View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Portsmouth University Research Portal (Pure)

Mon. Not. R. Astron. Soc. 000, 1–13 (2009) Printed 30 April 2010 (MN LATEX style file v2.2)

The effect of redshift-space distortions on projected 2-pt clustering measurements

Kelly Nock⋆, Will J. Percival, Ashley J. Ross

Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth, PO1 3FX, UK

arXiv:1003.0896v2 [astro-ph.CO] 29 Apr 2010

30 April 2010

ABSTRACT

Although redshift-space distortions only affect inferred distances and not angles, they still distort the projected angular clustering of galaxy samples selected using redshift dependent quantities. From an Eulerian view-point, this effect is caused by the apparent movement of galaxies into or out of the sample. From a Lagrangian view-point, we find that projecting the redshift-space overdensity field over a finite radial distance does not remove all the anisotropic distortions. We investigate this effect, showing that it strongly boosts the amplitude of clustering for narrow samples and can also reduce the significance of baryonic features in the correlation function. We argue that the effect can be mitigated by binning in apparent galaxy pair-centre rather than galaxy position, and applying an upper limit to the radial galaxy separation. We demonstrate this approach, contrasting against standard top-hat binning in galaxy distance, using sub-samples taken from the Hubble Volume simulations. Using a simple model for the radial distribution expected for galaxies from a survey such as the Dark Energy Survey (DES), we show that this binning scheme will simplify analyses that will measure baryon acoustic oscillations within such galaxy samples. Comparing results from different binning schemes has the potential to provide measurements of the amplitude of the redshift-space distortions. Our analysis is relevant for other photometric redshift surveys, including those made by the Panoramic Survey Telescope & Rapid Response System (Pan-Starrs) and the Large Synoptic Survey Telescope (LSST).

Key words: cosmology: observations, distance scale, large-scale structure

1 INTRODUCTION

The late-time acceleration of the expansion of the Universe has been one of the most exciting cosmological discoveries in recent years (Riess et al. 1998; Perlmutter et al. 1999). Understanding the nature of this acceleration is one of the main challenges facing cosmologists. One of the key observational methods that will be used to help meet this challenge involves using Baryonic Acoustic Oscillations (BAO) in the 2-point galaxy clustering signal as a standard ruler to make precise measurements of cosmological expansion. The acoustic signature has now been convincingly detected (Percival et al. 2001; Cole et al. 2005; Eisenstein et al. 2005) using the 2dF Galaxy Redshift Survey (2dFGRS; Colless et al. 2003) and the Sloan Digital Sky Survey (SDSS; York et al. 2000). The detection has subsequently been refined using more data and better techniques, and is now producing interesting constraints on cosmological models (Percival et al. 2007a,b; Gaztanaga et al. 2008; Sanchez et al. 2009; Percival et al. 2009).

Some of the next generation of sky surveys, including the Dark Energy Survey (DES `www.darkenergysurvey.org`), the Panoramic Survey Telescope and Rapid Response System (PanStarrs `pan-starrs.ifa.hawaii.edu`), and the Large Synoptic Survey Telescope (LSST `www.lsst.org`), will use photometric techniques to estimate galaxy redshifts, rather than more precise estimates from spectroscopic emission lines. The larger uncertainties on galaxy redshifts induce errors on inferred distances in the radial direction. The amplitude of the power spectrum and correlation function is reduced in the radial direction by this smoothing, removing information. In this scenario, where little information remains from fluctuations in the radial direction, it makes sense to use the projected 2-pt functions in photometric-redshift slices as the statistics to compare with models (Padmanabhan et al. 2007; Blake et al. 2007). The projection does not completely remove problems caused by inferring distances from velocity data (i.e. working in redshift-space).

The distribution of galaxies that we observe in sky surveys, where we measure radial distances from spectroscopic or photometric redshifts, is not a true 3D picture. We observe an apparent clustering pattern in *redshift-space*, which is systematically different from the true distribution in *real-space* because redshifts of galaxies are altered from their Hubble flow values by peculiar velocities. For example, on large scales, the infall of galaxies onto

⋆ e-mail: kelly.nock@port.ac.uk (KN)

