \int (1 + \xi(r))\, &\frac{1}{\sqrt{2\pi\sigma_\parallel^2(r,\mu)}} \\ &\exp\left\{ -\frac{\left[v_\parallel - v_r(r)\mu\right]^2}{2\sigma_\parallel^2(r,\mu)} \right\} \mathrm{d}v_\parallel, \end{aligned} \tag{2.4}$$

where $\mu = r_\parallel / r = \cos\theta$, and $v_r(r)$ is the pairwise velocity along the radial direction, also known as the mean streaming velocity. This was first derived by Fisher (1995), and is usually referred to as the Gaussian streaming model (GSM) (see also the derivations by Scoccimarro, 2004; Vlah & White, 2019, in Fourier space). Note that the velocity dispersion $\sigma_\parallel(r, \mu)$ depends on both r and μ. Eq. (2.3) can be used to predict the redshift-space correlation function, as long as the full distribution of pairwise velocities and the real-space correlation function are fully known. However, it is challenging to predict the distribution of pairwise velocities from first principles. Eq. (2.4) may work but only if the Gaussian assumption holds. Exploring the validity of this assumption is a focus for this study.

Another key ingredient for the GSM is the pairwise streaming velocity $v_r(r)$. In linear theory, it can be expressed in terms of the correlation function (Peebles, 1980; Sheth et al., 2001b)

$$v_r(r) = -\frac{2}{3}\frac{\beta a H r \bar{\xi}(r)}{[1+\xi(r)]}. \tag{2.5}$$

Here $\xi(r)$ is the galaxy real-space correlation function, $\beta = f/b$ where $f = \mathrm{d}\ln D/\mathrm{d}\ln a$ is the linear growth rate, D is the growth factor, b is the linear galaxy bias, and

$$\bar{\xi}(r) = \frac{3}{r^3}\int_0^r \xi(x)x^2\mathrm{d}x. \tag{2.6}$$

The linear coupling between density and velocity is not adequate for the quasi-nonlinear regime. From this point and throughout the thesis, my notations will not use the usual subscript g to refer to quantities related to galaxies. I will use $\xi(r)$ and $\bar{\xi}(r)$ to refer to the *galaxy* correlation function and its cumulative version. I will use the subscript m to refer to quantities about dark matter.

Modelling the exact coupling is another crucial step towards accurate modelling for RSD with the GSM. I adopt the empirical function introduced in Juszkiewicz et al. (1999), with one free parameter to model the coupling between the galaxy density and velocity field:

$$v_r(r) = -\frac{2}{3}aHr\beta\bar{\bar{\xi}}(r)\left[1+\nu\bar{\bar{\xi}}(r)\right] . \tag{2.7}$$

In the expression above, ν is a free parameter. When $\nu = 0$, the above goes back to the linear model. $\bar{\bar{\xi}}$ is defined by the following relation:

$$\bar{\xi}(r) \equiv \bar{\bar{\xi}}(r)[1 + \xi(r)]. \tag{2.8}$$

As I will describe in the following sections, I am also interested in the cross-correlation between randomly positioned centres ranked by their local number density of galaxies, and the entire galaxy number density field. In such a case, the equations for the streaming model outlined above still apply, but as there are no peculiar motions for the random centres themselves, the pairwise velocity profile becomes the stacked velocity profile, and the linear density-velocity coupling becomes

$$v_r(r) = -\frac{1}{3} aHr\beta\bar{\xi}(r), \tag{2.9}$$

where $\bar{\xi}(r)$ from Eq. (2.6) now refers to the cumulative cross-correlation function between fixed positions (e.g. voids or clusters) with the galaxy field, i.e. the stacked density profile, and so it is the same as the cumulative *galaxy* density contrast $\Delta(r)$ within r. This is commonly adopted in void RSD studies (e.g. Hamaus et al. 2014; Cai et al. 2016; Achitouv et al. 2017; Hawken et al. 2020; Hamaus et al. 2017; Nadathur et al. 2020). Again, I will go beyond the linear coupling by introducing

$$v_r(r) = -\frac{1}{3} \frac{aHr\beta\bar{\xi}(r)}{1 + \nu\xi(r)}, \tag{2.10}$$

where ν is a free parameter and $\nu = 0$ is the linear model. As we will see later, this empirical expression will allow me to not only fit the velocity profiles around voids, but also around high density regions.

Without causing confusion, I will use the same notations $[v_{\parallel},\, v_r,\, \sigma_{\parallel}, \xi, \bar{\xi}]$ to refer to variables for both two-point correlation function (2PCF) and cross-correlation function (CCF) throughout the thesis.

2.2 Geometrical distortions

In observations, when converting observed redshifts to distances using a cosmology that is different from the true underlying cosmology of the Universe, we artificially induce geometrical distortions in the clustering of galaxies, an effect that is also known as Alcock-Paczynski (AP) distortions (Alcock & Paczynski, 1979). We can parametrise these distortions by rescaling the transverse and line-of-sight separation vectors (Ballinger et al., 1996):

$$s_{\perp} = q_{\perp} s'_{\perp} \tag{2.11}$$

$$s_{\parallel} = q_{\parallel} s'_{\parallel} \ , \tag{2.12}$$

where the primes represent quantities in the fiducial cosmology. The scaling factors are related to cosmological parameters via

$$q_{\perp} = \frac{D_M}{D'_M} \tag{2.13}$$

$$q_{\parallel} = \frac{H'}{H} \ , \tag{2.14}$$

where D_M and H are the comoving angular diameter distance and the Hubble parameter at a, respectively. The redshift-space correlation function can then be rescaled as

$$\xi^s(s_{\perp}, s_{\parallel}) = \xi^s(q_{\perp} s'_{\perp}, q_{\parallel} s'_{\parallel}) \ . \tag{2.15}$$

The dynamical and geometrical distortions act at the same time on the observed redshift-space correlation function – the only observable at our disposal in this context. Adjusting the dynamical distortion parameters (RSD) and geometrical distortion parameters (AP) to fit for the observed redshift space clustering will in turn allow us to constrain those parameters of interests (e.g. Sánchez et al., 2017a; Beutler et al., 2017; Hou et al., 2018; Hamaus et al., 2020; Bautista et al., 2021).

2.3 Splitting densities

A crucial assumption for the Gaussian streaming model (Eq. 2.4) to work is that the PDF of the pairwise velocity needs to be Gaussian. This is only true in the linear regime, where non-linear gravitational evolution has not yet taken place significantly (either at high redshift or at very large scales). Reid & White (2011) have shown that this assumption can already lead to a 2 percent level bias for the quadrupole at $s = 25\,h^{-1}$Mpc. To improve the performance of the model, one obvious way is to go beyond the Gaussian assumption by modelling the full distribution of pairwise velocity. This is non-trivial from first principles and extra degrees of freedom are usually introduced (Kuruvilla & Porciani 2018; Cuesta-Lazaro et al. 2020).

I take an alternative approach to analyse the data by splitting the galaxy field into different density environments. The assumption is that the non-Gaussian PDF of the pairwise velocity at small scales can be decomposed into many Gaussian PDFs of different widths. This was elucidated in Tinker (2007), where it was shown that the PDF of pairwise velocities at a specific density environment is indeed close to Gaussian. The halo model was also adopted for the modelling in Tinker (2007). Examples for the velocity PDFs for several overdense environments were shown, but not for underdense environments. I will generalise this to low-density regions. I will use cross-correlations instead of auto-correlation functions. In this case, the relevant statistic is the distribution function of the velocity field, rather than the pairwise velocities. The main steps for this method are summarised as follows:

1. start from a galaxy sample in real space and apply a spherical top-hat filtering with a filter radius R on random positions.

2. rank order the filtered density contrasts $\Delta(r = R)$ (the same as $\bar{\xi}(R)$ as noted in Chapter 2) and split them into n density bins. For this study, I adopt

$R = 15\,h^{-1}$Mpc and $n = 5$.

3. cross-correlate the positions in each density bin, or quintile, with the entire galaxy sample in redshift space to obtain the CCF $\xi^i(s, \mu|\Delta^i)$, where $i = 1, 2, 3...n$[1]. These are in essence a series of conditioned correlation functions, with the condition being $\Delta^i(r = R)$ satisfying the density splitting criteria. I will write Δ^i instead of $\Delta^i(r = R)$ without loss of clarity.

The CCFs $\xi^i(r, \mu|\Delta^i)$ are the same as $\delta^i(r, \mu|\Delta^i)$, i.e. the stacked number densities of galaxies around the centres of spheres within the i-th density bin. Therefore, the CCF for the lowest density bin is similar to the void-galaxy cross-correlation function, and the highest density bin is similar to the cluster-galaxy cross-correlation function. The notation $\Delta(r)$ is also the same as $\bar{\xi}(r)$ from Eq. (2.6), with $\xi(r) = \delta(r)$, as mentioned in Chapter 2.

Instead of the 2PCF, the series of CCFs will be our main observable to be used to constrain cosmology. The streaming model and its Gaussian version can naturally be applied to model these void-galaxy, cluster-galaxy cross-correlations, and in general cross-correlations of different local densities with the galaxy field. In the GSM (Eq. 2.4), one can simply replace $\xi(r)$ by the real-space CCF (or the stacked density profile), and the pairwise velocity profile $v_r(r)$ by the stacked velocity profiles around different local densities. This can be seen by keeping the positions of the spherical regions fixed in real-space, i.e. one of the two points in the pair has zero velocity and thus the pairwise velocity becomes the radial velocity of a galaxy relative to its real-space spherical centre. Note that although the mathematical form is the same, the 2PCF is, fundamentally, a two-point statistic, while the cross-correlation between a set of centres with the entire galaxy field is a first-moment measurement.

The number of density quantiles chosen in this work is a balance between different factors. Firstly, a larger number of quantiles ensures that the distribution of densities

[1]Note that I have used $\mathbf{s}$ to denote the redshift-space distance vector in the earlier part of the paper. Here I use the same symbol to denote the distance between the DS centres (in real space) to galaxies (in redshift space). Their meanings are technically different.

in each bin is narrow, which results in a distribution of radial velocities that is more Gaussian, and thus a better performance for the GSM. Secondly, increasing the number of bins also increases the size of the covariance matrix that is needed for our likelihood analysis (see Sec. 3.2). A bigger covariance matrix will also require a larger number of mock realizations in order to be accurately estimated. Since I have 300 mocks at my disposal, 5 quantiles is a good compromise between these two factors.

Chapter 3

Validation on N-body simulations

Having listed the necessary ingredients to implement the RSD with split densities pipeline, I will proceed to validate this framework using a set of N-body cosmological simulations from which we can directly measure all necessary clustering statistics, both in real and redshift space. This will serve as a theoretical ground to quantify the accuracy of the GSM for these summary statistics and to study the cosmological constraining power that this method can offer, which will allow me to apply it to an observational galaxy sample in forthcoming chapters.

3.1 Performance of redshift-space distortions models

Following the steps described in the previous chapter, I will use mock galaxies to measure the series of $\xi^i(s, \mu|\Delta^i)$. I will also measure the 2PCF $\xi(s, \mu)$. I then use the GSM to extract cosmological information from the above two measurements from the same simulations. I will focus on asking: does the combination of the whole series of CCFs, i.e. all $\xi^i(s, \mu|\Delta^i)$, contain any different amount of cosmological information than the conventional 2PCF? Before doing this, it is crucial to validate the performance of RSD modelling with the Gaussian streaming model for CCFs from split densities and the 2PCF. This is the focus of this section. I will describe the mock galaxies catalogue (Sec. 3.1.1), followed by splitting densities with the mock galaxies (Sec. 3.1.2). I then analyse the velocity distribution functions with the split densities (Sec. 3.1.3), and cross-correlate each split density quintile with the entire galaxy field and compare them with the Gaussian streaming model (Sec. 3.1.4). I review and compare other RSD models for cross-correlations in Sec. 3.1.5.

3.1.1 Mock galaxies

I will use the Minerva simulation for the analysis (Grieb et al., 2016; Lippich et al., 2019). It consists of a set of 300 N-body simulations that represent different realisations of the same cosmology, which corresponds to the best-fitting flat ΛCDM model to the combination of CMB data (Planck Collaboration et al., 2016) and SDSS DR9 CMASS wedges, presented in Sánchez et al. (2013). The model is characterised by a matter density parameter of $\Omega_{\mathrm{m}} = 0.285$, a baryon physical density of $\omega_{\mathrm{b}} = 0.02224$, a present-day Hubble rate of $H_0 = 69.5\ \mathrm{kms^{-1}Mpc^{-1}}$, an amplitude of density fluctuations of $\sigma_8 = 0.828$ and an scalar spectral index of $n_{\mathrm{s}} = 0.968$. For each box, a total of 1000^3 dark matter particles were evolved with GADGET (Springel, 2005) in a cosmological box of 1.5 h^{-1}Gpc aside. The volume for each box is therefore $3.375(h^{-1}\mathrm{Gpc})^3$. This will be the default volume that sets the statistical errors for my predictions. Initial conditions were generated using second order-Lagrangian perturbation theory (Jenkins, 2010), starting from $z = 63$.

Dark matter haloes and their associated substructures are identified using subfind (Springel et al., 2001). These haloes were populated at the $z = 0.57$ snapshots of the simulations using the Halo Occupation Distribution method (HOD; Peacock & Smith 2000; Benson et al. 2000; Berlind & Weinberg 2002; Kravtsov et al. 2004). I adopt the HOD functional form presented in Zheng et al. (2007). The HOD parameters were calibrated to reproduce the clustering properties of the BOSS CMASS

DR9 galaxy sample for the same cosmology, as in Manera et al. (2013).

For each HOD catalogue I also construct a redshift-space analog by shifting the positions of galaxies along the line of sight with Eq. (2.1) (taken to be the z-axis of the simulation) using their peculiar velocities. In the remainder of the text, I will refer to each of these 300 simulations populated with HOD galaxies as mock realisations.

3.1.2 Splitting densities with mock galaxies

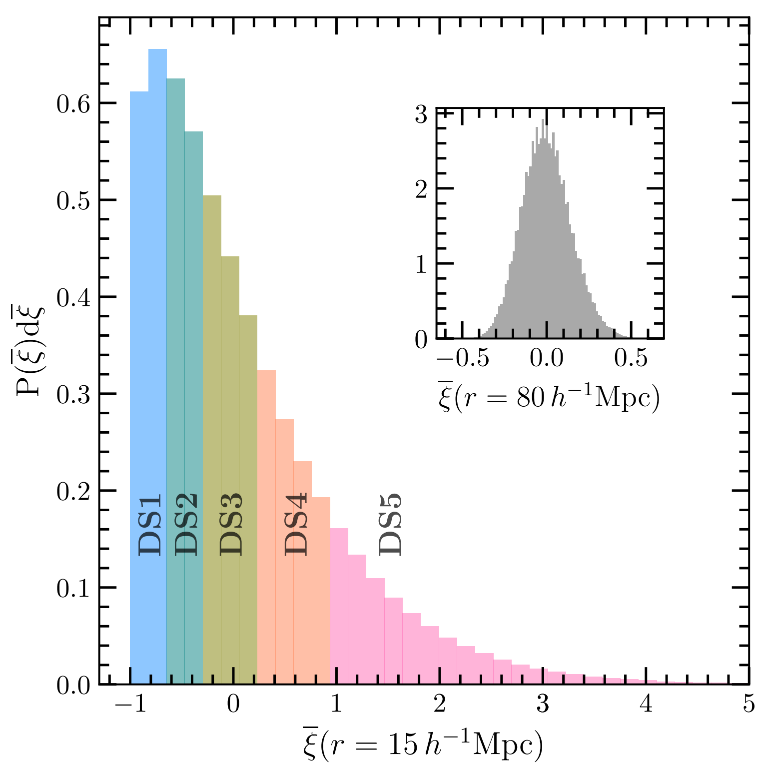

Figure 3.1: The probability distribution for the integrated galaxy number density $\bar{\xi}$, smoothed by a top-hat window function of $15\,h^{-1}$Mpc. The different colors delimit the split of the PDF into different quintiles, according to the value of $\bar{\xi}$, ranging from underdensities (DS1-2) to overdensities (DS4-5). The positions in each quintile will be cross-correlated with the entire galaxy field in redshift space. These are the main observables I will employ in this thesis. The inner subplot shows the PDF at a scale of $80\,h^{-1}$Mpc.

For each mock realisation, I place 1.3×10^6 (which is roughly the same as the number of mock galaxies) random points in the cosmological volume of the simulation, and measure the real-space integrated galaxy number densities $\bar{\xi}$ within spheres of radius $r = 15\,h^{-1}$Mpc centred on those points. I then rank the centres according to the value of $\overline{\xi}(r = 15\ h^{-1}\text{Mpc})$ in increasing order, and split them into 5 bins with equal number of centres, that I refer to as *Density Split* quintiles, labelling them as DS1 to DS5 from low to high density. All of this is done using the DENSITY SPLIT package that can be found in my GITHUB repository[1]. The coloured histogram in Fig. 3.1 shows the distribution of $\overline{\xi}(r = 15\ h^{-1}\text{Mpc})$ around the random positions in one of the mock realisations. DS1 is the lowest density quintile and has a negative $\overline{\xi}$, corresponding to voids. DS2-3 are also underdense, although their density contrasts are shallower than voids. DS4 marks the transition towards overdense regions, having a positive $\overline{\xi}$. DS5 has the largest positive amplitude, corresponding to cluster-like regions. As has been shown by previous studies, the full PDF of $\overline{\xi}(r = 15\,h^{-1}\text{Mpc})$ is highly non-Gaussian, with its peak corresponding to a negative density, i.e. a major fraction of the volume is in voids, and a high-density tail that extends to large positive values. It follows closely a lognormal distribution (e.g. Coles & Jones, 1991; Colombi, 1994; Bernardeau & Kofman, 1995; Scoccimarro, 2004; Uhlemann et al., 2016; Repp & Szapudi, 2018; Einasto et al., 2020). These features are expected, and are the result of the growth of structures via gravitational evolution. While initial overdensities collapsed due to gravitational attraction into small and highly overdense structures, initially underdense regions expanded and ended up covering a large volume, but could only reach moderate underdensities, as the density contrast is bound from below (i.e. $\delta_m = -1$). The PDF and its evolution in the non-linear regime follow spherical dynamics closely, and can be predicted accurately (Uhlemann et al., 2016; Repp & Szapudi, 2018; Jamieson & Loverde, 2020). On large scales where the density evolution is linear, the density PDF is expected to be Gaussian. An example is shown by the inset of Fig. 3.1. At the smoothing scale

[1] https://github.com/epaillas/densitysplit

of $r = 80\,h^{-1}$Mpc, the distribution is much more symmetric and can be well fitted by a Gaussian PDF.

3.1.3 Velocity distribution functions in different environments

With the above setup, I can measure the velocity distribution function at different scales[2].

On large scales, the PDFs of these radial velocities are expected to be Gaussian, but it is the most interesting at small scales where non-linear evolution is prominent. An example at $r = 15\,h^{-1}$Mpc is presented in Fig. 3.2. We can see that the PDFs for each quintile are well fitted by a Gaussian function of different widths (DS1-5 of Fig. 3.2). The mean of the Gaussian, $\bar{v}_r$, is usually offset from zero. It monotonically decreases from DS1 ($\bar{v}_r = 131.2$ km/s) to DS5 ($\bar{v}_r = -63.9$ km/s). This indicates that the radial velocity is turning from outflow at DS1 to infall at DS5. This is expected, because the local density increases from DS1 to DS5 i.e. DS1 is underdense, similar to the case of voids; DS5 is overdense, corresponding to clusters; and DS2-4 are their transition phases. This is clearly seen in the full radial velocity profiles around each density quintile Fig. 3.3 (data points), where the outflow for DS1-2, and the infall for DS5 continue up to large radii.

In contrary, at the same scale, the full pairwise velocity distribution, the quantity relevant for modelling the redshift-space 2PCF, is highly skewed (bottom-right panel of Fig. 3.2). The mean pairwise velocity is also highly negative, indicating that the pairs are predominantly sampling the high density regions. This is expected, as pair-counting at this small scale is heavily weighted by high density regions. This

[2] The relevant quantity that enters the streaming model equation is the line-of-sight velocity distribution, which receives contributions from both the radial and tangential components, and that is what I have used to plug into the GSM. For this part of the thesis, I choose to show the velocity PDFs for the radial component to make them better correspond to the streaming velocity profiles that I will show later. I have checked that line-of-sight velocity PDFs are qualitatively similar to the PDFs of the radial components.

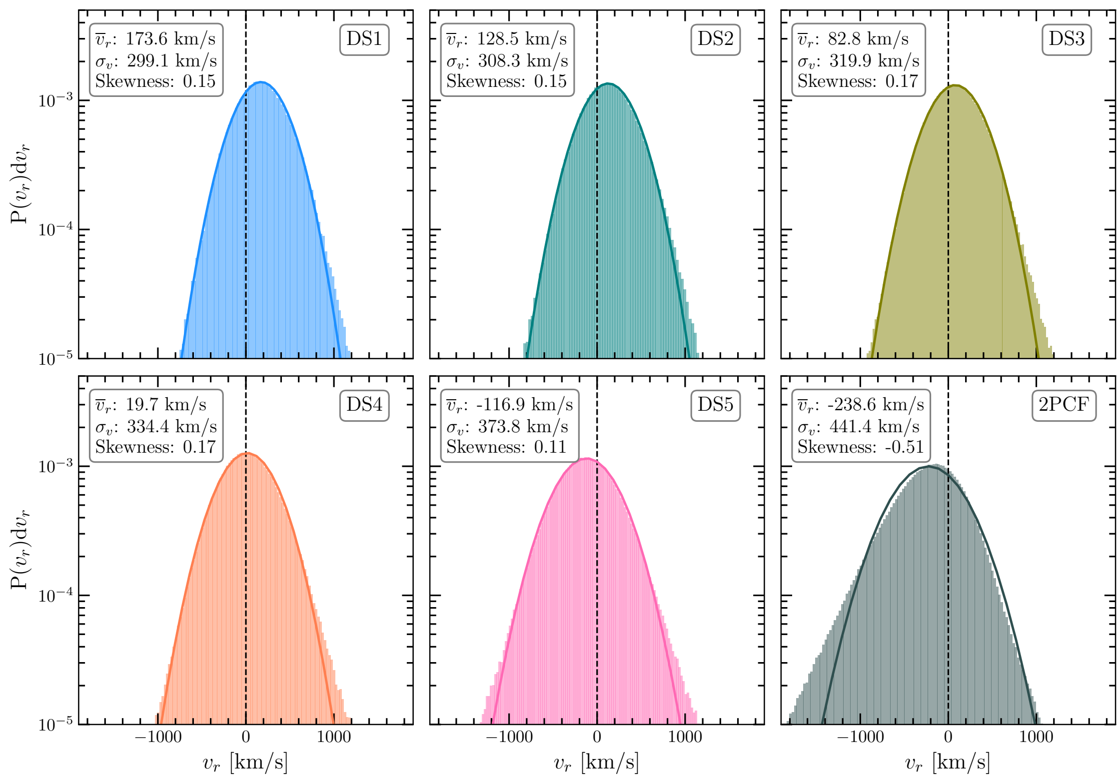

Figure 3.2: Distributions of radial velocities in density splitting (DS) quintiles and pairwise velocities for the two-point correlation function (2PCF) (different panels as labelled), measured from one of the mock realisations, at a scale of $r = 15\,h^{-1}\mathrm{Mpc}$. For DS, the colours and legends (DS1-5) are matched to those in Fig. 3.1 to represent densities of different depths, varying from voids (DS1) to clusters (DS5). The mean $\bar{v}_r$, standard deviation σ_v and skewness of the distributions are shown on the upper left corner of each panel. The skewness is the third standardised moment of a distribution, characterising the asymmetry of the distribution about its mean. The solid lines show a Gaussian distribution with the same mean and standard deviation as the data, after applying 3σ clippings. The PDFs for velocities in each DS are well-fit by a Gaussian profile, with their mean $\bar{v}_r$ varying from positive in DS1 to negative in DS5, corresponding to outflow and infall around those quintiles, respectively. The PDF for the pairwise velocities (2PCF) is significantly skewed towards negative values, strongly deviating from the Gaussian distribution.

leads to a large velocity dispersion as the dominant feature of the pairwise velocity distribution, which manifests itself as Fingers-of-God (FoG, Jackson, 1972) at those scales in the redshift-space 2PCF. This is naturally avoided in DS, as we will see in the next sub-section.