© 2008 RAS

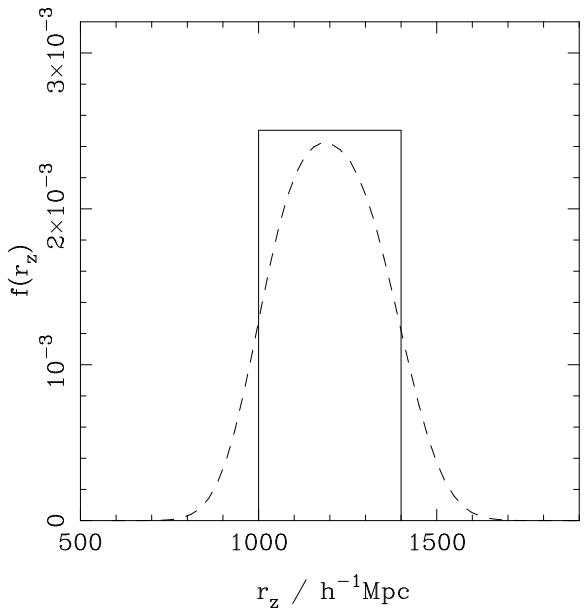

Figure 1. The radial distribution of galaxies selected in a bin of width $400\,h^{-1}\,\mathrm{Mpc}$, calculated using photometric redshifts to estimate distances (solid line). This is compared against the distribution of true distances to these galaxies (dashed line) assuming a photometric redshift error of $\sigma_z = 0.03(1+z)$. If the photometric redshifts of different galaxies are independent, then the expected projected correlation function of the photo-z selected sample, and a sample selected applying the dashed line as a selection function based on the true distances, are the same.

collapsed objects leads to an apparent enhancement of clustering in the radial direction as galaxies are projected along their velocity vectors (Kaiser 1987; Hamilton 1998). When we infer galaxy distances assuming that the total velocity relative to the observer comes from the Hubble expansion flow, the result is that we see a distorted (redshift-space) density field.

For angular measurements, these redshift-space distortions can alter the angular clustering in a redshift slice because the distortions are correlated across the direction of projection. Although redshift-space distortions are sub-dominant compared with photometric redshift uncertainties, they give rise to a systematic effect, which needs to be included when photometric redshift surveys are analysed (Padmanabhan et al. 2007; Blake et al. 2007). This can complicate the analysis as the size of the redshift-space distortions, and therefore of this effect, is dependent on the cosmological model. Consequently, for every model to be tested against the data, we need to make a revised estimate of the redshift-space effect.

In this paper, we consider the simplified problem in the plane-parallel approximation, and only consider linear redshift-space distortions. Both photometric redshift errors and the random motion of galaxies in clusters provide an additional convolution of the overdensity field along the radial direction. While these effects need to be corrected in any analysis, the required correction is easily modelled and can be separated from the linear redshift-space distortion effects. For a measurement of the projected clustering, including such effects is equivalent to simply broadening the radial window function with which the galaxies were selected. This is demonstrated in Fig. 1. A top-hat bin in photometric redshift gives the same expected projected correlation function as simply applying the convolved version of the bin as a selection function for the true distances. As we have to include a window function anyway, we simply assume in this paper that this window already includes the effects of both photometric redshift errors and the random motion of galaxies in clusters. In the following analysis, we therefore assume that there are no redshift errors without loss of generality.

The layout of our paper is as follows. In Section 2 we analyse the projected overdensity field and redshift-space effects upon it, both analytically (Section 2.1 & 2.3) and using Monte-Carlo simulations (Section 2.4). We then consider how the recovered correlation function depends on galaxy binning (Section 3). Mock catalogues drawn from the Hubble Volume simulation are constructed and analysed in Section 4 in order to validate this analytic work. We incorporate hybrid selection functions based on both real and redshift-space boundaries into the analysis in Section 5. In Section 6 we consider a non-uniform redshift distribution similar to that of future sky survey DES, and the realistic implementation of our work is discussed in Section 7.

2 PROJECTED 2-POINT STATISTICS OF THE OVERDENSITY FIELD

2.1 Correlation Function

In order to simplify the problem, we assume that the clustering strength does not change across the samples under consideration and make the plane-parallel (distant observer) approximation, with redshift-space distortions along the z-axis of a Cartesian basis. In the absence of redshift distortions the projected correlation function is given by:

$$\xi_p(d_p) = \langle \delta_p(\mathbf{r_p})\delta_p(\mathbf{r'_p})\rangle, \quad (1)$$

$$= \int\int dr_z dr'_z \phi(r_z)\phi(r'_z)\xi\left[d(r_z, r'_z, d_p)\right] \quad (2)$$

where $d(r_z, r'_z, d_p) = \sqrt{(r_z - r'_z)^2 + d_p^2}$, and subscripts x, y and z denote the direction along each Cartesian axis, and p denotes projected quantities $p \equiv xy$. $\phi(r_z)$ is the radial galaxy selection function, normalised such that $\int dr_z\ \phi(r_z) = 1$, and $\xi(d) = \langle \delta(\mathbf{r})\delta(\mathbf{r'})\rangle$, where $\delta(\mathbf{r})$ is the overdensity of galaxies at real-space position $\mathbf{r}$. Throughout our paper we use $\mathbf{r}$ to describe a galaxy position and $\mathbf{d}$ to describe the distance between two galaxies, so, for example, r_z is the position of a galaxy along the z-axis, while d_p is amplitude of the separation between two galaxies when projected into the x, y-plane.

In reality, our radial position is determined via a redshift. In this case, Eq. (2) must be altered to

$$\xi^s_p(d_p) = \langle \delta_p(\mathbf{s_p})\delta_p(\mathbf{s'_p})\rangle. \quad (3)$$

The weighted, projected overdensity field $\delta_p(r_p)$ can now be written

$$1 + \delta_p(\mathbf{r_p}) = \int ds_z\ \phi(s_z)[1 + \delta(\mathbf{s})], \quad (4)$$

where $\mathbf{s} = (\mathbf{r_p}, s_z)$ is the redshift-space position of each galaxy and $\phi(s_z)$ gives the galaxy selection function along the line of sight corresponding to s_z (e.g. Peebles 1980).

The difference between the projection in redshift-space and real-space is shown schematically in Fig. 2. An edge to a window function (or a contour of constant galaxy density) that is straight in redshift-space is systematically distorted in real-space. The edge of the bin is itself clustered with a non-negligible projected correlation function, i.e. the real-space boundary has a correlation function that depends on $\mathbf{r_p}$. The inclusion or exclusion of galaxies is balanced in terms of the 3D correlation function within the boundary; while

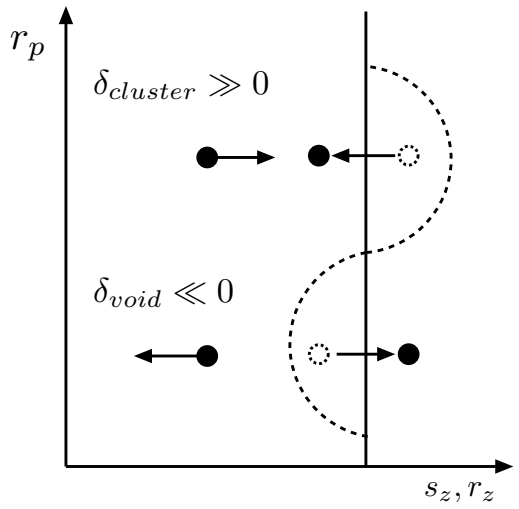

Figure 2. Schematic showing the boundary of a region selected in redshift-space (solid line) compared with the boundary of the same region in real-space (dashed line). The boundary is distorted in real-space around an overdensity and an underdensity. The positions of two galaxies whose apparent motion crosses the boundary are shown in redshift-space (solid circles) and in real-space (dashed circles). Note that, in this simplified picture where the under and overdensities have the same amplitude, the galaxy pair lost and the galaxy pair gained would contribute the same amount to the 3D real-space correlation function, following the dashed boundary. However, the projected clustering is different because we do not know the shape of the dashed line, and instead assume that the projection length is the same for all r_x, r_y. It is the 2D clustering strength of the boundary, and its correlation with the density field that is important, rather than the loss or gain of particular galaxy pairs.

we lose voids, we gain clusters and these give the same clustering signal. However, we assume that the projected field has a constant projection length, and this implies that the underdensity of the void will become larger (since we include less of the galaxies) and the overdensity of the cluster becomes larger (since we will include more of its galaxies). Thus the overall clustering signal becomes stronger.

The apparent shift in galaxy positions caused by moving from real to redshift space $(s_z - r_z)$ can be treated by Taylor expanding the selection function (Fisher et al. 1993), which gives to first order

$$\phi(s_z) = \phi(r_z) + \frac{d\phi(r_z)}{dr_z}(s_z - r_z). \quad (5)$$

We consider this to be an Eulerian picture as it is based on apparent galaxy motions. We can write

$$\delta_p(\mathbf{r_p}) = \int dr_z \left[\phi(r_z)\delta(\mathbf{r}) + (s_z - r_z)\frac{\partial\phi(r_z)}{\partial r_z}\right] \quad (6)$$

to first order in $\delta(\mathbf{r})$. Following linear theory, $(s_z - r_z)$ can be written as a function of the overdensity field,

$$(s_z - r_z) = -\beta\frac{\partial}{\partial r_z}\nabla^{-2}\delta(\mathbf{r}), \quad (7)$$

where $\beta \equiv f/b$, with f being the logarithmic derivative of the linear growth rate with respect to the logarithm of the scale factor, and b the galaxy bias. We therefore have that

$$\delta_p(\mathbf{r_p}) = \int dr_z \left[\phi(r_z) - \beta\frac{\partial\phi(r_z)}{\partial r_z}\frac{\partial}{\partial r_z}\nabla^{-2}\right]\delta(\mathbf{r}). \quad (8)$$