The non-Gaussianity of the PDF for the pairwise velocity violates the Gaussian

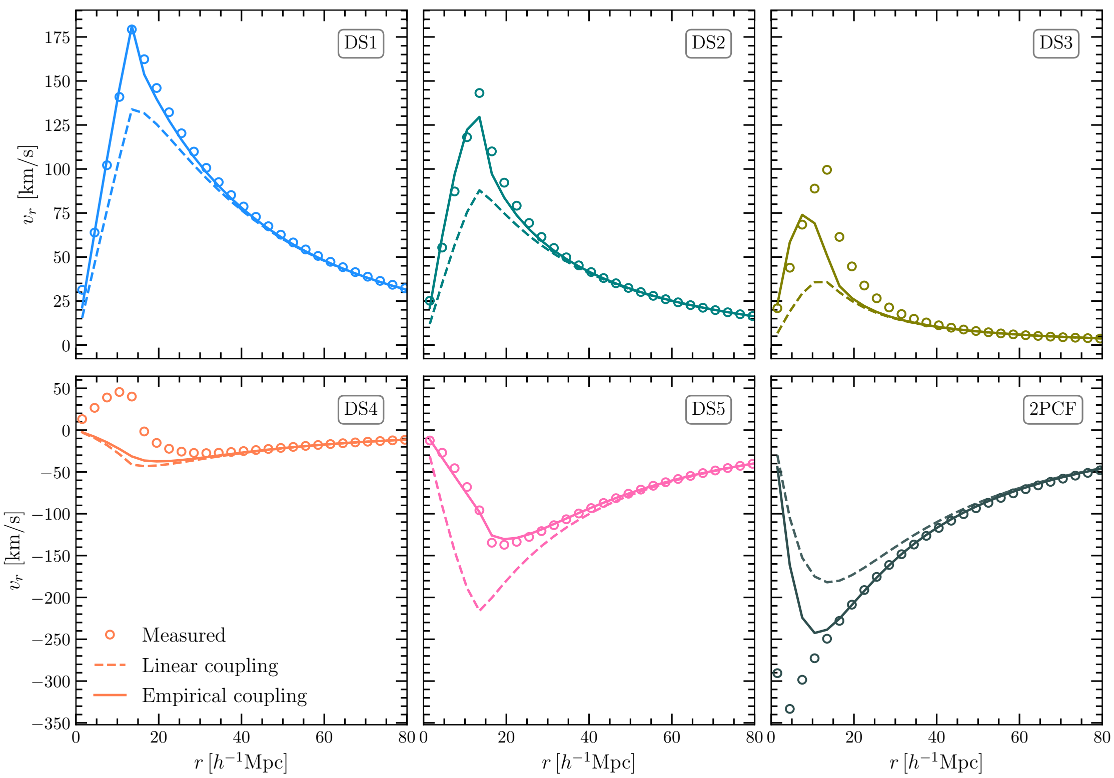

Figure 3.3: Galaxy radial velocity profiles around regions corresponding to different density quintiles (DS1-5) and the pairwise radial velocity profiles (2PCF), as indicated in each panel. The circles show measured velocities from the simulations. The dashed lines show the predictions from linear theory (Eqs. 2.5 & 2.9), whereas the solid lines show the best-fit results from our empirical model (Eqs. 2.7 & 2.10) using the redshift-space galaxy correlation functions as observables. The linear coupling model fails to reproduce the observed velocities on small scales for all cases, where non-Gaussianity becomes important. The empirical model accurately captures density-velocity velocity coupling for DS1 and DS5 (voids and clusters), but it underestimates the velocity profiles for intermediate quintiles and for the pairwise velocity profiles at small scales.

assumption necessary for modelling the 2PCF with the GSM. It is therefore unsurprising that the model becomes inaccurate for the 2PCF at small scales (e.g. Reid & White, 2011). The Gaussian nature of the velocity PDFs for DS1-DS5 ensures that such key condition for the GSM is met, thus promises a better performance, as I will show in the next sub-section.

Perhaps more importantly, by splitting the non-Gaussian density PDF, and cross-correlating local densities with the entire galaxy field, we will naturally capture the non-Gaussianity of the density field. This lays the foundation for a possible gain of cosmological information over standard two-point statistics, as I will demonstrate in Sec. 3.2.

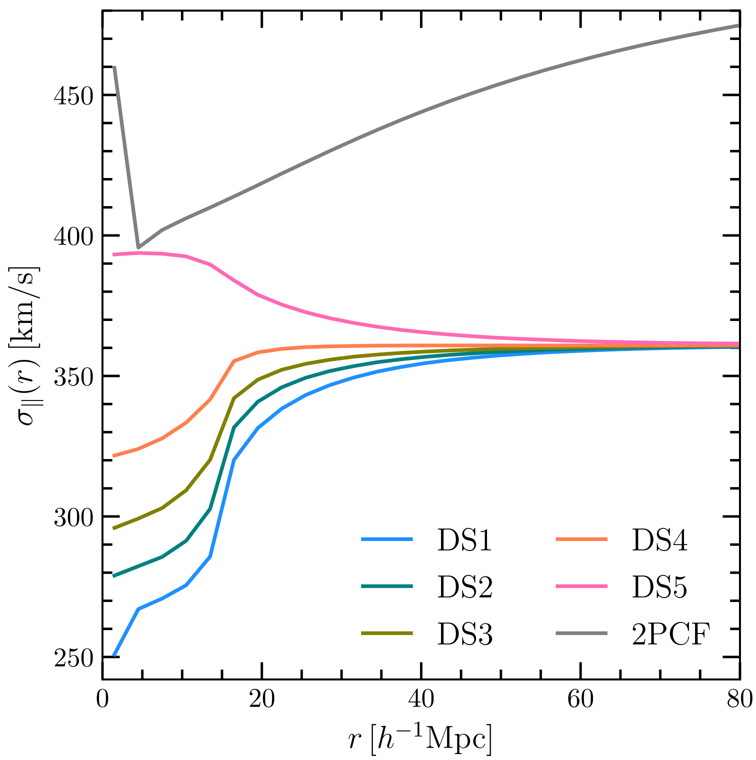

Figure 3.4: The line-of-sight velocity dispersion as a function of r (after integrating over the μ dependence). These are direct measurements from the mock galaxies. Different colours correspond to the different DS quintiles and to the pairwise galaxy velocities, as indicated in the legend.

For completeness, I show the velocity dispersion profiles in Fig. 3.4. Note that $\sigma_\parallel(r,\mu)$ is necessarily a function of r and μ. Its variation with μ is relatively strong for the pairwise velocities in the 2PCF and should not be neglected for the modelling. After averaging over μ, all the $\sigma_\parallel$'s in DS appear to be flattened on large scales and converge to the same value. This offers possibilities to use a single free parameter to capture their amplitudes on large scales (see Sec. 3.2). The dispersion decreases with scale for DS1-4 as these are underdensities at small scales. Such a trend is the opposite for DS5, which corresponds to over-densities. For the 2PCF, the dispersion decreases with scale until the 1-halo term kicks in at the smallest scales ($\sim 1\,h^{-1}$Mpc), which causes the velocity dispersion to increase sharply. This is the main drive for the FoG for the 2PCF. The velocity distribution function is expected to be highly non-Gaussian, as already seen at $r = 15\ h^{-1}$Mpc in Fig. 3.2.

3.1.4 Performance of the GSM for splitting densities & the 2PCF

With the promising behavior of the velocity PDFs in the different DS quintiles, I will test the performance of the GSM in the simulation in this section. At this stage, I will take all the ingredients of the model [i.e. $\xi(r|\Delta^i)$, $v_r(r|\Delta^i)$, $\sigma_{\parallel}(r, \mu, |\Delta^i)$ for DS and $\xi(r)$, $v_r(r)$, $\sigma_{\parallel}(r, \mu)$ for the 2PCF] from our mock galaxies. This allows us to focus on testing the validity of the Gaussian assumption. I will relax some of these model conditions in the next section for cosmological constraints. All of the clustering measurements are carried out with the CONTRAST package that is available on my GITHUB repository[3].

In simulations with periodic boundary conditions, the cross-correlation functions can be estimated without the use of a random catalogue as

$$\xi(s, \mu|\Delta^i) = \frac{D_1 D_2(s, \mu)}{N_1 N_2} \frac{V_{\text{box}}}{\delta V(s, \mu)} - 1, \tag{3.1}$$

where $D_1 D_2(s, \mu)$ are the pair counts between DS centres and galaxies, N_1 and N_2 are the total number of DS centres and galaxies, respectively, V_{box} is the volume of the simulation box, and $\delta V(s, \mu)$ is the volume of a bin centred on (s, μ) with radial thickness ds, which can be calculated analytically as

$$\delta V = \frac{4\pi}{3} \frac{(s + \mathrm{d}s/2)^3 - (s - \mathrm{d}s/2)^3}{N_\mu}, \tag{3.2}$$

[3] https://github.com/epaillas/contrast

where N_μ is the total number of μ bins. For each DS quintile, I measure the real-space quantities $\xi(r|\Delta^i)$, $v_r(r|\Delta^i)$, $\sigma_\parallel(r,\mu,|\Delta^i)$, as well as the redshift-space CCF $\xi^s(s,\mu|\Delta^i)$ from our mock galaxies, taking the average over the 300 simulation boxes. I use radial bins of $3\,h^{-1}$Mpc, $9\,h^{-1}$Mpc and $6\,h^{-1}$Mpc widths at scales of $0-39\,h^{-1}$Mpc, $39-84\,h^{-1}$Mpc and $84-150\,h^{-1}$Mpc, respectively, as well as μ bins of a constant width of 0.02 between -1 and 1. I try to minimise the number of radial bins while being able to sample important features in the correlation function such as the BAO. This is necessary to make sure that the data vector is sufficiently smaller than number of independent mocks used to construct the covariance matrix (see also the discussions about covariance matrix in Sec. 5). I plug the model ingredients into Eq. (2.3) to obtain the predicted redshift-space CCF. These are compared with direct measurements in redshift space in Figs. 3.5 & 3.6. The CCF for DS1 resembles the void-galaxy CCF (see e.g. Paz et al., 2013; Paillas et al., 2017; Correa et al., 2019). This is not a surprise, since the algorithms designed for identification of spherical underdensities in the cited works use a top-hat filter for the definition of the voids, in a similar fashion as I have done by using a top-hat filter for the density splitting. DS5 shows an overdensity of galaxies near the centre, similar to the case of cluster-galaxy CCF (e.g. Seldner & Peebles, 1977; Lilje & Efstathiou, 1988, 1989; Croft et al., 1999; Yang et al., 2005; Zu & Weinberg, 2013; Mohammad et al., 2016).

It is striking to see that despite the big variation of distortion patterns for DS across different density quintiles and scales, the agreement between the model (dashed lines) and simulation (solid lines) is nearly perfect at all scales. In contrary, for the 2PCF, deviations between the model and simulation are obvious even at scales of a few tens of h^{-1}Mpc when μ is near unity, i.e. along the line of sight direction (Fig. 3.6) .

The distortion patterns are similar between the CCFs and the 2PCF on large scales, where both appear to be flattened. At scales $s \simeq 30\ h^{-1}$Mpc, however, the CCF becomes elongated for DS1 and DS2, while the 2PCF remains flattened at

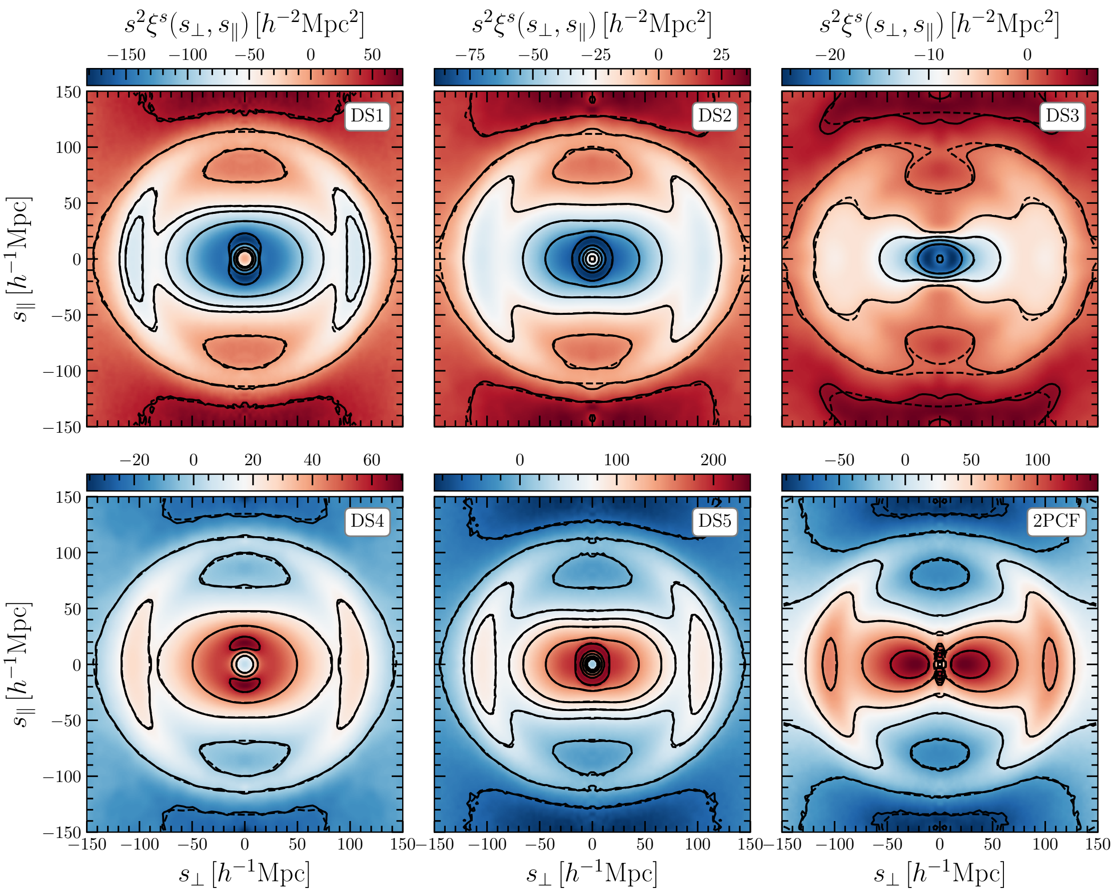

Figure 3.5: Cross-correlations between centres of density quintiles in DS with the entire galaxy sample in redshift space (labels as DS1-5), and the galaxy two point correlation function (2PCF), averaged over the 300 mock galaxy catalogues. The colourbars indicate the clustering amplitude, scaled by a factor of s^2 to highlight large-scale features. The solid lines show contours of constant amplitude in the simulation. The predictions from the Gaussian streaming model (Eq. 2.4), with all its ingredients measured from simulations, are shown in dashed lines. Contours for DS quantiles 2, 3 and 4 have been smoothed using a Gaussian kernel with a Gaussian σ of 1, 2 and 1 pixels, respectively.

the same scale, with the exception of the FoG feature near $\mu = 1$, which causes deviations between the model and the simulation results. At even smaller scales, the distortion of CCF for DS1 and DS5 becomes very weak, while the 2PCF is dominated by the elongated FoG. The GSM clearly fails to capture the FoG feature (bottom-right panel of Fig. 3.6). This again indicates that the dispersion is insufficient to describe the non-Gaussian PDF of the pairwise velocities. In contrast, there is no

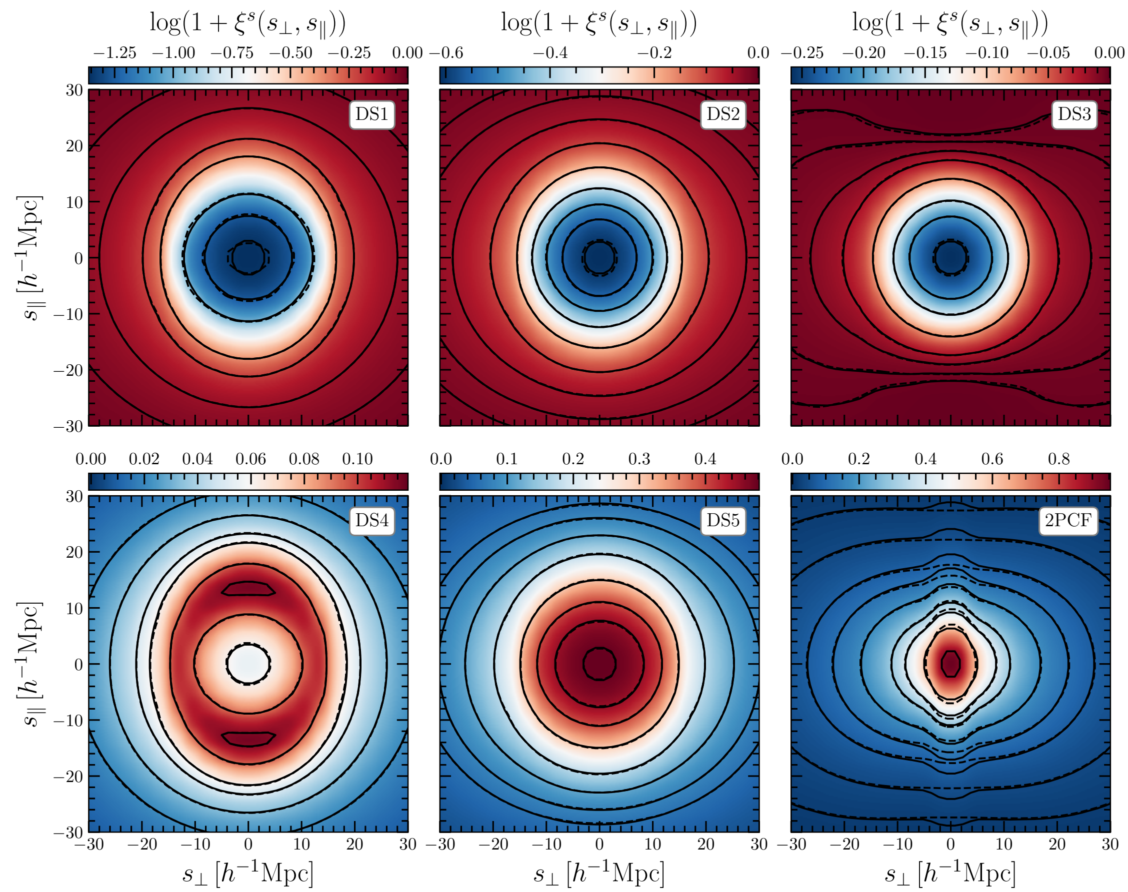

Figure 3.6: Similar to Fig. 3.5, but showing clustering at small scales. The Gaussian streaming model (in dashed contours) lays on top of the measurements from mock galaxies (solid contours) for CCFs from all density quintiles (DS1-5). All the CCFs show no sign of the Fingers-of-God effect. Deviations for the 2PCF from the same model can be seen close to the line-of-sight direction, mainly due to the strong Fingers-of-God effect. At these relatively small scales, linear theory is not as accurate, but its prediction agrees reasonably well for DS1 and DS2 (underdense environments), as I have checked.

FoG feature in the CCF of the different DS quintiles, and this allows the GSM to perform better at the same scale. This is expected, as the centres of the CCFs do not usually correspond to a galaxy. The selection of centres based on the top-hat smoothed density field at 15 h^{-1}Mpc naturally averages out the FoG. This is also consistent with the fact that the velocity PDF for each quintile of DS is very close to Gaussian (Fig. 3.2).

To characterize the performance of the GSM more quantitatively, I extracted the monopole, quadrupole, and hexadecapole [$\xi_{0,2,4}(s)$] of the correlation functions inferred from the simulations following

$$\xi_\ell(s) \equiv \frac{2\ell+1}{2}\int_{-1}^{1} L_\ell(\mu)\xi^s(s,\mu)\mathrm{d}\mu \ , \tag{3.3}$$

where $L_\ell(\mu)$ is the Legendre polynomial of order ℓ. Fig. 3.7 shows a comparison of the simulation results (symbols) and the corresponding model predictions (solid lines). The agreement for $\xi_{0,2,4}(s)$ is excellent at all scales for the DS results, with small deviations near the smoothing scale where the slopes of the multipoles are steep. Instead, the multipoles of the standard 2PCF show noticeable deviations from the theory predictions at scales below 20 h^{-1}Mpc. This comparison becomes clearer when quantifying the agreement by taking the fractional difference between the model predictions and the simulations in terms of the 1-σ errors corresponding to a volume of $(1.5h^{-1}$ Gpc$)^3$. We can see that the ratios at around ~ 15 h^{-1}Mpc for DS may sometimes fluctuate beyond the 1-σ level, but overall stay within the errors. For 2PCF however, the differences between the model and the results from the simulations are noticeable at scales $s < 30\,h^{-1}$Mpc for ξ_2, and become stronger at smaller scales.