If we think of $\phi(s_z)$ as setting up boundaries in s_z, then substituting Eq. (8) into Eq. (3) shows that we can expect coherent apparent galaxy motion across these boundaries. Correlations between galaxies moved into the sample by the redshift-space distortions, and those already within the sample, give rise to cross terms from the two terms in Eq. (8). The second term in Eq. (8) also adds a component to the projected correlation function from the coherence of the velocities at different points on the boundary. We see that, even with constant $\phi(s_z)$ within a fixed interval, redshift-space distortions can still affect the correlation function of the volume within the sample due to the motion of galaxies across the boundary. Modelling the effect of redshift-space distortions based on predicting galaxy motions (e.g. Regos & Szalay 1995) is difficult because we need to correlate multiple points on the boundary and internal locations within the bin.

In addition to the Eulerian picture given by Eq. (8), we can also consider a Lagrangian picture based on the redshift-space overdensity field that we wish to project. Following this equivalent picture, we can work directly with redshift-space overdensities using Eq. (3),

$$\xi^s_p(d_p) = \int\int ds_z ds'_z \phi(s_z)\phi(s'_z)\xi^s\left[d(s_z, s'_z, d_p)\right]. \quad (9)$$

In the plane-parallel approximation, we can use the redshift-space correlation function of equation 5 of Hamilton (1992) as input into the projection equation.

$$\xi^s(\mathbf{d}) = \xi_0(d)P_0(\mu) + \xi_2(d)P_2(\mu) + \xi_4(d)P_4(\mu), \quad (10)$$

where

$$\xi_0(d) = (b^2 + \frac{2}{3}bf + \frac{1}{5}f^2)\xi(d), \quad (11)$$

$$\xi_2(d) = (\frac{4}{3}bf + \frac{4}{7}f^2)[\xi(d) - \xi'(d)], \quad (12)$$

$$\xi_4(d) = \frac{8}{35}f^2[\xi(d) + \frac{5}{2}\xi'(d) - \frac{7}{2}\xi''(d)], \quad (13)$$

P_i are the standard Legendre polynomials, and

$$\xi'(d) \equiv 3d^{-3}\int_0^d \xi(d')(d')^2 dd', \quad (14)$$

$$\xi''(d) \equiv 5d^{-5}\int_0^d \xi(d')(d')^4 dd'. \quad (15)$$

b is the large-scale bias of the galaxy population being considered, f is the standard dimensionless linear growth rate, ξ is the 3-dimensional real-space correlation function, and μ is the cosine of the angle between the separation along the line of sight and the transverse separation, $\mu \equiv |s_z - s'_z|/d$. One strong advantage of the Lagrangian framework is that it is straightforward to determine the projected correlation function, even when the galaxy selection function is discontinuous. This allows simple comparison between the results one expects to obtain with and without redshift-space distortions.

2.2 The Limber approximation

For pairs of galaxies, we can define the mean $m_z \equiv (r_z + r'_z)/2$ and separation along the z-axis $d_z \equiv r_z - r'_z$. For a survey whose depth is larger than the correlation length, and with a slowly vary-

ing selection function, so that $\phi(r_z) \simeq \phi(r'_z) \simeq \phi(m_z)$, Eq. (2) reduces to the Limber equation in real-space ($s_z - r_z = 0$)

$$\xi_p(d_p) = \int_{-\infty}^{+\infty} dm_z\ \phi^2(m_z) \int_{-\infty}^{+\infty} dd_z\ \xi\left(\sqrt{d_p^2 + d_z^2}\right). \quad (16)$$

We see that, for the Limber approximation, ϕ is a function of m_z alone, and the integrals over dm_z and dd_z in Eq. (17) are separable. In redshift-space, a similar reduction of Eq. (9) gives

$$\xi_p^s(d_p) = \int_{-\infty}^{+\infty} dm_z \int_{-\infty}^{+\infty} dd_z \left[\phi(m_z) - \beta \frac{\partial \phi(m_z)}{\partial m_z} \frac{\partial}{\partial r_z} \nabla^{-2}\right]^2 \times\ \xi\left(\sqrt{d_p^2 + d_z^2}\right), \quad (17)$$

if we expand redshift-space distortions in $(s_z - r_z)$, or

$$\xi_p^s(d_p) = \int_{-\infty}^{+\infty} dm_z\ \phi^2(m_z) \int_{-\infty}^{+\infty} dd_z\ \xi^s\left(\sqrt{d_p^2 + d_z^2}\right), \quad (18)$$

in the Lagrangian picture. Because no galaxies are lost or gained moving from real-space to redshift-space, the result of the integral over d_z is the same in real or redshift space, so we see that in this approximation there are no redshift-space effects. But, as we show later, this picture is too simplistic to be applied to the analysis of future data sets.