These results for the 2PCF are similar to those reported in Reid & White (2011) with similar mocks, where a few percent accuracy was achieved at ~ 25 h^{-1}Mpc for the quadrupole (see also Chen et al., 2020). The performance of the GSM is better for most density quintiles, and for almost all scales. The exceptions are: 1. for the quadrupole in DS1 and DS5 at $s \sim 15$ h^{-1}Mpc (solid blue lines in the lower panels corresponding to DS1 and DS5 in Fig. 3.7). This is likely due to taking ratios with respect to $\xi_2(s)$ which crosses zero at those scales; 2. the monopole in DS1 also fluctuates outside the 1σ error at $s < 20$ h^{-1}Mpc (solid orange line at bottom panel corresponding to DS1). This is possibly due to the sparsity of galaxies for DS1 at

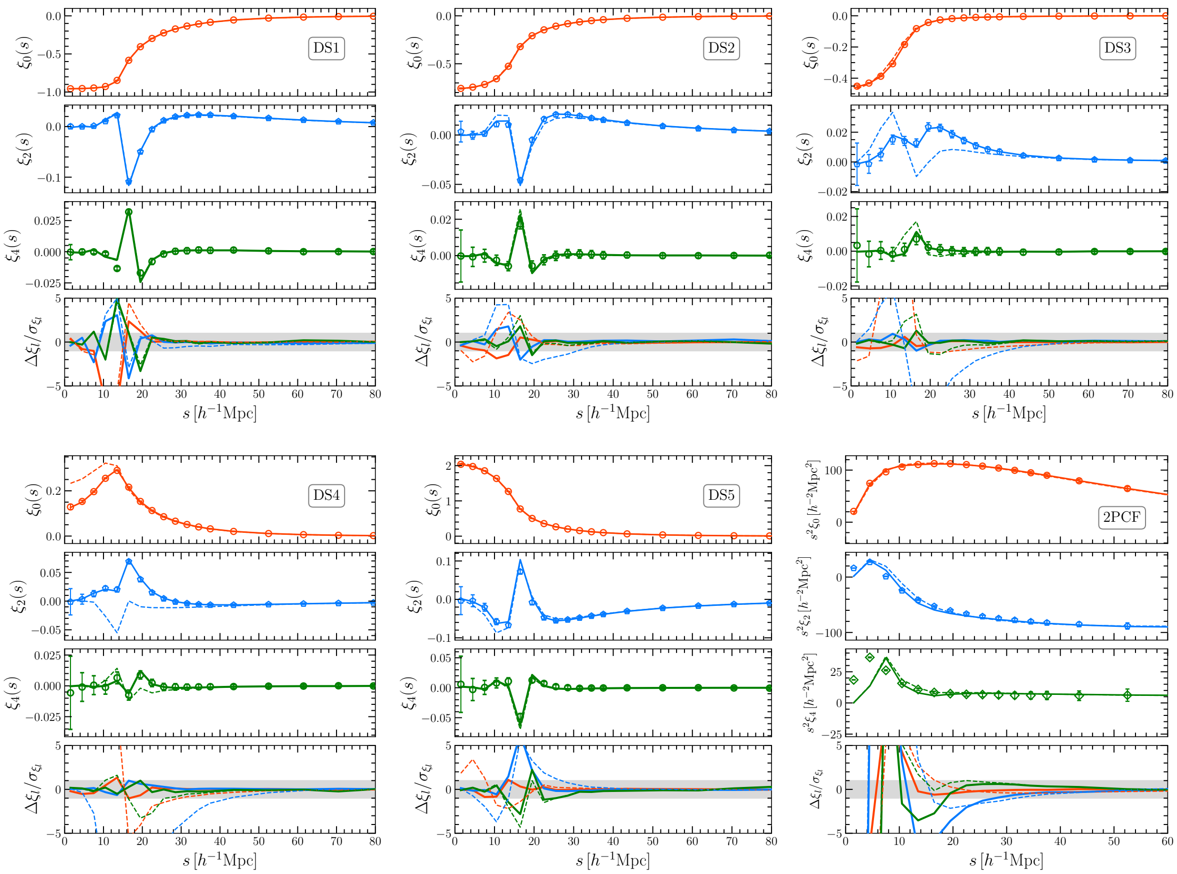

Figure 3.7: Comparison of monopoles (ξ_0, orange), quadrupoles (ξ_2, blue) and hexadecapole (ξ_4, green) between models (lines) and simulations (data points with errors). The first five main panels (DS1-5) show the CCF multipoles of each DS quintile, whereas the last panel at the bottom right (labeled 2PCF) shows the multipoles for the galaxy two point correlation function- Solid lines represent model predictions with the Gaussian streaming model with all its ingredients measured from simulation. Dashed lines shows our best-fit model with the free parameters that account for the density-velocity coupling and velocity dispersion. The error bars represent 1σ standard deviations for 1 simulation box of $(1.5\,h^{-1}\mathrm{Gpc})^3$ from the 300 mock realisations. The bottom sub-panels show the deviation between the model and the measurements from simulations, in units of the dispersion in the simulation. The grey-shaded area shows the 1σ region. The amplitudes of the galaxy 2PCF multipoles have been scaled by a factor of s^2 for clarity.

those scales.

As can be seen in Fig. 3.5, the BAO feature is clearly present in the CCF of all quintiles at $s \sim 105\ h^{-1}\mathrm{Mpc}$, especially when $\mu \sim 0$, i.e. the direction perpendicular to the line of sight. To better visualise this large-scale feature, Fig. 3.8 shows the

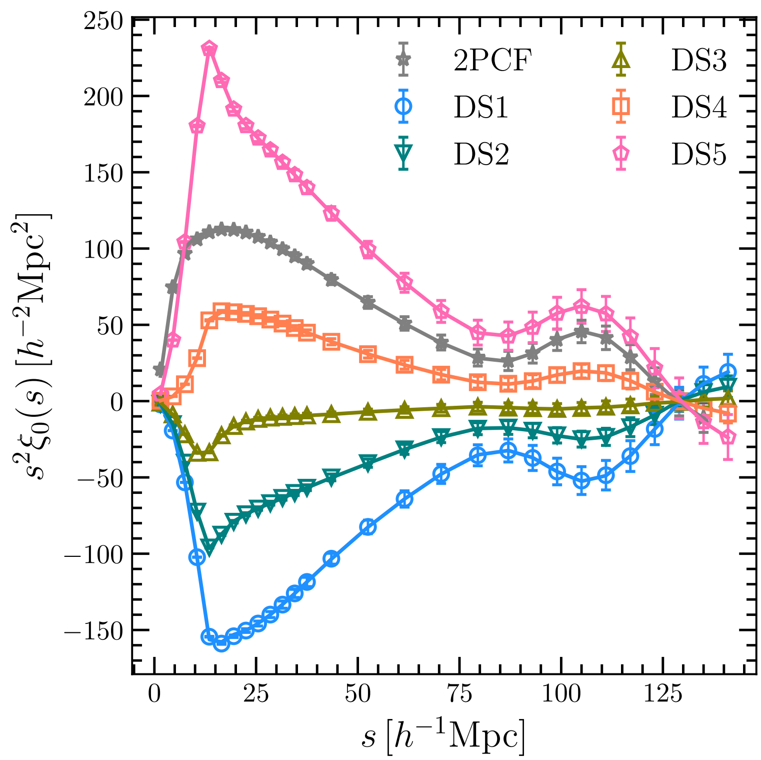

Figure 3.8: Monopoles $s^2\xi_0$ of the redshift-space cross-correlation functions between positions in different density quintiles (labeled as DS1-5) and the entire galaxy sample. The BAO feature is seen as prominent peaks or dips around $\sim 105\,h^{-1}$Mpc for all quintiles. The case for the two-point correlation function is labeled 2PCF. Data points with errors are measurements from simulations. Solid lines are predictions from the Gaussian streaming model, with all its ingredients measured from the simulation (see also Fig 3.5 for their 2D versions).

monopoles of the CCFs of all DS and the standard 2PCF, rescaled by s^2. The BAO feature is clearly seen for all quintiles as a peak (DS4-5) or dip (DS1-3), with an amplitude proportional to that of the overall CCF. As we will see later, accessing this information significantly improves the constraints on the geometrical (AP) distortion parameters. While the information from the BAO feature has been commonly used in RSD analyses of the galaxy 2PCF (e.g., Sánchez et al. 2017a; Beutler et al. 2017; Bautista et al. 2021), it has often been ignored in void RSD studies (but see Zhao et al., 2020, for a demonstration of how voids can help to optimise the BAO measurement in a galaxy sample). This is partly due to the fact that many authors choose to re-scale the void-galaxy profiles by the void radius (e.g. Cai et al., 2016; Hamaus et al., 2020). This rescaling helps to boost the signal from the void ridge, which is useful for gravitational lensing analyses around voids (e.g. Higuchi et al., 2013; Krause et al., 2013; Melchior et al., 2014; Clampitt & Jain, 2015; Cautun et al., 2018; Cai et al., 2017; Sánchez et al., 2017b; Raghunathan et al., 2020). It is also convenient when applying the Alcock-Paczynski test for voids (Sutter et al., 2012). However, it effectively washes out any BAO signal present on large scales.

3.1.5 Other RSD models suitable for split densities

The modelling of RSD around different density quantiles has no fundamental difference from the modelling of the void-galaxy or cluster-galaxy CCFs (Lilje & Lahav, 1991; Croft et al., 1999; Zu & Weinberg, 2013; Mohammad et al., 2016). Hamaus et al. (2015) employed the Gaussian streaming model for modelling RSD around voids and, together with the assumption that the density and streaming velocity are linearly coupled, they found that the model is biased at the $\sim$10 percent level for both AP and growth parameters, but their application of the model to BOSS CMASS mock galaxies and real data seems to be consistent with their fiducial cosmology (Hamaus et al., 2016). Accounting only for the streaming velocity, Cai et al. (2016) derived the full expression for RSD around voids, and for spherical density profiles in general (Eqs. 1-4 of their paper), which I summarise here [4]:

$$1 + \xi^s(\mathbf{s}) = [1 + \xi^r(r)] \left[1 + (1-\mu^2)\frac{1}{aH}\frac{v(r)}{r} + \mu^2 \frac{1}{aH}\frac{\partial v}{\partial r}\right]^{-1} . \tag{3.4}$$

[4](Cai et al., 2016) noted ξ as δ because the cross-correlation function is indeed the same as the stacked density profile, which is usually written as $\delta(r)$. There was a typo where they should have noted the redshift-space coordinate as $\mathbf{s}$, rather than $\mathbf{r}$, which was pointed out in Nadathur et al. (2019b).

The above can be applied to map between the real and redshift-space CCFs around spherical regions. When expanding to the linear order, we have

$$\xi^s(\mathbf{s}) = \xi(r) - (1-\mu^2)\frac{1}{aH}\frac{v(r)}{r} - \mu^2\frac{1}{aH}\frac{\partial v}{\partial r}. \tag{3.5}$$

This takes the same form as expanding the Gaussian streaming model to the linear order, keeping only the first derivative, i.e. neglecting terms that are related to the velocity dispersion [e.g. Page 289 of Peebles (1980), Eqs. 24-25 of Fisher (1995), Eqs. 16-17 of Reid & White (2011)]. When further assuming linear coupling between the density and velocity, i.e. Eq. (2.9), we have

$$\xi^s(\mathbf{s}) = \xi(r) + \frac{1}{3}f\bar{\xi}(r) + f\mu^2[\xi(r) - \bar{\xi}(r)]. \tag{3.6}$$

This is the linear theory expression for CCFs (Kaiser, 1987; Cai et al., 2016)[5].

All the above expressions account for only the streaming velocity. To include velocity dispersion, Nadathur et al. (2019b) convolve Eq. (3.4) with the velocity dispersion as

$$1 + \xi^s(s_\perp, s_\parallel) = \int \left\{1 + \xi^s\left[s_\perp, s_\parallel - v_\parallel/(aH)\right]\right\} \mathcal{P}(v_\parallel, \mathbf{r})\mathrm{d}v_\parallel \ , \tag{3.7}$$

where $\xi^s(\mathbf{s})$ takes the expression from Eq. (3.4), and $\mathcal{P}(v_\parallel, \mathbf{r})$ is assumed to be Gaus-

[5]Note that, although Nadathur et al. (2019b) assume the linear coupling of Eq. (2.9), they expand Eq. (3.4) and keep terms such as $\xi\bar{\xi}$ and $\xi\xi$ ($\xi\Delta$ and $\xi\delta$ in their notation). This to us is 2nd order in mathematical term, and is no longer a linear model.

It was also noted in Nadathur et al. (2019b) and Nadathur et al. (2020) that the GSM does not reduce to linear theory at the limit when the dispersion is small. However, it has been shown in Fisher (1995) that the GSM naturally reduces to linear theory when allowing scale-dependent velocity dispersion. I have also verified numerically that when $\sigma_\parallel$ is very small, the GSM reduces to the case where there is only streaming velocity, i.e. Eq (3.4). Further, if a linear coupling between the density and velocity is assumed, i.e. a pure linear system, then the GSM does get back to it.

sian with a zero mean, i.e.

$$\mathcal{P}(v_\parallel, r, \mu) = \frac{1}{\sqrt{2\pi}\sigma_\parallel(r,\mu)} \exp\left[-\frac{v_\parallel^2}{2\sigma_\parallel^2(r,\mu)}\right] . \tag{3.8}$$

This is similar to the well-known Gaussian streaming model (i.e. Eq. 2.3), except that the streaming velocity is explicitly taken off from the exponential part, and being left to be accounted for by the Jacobian, i.e. the mapping from real to redshift-space CCF in Eq. (3.4).[6]

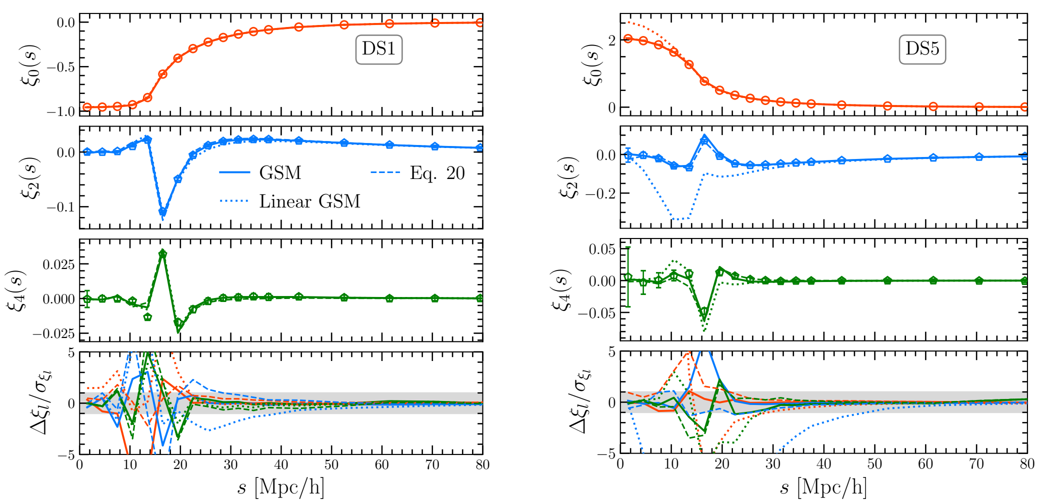

Figure 3.9: Similar to Fig. 3.7, showing multipoles for the cross-correlation functions in the extreme DS quintiles (voids on the left and clusters on the right-hand side panels, respectively). The data points and errors are measurements from the mock galaxy catalogues. The solid and dashed lines show predictions from the Gaussian streaming model (Eq. 2.3) and Eq. (3.7) (with Eqs. 3.4 & 3.8), where all model ingredients are fully known. The dotted lines show predictions from the Gaussian streaming model but with the linear density-velocity coupling assumption (Eq. 2.9).

I used the simulations to compare the performance of the this prescription (i.e. Eq. 3.7 with Eqs. 3.4 & 3.8) with the GSM for DS, taking all the ingredients of the model from our mock galaxies, i.e. $\xi(r|\Delta^i)$, $v_r(r|\Delta^i)$ and $\sigma_\parallel(r,\mu|\Delta^i)$. Fig. 3.9 presents

[6] We notice that the μ-dependence was missed out in the expression of $\mathcal{P}(v_\parallel)$ in (Nadathur et al., 2019b, 2020). Without the explicit μ-dependence for the dispersion term, the impact of $\sigma_\parallel$ is slightly weakened.

this comparison for DS1 and DS5. We can see that, although small deviations from the quadrupole can be spotted between 20-30 h^{-1}Mpc, and from the measured monopole in DS5 below 10 h^{-1}Mpc, their overall predictions are very similar across all scales for both DS1 and DS5.

Note that the solid and dashed lines are the most optimistic scenario where all the ingredients of the model are assumed to be known. Many studies have assumed linear coupling between the density and velocity (Eq. 2.9) (e.g. Hamaus et al., 2016; Hamaus et al., 2017, 2020; Nadathur & Percival, 2019; Nadathur et al., 2020). To test the accuracy of this assumption, I show in dotted lines the GSM with this approximation, labeled as 'Linear GSM'. We can see that for DS1, Linear GSM shows larger deviation than GSM at small scales for the monopole, and slightly underpredicts the quadrupole at intermediate scales, but the overall agreement with the simulation is fairly good. This is expected, as DS1 is similar to the void-galaxy cross-correlation function, where non-linearity around voids is thought to be weaker. This does justify the linear coupling assumption for voids so far, as has been adopted in the literature, but at the per cent-level precision promised by future surveys, it is important to model this coupling relationship more accurately.

For DS5, however, the deviations between the linear models and the measurements are much larger at a few tens of h^{-1}Mpc. This is because the density contrast for DS5 is larger, similar to the cluster-galaxy CCF, thus the growth of structure is expected to be more non-linear. The failure of the linear coupling model is also evident from Fig. 3.3, where the linear velocity profiles over predict the version from the simulations at small scales. It is also worth nothing that even though DS2-4 are transition stages between voids and clusters, thus having more moderate density constrasts, neither the linear nor the empirical density-velocity couplings are sufficient to correctly describe their radial velocity profiles in Fig. 3.3. After having calculated the radial velocity profiles directly from the dark matter distribution, I have found that the velocities sampled by galaxies around DS centres are biased against that of dark matter at the level of tens of km/s. I believe that this is due to the sparsity of our galaxy samples, as a galaxy number density that is too low might result in a misestimation of the measured velocities. While this velocity bias is sub-dominant for DS1 and DS5, it becomes relatively more severe for DS3 and DS4, which have smaller velocity magnitudes at small scales. A galaxy sample with a higher number density might be able to reduce this velocity bias and help reconcile the empirical model with the measurements. Generally speaking, I believe that there is substantial room for improving the prediction of velocity statistics for DS. In a sense, the free parameters from the empirical model I introduced are absorbing some of the uncertainties due to the assumptions in the modelling (e.g. the linear bias assumption). It is likely that many of the developments for the calculation of pairwise velocity statistics for the 2PCF (e.g. Vlah et al., 2016; Chen et al., 2020) will also be useful for DS.

In summary, I have shown that:

1. given the same conditions, and assuming that the model ingredients are fully known, the GSM and Eq. (3.7) lead to almost equivalent predictions. These results differs from what was reported in Nadathur et al. (2019b, 2020) for voids, where significant deviations for the GSM model were reported.
2. the assumption of a linear coupling between density and velocity is not accurate for the statistical errors of our concern, i.e. a volume of $(1.5h^{-1}\ \mathrm{Gpc})^3$.

I therefore will use the GSM model as default for our analysis. Eq. (3.7) should perform similarly, as demonstrated in this section.

3.2 Constraining cosmology with density split RSD

I have demonstrated in the previous section that the GSM works well for all CCFs in DS, and better than for the standard galaxy 2PCF. The main reason is that the key condition for the model – the Gaussianity of the density and velocity field, is not valid at small scales for the 2PCF. This implies that the 2PCF is insufficient to capture all the information in this non-Gaussian regime. With DS, however, the Gaussianity condition continues to be valid. It is compelling to ask if the combination of the whole series of cross-correlations, i.e. $\xi(s, \mu|\Delta^{1,2,3,4,5})$, contains any more cosmological information than the 2PCF, $\xi(s, \mu)$. Using the same model (GSM) for both the DS and 2PCF sets a common ground for us to make a fair comparison for the cosmological information contained in these statistics.

Three ingredients are needed for extracting cosmological information with the GSM using redshift-space auto- and cross-correlation functions:

- The real-space two-point correlation function $\xi(r)$; or the conditioned correlation functions $\xi(r|\Delta^i)$.
- The real-space pairwise velocity profile $v_r(r)$; or conditioned (stacked) velocity profiles $v_r(r|\Delta^i)$.
- The real-space pairwise velocity dispersion $\sigma_{\parallel}(r, \mu)$; or the conditioned (stacked) velocity dispersions $\sigma_{\parallel}(r, \mu|\Delta^i)$.

Significant efforts have been made to predict these three profiles. Among them, the real-space 2PCF $\xi(r)$ is the key. It can be predicted with the halo model, perturbation theory, or emulators with simulations (e.g. Peacock & Smith, 2000; Seljak, 2000; Zentner et al., 2005; Angulo & White, 2010; Zhai et al., 2019). Once $\xi(r)$ is known, it can be used to compute $\xi(r|\Delta^i)$ if the PDF of Δ is known. There are existing model frameworks that allow us to do this (Abbas & Sheth, 2005; Friedrich et al., 2018; Gruen et al., 2016; Neyrinck et al., 2018). With spherical dynamics, Uhlemann et al. (2016) was able to predict the full PDF of Δ accurately at $10\,h^{-1}$Mpc, which would be sufficient for our purpose. The velocity profiles can be predicted also from $\xi(r)$ and $\xi(r|\Delta^i)$, once the coupling relationship between density and velocity is known. In addition, when galaxies are used as tracers, a bias model is needed to connect the clustering of galaxies to the clustering of dark matter.

In this chapter, with the aim to compare the constraining power on cosmology of the 2PCF versus DS, I will use the real-space 2PCF and CCF measured from the simulations as my model inputs. I then employ the empirical model with one free parameter to model the coupling between the densities and velocities, i.e. Eqs. (2.7) & (2.10). I further adopt the shape of the velocity dispersion profiles from the simulations, and allow their overall amplitudes to vary with one single free parameter σ_v. I will also adopt a linear bias scheme, i.e. $\xi(r) = b^2\xi_m(r)$; and $\xi(r|\Delta^i) = b\xi_m(r|\Delta^i)$, where $\xi_m(r)$ is the two-point correlation function of matter (the condition on the spherical top-hat density is the same one I introduced in Chapter 2).

There may be considerations for another bias parameter related to the densities smoothed at 15 h^{-1}Mpc, which I will call $b_{\rm DS}$, and this is also relevant to the discussion for the possible peculiar motions of these 15 h^{-1}Mpc spheres (Cai et al., 2016; Massara & Sheth, 2018). Indeed, if one measures the ratio between $\xi(r|\Delta^i)$ and the matter 2PCF $\xi_m(r)$, I find that on large scales, the result is consistent with the product of $b_{\rm DS}$ and the galaxy bias b, with $b_{\rm DS}$ varying from ~ -3 to ~ 3 from DS1 to DS5. However, $b_{\rm DS}$ is irrelevant for this modelling, as it is only the ratio between $\xi(r|\Delta^i)$ and $\xi_m(r|\Delta^i)$, which is the linear galaxy bias b, that matters. Likewise, the possible peculiar motion of those 15 h^{-1}Mpc spheres does not seem to have any meaningful impact on the mapping between real and redshift-space CCFs, as evident by the success of the modelling for the redshift-space CCFs shown in Fig. 3.5 & Fig. 3.6.

Note that the density splitting and model ingredients for the GSM have been obtained in real space. Even with the PDF of the density and all the ingredients of the GSM model being predictable analytically, as mentioned above, we still have to face the issue that density splitting may have to be done in redshift space in observation. As the ranks of the smoothed densities by a 15 h^{-1}Mpc top-hat filter may be affected by redshift-space distortions, splitting density in redshift space will inevitably cause mixing between quantiles from their real space version. This issue exists similarly for identifying voids in redshift space. There are at least three known approaches to tackle this. 1. Nadathur et al. (2019b) and their later work use reconstruction to re-install the positions of galaxies in real space, and perform void finding in the reconstructed galaxy field. This seems to work well and one can do the same for DS in principle. 2. Hamaus et al. (2020) uses the Abel transformation to reconstruct the real-space profiles of voids from the projected profiles along the transverse direction. Using this approach for DS, one may not need analytical predictions for the PDF and real-space profiles. 3. Repp & Szapudi (2020) has also developed a method to map counts-in-cells between real and redshift space. This may also be applied to reinstall the real-space PDF of the smoothed galaxy number densities, and hence allow the subsequent density splitting in real space. In Chapter 4, I will apply the DS framework to an observational dataset, using reconstruction to remove RSD from the galaxy samples.