2.3 Power Spectra

In Padmanabhan et al. (2007), the projection of the 2-pt clustering was analysed through the power spectrum. We now consider such an approach in the plane-parallel approximation and for a Cartesian basis. Taking the Fourier transform of $\delta(\mathbf{s})$ in Eq. (4) gives

$$\delta_p(\mathbf{r_p}) = \int ds_z\ \phi(s_z) \int \frac{d^3k}{(2\pi)^3} \delta(\mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{s}}. \quad (19)$$

We now define a window function

$$W(k_z) = \int ds_z\ \phi(s_z) e^{-ik_z s_z}, \quad (20)$$

and use statistical isotropy and homogeneity within the definition of the power spectrum $\langle \hat{\delta}(\mathbf{k}) \hat{\delta}^*(\mathbf{k}') \rangle = P(k) \delta_D(\mathbf{k} - \mathbf{k}')$, where δ_D is the Dirac delta function. We assume that the power spectrum does not evolve over the volume covered by the window[1]. Taking the 2-point function of the projected overdensity (Eq. 19) gives

$$\xi_p(d_p) = \langle \hat{\delta}_p(\mathbf{r_p}) \hat{\delta}(\mathbf{r'_p}) \rangle \quad (21)$$

$$= \int \frac{dk^3}{(2\pi)^3} W^2(k_z) P(\mathbf{k}) e^{-i\mathbf{k_p}\cdot(\mathbf{r_p} - \mathbf{r'_p})}. \quad (22)$$

The projected overdensity can be written in terms of a 2D power spectrum $P_p(k_p)$,

$$\xi_p(d_p) = \int \frac{dk_x\, dk_y}{(2\pi)^2} P_p(k_p) e^{-i\mathbf{k_p}\cdot(\mathbf{r_p} - \mathbf{r'_p})}. \quad (23)$$

If we compare Eqns. (22) & (23), we see that

$$P_p(k_p) = \int \frac{dk_z}{(2\pi)} W(k_z)^2 P\left(\sqrt{k_p^2 + k_z^2}\right). \quad (24)$$

Note that the power $P(\mathbf{k})$ depends on the amplitude of the full 3-dimensional wavevector, and so is dependent on k_p.

Using Eq. (5) to include redshift-space distortions, the window $W(k_z)$ has an extra term,

$$W(k_z) = \int dr_z\ \left[\phi(r_z) + (s_z - r_z) \frac{d\phi(r_z)}{dr_z}\right] e^{-ik_z r_z}. \quad (25)$$

In Fourier space, $(s_z - r_z) = -\beta(k_z^2/k^2)\delta(\mathbf{r})$, so we can expand $\delta(\mathbf{s})$ to 1st order in $\delta(\mathbf{r})$, leaving a new window function for Eq. (24)

$$W(k_z) = \int dr_z\ \left[\phi(r_z) - \beta\left(\frac{k_z}{k}\right)^2 \frac{d\phi(r_z)}{dr_z}\right] e^{-ik_z r_z}. \quad (26)$$

If we drop the plane-parallel approximation and expand in Spherical Harmonics, the standard result (Peebles 1973) is

$$\langle |a_{lm}|^2 \rangle = \frac{1}{2\pi^2} \int dk\ k^2 P(k) W^2(k), \quad (27)$$

where

$$W(k) = \int dr\ \phi(r) j_l(kr) + \frac{\beta}{k} \frac{d\phi(r)}{dr} j'_l(kr). \quad (28)$$

Here the l dependence is contained within $W(k)$, while in Eq. (24), it was the power that depended on k_p. Eq. (24) could have been rewritten by changing the variable of the convolution integral to k to match.

2.4 Monte-Carlo simulations of the projection effect

In order to test the projection formulae presented in Sections 2.1 & 2.3 without redshift-space distortions, we have used Monte-Carlo realisations of δ-function real-space correlation functions in a similar vein to that of Simpson et al. (2009). We work in a plane parallel approximation throughout and construct a real-space 3D δ-function correlation function at an arbitrary location d_0 such that

$$\xi(d) = \delta_D(d - d_0)\ \xi_0, \quad (29)$$

where δ_D is the standard Dirac delta function. We do this by introducing a pre-determined excess of data pairs at the location d_0. The number of excess pairs we introduce depends on the value of $\xi(d_0)$ we require and is determined using the natural estimator $\xi = D/R - 1$. For example, if we have a uniform distribution of data and random pairs with $100,000$ pairs per bin of separation, we would require an excess of $10,000$ data pairs at the location d_0 for $\xi(d_0) = 0.1$. In doing this we create an unnormalised 3D δ-function correlation function.

Changing the variables in the inner integral of Eq. (16) to be a function of 3D pair separation d gives

$$\xi_p(d_p) = \int\int_V dm_z\, dd\ \ \phi^2(m_z)\ \frac{2\xi(d)\, d}{\sqrt{d^2 - d_p^2}}, \quad (30)$$

and is simplified for the δ-function case such that

$$\xi_p(d_p) = \frac{2}{\pi d_0} \int dm_z\ \phi^2(m_z)\, \xi_0 \frac{d_0}{\sqrt{d_0^2 - d_p^2}}. \quad (31)$$

The factor $1/\pi d_0$ accounts for the fact that the δ-function real-space correlation function was unnormalised. By introducing a radial window, we are preferentially selecting pairs of galaxies from the sample. A further volume reduction normalisation is required in Eq. (30) to account for this. The excess probability of finding two galaxies in areas δA_1 and δA_2 with a 2D projected separation d_p is the sum of all the probabilities of finding two galaxies in volumes δV_i and δV_j along the radial axis at *all* 3D separations d. That is,

$$1 + \xi_p(d_p) = \frac{\bar{n_V}^2}{\bar{n_A}^2} \frac{1}{\delta A_1 \delta A_2} \left(\sum_i \sum_j [1 + \xi(d_{ij})] \delta V_i \delta V_j\right). \quad (32)$$

[1] This is true if analysing a single time slice from a simulation

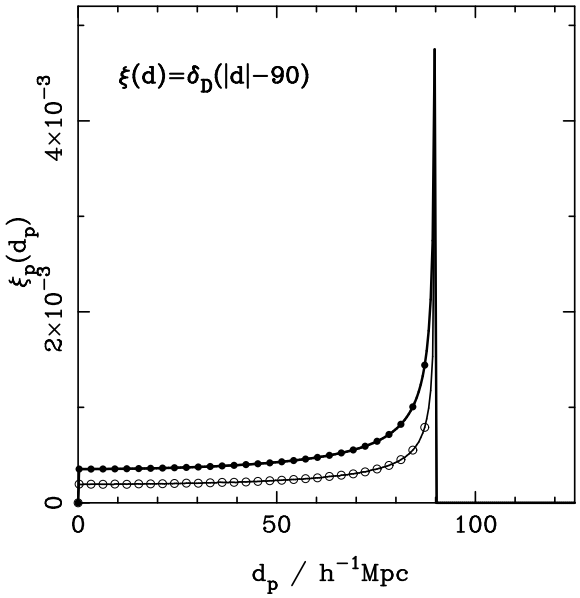

Figure 3. Projected correlation functions calculated for a 3-dimensional δ-function correlation function centred on $d_{3D} = 90\,h^{-1}\,\mathrm{Mpc}$, with no radial window (solid symbols) and with a top-hat window in radial distribution of width $100\,h^{-1}\,\mathrm{Mpc}$ (open symbols). Models calculated using Eq. (31) are shown by the solid lines.

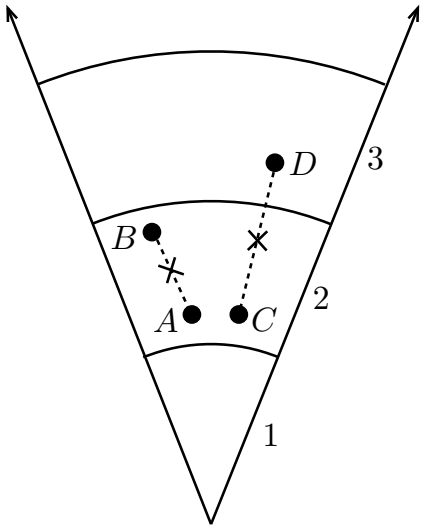

Figure 4. A schematic representation showing how galaxy pairs are selected using top-hat and pair-centre binning schemes. Using a top-hat binning scheme, where galaxy pairs are selected according to the position of each individual galaxy, pair AB would be placed in redshift bin 2, whereas pair CD would not be placed in any bin, and would simply not be counted in an analysis. In contrast, the pair-centre binning scheme would place both pairs in bin 2.