Similarly, density splitting may also be affected by geometry distortions, the AP effect. In principle, for each $q_\perp$, $q_\parallel$ combination, the smoothing radius should be rescaled by $q = q_\perp^{2/3} q_\parallel^{1/3}$, and thus the density splitting would need to change at each iteration of the MCMC. However, I have explicitly tested this effect by performing the density splitting with different smoothing radii, using values that are within the range of our priors for the AP parameters. I have found that the inferred real-space profiles do not change significantly in shape or amplitude within the scale ranges that are used in the fits, which means that using a fixed smoothing radius is a good approximation under these conditions. One way around this issue for future analysis with observational data would consist in predicting the real-space density PDF, together with the real-space stacked density and velocity profiles, for any given cosmology and without having to rely on simulations, thus effectively generating the density splitting model ingredients on-the-fly for each step of the MCMC chain.

Our observables/data vectors would be the combination of all five $\xi(s, \mu|\Delta^i)$ in redshift space for DS, and $\xi(s, \mu)$ for the 2PCF. In practice, I use Eq. (3.3) to extract the monopole, quadrupole and hexadecapole contributions, and concatenate the multipoles into a single vector:

$$\boldsymbol{\xi} = (\xi_0, \xi_2, \xi_4) \,. \tag{3.9}$$

In this way, for DS our data vector is the concatenation of 3 multipoles and 5 quintiles, $\xi^{1+2+3+4+5}(s)$. For the 2PCF, it is simply the combination of 3 multipoles $\xi(s)$.

The model parameters for the DS fits are

$$\boldsymbol{\theta}_{\mathrm{DS}} = \left(f\sigma_{12}, b\sigma_{12}, \sigma_v, q_\perp, q_\parallel, \nu^1, ..., \nu^n\right), \tag{3.10}$$

where n is the number of quintiles to be used in the analysis, $f(z)$ is the growth rate of structure parameter; σ_{12} is the rms mass fluctuation in spheres with a radius of 12 Mpc, expressed in h-independent units (see Sánchez, 2020, for a discussion on why $f\sigma_{12}$ is a more adequate parameter than the standard $f\sigma_8$ parameter combination to describe the information content of RSD), b is the linear galaxy bias, σ_v is the amplitude parameter for the velocity dispersion (one free parameter σ_v is sufficient to fit all quintiles simultaneously, as I have explicitly verified that the individual DS quintiles show $\sigma_\parallel(r,\mu)$ functions that converge at approximately the same value on large scales), $q_\parallel$ and $q_\perp$ are the Alcock-Paczynski scaling parameters that vary independently (Eq. 2.15) to re-scale the corresponding parallel and transverse distance separation vectors. As the ratio of these AP parameters is of cosmological interest, I will also show the derived parameter $q_\perp/q_\parallel$. ν^i are the parameters specifying the density-velocity couplings from Eqs. (2.7) & (2.10). For DS, I use one ν parameter per quintile, as the radial velocity profiles can have notoriously different shapes depending on the environment. The analysis of the 2PCF requires the same parameters, with the exception of a single coupling parameter ν, that is

$$\boldsymbol{\theta}_{\mathrm{2PCF}} = \left(f\sigma_{12}, b\sigma_{12}, \sigma_v, q_\perp, q_\parallel, \nu\right) . \tag{3.11}$$

This means that the analysis of DS has additional degrees of freedom that could potentially degrade its constraining power for cosmology.

Since the galaxy monopole was measured from a simulation with a fixed $b\sigma_{12}$ in the mock catalogues, I allow for DS a rescaling of its amplitude as

$$\xi(r|\Delta^i) = b\sigma_{12}\frac{\xi(r|\Delta^i)^{\mathrm{mock}}}{(b\sigma_{12})^{\mathrm{mock}}}, \tag{3.12}$$

and for the 2PCF as

$$\xi(r) = (b\sigma_{12})^2\frac{\xi(r)^{\mathrm{mock}}}{[(b\sigma_{12})^2]^{\mathrm{mock}}}, \tag{3.13}$$

where the subscript 'mock' denotes the quantity measured from the simulations. Note that for DS, linear galaxy bias b and the amplitude parameter σ_{12} enter Eq. (3.12) linearly, rather than quadratically, as $\xi(r|\Delta^i)$ is a measure of density profiles, and not a two-point statistic. This is also evident from Fig. 3.3 where the density-velocity coupling in the linear regime for cases in DS follow Eq. (2.9) and not Eq. (2.5).

I explore the parameter space by means of a Markov Chain Monte Carlo (MCMC) procedure, using the EMCEE software[7]. I set flat, non-informative priors for all parameters, as specified in Table 3.1. At each step of the chain, I concatenate the model predictions for the monopole, quadrupole and hexadecapole from each DS quintile (or a single measurement in the case of the 2PCF) into a single theory vector

[7]emcee.readthedocs.io/

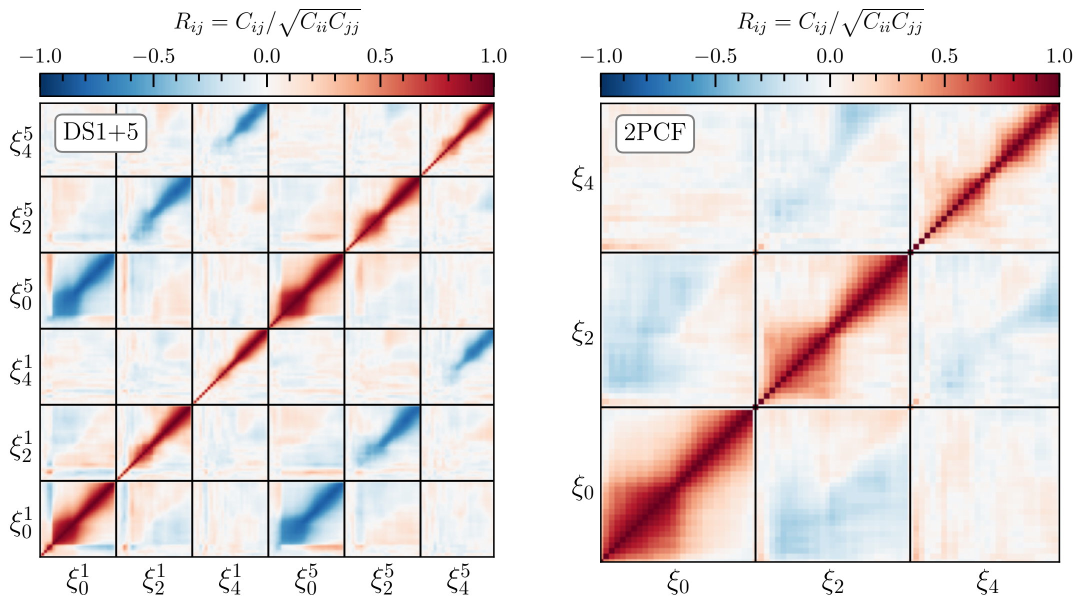

Figure 3.10: Cross-correlation coefficients for the covariance matrix defined by Eq. (3.16), for the Density Split (left) and the 2PCF (right). The solid lines divide the regions where each multipole contributes. The superindices in the DS covariance indicate the contributing quintile. Significant contributions can be seen from the first (ξ_l^1) and fifth (ξ_l^5) quintiles, respectively. There are no strong correlations among ξ_0, ξ_2 and ξ_4 within each quintile, but there is significant anti-correlation between the multipoles from the first and fifth quintiles (voids and clusters) in DS.

$\xi^{\rm theory}$, and quantify its deviation from the measured data vector $\xi^{\rm data}$ by computing

$$\chi^2 = \left(\boldsymbol{\xi}^{\rm theory} - \boldsymbol{\xi}^{\rm data}\right) \mathbf{C}^{-1} \left(\boldsymbol{\xi}^{\rm theory} - \boldsymbol{\xi}^{\rm data}\right)^T , \tag{3.14}$$

where $\mathbf{C}$ is the covariance matrix of the data vector. Then, the log-likelihood can be expressed as

$$\log \mathcal{L} = -\frac{1}{2}\chi^2 . \tag{3.15}$$

The covariance matrix is estimated by measuring $\xi^{\rm data}$ on all the mocks, as

$$\mathbf{C} = \frac{1}{N-1} \sum_{k=1}^{N} \left(\boldsymbol{\xi}_k^{\rm data} - \overline{\boldsymbol{\xi}^{\rm data}}\right) \left(\boldsymbol{\xi}_k^{\rm data} - \overline{\boldsymbol{\xi}^{\rm data}}\right) , \tag{3.16}$$

where $N = 300$, and $\overline{\xi^{\mathrm{data}}}$ is the mean data vector. As this covariance is measured from a finite number of mocks, its inversion provides a biased estimator of the true precision matrix. In order to account for this effect, I multiply $\mathbf{C}^{-1}$ by a correction factor (Hartlap et al., 2007), given by

$$\alpha = \left(1 - \frac{N_{\mathrm{b}} + 1}{N - 1}\right) \ , \tag{3.17}$$

where N_{b} is the number of bins of the data vector, and N is the number of mocks (see also Sellentin & Heavens, 2016, for an alternative approach to robustly estimate the uncertainty in the covariance matrix). After including this correction factor, the estimate of the precision matrix is unbiased, although it can still be affected by noise due to the finite number of simulations that are used, which can propagate further into the cosmological constraints (Percival et al., 2014). However, I have explicitly verified that when using a theoretically-derived Gaussian covariance matrix that is free of noise (Grieb et al., 2016), the inferred cosmological parameters from the 2PCF are recovered to better than 1 per cent with respect to the case where the covariance is estimated from the mocks. Currently, it is not possible to perform such a test for DS, as I lack a method to generate a noise-free Gaussian covariance for the DS multipoles. Nevertheless, I expect such a correction to be of similar order to the 2PCF case. An explicit verification of this is left for future work. Examples for the full covariance matrices for DS1+5 and 2PCF are shown in Fig. 3.10. There are significant anti-correlations between the multipoles in DS1 and DS5, as these correspond to voids and peaks of the density field.

3.2.1 Parameter constraints

Before presenting the main cosmological constraints, I want to stress-test the numerical accuracy of the model and our simulation set up. There are at least two factors to consider: 1. with a large covariance matrix to invert, and a finite num-

Table 3.1: Comparing cosmological parameter constraints from two-point correlation functions (2PCF) and cross-correlation functions from density splitting (DS). All quoted errors (noted as $\sigma_{\rm stat}$) correspond to 68 per cent confidence levels around the best fit values μ.

Parameter	Prior	2PCF ($\mu \pm \sigma_{\rm stat}$)	2PCF ($\sigma_{\rm stat}/\mu$)	DS ($\mu \pm \sigma_{\rm stat}$)	DS ($\sigma_{\rm stat}/\mu$)	$(\sigma_{\rm 2PCF} - \sigma_{\rm DS})/\sigma_{\rm DS}$
$s_{\rm min} = 20\,h^{-1}{\rm Mpc}$						
$f\sigma_{12}$	$[0.1, 2.0]$	0.445 ± 0.049	0.110	0.496 ± 0.037	0.074	33.6%
$b\sigma_{12}$	$[0, 3]$	1.210 ± 0.017	0.014	1.195 ± 0.028	0.023	-39.6%
$q_\perp/q_\parallel$		1.007 ± 0.029	0.029	1.012 ± 0.013	0.013	121.2%
$q_\perp$	$[0.8, 1.2]$	0.999 ± 0.012	0.012	1.001 ± 0.009	0.009	37.9%
$q_\parallel$	$[0.8, 1.2]$	0.991 ± 0.025	0.025	0.989 ± 0.014	0.014	76.1%
$s_{\rm min} = 15\,h^{-1}{\rm Mpc}$						
$f\sigma_{12}$	$[0.1, 2.0]$	0.477 ± 0.036	0.079	0.456 ± 0.028	0.062	33.7%
$b\sigma_{12}$	$[0, 3]$	1.209 ± 0.016	0.013	1.215 ± 0.014	0.011	14.1%
$q_\perp/q_\parallel$		1.006 ± 0.028	0.027	0.998 ± 0.004	0.004	596.8%
$q_\perp$	$[0.8, 1.2]$	1.000 ± 0.012	0.012	0.995 ± 0.002	0.002	502.2%
$q_\parallel$	$[0.8, 1.2]$	0.994 ± 0.023	0.023	0.997 ± 0.005	0.005	378.1%

ber of simulation boxes (300), we want to reduce the size of the covariance matrix where possible to make sure that it is invertible and accurate. 2. the precision of constraints increases rapidly when the minimum scale included in the fits, $s_{\rm min}$, is decreased. Despite the excellent agreement between the GSM and the two-dimensional correlation functions measured from the simulations (Figs. 3.5 & 3.6), I need to find a balance between precision and accuracy, which is likely to depend on $s_{\rm min}$.

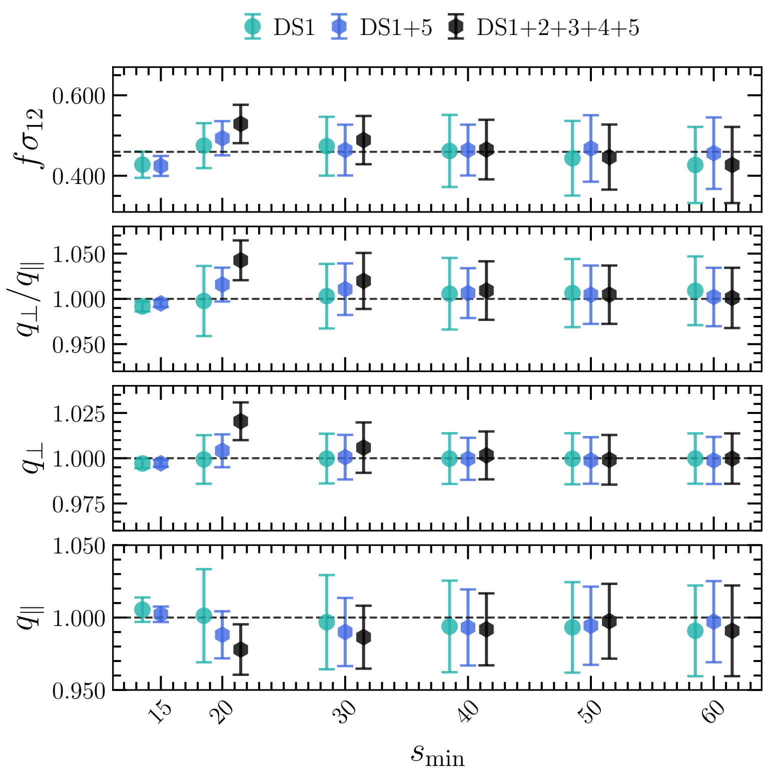

Figure 3.11: Cosmological parameter constraints obtained by combining CCFs from different DS quintiles (labeled in different colours and deliberately offset from each other for clarity). The markers show the median value of the marginalised posterior distribution for each parameter, while the error bars represent 68 per cent confidence levels. The horizontal dashed lines show the values from the cosmology of the simulations. Note that in order to reduce the size of the data vector, necessary to construct a reliable covariance matrix that includes all quintiles, given the limited number of simulation boxes, ξ_4's are not used for the fitting in this figure. This only causes a minor increase of the errors.

I start by running MCMC for a wide range of $s_{\rm min}$, different combinations of density quintiles and a fixed maximum fitting scale $s_{\rm max} = 141\,h^{-1}$Mpc. The constraints on cosmological parameters after marginalising over all the nuisance parameters (i.e. the velocity dispersion σ_v and the couplings ν^i) are shown in Fig. 3.11. We can see that on scales $s_{\min} \geq 50\,h^{-1}$Mpc, using only DS1 (equivalent to the void-galaxy CCF) yields almost the same constraints as DS1+5 (equivalent to the void-galaxy CCF plus the cluster-galaxy CCF), or DS1+2+3+4+5. This is expected in the Gaussian linear regime. At $s_{\min} = 30\ h^{-1}$Mpc, the combination of DS1+5 starts to yield better constraints than DS1 alone, and DS1+2+3+4+5 is slightly better than DS1+5. At $s_{\min} = 20\,h^{-1}$Mpc, the combinaton DS1+5 leads to tighter constraints than DS1 alone, indicating the value of combining the void-galaxy and cluster-galaxy CCFs. Although at these scales the uncertainties obtained from the joint analysis of DS1+2+3+4+5 (black points) are comparable to those obtained in the DS1+5 case, the results are clearly biased. Therefore, instead of the full combination, I will use DS1+5 as our default case for DS, knowing that they can already yield similar constraints as the case of DS1+2+3+4+5 in the range of scales where the model is unbiased. These tests also inform us that with the volume that is being simulated, $(1.5\,h^{-1}\mathrm{Gpc})^3$, the GSM provides unbiased constraints down to scales $s \simeq 15\,h^{-1}$Mpc. Therefore, I will set $s_{\min} = 15\,h^{-1}$Mpc as the default case and consider also the case of $s_{\min} = 20\,h^{-1}$Mpc for comparison. Note that all data points in Fig. 3.11 were obtained by fitting only the monopole and quadrupole. This is the case for DS1, DS1+5 and DS1+2+3+4+5. The motivation behind this is that I wanted to make a fair comparison between the different quantile combinations. Since the covariance matrix of DS1+2+3+4+5 is prohibitively large when using all 3 multipoles, I stick to using monopole and quadrupole for all combinations, which are thought to encapsulate most of the cosmological information of interest. This criterion was only applied for this figure. Throughout the rest of the paper, DS1+5 always uses all 3 multipoles.

The posterior distributions of the parameters that contain cosmological information are shown in Fig. 3.12 with $s_{\min} = 15\,h^{-1}$Mpc and $s_{\min} = 20\,h^{-1}$Mpc,

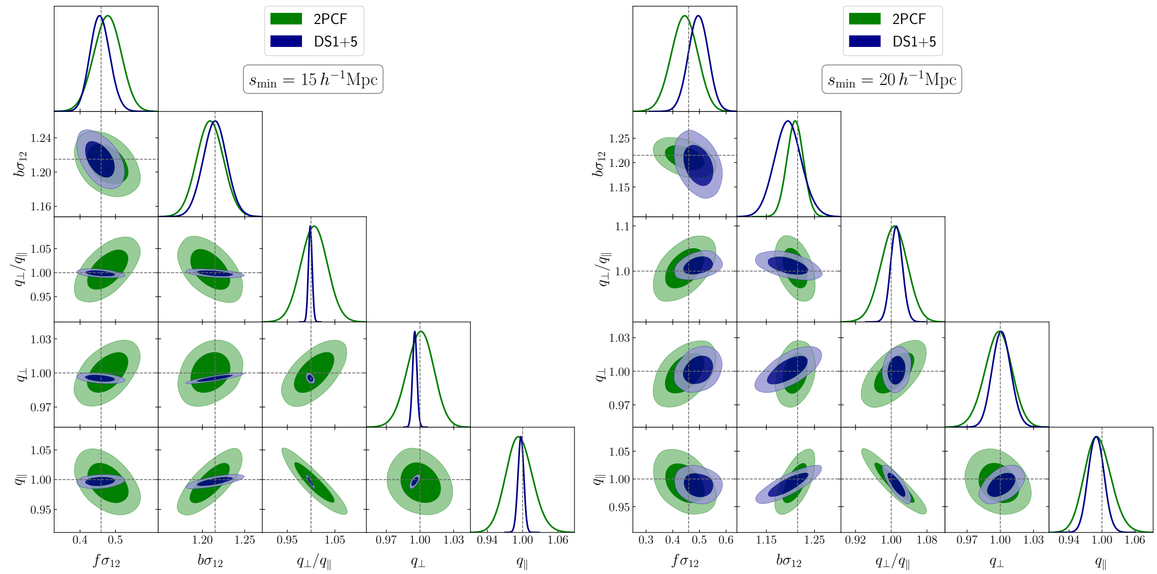

Figure 3.12: Posterior distributions of the cosmological parameters inferred from two-point correlation functions (2PCF) and cross-correlation functions from density splitting (DS) (blue), with DS1+5 combined, using scales between 15-141 h^{-1}Mpc (left-hand side side panel) and 20-141 h^{-1}Mpc (right-hand panel). Darker and lighter shades show the 68 and 95 per cent confidence levels around the best-fit values, respectively. The dashed lines show values corresponding to the cosmology from the simulations.

respectively (see also Fig. A1 & A2 where all other parameters are presented). For $s_{\rm min} = 20\,h^{-1}$Mpc (right-hand side panel), the marginalised constraints for the growth rate parameter at the 68 per cent CL are $f\sigma_{12} = 0.445 \pm 0.049$ for the 2PCF, and $f\sigma_{12} = 0.496 \pm 0.037$ for DS. The result recovered from the DS 1+5 combination represents a $\sim$ 33 per cent improvement in precision over 2PCF, even though this case uses 60 per cent fewer sampling points than the full 2PCF. The advantage of DS is stronger for AP parameters. While the 2PCF yields a constraint for the AP ratio of $q_\perp/q_\parallel = 1.007 \pm 0.029$, DS has $q_\perp/q_\parallel = 1.012 \pm 0.013$, a 121% reduction of the error, and reaching a $\sim$ 1% precision. Previous studies have already reported that the multipoles of the void-galaxy CCF are particularly good at measuring this ratio (Nadathur & Percival, 2019; Nadathur et al., 2020; Hamaus et al., 2020), which is similar to having DS1. Here we can see that adding DS5, similar to the cluster-galaxy CCF, helps to beat down the errors for AP parameters (Fig. 3.11). The galaxy bias parameter encapsulated in $b\sigma_{12}$ is recovered at a precision of $\sim 2.3\%$ in DS, with $b\sigma_{12} = 1.195 \pm 0.028$. This is slightly worse than 1.4% constraint from 2PCF. This perhaps indicates that on this scale, the constraints with DS are still partially penalised by having the extra free parameter. It is also interesting to see that the constraints on some parameter combinations from DS, such as $f\sigma_{12}$ against $q_\perp/q_\parallel$, show contours that are rotated with respect to the 2PCF, implying that there is possibility for complementing the information from these two probes in galaxy surveys, although this is beyond the scope of this work. A summary of the quoted constraints with their corresponding precision and figures of merit can be found in Table 3.1.