In Fig. 3 we show the clustering expected for a projection of a density field created from a δ_D-function 3D correlation function in the case where there is no window function (solid symbols) and for a window function of width $100\,h^{-1}\,\mathrm{Mpc}$ (open symbols). The excess of pairs that exists at a single scale in 3D is projected onto a range of scales, up to and including this scale, in 2D. The projection window leads to a damping of power on all scales. This effect depends upon the window size; as we move to smaller projection windows the effect of the projection is decreased and $\xi_{2D} \rightarrow \xi_{3D}$. The projection of a more general density field, where there is clustering on a range of scales, can be considered as the linear combination of the projections of a series of δ_D-function 3D correlation functions. The trends observed in this analysis will help us to interpret the behaviour of the projected correlation function in the more general situation analysed in later sections.

3 BINNING GALAXY SAMPLES

Future surveys will automatically have a standard selection function caused by the changing cosmological volume, the number density of galaxies as a function of redshift, and selection effects such as a magnitude limit below which we cannot observe galaxies or obtain accurate photometric redshifts. In addition to this distribution we will wish to bin galaxies based on their photometric redshifts in order to analyse the evolution of galaxy properties and/or cosmology across the sample. We now consider how the way in which this sub-division is applied affects the importance of redshift-space distortions.

One simple approach would be to bin galaxy positions in redshift, equivalent to a **top-hat binning**. Such galaxy selection means that galaxy pairs, where galaxies lie in different bins, are not included in the estimate of the correlation function. This exclusion of pairs leads to the observed difference between the projected real-space and redshift-space correlation function, as described in Section 2.1. An alternative to this approach, considered here, would be to bin galaxy pairs rather than individual galaxies.

A simple argument shows that in an ideal situation, applying a binning based on the centre of galaxy pairs in the radial direction, which hereafter we refer to as **pair-centre binning**, can completely remove the effect of redshift-space distortions while retaining information about the evolution of the correlation function. A schematic representation of this binning scheme is shown in Fig. 4. Consider the galaxy pair defined by galaxies A and B: the positions of both galaxies and their pair-centre are within redshift slice 2. This pair would therefore be included in analyses conducted on this slice in both top-hat and pair-centre binning schemes. The positions of galaxies C and D span two separate redshift slices and therefore the pair they define would not be included in an analysis of either slice 2 or 3 when using a top-hat binning scheme. However, this pair would be included in an analysis of slice 2 when using the pair-centre binning scheme. This schematic demonstrates both the pair-centre binning scheme and the fact that such a scheme includes all pairs within an analysis.

Suppose that we have a clustered distribution of D galaxy pairs of separation r with a uniform sampling function along the z-axis in a large volume that would contain R pairs if galaxies were randomly distributed. Because of the large volume assumption, we can assume that boundary effects for this sample are negligible. Therefore, redshift-space distortions have no effect for the full catalogue for which our estimate of $\xi_p(d_p)$ is $\hat{\xi_p}(d_p) = D/R - 1$. Now suppose the sample is split into n sub-samples, based on the redshift-space positions of the centres of the pairs within equal volumes, chosen independently of the observed galaxy distribution. Then all pairs are still counted in some bin; none are lost or gained as opposed to galaxy based selection functions. For the subsamples, $\langle D' \rangle = D/n$, $R' = R/n$, and $\langle \xi_p(d_p) \rangle$ is unchanged from the value for the full sample. This is true regardless of bin size. The key difference here, compared with considering a set of bins based

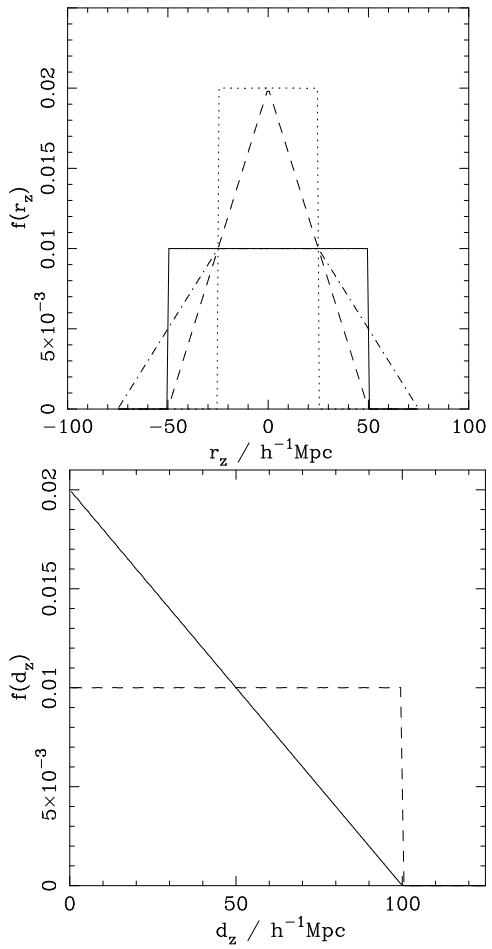

Figure 5. Top panel: the normalised radial distribution of galaxies (solid line) and pair-centres (dashed line) for the distribution of galaxies in a top-hat bin of width $100\,h^{-1}\,\mathrm{Mpc}$. These are compared with the distributions of galaxies (dot-dash line) and pair-centres (dotted line) for galaxies whose pair-centre is within a $50\,h^{-1}\,\mathrm{Mpc}$ bin, and with $d_z < 100\,h^{-1}\,\mathrm{Mpc}$. Bottom panel: comparison of the radial pair separations (d_z), between top-hat (solid line) and pair-centre (dashed line) binning.

on galaxy selection, is that no pairs are left out, so the expected correlation function has to be the same for all bins.