At $s_{\min} = 15\,h^{-1}\mathrm{Mpc}$, the parameters $f\sigma_{12}$ and $b\sigma_{12}$ are constrained to $\sim 6\%$ and $\sim 1\%$ respectively for DS. All the AP parameters are down at the sub-per cent level. The improvement for DS1+5 over 2PCF is much more significant (with the exception of the growth rate), as shown on the left of Fig. 3.12 and in Table 3.1, with reductions of errors on the $f\sigma_{12}$ by 33%, $b\sigma_{12}$ by 14%, and AP parameters typically 400-600% smaller. This is qualitatively expected. At smaller scales, the density and velocity fields become even more non-Gaussian. Less information is extracted from the two point statistics, and the cross-correlations between voids and peaks with the galaxy field (DS1+5) seem capable of uncovering it (White, 1979; Saslaw & Hamilton, 1984; Fry, 1985, 1986). However, note that the best-fit constraint for $q_\perp$ is 2.4σ from the fiducial value, which is at the margin of being unbiased. Nevertheless, it is also important to realise that the error on $q_\perp$ is at the sub-per cent level, and is a factor of $\sim$6 smaller than that recovered from the 2PCF. Even if I take the model to be biased as the 3-σ level, and add this as the systematic error, the total error is still significantly smaller (a factor of $\sim$2) than that from the 2PCF. It is likely that with decreasing $s_{\min}$, the advantage of using DS for cosmological constraints over the 2PCF will continue to increase. From the left-hand side panel of Fig. 3.12, the rotation of the contours for DS relative to the 2PCF becomes very obvious. This again indicates the potential of even better cosmological constraints with DS and the 2PCF combined.

The gain for the AP parameters $q_\perp$ and $q_\parallel$, as well as their ratio $q_\perp/q_\parallel$ with DS over the 2PCF is more obvious than for the growth-rate parameter ($f\sigma_{12}$). This is likely because the CCFs in DS have relatively steep profiles around the scales of the top-hat filter. In particular, DS1 and DS5 have very different slopes, sometimes of the opposite sign from each other, while the 2PCF is relatively featureless at small scales. This makes those CCFs more sensitively to the change of AP parameters via Eq. (2.15), i.e. the derivatives of the CCFs with respect to the re-scaling parameters q are larger.

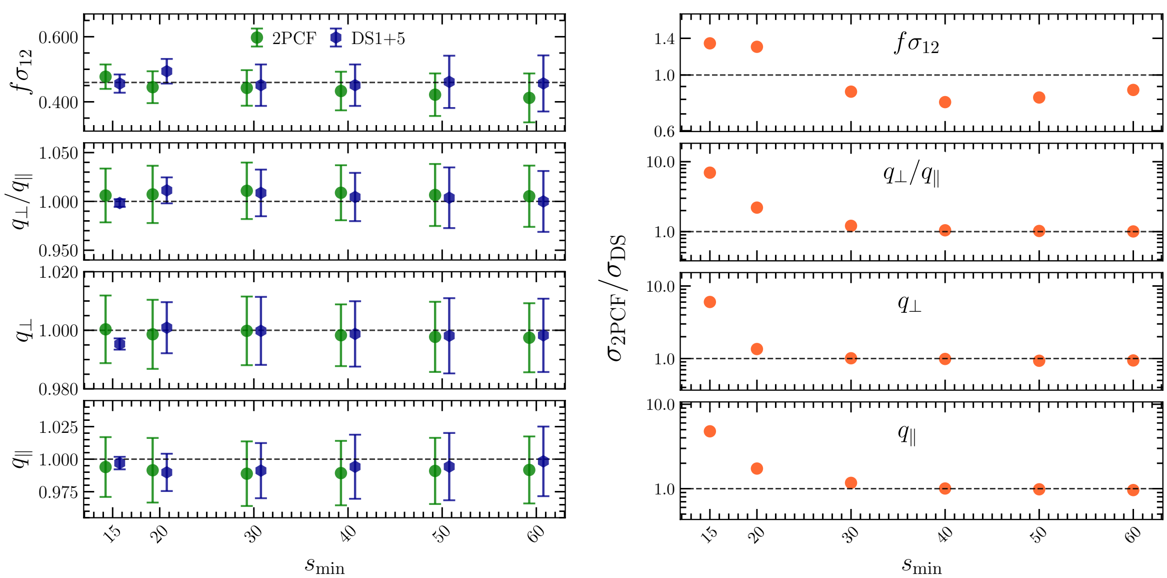

Figure 3.13: Left-hand side: parameter constraints with the 2PCF (green) and DS1+5 (blue) as a function of the minimum scale used in the fit. The error bars show the 1σ spread around the best-fit values. The dashed-horizontal lines indicate values from the cosmology of the simulation. Right-hand side: The ratio of the 1σ errors (from the left) between DS and the 2PCF. The y-axis is shown in a logarithmic scale.

More generally, I also compare the constraining power between DS and the 2PCF for a wide range of minimal scales $s_{\rm min}$, shown in Fig. 3.13. On large scales, where the density field is nearly Gaussian, DS will not reveal information that is not already contained in the 2PCF. In fact, DS is expected to deliver slightly worse constraints, since the fit is penalised by the inclusion of an extra nuisance parameter with respect to the 2PCF (ν^1 & ν^5 versus ν). This is seen for the $f\sigma_{12}$ parameter for $s_{\rm min} > 30\,h^{-1}$Mpc, but the constraints for AP parameters are virtually the same between the two cases for those large scales.

As the minimum fitting scale progressively gets smaller, the density and velocity fields deviate from a Gaussian distribution, and the inclusion of the extra free parameter is compensated by the additional information that is captured by DS. This happens at around $s_{\rm min} = 30\,h^{-1}$Mpc. At $s_{\rm min} = 20\,h^{-1}$Mpc, the constraints for the growth parameter and AP parameters are all tighter with DS. This trend continues towards smaller scales and the improvement for DS over the 2PCF continues until $s_{\rm min} < 15\,h^{-1}$Mpc, where the model for DS starts to show systematic biases larger than the statistical errors for our sample.

3.2.2 Inclusion of BAO information in DS fits

As discussed earlier, the BAO information encoded at very large scales allows us to constrain $q_\perp$ and $q_\parallel$ individually. This is demonstrated in Fig. 3.14, where I compare the cases where the DS model fit is performed with and without including scales between 80-141 h^{-1}Mpc, where the BAO feature is found. The exclusion of the BAO scales leads to a strong degeneracy between $q_\perp$ and $q_\parallel$. Since the correlation function in those intermediate scales is smooth and featureless, the model is not particularly sensitive to the angle-averaged parameter combination $q = q_\perp^{2/3} q_\parallel^{1/3}$, and can only constrain the AP ratio precisely. The inclusion of the BAO peak helps to break this degeneracy, allowing constraints at the 0.8% and 1.4% level for $q_\perp$ and $q_\parallel$, respectively. This information is valuable, as $q_\perp$ and $q_\parallel$ are defined in terms of the comoving angular diameter distance $D_{\rm M}(z)$ and the Hubble rate $H(z)$, respectively, which in turn allows us to put constraints on $\Omega_{\rm m}$. On the other hand, the constraint on $f\sigma_{12}$ is virtually unchanged with the addition of scales of 80-141 h^{-1}Mpc. This is expected, since $f\sigma_{12}$ only benefits from the additional perturbation modes at those scales, which is a negligibly small fraction compared to those at small scales.

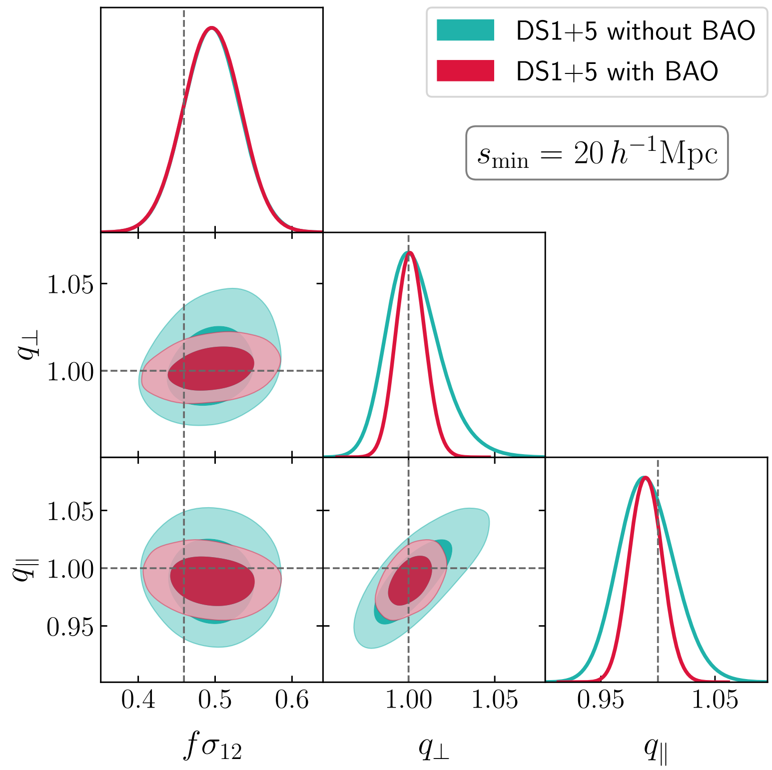

Figure 3.14: Comparing cosmological constraints with cross-correlation functions from DS, using scales between 20-80 h^{-1}Mpc (red) and 20-141 h^{-1}Mpc (turquoise). The latter case includes scale of the BAO. The BAO feature in the cross-correlation function helps to break degeneracies for the AP parameters ($q_\perp$ and $q_\parallel$) and improves their constraints, but it has negligible impact on the growth parameter $f\sigma_{12}$.

In summary, I have shown that when applying the same GSM model to the same data with the same ranges of scale, DS tends to tighten the constraints for both the growth parameter ($f\sigma_{12}$) and AP parameters ($q_\parallel$ and $q_\perp$) over the 2PCF. The improvement of DS over the 2PCF is relatively more significant for AP parameters. I believe that there are two main reasons behind this:

(1) the density PDF becomes non-Gaussian at small scales. The 2PCF, which is essentially a measure for the variance of the density field, becomes incomplete. The combination of a series of CCFs in DS allows us to sample the non-Gaussian dis-

tribution of the density PDF, hence recovering some information lost in the 2PCF. This appears to be effective even with the combination of only DS1 and DS5, the equivalent of voids plus clusters.

(2) the diverse and steep slopes of CCFs in DS near the top-hat smoothing scale makes the use of DS more sensitive to AP parameters at small scales than for the 2PCF. This allows DS to break the degeneracy between AP parameters and the growth parameter with just the small-scale information, and without employing the BAO. On the other hand, the 2PCF needs the BAO scale to be included to break such degeneracy. Therefore, when the BAO scale is not included in both cases, i.e. with $s < 80\,h^{-1}$Mpc, DS provides constraints on those cosmological parameters many times better than the 2PCF, as I have checked explicitly. This suggests that DS will be particularly powerful for RSD analysis with galaxy surveys covering a relatively small volume where the BAO peaks are not well constrained.

Chapter 4

Application to observations

This chapter will be dedicated to the application of the redshift-space distortions with split densities method to an observational galaxy sample. Having validated this framework on the Minerva cosmological simulations, and given the promising prospects for precision cosmology using these clustering statistics, I wish to provide a proof-of-concept of how to implement this methodology using galaxies from SDSS DR12, highlighting the different sources of systematic errors that arise in observations and add complexity to this clustering analysis.

4.1 The BOSS CMASS and LOWZ galaxy samples

The Sloan Digital Sky Survey[1] (SDSS) has been pivotal for our understanding of the Cosmos at the very large scales. It has granted us one of the most detailed three-dimensional maps of the Universe ever made, with deep multi-color images of one third of the sky, and spectra for more than three million astronomical objects.

The execution of SDSS has been separated in different stages and projects that

[1] https://www.sdss.org/

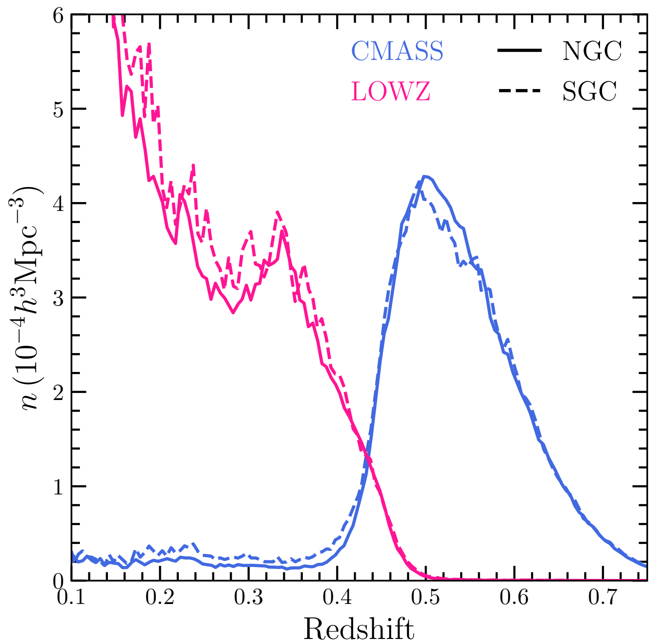

Figure 4.1: Number density as a function of redshift for the LOWZ and CMASS galaxy samples across the two Galactic caps.

aim to tackle different scientific questions. SDSS-III, for instance, consists of four surveys executed on the same 2.5m telescope: the Apache Point Observatory Galactic Evolution Experiment (APOGEE) the Baryon Oscillation Spectroscopic Survey (BOSS), the Multi-Object APO Radial Velocity Exoplanet Large-area Survey (MARVELS), and the Sloan Extension for Galactic Understanding and Exploration 2 (SEGUE-2).

I use the final galaxy catalogues from BOSS, corresponding to SDSS DR12 (Alam et al., 2015). The catalogue is divided into two spectroscopic galaxy samples, LOWZ and CMASS, which cover a redshift range $0.15 < z < 0.75$ and were selected based on SDSS multicolour photometric observations (Gunn et al., 1998, 2006).

The CMASS sample is dominated by early type galaxies, and it is nearly complete down to a stellar mass of $M_* \simeq 10^{11.3}\,\mathrm{M}_\odot$ for $z > 0.45$ (Maraston et al., 2013). The LOWZ sample is primarily composed of red galaxies that live in massive haloes (Parejko et al., 2013). Both samples were observed across the two Galactic hemispheres, referred to as the Northern and Southern galactic caps (NGC and SGC, respectively). Fig. 4.1 shows how is the number density of each of these samples of galaxies distributed across redshift.

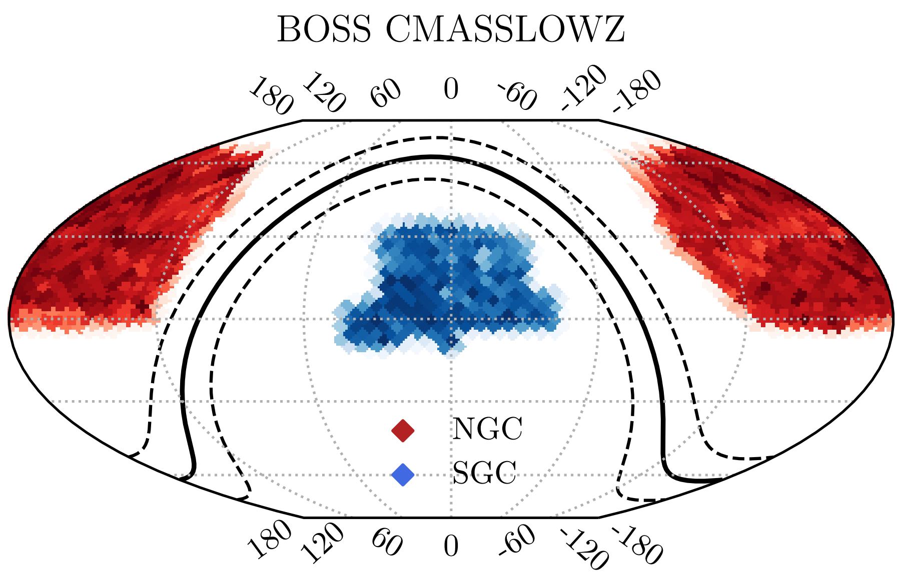

Figure 4.2: Footprint of the DR12 CMASSLOWZ galaxy sample across the northern (red) and southern (blue) Galactic caps. The black lines show the plane of the Milky Way across the sky.

I follow the procedure of DR12 clustering analyses and combine the LOWZ and CMASS samples into a single catalogue that we call CMASSLOWZ. I further divide this sample into two non-overlapping redshift bins of roughly equal volume, from $0.2 < z < 0.5$ and $0.5 < z < 0.75$. Fig. 4.2 shows how are these two subsets distributed across the sky. The Milky Way disk is represented by the solid and dashed black lines, while the NGC and SGC galaxy distributions are shown in red and blue, respectively. As the main purpose of this chapter is to provide a proof of concept of the application of the DS algorithm to observations, the analysis will be restricted to the NGC. However, I plan to carry out a more detailed analysis including the SGC in future work.

Throughout this observational analysis, I will assume a flat ΛCDM model, adopting $\Omega_m = 0.307$ (Planck Collaboration et al., 2016) to convert redshifts to radial comoving distances. This will facilitate the comparison with other studies in the literature.

4.2 Density field reconstruction

As explained in Chapter 2, the DS algorithm relies on calculating cross-correlation functions between split densities identified in real space, and redshift-space galaxies. While real-space galaxy positions are always known in N-body simulations, this is not the case in observations, where the galaxy samples are naturally observed in redshift space. To overcome this, we use the method proposed in Nadathur & Percival (2019) to reconstruct the approximate real-space galaxy positions by removing effects of large-scale velocity flows. This way, we can identify the split density centres in the reconstructed galaxy field, which works as a pseudo-real space, and calculate the cross-correlation functions between those centres and the redshift-space galaxy field.

Let us place ourselves in a Lagrangian framework, in which the Eulerian position $\mathbf{x}$ at time t can be described in terms on the initial Lagrangian position $\mathbf{q}$ and a non-linear displacement field $\mathbf{\Psi}(\mathbf{q}, t)$:

$$\mathbf{x}(\mathbf{q}, t) = \mathbf{q} + \mathbf{\Psi}(\mathbf{q}, t)\,. \tag{4.1}$$

The galaxy overdensity field $\delta_g(\mathbf{x}, t)$, can be related to the displacement field by (Nusser & Davis, 1994)

$$\nabla \cdot \mathbf{\Psi} + \frac{f}{b} \nabla \cdot (\mathbf{\Psi} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} = -\frac{\delta_g}{b}\,, \tag{4.2}$$

where b is the linear bias parameter that relates the galaxy density field to the underlying matter distribution. The full solution to Eq. 4.2 includes contributions to the velocity flow coming from galaxy peculiar velocities at the corresponding redshift, as well as additional non-linear evolution that can be traced back to earlier epochs. In BAO analyses (e.g. Alam et al., 2017), in an attempt to undo all effects of non-linear clustering to sharpen the BAO feature to the best extent possible, the full displacement field obtained by solving Eq. 4.2 is used. In my analysis, I am only concerned about removing the RSD coming from galaxy peculiar velocities at a certain epoch, so the part of the solution I am interested in is

$$\mathbf{\Psi}_{\mathrm{RSD}} = -f(\mathbf{\Psi} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}}\,. \tag{4.3}$$

By shifting the redshift-space galaxy positions by $-\mathbf{\Psi}_{\mathrm{RSD}}$, I obtain a pseudo-real space galaxy catalogue that can be used to apply the density split framework. To apply reconstruction to the CMASSLOWZ galaxy sample, I use the REVOLVER[2] code (Nadathur & Percival, 2019), which solves Eq. 4.2 by using an iterative fast Fourier transform procedure. To run the code, we need to provide the redshift-space galaxy catalogue, as well as the corresponding random catalogue that can be used to estimate the density in each region of the survey volume. As the code performs calculations of the density field on a rectangular grid, we also need to specify a grid size to optimize the efficiency of calculations, which I take to be 512^3, following BOSS reconstruction analyses (Alam et al., 2017). The galaxy density field δ_g is smoothed with a Gaussian kernel of width R_s, in order to reduce the sensitivity of the algorithm to small scale density modes, for which the assumptions of the equations that were presented above are not valid. I adopt $R_s = 10\,h^{-1}$Mpc, which has been shown to be a robust smoothing scale for previous analysis concerning RSD around voids (Nadathur & Percival, 2019).

In terms of the cosmological dependence of the algorithm, Eq. 4.2 shows that reconstruction is sensitive to the linear growth rate of structure f and the linear bias

[2] https://github.com/seshnadathur/Revolver

parameter b. In practice, the RSD solution that I am concerned about has a full degeneracy between f and b. In other words, it is only sensitive to the linear growth parameter $\beta = f/b$. Therefore, as the reconstructed galaxy field itself now depends on β, all summary statistics that are derived from this catalogue, including the DS CCFs, will inherit this cosmological dependence. This point will need to be taken into consideration in the likelihood analysis presented in Sec. 4.4.

4.3 Clustering measurements

I calculate the CCFs between DS centres and galaxies using the Davis & Peebles estimator (Davis & Peebles, 1983):

$$\xi(s, \mu|\Delta^i) = \frac{D_1D_2(s,\mu)}{D_1R_2(s,\mu)} - 1, \tag{4.4}$$

where $D_1D_2(s,\mu)$ are the pair counts between DS centres and galaxies, while $D_1R_2(s,\mu)$ are the pair counts between DS centres and randoms. I assign weights to each galaxy in our catalogues before pair counting. The first weight is designed to minimize the variance in the measurments (Feldman et al., 1994) and is given by

$$w_{\mathrm{FKP}}(\mathbf{x}) = \frac{1}{1 + P_w\overline{n}(\mathbf{x})}\,, \tag{4.5}$$

where $\overline{n}(\mathbf{x})$ is the galaxy number density at a position $\mathbf{x}$, and $P_w = 10^4\,h^{-3}\mathrm{Mpc}^3$. Additionally, I assign weights for close pairs affected by fiber collisions, w_{cp}, failures in the redshift determination, w_{noz}, and additional systematic effects in observations, w_{sys}. The total weight for each galaxy is then given by

$$w_{\mathrm{tot}} = w_{\mathrm{FKP}}w_{\mathrm{sys}}(w_{\mathrm{cp}}w_{\mathrm{noz}} - 1)\,. \tag{4.6}$$

I use radial bins of $5\,h^{-1}$Mpc and μ bins of a constant width of 0.02 between -1 and 1. To estimate the covariance matrix for our clustering measurements, I use the Multidark-Patchy mocks described in Kitaura et al. (2016), a suite of 2048 mock galaxy catalogues that were designed to match the clustering, footprint and observational systematics of the BOSS DR12 galaxy samples. These mocks were generated using the MD-Patchy algorithm (Kitaura et al., 2016), and use the best-fit ΛCDM cosmology to the Planck 2013 CMB measurements (Planck Collaboration et al., 2014). While the MD-Patchy mocks were calibrated to accurately reproduce the monopole and quadrupole of the redshift-space galaxy 2PCF, it is not guaranteed that they will reproduce other observables, such as the void-galaxy or cluster-galaxy CCFs, with the same level of precision. This point will be of important when considering potential sources of systematic errors in the analysis.

The DS method relies on measuring CCFs between centres in real space and galaxies in redshift space. To generate a pseudo-real space galaxy catalogue from observations, we reconstruct the CMASSLOWZ galaxy density field using the method discussed in Sec. 4.2. I apply the same procedure in the MD-Patchy mocks to generate a reconstructed mock catalogue that can be used for the covariance matrix estimation and model predictions. As explained in Sec. 4.2, the reconstruction algorithm has a dependence on the linear growth rate parameter β. Since I am interested in providing constraints for the growth rate of cosmic structure, I need to incorporate this cosmological dependence into the likelihood analysis. In practice, this means that the measurements and the model predictions (which are based on the reconstructed density field) should change for each value of β. Since it is computationally unfeasible to run reconstruction and calculate the necessary clustering statistics for all the parameter space that will be explored, I instead reconstruct the galaxy field and perform clustering measurements for 30 fixed β values, and interpolate over these profiles in the likelihood analysis.

For 30 linearly spaced β values in the range $[0.162, 0.649]$, I run the following algorithm, both for the CMASSLOWZ sample as well as for the MD-Patchy mocks:

1. Remove RSD from galaxy positions by applying reconstruction on the redshift-space galaxy density field.

2. Calculate top-hat filtered density contrasts $\overline{\xi}(r = R)$ around 1 million random points within the survey volume using the reconstructed galaxy catalogues and a filter radius $R = 25\,h^{-1}$Mpc.

3. Rank the filtered density contrasts in increasing order and split them into 5 bins or quintiles.

4. Cross-correlate the positions in each quintile with the galaxy field in redshift space to obtain $\xi(s, \mu|\Delta^i)$.

5. Cross-correlate the positions in each quintile with the reconstructed galaxy field to obtain the real-space monopole $\xi(r|\Delta^i)$.

As $\xi(s, \mu|\Delta^i)$ now implicitly depends on β through reconstruction, the covariance matrix should also inherit this dependence. However, I have explicitly checked that this dependence is very weak, and therefore I only calculate the covariance matrix for the fiducial cosmology. This procedure has been shown to be a very good approximation for cosmological analyses (Kodwani et al., 2019).

Fig. 4.3 shows the PDF of the integrated galaxy density contrasts for the CMASS-LOWZ NGC sample, using $\beta = 0.389$, at the two redshift bins. The distributions look fairly smooth and closely resemble what we had seen for the tests in N-body mock catalogues. I have chosen an slightly larger filter radius for the top-hat filter than in the N-body case, as I have found that using $R = 15\,h^{-1}$Mpc results in a dominant fraction of spheres being completely empty at that scale, presumably due to the larger mean galaxy separation for this joint sample. Dividing the filtered density contrasts in 5 bins results in 2×10^5 points per quintile, which is sufficient for

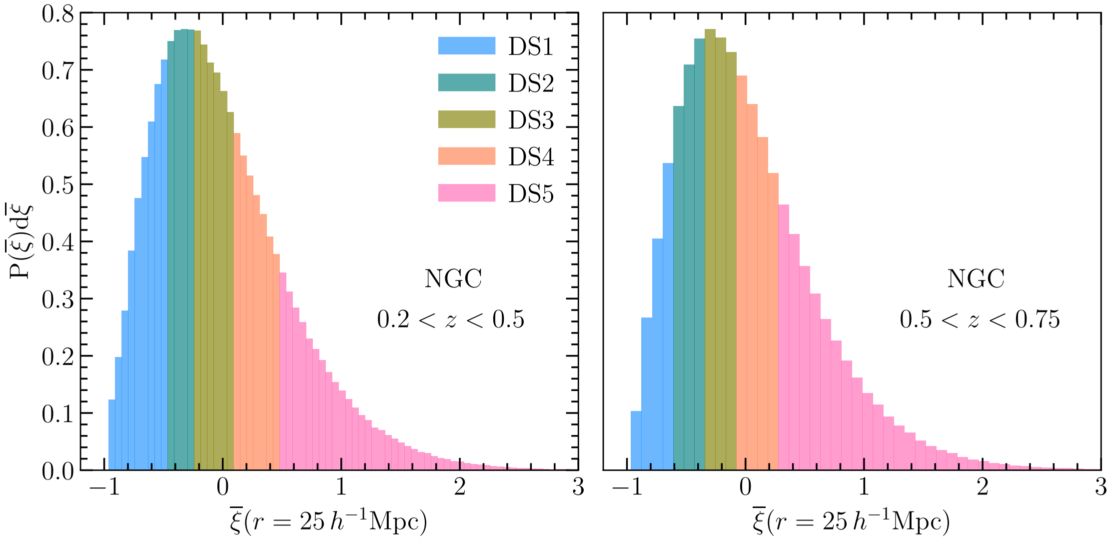

Figure 4.3: PDF of integrated galaxy density contrasts around random positions, using a top-hat smoothing filter of radius $R = 25\,h^{-1}$Mpc. The left-hand side and right-hand side panels show the measurement in the lower and higher redshift bins of the NGC CMASS-LOWZ samples, respectively. The different colours show the division of the PDFs in different density quantiles.

the demonstration of the application of the DS algorithm on this dataset, but can be increased in the future if one wishes to reduce the noise in the measurements, given that enough computational resources are available. Fig. 4.4 shows the monopole of the redshift-space CCF for the same samples. Again, the profiles resemble the simulation-based scenario very closely, although the profiles are slightly noisier at large scales. The BAO features are clearly present in each of the quintiles as either a peak or depression around $105\,h^{-1}$Mpc. The profiles from the lower and higher redshift bins are qualitatively similar, although the amplitude of the monopoles at large scales appear to be slightly lower in amplitude for the lower redshift sample. Figs. 4.5 & 4.6 show the decomposition of the redshift-space CCF into multipoles for the lower and higher redshift bins, respectively. We can see that all of the features that were previously spotted in the N-body simulations are also present in the CMASSLOWZ multipoles, such as the change of sign and slope for some of the quadrupoles when going from small to large scales. The features in the hexadecapole

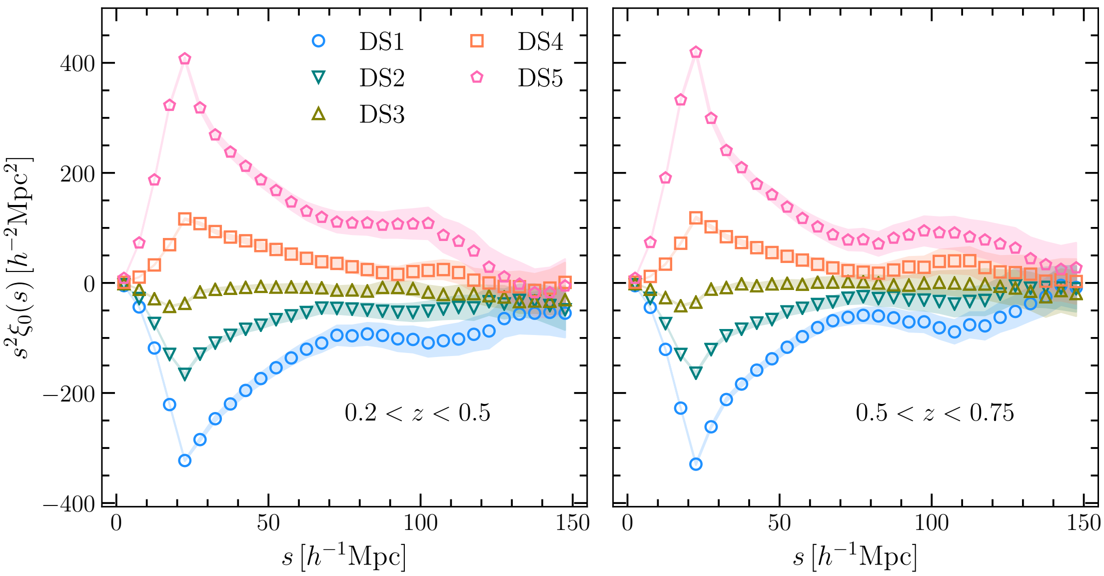

Figure 4.4: Monopole of the redshift-space CCF for different DS quintiles. The markers show measurements from the CMASSLOWZ NGC sample from $0.2 < z < 0.5$ (left-hand side panel) and $0.5 < z < 0.75$ (right-hand side panel). The shaded regions represent 1-σ errors estimated from the MD-Patchy mocks. The amplitude of the profiles are scaled by the square of the separation distance to highlight features at larges scales.

are much more subtle now due to the lower signal to noise, so it is expected that they won't contain significant cosmological constraining power. Nevertheless, I will include them in the analysis for the sake of completeness and consistency with the previous analysis on MINERVA.

As discussed in Chapter 2, the redshift-space CCF prediction coming from the Gaussian streaming model requires 2 ingredients: the real-space $\xi(r|\Delta^i)$ and the line-of-sight velocity dispersion profile $\sigma_v(s,\mu|\Delta^i)$. As we have run reconstruction on both the observational sample and the MD-Patchy mocks, we have two $\xi(r|\Delta^i)$ estimates at our disposal. For our model predictions, we use the one estimated from the Patchy mocks, as it is less noisy due to the possibility of averaging this measurement over multiple realizations of the survey. For the LOS velocity dispersion, $\sigma_v(s,\mu|\Delta^i)$, it is not possible to measure it from the observed galaxy sample nor the MD-Patchy mocks. A possibility would be adopt the profiles measured from

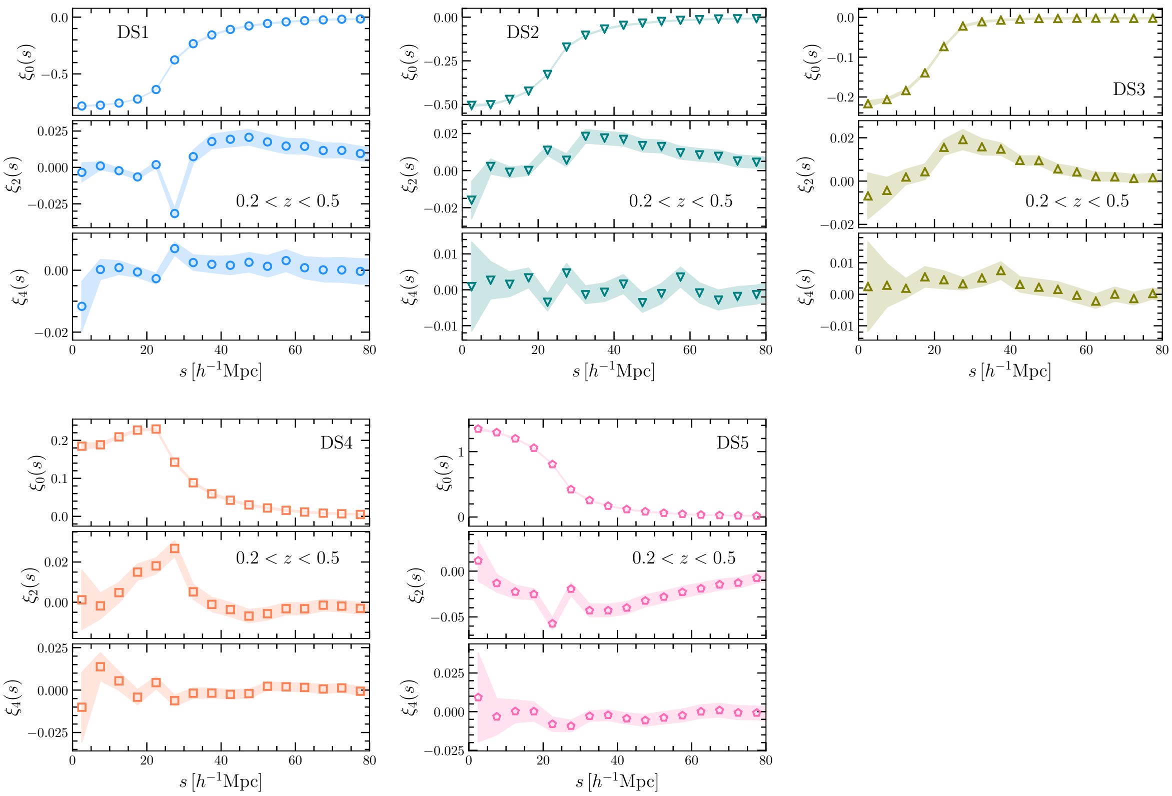

Figure 4.5: Each panel shows the multipoles of the redshift-space cross-correlation function of different DS quintiles at $0.2 < z < 0.5$. The markers show measurements from observations, while the shaded regions show the 1-σ error bars estimated from the MD-Patchy mocks.

the MINERVA simulations presented in Sec. 3.1.1. However, the HOD samples that were built for Minerva only match the clustering of the CMASS sample, and were generated at an effective redshift that differs from that of the redshift slices we use here. We instead assume a constant velocity dispersion profile, which is not a bad approximation for large scales (see Fig. 3.4).

Figs. 4.7 & 4.8 show the multipole decomposition of $\xi(s, \mu|\Delta^i)$ from one of the realizations of the MD-Patchy mocks. The solid line shows the predicted multipoles from the GSM, using the true $f\sigma_{12}$ and $b\sigma_{12}$ from the mocks, as well as the best-fit ν and σ_v parameters, assuming there are no AP distortions in the measurement. The agreement between model and data for DS1 is quite good across a wide range

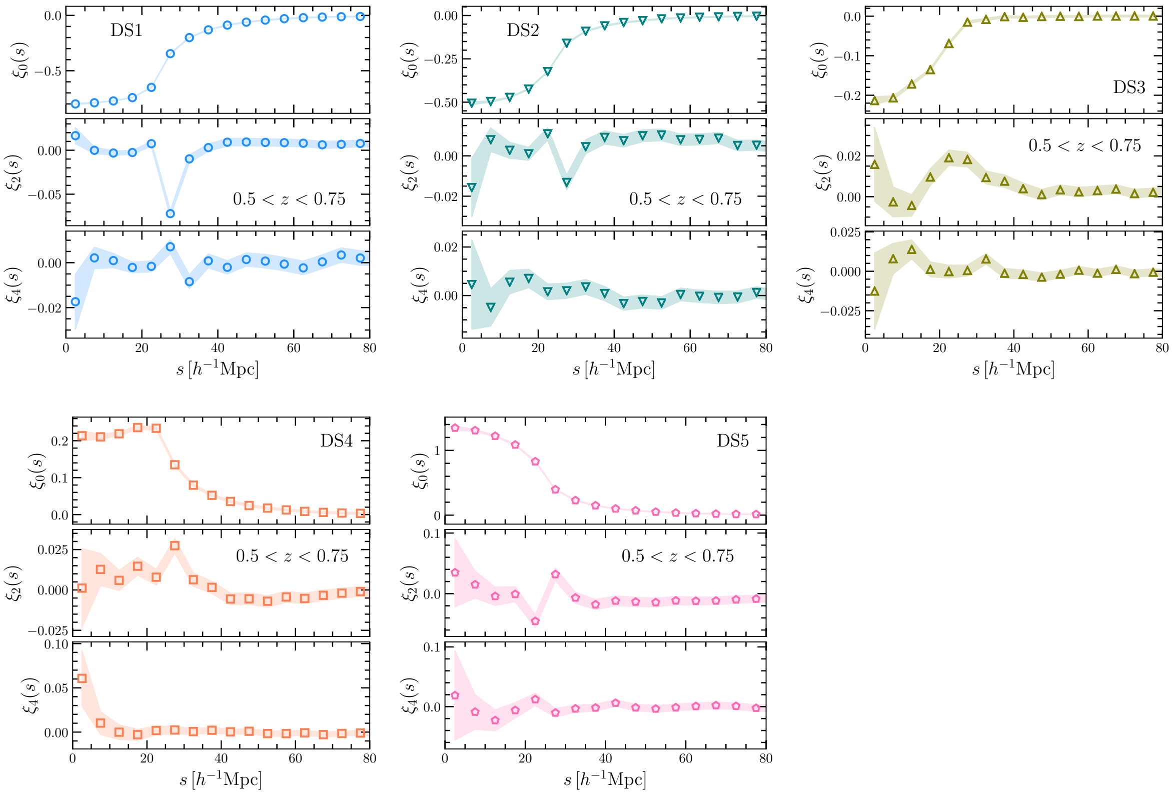

Figure 4.6: Same as Fig. 4.5, but showing multipoles from $0.5 < z < 0.75$.

of scales, even down to $\approx 1\,h^{-1}$Mpc, showing the good performance of the model at predicting the dynamics of underdense regions. The fit to the higher redshift voids appears to be better than its lower redshift counterpart at small scales. We notice that the DS1 quadrupole seems to be slightly under-predicted at large scales, although this deviation is still consistent within the error bars. For DS5, the model prediction is accurate for the monopole across most scale ranges, but the agreement with the quadrupole severely breaks down around $30\,h^{-1}$Mpc, which corresponds to a larger scale than what we had seen earlier with the N-body simulations. This discrepancy could be due to the assumption of a constant LOS velocity dispersion profile, which starts to have a pronounced slope around this scale. The intermediate quintiles, on the other hand, break down on much larger scales, around $60\,h^{-1}$Mpc, similar to what we had observed with the N-body mocks. This is likely due to the inability of the empirical model to accurately predict the streaming velocity for intermediate quantiles, an issue that could be related to DM-galaxy velocity bias for these samples, as discussed in Sec. 3.1.5. The hexadecapole has a low signal to noise for all quintiles and redshift bins, with the model prediction essentially being consistent with zero across most scales. The measured ξ_4 for DS5 at $0.2 < z < 0.5$ appears to be systematically offset from the model prediction across all scales. As mentioned in earlier, the MD-Patchy mocks are not guaranteed to match all multipole moments of the void-galaxy or cluster-galaxy CCF from the BOSS DR12 sample, as they were not calibrated considering these summary statistics. In fact. going back to Fig. 4.5, we see that the measured DS5 hexadecapole from the CMASSLOWZ sample is consistent with zero. This potential artifact in the mocks could also explain the small under-prediction of the voids quadrupole at large scales, which should be very easy to fit with the GSM if the true cosmology is known.

As mentioned earlier, I have measured all the clustering statistics for a grid of β for which I ran reconstruction. In order to highlight how this helps to constrain this parameter, Fig. 4.9 shows the measured and predicted quadrupoles for DS1, taken directly from the MD-Patchy mocks. The different coloured lines show the quadrupole that is obtained when using a reconstructed catalogue that was generated adopting a particular β value. We can see that the amplitude of the features in the measured quadrupole increases for lower β values, while the opposite trend is seen for the model prediction. As the model will only be a good fit to the measurement for values of β that are close to the true cosmology (such as in Fig. 4.8), this helps to put constraints on the parameter combinations that will be explored in the likelihood analysis.

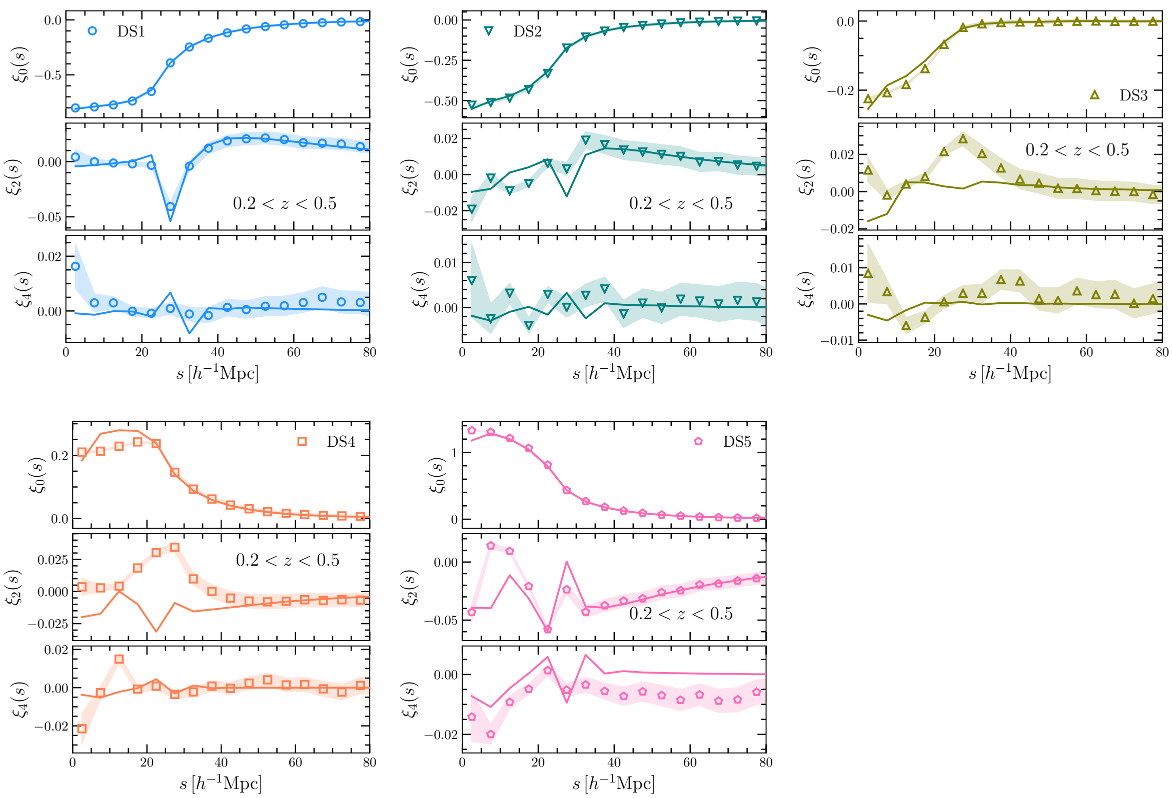

Figure 4.7: Multipoles of the DS redshift-space CCFs from the MD-Patchy mocks. The markers show the measurements from the mock catalogues, while the solid lines show the predictions from the GSM, using the true cosmology from the simulations.

4.4 Likelihood analysis

The setup for the likelihood analysis is similar to the one described in Sec. 3.2. The parameter space is given by

$$\theta_{\mathrm{DS}} = \left(f\sigma_{12}, b\sigma_{12}, \sigma_v, q_\perp, q_\parallel, \nu^1, ..., \nu^n\right). \tag{4.7}$$

As I measure $\xi(r|\Delta^i)$ from the MD-Patchy mocks, I allow for a rescaling of its amplitude at each MCMC step as

$$\xi(r|\Delta^i) = b\sigma_{12}\frac{\xi(r|\Delta^i)^{\mathrm{mock}}}{(b\sigma_{12})^{\mathrm{mock}}}, \tag{4.8}$$

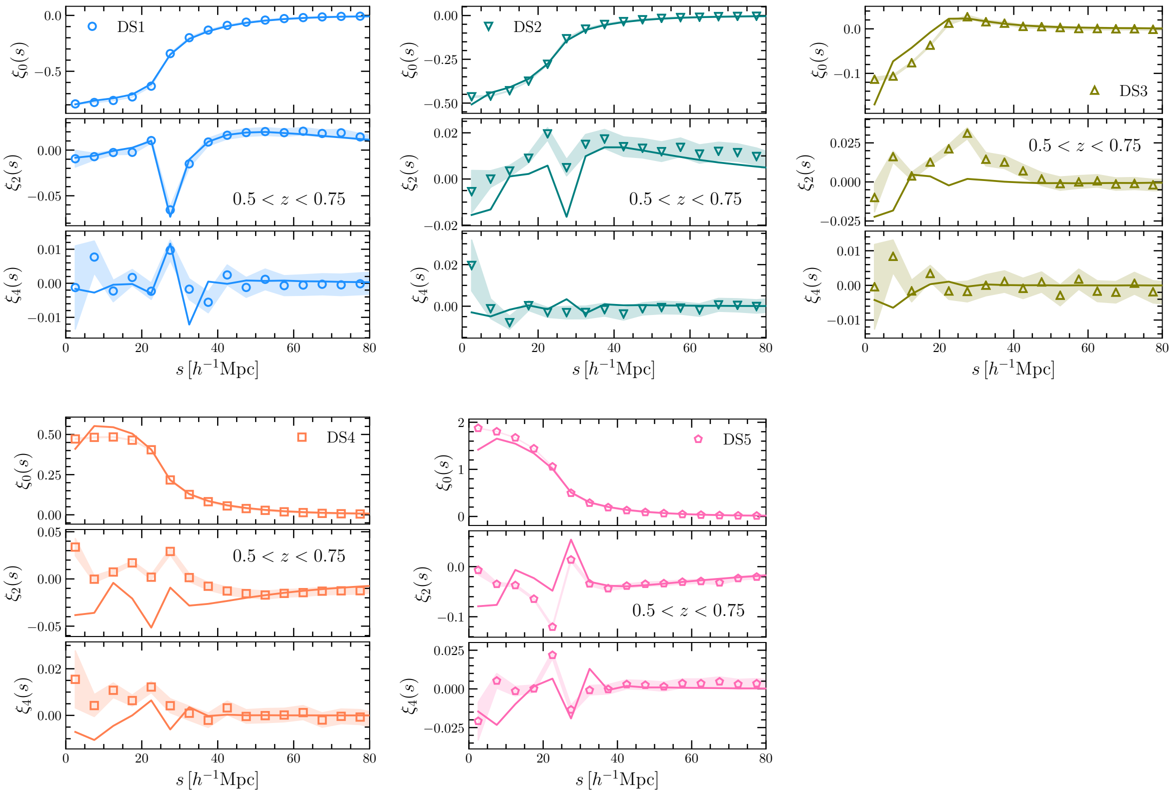

Figure 4.8: Sames as Fig. 4.7, but showing results from $0.5 < z < 0.75$.

where 'mock' denotes the quantity measured from the mock catalogues.

As an intermediate step between the likelihood analysis from N-body mocks and the CMASSLOWZ sample, I first run the MCMC pipeline on the MD-Patchy mocks, in order to verify if my model is able to recover the true cosmology of these mock catalogues that not only match the clustering and number density of CMASSLOWZ, but also try to mimic the observational systematics present in the sample. To do so, I average the clustering measurements over multiple MD-Patchy realizations, in a similar fashion as was done for the Minerva simulations. Motivated by the agreement between the measured and predicted multipoles for the MD-Patchy data in the previous chapter, I perform the analysis on the DS1+2+4+5 combined dataset. I adopt a minimum fitting scale of $s_{\min} = 0$, 60, 60 and 30 h^{-1}Mpc for each quintile, respectively, whereas the maximum fitting scale is 150 h^{-1}Mpc for all quantiles.

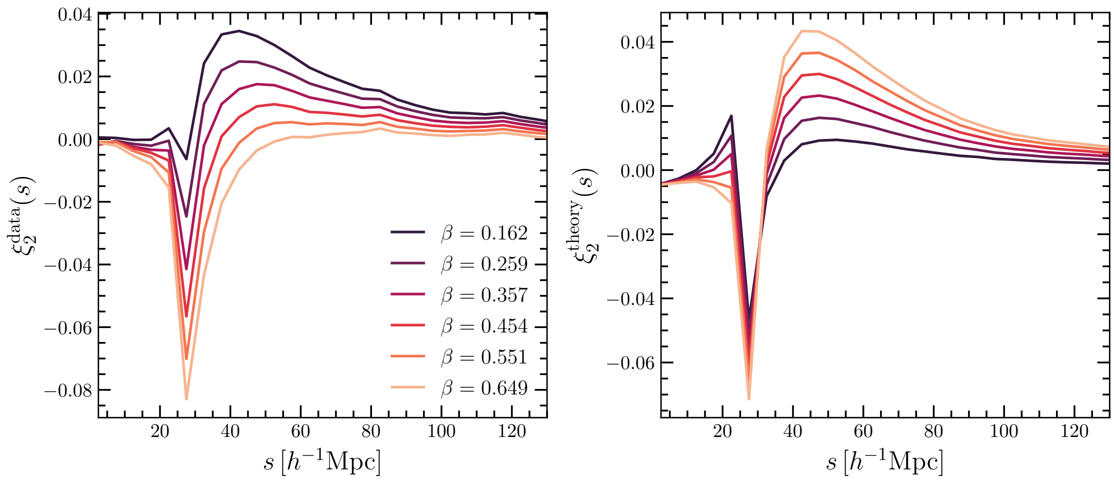

Figure 4.9: The measured (left) and predicted (right) quadrupole of the DS1 redshift-space CCF from the MD-Patchy mocks. The different curves show the dependence of the quadrupole with the adopted value of β in reconstruction. Only at those values close to the true β, the model will be a good fit to the measured ξ_2.

Fig. 4.10 shows the posterior distributions for the main model parameters for the two redshift bins, obtained by fitting the MD-Patchy mocks. For $0.5 < z < 0.75$, the precision for $f\sigma_{12}$ is around 2 per cent, and the true cosmology from the mocks perfectly agrees with the model fit. For $0.2 < z < 0.5$, the precision is close to 3 per cent, but the true $f\sigma_{12}$ lies more than 2-σ away from the best-fit value. This hints at the possibility that there are different sources of systematic errors affecting these independent redshift bins. The constraints for the AP parameters are below 1 per cent for both redshift bins. While the posterior for $q_\perp$ agrees with the expected value of 1 to within 1-σ (there should be no AP distortions under this setup, where we used the true simulation cosmology to map redshifts to distances), the best fit values for $q_\perp$ are biased high with respect to the true value. However, we have to consider that the precision for $q_\perp$ is around 0.2 per cent, greatly exceeding the recovered precision by most other cosmological probes in the literature. As we did not observe a strong level of bias in the model for the constraints of $f\sigma_{12}$ or the AP parameters when testing our methodology on the N-body Minerva mocks, we also need to consider the possibility

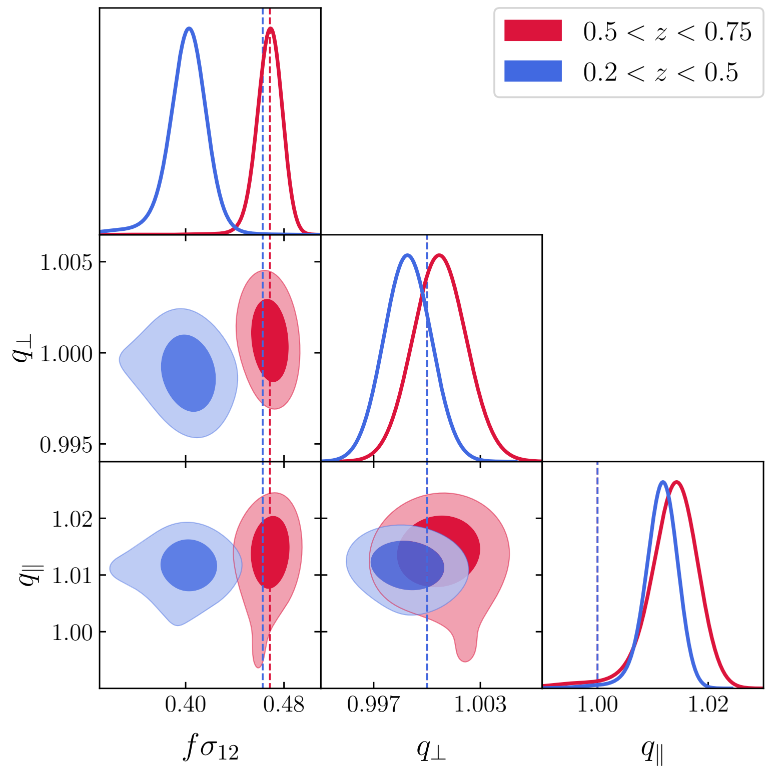

Figure 4.10: Parameter constraints obtained by running the likelihood analysis on the MD-Patchy mocks. Blue and red colours show results for $0.2 < z < 0.5$ and $0.5 < z < 0.75$, respectively. Darker and lighter shades show 1-σ and 2-σ confidence regions around the best-fit values, respectively.

of potential imperfections in the MD-Patchy mocks playing a role here. The MD-Patchy mocks were created using the augmented Lagrangian perturbation theory simulation method (ALPT, Kitaura & Hess, 2013), and therefore are not guaranteed to reproduce the correct dynamics on small scales for some of the clustering statistics we have used in our likelihood analysis. This scenario seems likely, as I have verified that I obtain a similar level of bias between model and data when using larger minimum fitting scales $s_{\rm min}$, for which other sources of systematic errors, such as the assumption of a constant LOS velocity dispersion profile, should be less important. Another potential source of uncertainty is the use of reconstruction to remove RSD from the galaxy catalogues, and in particular, the use of iterative reconstruction aided

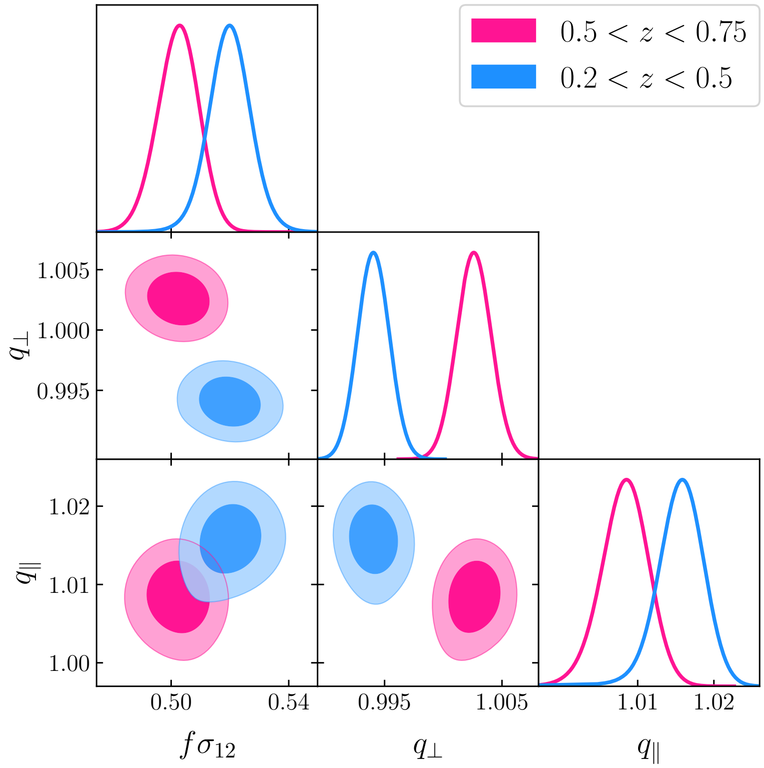

Figure 4.11: Similar to Fig. 4.10, but showing results for the CMASSLOWZ observational sample.

by interpolation, as explained in Sec. 4.3. While I have verified with the MINERVA simulations that the model predictions for the multipoles of the DS CCFs are still very accurate when using a reconstructed instead of a real-space galaxy catalogue, I have not yet examined how is the likelihood analysis in those mocks affected by this implementation.

Having observed the level of agreement between model and data for the higher redshift bin, and keeping in mind the systematic offset of the fit for the lower redshift sample, I proceed to fit the model to the observational dataset, using the combined DS1+2+4+5 measurements under the same setup presented above. Fig. 4.11 shows the parameter constraints when fitting the model to the CMASSLOWZ sample. For the lower redshift bin, I obtain $f\sigma_{12} = 0.519 \pm 0.006$, which corresponds to a precision of 1.2 per cent. For the AP parameters, I obtain $q_\perp = 1.002 \pm 0.001$ and

$q_{\parallel} = 1.008 \pm 0.002$, which translates to a 0.1 and 0.2 per cent precision, respectively. The higher redshift bin, on the other hand, yields $f\sigma_{12} = 0.468 \pm 0.008$, which is close to a 2 per cent precision, while the AP parameters are $q_{\perp} = 1.0 \pm 0.001$ and $q_{\parallel} = 1.014 \pm 0.003$.

To obtain the final constraints for $f\sigma_{12}$, I add the statistical error found in the MCMC fit to the CMASSLOWZ sample, $\delta f\sigma_{12,\mathrm{STAT}}$, to the systematic error found in the fit MD-Patchy mocks, $f\sigma_{12,\mathrm{SYS}}$ defined as the difference between the best-fit value and the true cosmology from the mocks, as

$$\delta f\sigma_{12,\mathrm{tot}} = \sqrt{\delta f\sigma_{12,\mathrm{stat}}{}^2 + \delta f\sigma_{12,\mathrm{sys}}{}^2}\,. \tag{4.9}$$

I plot my estimate of $f\sigma_{12}$ against values reported in other studies at different redshifts in Fig. 4.12. I also extrapolate the best-fit value from Planck (Planck Collaboration et al., 2016) to different redshifts, assuming a ΛCDM model, for comparison. The precision of the constraint at $z_{\mathrm{eff}} = 0.38$ is similar to that of other studies at a similar redshift, and is consistent with Planck at the 1-σ level. The size of this error bar is largely driven by the systematic error that was found in the model when fitting the MD-Patchy mocks. One could argue that since we did not see a significant bias in the modelling when analyzing the data from N-body simulations, it is not advisable to estimate the systematic uncertainty from the MD-Patchy mocks alone, which do not necessarily reproduce the correct dynamics of the statistics I am probing in this study. However, my calculations in the MINERVA simulations were done at $z_{\mathrm{eff}} = 0.54$, so it is difficult to extrapolate those results to the low redshift bin of CMASSLOWZ. A more thorough analysis, that unfortunately lies beyond the timeline of this thesis, would be benefited from the construction of HOD catalogues based from N-body simulations at a simulation snapshot that is closer to $z_{\mathrm{eff}} = 0.38$, which would allow us to examine if the biased estimation for $f\sigma_{12}$ at this redshift is a failure of the RSD model, or if it as artifact of the mocks themselves. The measurement at $z_{\text{eff}} = 0.61$, on the other hand, has a much higher precision than other RSD studies around the same redshift, and predicts a higher rate of structure formation. This estimate only agrees with Planck at the 2-σ level. Since all data points in this figure adopt ΛCDM as their baseline model, their dispersion highlights how the different assumptions in the modelling, as well as the characteristics of the surveys themselves, introduce different systematic uncertainties in the same measured parameter.

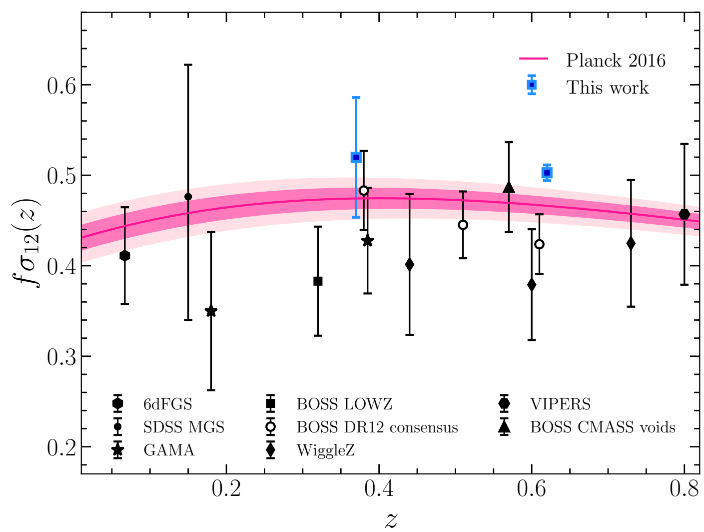

Figure 4.12: Comparison of the constraint on $f\sigma_{12}$ between this work and previous analyses in the literature, including the 6dFGS (Beutler et al., 2012), GAMA (Blake et al., 2013), WiggleZ (Blake et al., 2011), VIPERS (Pezzotta et al., 2017) (Howlett et al., 2015) and BOSS DR12 analyses (Gil-Marín et al., 2016; Alam et al., 2017; Nadathur et al., 2019a).

Chapter 5

Discussion and conclusions

I have presented a new method to analyse the signature of RSD in clustering measurements inferred from galaxy redshift surveys. Instead of the conventional 2PCF, I propose using the combination of a series of CCFs. These are CCFs obtained by splitting random positions according to the local galaxy density, and cross-correlating those positions with the entire galaxy field, thus using the exact same galaxy sample as in the 2PCF. This method builds upon the idea of DS statistics from weak lensing analysis (Gruen et al., 2016, 2018; Friedrich et al., 2018) and the density-dependent halo clustering presented in Tinker (2007), and generalises the modelling of RSD around voids to environments of different local densities. The algorithm can be summarised as follows:

1. Smooth the real-space galaxy number density field with a spherical top-hat window in a number of randomly selected locations, and split the positions into quintiles according to the smoothed densities.

2. Cross-correlate positions from each quintile with the galaxy number density field in redshift space, and model the RSD in their clustering pattern using the GSM.

3. Use MCMC to jointly fit RSD and Alcock-Paczynski distortions for each quin-

tile, and we combine their information to obtain cosmological constraints.

I have started by testing and validating the model using a suite of N-body mock galaxy catalogues that mimic the number density and clustering properties of the BOSS CMASS galaxy sample. The main conclusions from this framework validation can be listed as follows:

- The velocity field within each quintile of split density is close to Gaussian, and this meets the key condition necessary for the GSM, providing the physical reason for it to perform well at small scales.

- The GSM can fit the CCFs for every DS quintile with a remarkable accuracy. It appears to work well at small scales (below 20-30 h^{-1}Mpc) where the same model, when applied to 2PCF, has stronger deviations from simulation. I have also shown that recent models for the void-galaxy CCF (Cai et al., 2016; Nadathur et al., 2019b) have no fundamental difference from the Gaussian streaming model, and can work equally well for DS.

- By comparing the parameter constraints obtained with combining different DS quintiles, I find that most of the cosmological information is contained in the cross-correlations between the lowest and highest density quintiles (voids and clusters) with the galaxy field. Consequently, there is no major loss of constraining power when discarding the intermediate quintiles.

- On large scales, where the density field is is nearly Gaussian, there is no additional gain of information when using DS, compared to the traditional RSD analysis using the galaxy 2PCF.

- On scales below $\sim 30\,h^{-1}$Mpc, the non-Gaussian features of the density and velocity field start to become important, and DS outperforms the constraints from the galaxy 2PCF for $f\sigma_{12}$, yielding a $\sim$7 per cent constraint with $s_{\min} =$

$20\,h^{-1}$Mpc. It is approximately 30 per cent better in precision than the 2PCF. DS is most sensitive to geometrical distortions and can constrain the AP distortion parameters at the sub-per cent level, and up to ~ 6 times better than the 2PCF with $s_{\rm min} = 15\ h^{-1}$Mpc.

Having tested the proposed methodology with N-body mocks, I applied the density split algorithm to the combined BOSS CMASS and LOWZ sample from SDSS DR12. I split the galaxy sample into two non-overlapping redshift bins, from $0.2 < z < 0.5$ and $0.5 < z < 0.75$. I removed the RSD from the galaxy samples by using a density field reconstruction algorithm, and used the reconstructed galaxy field as a pseudo-real space galaxy catalogue, which is necessary for the application of the density split algorithm. The main conclusions from the observational analysis can be summarised as follows:

- Reconstruction is able to accurately remove RSD from the observational galaxy catalogues, and the multipoles of the DS CCFs closely resemble what was found in the N-body simulations. The dependence of the reconstruction algorithm on the linear growth parameter β helps to constrain this parameter in the joint RSD and AP likelihood analysis.

- The multipoles of the redshift-space CCF of split densities can be accurately predicted for the lowest density quantile (i.e. cosmic voids) across all scale ranges, and down to $\sim 30\,h^{-1}$Mpc for the highest density quantile (i.e. cluster-like regions). The intermediate quantiles are well-fit down to a scale of $\sim 60\,h^{-1}$Mpc.

- Using a suite of N-body mocks that mimic the clustering, number density and selection function of the CMASSLOWZ sample, I verified that the GSM is able to recover the true $f\sigma_{12}$ value from the mocks without bias and with a ~ 2 per cent precision for $0.5 < z < 0.75$. However, the estimate for $0.2 < z < 0.5$

is biased low with respect to the mocks cosmology. A possible explanation for this offset is a potential artifact in the mock catalogues, which might be unable to reproduce the dynamics that drive the main clustering statistics this study. There are also uncertainties introduced by reconstructing the galaxy density field, and the assumption of a constant line-of-sight galaxy velocity dispersion profile in the modelling. I add this offset to the systematic error budget of the DS method for this redshift bin.

- By jointly fitting RSD and geometrical distortions to the CMASSLOWZ data, DS is able to constrain $f\sigma_{12}$ at the 2 per cent level for $0.5 < z < 0.75$, which constitutes an unprecedented precision compared to other probes at a similar redshift. This estimate is in a 2-σ tension with Planck, and lies above the constraints from other RSD studies that use the same data. The DS constraint at $0.2 < z < 0.5$ has a similar precision as other RSD studies around the same redshift (around 12 per cent), and the error is mainly driven by the systematic offset between the model and data found in the mock analysis.

As seen both in the theoretical and observational analysis of this thesis, the advantage of DS over the galaxy 2PCF is most significant for AP parameters. DS can reduce the errors on $q_\perp$ and $q_\parallel$ by a few times compared to the 2PCF at $s_{\rm min} \sim 15\,h^{-1}$Mpc. This is likely due to the diverse and steep slopes for the monopoles of CCFs in DS near the smoothing scale of the top-hat filter. The constraints on AP parameters can be translated into constraints on parameters of the standard ΛCDM model. For example, varying only $\Omega_{\rm m}$ and σ_{12}, and all other cosmological parameters fixed in the model, I find the marginalised constraints with DS on $\Omega_{\rm m}$ and σ_{12} under these somewhat strict assumptions to be a factor of ~ 3 and $\sim 40\%$ tighter over the 2PCF, respectively.

The sensitive response to cosmology of the tails of the density distribution has already been suggested in previous papers (e.g. White, 1979; Saslaw & Hamilton, 1984; Fry, 1985, 1986; Zorrilla Matilla et al., 2020). Given that the most significant constraining power is coming from the extreme DS quintiles, it would be interesting to address whether the voids and clusters that are identified with the DS method provide tighter constraints than those identified with dedicated void-finding algorithms, such as the Spherical Void Finder (Padilla et al., 2005) and the Watershed Void Finder (van de Weygaert & Platen, 2011), or group-finding algorithms, such as redMaPPer (Rykoff et al., 2014). The selection criteria for these regions could play an important role in the modelling side.

RSD with split densities is general. Although it is similar to the combination of void-galaxy and cluster cross-correlations, it does not depend on a specific void or cluster finding algorithm, and can be predicted from first principles. It is essentially a series of conditioned correlation functions, which can be predicted once the PDF of the density field and the real-space correlation function are known (Abbas & Sheth, 2005; Shi & Sheth, 2018; Neyrinck et al., 2018; Friedrich et al., 2018). Given the success of models that allow us to predict the PDF of a smoothed density field (Uhlemann et al., 2016; Repp & Szapudi, 2018; Jamieson & Loverde, 2020), and the existing model framework for predicting CCFs, it is hopeful that the model prediction can be achieved with even higher accuracy in future studies.

When analyzing the N-body mock catalogues, I found that at small scales, e.g. below $15\,h^{-1}$Mpc, the GSM starts to be biased at the precision of interest. I suspect that multiple reasons can lead to the complication at those small scales: 1. the linear galaxy bias model starts to break down; .2 higher order terms in the streaming model start to become important, especially for DS5 (cluster-galaxy CCF), where the velocity dispersion and its derivative may start to be significant (Fisher, 1995; Reid & White, 2011); 3. as a long positive tail in the density PDF is expected, splitting the density with 5 quantiles may not be sufficient, as the highest density quantile will have a wide range of densities, and this will violate the Gaussian condition for the velocities.

The linear coupling between the density and velocity is insufficient. Forthcoming large-area surveys, such as Euclid (Laureijs et al., 2011) and DESI (Levi et al., 2019) will generate data sets that will be several factors larger in volume than current surveys, requiring per cent-level precision on the modelling side. It is certainly needed to go beyond the linear coupling. I have tested employing spherical dynamics for modelling the velocity profiles, finding some improvements over the linear coupling model, as also elucidated by previous work for voids (e.g. Demchenko et al., 2016), but the accuracy is still insufficient for this purpose. The empirical model for the coupling allows me to apply the GSM to a relatively small scale, at the cost that the cosmological constraints have been penalised by marginalising over nuisance parameters. Given the potential gain of cosmological information at small scales, improving the modelling for the velocity and density coupling model may be very rewarding.

Throughout this work, I have not yet attempted the combination of DS CCFs with the 2PCF, nor have I involved the PDFs of the density field for cosmological constraints. Given the apparent different orientations for contours of the posteriors between DS and the 2PCF, e.g. Figs. A1 & A2, it is hopeful that their combination would offer even tighter constraints for cosmology. This can be explored when a larger suit of N-body simulations is employed, allowing a more reliable and accurate covariance matrix to be constructed; or having an accurate analytical covariance matrix (Grieb et al., 2016). Likewise, the 3D PDFs of the density field, i.e. the counts-in-cells statistics, have been demonstrated to be complementary for parameter constraints (e.g. Uhlemann et al., 2020), especially for breaking degeneracy between b and σ_{12} (e.g. Repp & Szapudi, 2020). It is compelling to combine DS, the 2PCF, and counts in cells all together, which may one of the best ways to extract cosmological information from the initial conditions at small scales without explicitly employing higher order statistics. Developing the joint covariance for the three different cosmological probes is planned as a continuation of this work.

Appendix

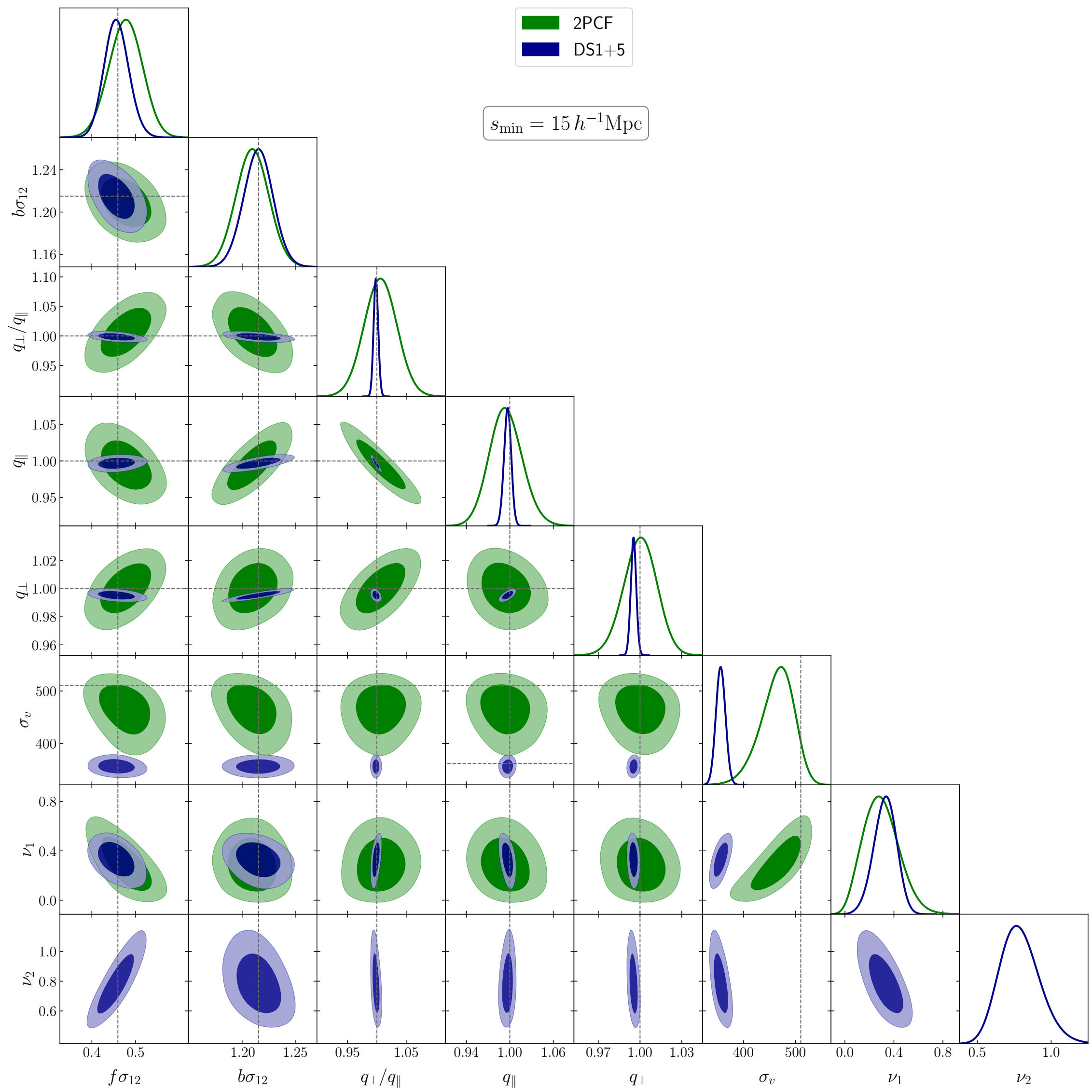

Figure A1: Similar to Fig. 3.12, but showing the posterior distributions for all parameters included in the fit, with the exception of the second nuisance parameter ν_2 in DS. The minimum fitting scale used was $s_{\min} = 15\,h^{-1}$Mpc.

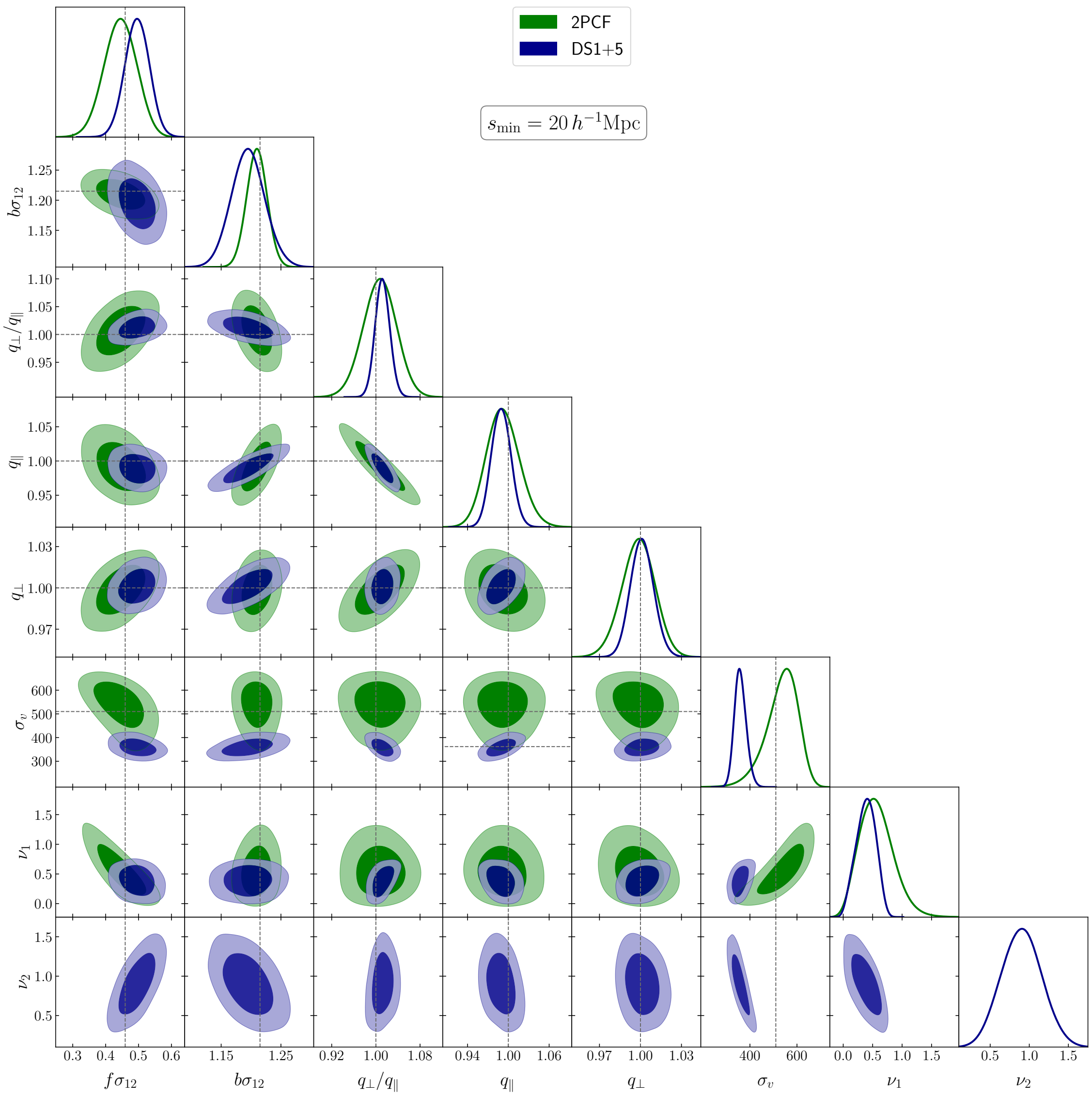

Figure A2: Similar to Fig. A1, but showing the posterior distributions obtained using a minimum fitting scale of $s_{\rm min} = 20\,h^{-1}{\rm Mpc}$.

Bibliography

Abbas U., Sheth R. K., 2005, MNRAS, 364, 1327

Achitouv I., 2019, Phys. Rev. D, 100, 123513

Achitouv I., Blake C., Carter P., Koda J., Beutler F., 2017, Phys. Rev. D, 95, 083502

Alam S., et al., 2015, ApJS, 219, 12

Alam S., et al., 2017, MNRAS, 470, 2617

Alcock C., Paczynski B., 1979, Nature, 281, 358

Angulo R. E., White S. D. M., 2010, MNRAS, 405, 143

Ballinger W. E., Peacock J. A., Heavens A. F., 1996, MNRAS, 282, 877

Bautista J. E., et al., 2021, MNRAS, 500, 736

Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 2000, MNRAS, 311, 793

Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587

Bernardeau F., Kofman L., 1995, ApJ, 443, 479

Beutler F., et al., 2012, MNRAS, 423, 3430

Beutler F., et al., 2017, MNRAS, 466, 2242

Bianchi D., Percival W. J., Bel J., 2016, MNRAS, 463, 3783

Blake C., et al., 2011, MNRAS, 415, 2876

Blake C., et al., 2013, MNRAS, 436, 3089

Cai Y.-C., Taylor A., Peacock J. A., Padilla N., 2016, MNRAS, 462, 2465

Cai Y.-C., Neyrinck M., Mao Q., Peacock J. A., Szapudi I., Berlind A. A., 2017, MNRAS, 466, 3364

Cautun M., Paillas E., Cai Y.-C., Bose S., Armijo J., Li B., Padilla N., 2018, MNRAS, 476, 3195

Chen S.-F., Vlah Z., Castorina E., White M., 2020, arXiv e-prints, p. arXiv:2012.04636

Chiang C.-T., Wagner C., Sánchez A. G., Schmidt F., Komatsu E., 2015, JCAP, 2015, 028

Clampitt J., Jain B., 2015, MNRAS, 454, 3357

Coles P., Jones B., 1991, MNRAS, 248, 1

Colombi S., 1994, ApJ, 435, 536

Copeland E. J., Sami M., Tsujikawa S., 2006, International Journal of Modern Physics D, 15, 1753

Correa C. M., Paz D. J., Padilla N. D., Ruiz A. N., Angulo R. E., Sánchez A. G., 2019, MNRAS, 485, 5761

Correa C. M., Paz D. J., Sánchez A. G., Ruiz A. N., Padilla N. D., Angulo R. E., 2020, MNRAS,

Croft R. A. C., Dalton G. B., Efstathiou G., 1999, MNRAS, 305, 547

Cuesta-Lazaro C., Li B., Eggemeier A., Zarrouk P., Baugh C. M., Nishimichi T., Takada M., 2020, arXiv e-prints, p. arXiv:2002.02683

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

Demchenko V., Cai Y.-C., Heymans C., Peacock J. A., 2016, MNRAS, 463, 512

Einasto J., Hütsi G., Liivamägi L. J., Einasto M., 2020, arXiv e-prints, p. arXiv:2011.13292

Feldman H. A., Kaiser N., Peacock J. A., 1994, ApJ, 426, 23

Fisher K. B., 1995, ApJ, 448, 494

Friedrich O., et al., 2018, Phys. Rev. D, 98, 023508

Fry J. N., 1985, ApJ, 289, 10

Fry J. N., 1986, ApJ, 306, 358

Gil-Marín H., Wagner C., Verde L., Porciani C., Jimenez R., 2012, JCAP, 2012, 029

Gil-Marín H., et al., 2016, MNRAS, 460, 4210

Grieb J. N., Sánchez A. G., Salazar-Albornoz S., Dalla Vecchia C., 2016, MNRAS, 457, 1577–1592

Gruen D., et al., 2016, MNRAS, 455, 3367

Gruen D., et al., 2018, Phys. Rev. D, 98, 023507

Gunn J. E., et al., 1998, AJ, 116, 3040

Gunn J. E., et al., 2006, AJ, 131, 2332

Hamaus N., Sutter P. M., Wandelt B. D., 2014, Phys. Rev. Lett., 112, 251302

Hamaus N., Sutter P. M., Lavaux G., Wandelt B. D., 2015, JCAP, 1511, 036

Hamaus N., Pisani A., Sutter P. M., Lavaux G., Escoffier S., Wandelt B. D., Weller J., 2016, Phys. Rev. Lett., 117, 091302

Hamaus N., Cousinou M.-C., Pisani A., Aubert M., Escoffier S., Weller J., 2017, JCAP, 2017, 014

Hamaus N., Pisani A., Choi J.-A., Lavaux G., Wandelt B. D., Weller J., 2020, arXiv e-prints, p. arXiv:2007.07895

Hartlap J., Simon P., Schneider P., 2007, A&A, 464, 399

Hawken A. J., Aubert M., Pisani A., Cousinou M.-C., Escoffier S., Nadathur S., Rossi G., Schneider D. P., 2020, JCAP, 2020, 012

Higuchi Y., Oguri M., Hamana T., 2013, MNRAS, 432, 1021

Hou J., et al., 2018, MNRAS, 480, 2521

Howlett C., Ross A. J., Samushia L., Percival W. J., Manera M., 2015, MNRAS, 449, 848

Jackson J. C., 1972, MNRAS, 156, 1P

Jamieson D., Loverde M., 2020, arXiv e-prints, p. arXiv:2010.07235

Jenkins A., 2010, MNRAS, 403, 1859

Jennings E., Baugh C. M., Pascoli S., 2011, MNRAS, 410, 2081

Joyce A., Jain B., Khoury J., Trodden M., 2015, Phys. Rept., 568, 1

Juszkiewicz R., Springel V., Durrer R., 1999, ApJ, 518, L25

Kaiser N., 1987, MNRAS, 227, 1

Kitaura F. S., Hess S., 2013, MNRAS, 435, L78

Kitaura F.-S., et al., 2016, MNRAS, 456, 4156

Klypin A., Prada F., Betancort-Rijo J., Albareti F. D., 2018, MNRAS, 481, 4588

Kodwani D., Alsono D., Ferreira P., 2019, The Open Journal of Astrophysics, 2

Koyama K., 2016, Rep. Prog. Phys., 79, 046902

Krause E., Chang T.-C., Doré O., Umetsu K., 2013, ApJ, 762, L20

Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, ApJ, 609, 35

Kuruvilla J., Porciani C., 2018, MNRAS, 479, 2256

Laureijs R., Amiaux J., Arduini S., Augueères J. L., Brinchmann J., Cole R., et al., 2011, preprint (arXiv:1110.3193)

Levi M., et al., 2019, in Bulletin of the American Astronomical Society. p. 57 (arXiv:1907.10688)

Lilje P. B., Efstathiou G., 1988, MNRAS, 231, 635

Lilje P. B., Efstathiou G., 1989, MNRAS, 236, 851

Lilje P. B., Lahav O., 1991, ApJ, 374, 29

Linder E. V., Cahn R. N., 2007, Astroparticle Physics, 28, 481

Linder E. V., Polarski D., 2019, Physical Review D, 99

Lippich M., et al., 2019, MNRAS, 482, 1786

Mandal A., Nadkarni-Ghosh S., 2020, MNRAS, 498, 355

Manera M., et al., 2013, MNRAS, 428, 1036

Maraston C., et al., 2013, MNRAS, 435, 2764

Massara E., Sheth R. K., 2018, arXiv e-prints, p. arXiv:1811.03132

Melchior P., Sutter P. M., Sheldon E. S., Krause E., Wandelt B. D., 2014, MNRAS, 440, 2922

Mohammad F. G., de la Torre S., Bianchi D., Guzzo L., Peacock J. A., 2016, MNRAS, 458, 1948

Nadathur S., Percival W. J., 2019, MNRAS, 483, 3472

Nadathur S., Carter P. M., Percival W. J., Winther H. A., Bautista J. E., 2019a, Phys. Rev. D, 100, 023504

Nadathur S., Carter P., Percival W. J., 2019b, MNRAS, 482, 2459

Nadathur S., et al., 2020, MNRAS, 499, 4140

Neyrinck M. C., 2011, ApJ, 742, 91

Neyrinck M. C., Szapudi I., Szalay A. S., 2009, ApJ, 698, L90

Neyrinck M. C., Szapudi I., Szalay A. S., 2011, ApJ, 731, 116

Neyrinck M. C., Szapudi I., McCullagh N., Szalay A. S., Falck B., Wang J., 2018, MNRAS, 478, 2495

Nusser A., Davis M., 1994, ApJ, 421, L1

Okumura T., et al., 2016, PASJ, 68, 38

Padilla N. D., Ceccarelli L., Lambas D. G., 2005, MNRAS, 363, 977

Paillas E., Lagos C. D. P., Padilla N., Tissera P., Helly J., Schaller M., 2017, MNRAS, 470, 4434

Paillas E., Cautun M., Li B., Cai Y.-C., Padilla N., Armijo J., Bose S., 2019, MNRAS, 484, 1149–1165

Paillas E., Cai Y.-C., Padilla N., Sánchez A. G., 2021, MNRAS

Parejko J. K., et al., 2013, MNRAS, 429, 98

Paz D., Lares M., Ceccarelli L., Padilla N., Lambas D. G., 2013, MNRAS, 436, 3480

Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144

Peebles P. J. E., 1980, The large-scale structure of the universe

Percival W. J., et al., 2014, MNRAS, 439, 2531

Pezzotta A., et al., 2017, A&A, 604, A33

Planck Collaboration et al., 2014, A&A, 571, A16

Planck Collaboration et al., 2016, A&A, 594, A11

Pollina G., Hamaus N., Dolag K., Weller J., Baldi M., Moscardini L., 2017, MNRAS, 469, 787

Raghunathan S., Nadathur S., Sherwin B. D., Whitehorn N., 2020, ApJ, 890, 168

Reid B. A., White M., 2011, MNRAS, 417, 1913

Repp A., Szapudi I., 2018, MNRAS, 473, 3598

Repp A., Szapudi I., 2020, MNRAS, 498, L125

Rykoff E. S., et al., 2014, ApJ, 785, 104

Sánchez A. G., 2020, Phys. Rev. D, 102, 123511

Sánchez A. G., et al., 2013, MNRAS, 433, 1202

Sánchez A. G., et al., 2017a, MNRAS, 464, 1640

Sánchez C., et al., 2017b, MNRAS, 465, 746

Saslaw W. C., Hamilton A. J. S., 1984, ApJ, 276, 13

Scoccimarro R., 2004, Phys. Rev. D, 70, 083007

Sefusatti E., Crocce M., Pueblas S., Scoccimarro R., 2006, Phys. Rev. D, 74, 023522

Seldner M., Peebles P. J. E., 1977, ApJ, 215, 703

Seljak U., 2000, MNRAS, 318, 203

Seljak U., McDonald P., 2011, JCAP, 2011, 039

Sellentin E., Heavens A. F., 2016, MNRAS, 456, L132

Sheth R. K., Hui L., Diaferio A., Scoccimarro R., 2001a, MNRAS, 325, 1288

Sheth R. K., Diaferio A., Hui L., Scoccimarro R., 2001b, MNRAS, 326, 463

Shi J., Sheth R. K., 2018, MNRAS, 473, 2486

Slepian Z., Eisenstein D. J., 2015, MNRAS, 448, 9

Slepian Z., Eisenstein D. J., 2017, MNRAS, 469, 2059

Springel V., 2005, MNRAS, 364, 1105

Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726

Sutter P. M., Lavaux G., Wandelt B. D., Weinberg D. H., 2012, ApJ, 761, 187

Szapudi I., Pan J., 2004, ApJ, 602, 26

Tinker J. L., 2007, MNRAS, 374, 477

Uhlemann C., Codis S., Pichon C., Bernardeau F., Reimberg P., 2016, MNRAS, 460, 1529

Uhlemann C., Friedrich O., Villaescusa-Navarro F., Banerjee A., Codis S. r., 2020, MNRAS, 495, 4006

Vlah Z., White M., 2019, JCAP, 2019, 007

Vlah Z., Castorina E., White M., 2016, JCAP, 2016, 007

Wagner C., Schmidt F., Chiang C. T., Komatsu E., 2015, MNRAS, 448, L11

Wang X., Neyrinck M., Szapudi I., Szalay A., Chen X., Lesgourgues J., Riotto A., Sloth M., 2011, ApJ, 735, 32

White S. D. M., 1979, MNRAS, 186, 145

Yang X., Mo H. J., van den Bosch F. C., Weinmann S. M., Li C., Jing Y. P., 2005, MNRAS, 362, 711

Zentner A. R., Berlind A. A., Bullock J. S., Kravtsov A. V., Wechsler R. H., 2005, ApJ, 624, 505

Zhai Z., et al., 2019, ApJ, 874, 95

Zhao C., et al., 2020, MNRAS, 491, 4554

Zheng Z., Coil A. L., Zehavi I., 2007, ApJ, 667, 760

Zorrilla Matilla J. M., Sharma M., Hsu D., Haiman Z., 2020, arXiv e-prints, p. arXiv:2007.06529

Zu Y., Weinberg D. H., 2013, MNRAS, 431, 3319–3337

eBOSS Collaboration et al., 2020, arXiv e-prints, p. arXiv:2007.08991

van de Weygaert R., Platen E., 2011, Int. J. Mod. Phys., 1, 41