For a sample where we do not know the true distance to each galaxy, but instead rely on photometric redshifts, binning based on apparent pair centre will also remove redshift-space distortions. The above argument based on pair conservation will also hold in this situation.

The radial distributions of galaxies and pair-centres along with the distribution of radial pair separations for top-hat and pair-centre binning schemes are compared in Fig. 5. For measuring the radial evolution of clustering through binned projected correlation function measurements, there is no obvious advantage to either scheme. This is particularly true when photometric redshifts are used to estimate radial positions, as it is then impossible to select galaxies from non-overlapping radial bins (see Section 7).

We therefore see that we can add boundaries based on pair-centres and analyse projected clustering in bins without being affected by redshift-space distortions. However, there are two problems with applying this approach in practise:

(i) galaxy pairs of wide separation now have to be included,

(ii) galaxy surveys typically have flux limited boundaries, which cause redshift dependent effects that cannot be removed by any binning. However, this effect can be removed by k-correcting the observed luminosities and cutting the sample at a more stringent k-corrected luminosity limit. We now investigate this further.

3.1 Flux-limited Selection Functions

Peculiar velocities can directly influence galaxy brightness through relativistic beaming, but such effects are small for typical galaxy peculiar velocities. Redshift distortions would additionally change the apparent magnitudes through the k-correction, potentially causing galaxies to either enter or exit flux-limited samples. The change in apparent magnitude will correlate with bulk-flow motions and thus the boundary of the survey in real-space will fluctuate in a manner analogous to that described in Fig. 2. In this situation, the amplitude of the effect and whether it enhances or reduces the real-space clustering signal will depend on galaxy type and the band used for detection, but for a homogeneous sample of galaxies (e.g. Luminous Red Galaxies) one would expect that this effect will be significant.

This redshift-space effect is simple to remove - k-corrections derived by fitting to galaxy spectra will correct for spectral shifts caused by both the Hubble flow and any peculiar velocities. It therefore makes sense to select galaxy samples after applying the k-correction, and cutting back from survey boundaries based on apparent magnitude, until no galaxies outside the original sample would be expected to pass the revised boundary. This is not as onerous as it sounds as one has to do this to create true volume-limited catalogues.

Given purely photometric data, k-corrections can only be estimated given a photometric redshift and spectral-type fit, and are therefore unreliable for individual galaxies. For this reason, and the fact that cutting back from the survey boundary removes a large amount of data, k-corrections have not always been applied to apparent magnitude limits (e.g. Ross & Brunner 2009 select galaxies with de-reddened $r < 21$ for their parent sample). We therefore now consider the amplitude of the effect. One can express the fluctuation in magnitude, δm, as

$$\delta m = \mathrm{d}k_{corr}/\mathrm{d}z\delta z \tag{33}$$

where δz is the magnitude of the redshift distortion. This will cause fluctuations in the effective depth of the survey such that

$$DM(z_{eff}) - DM(z) = \delta_m \tag{34}$$

where $DM(z)$ is the distance modulus, z_{eff} is the effective depth and z would be the predicted depth. The SDSS DR7 photometric redshift table includes estimated r-band k-corrections for every galaxy. Studying galaxies with type-value equal to 0 (the most early-type), one can determine that $\mathrm{d}k_{corr}/\mathrm{d}z \sim 3.3$ at $z = 0.4$. For an arbitrary δz, this $\mathrm{d}k_{corr}/\mathrm{d}z$ yields $z_{eff} - z = 0.5\delta z$. For example, assuming bulk flows have a velocity $\sim 10^3$km/s — thereby imparting redshift distortions at the $\sim 1\%$ level ($\delta z = 0.004$) — they impart coherent fluctuations in apparent magnitude equivalent to 0.013 magnitudes (in the r-band). At $z = 0.4$, these fluctuations in magnitude imply a change in the survey depth of $z_{eff} - z = 0.002$ (0.5%). Thus, the redshift distortions caused by selecting a flux-limited sample of galaxies can be as large as 50% of those caused by selecting a sample in redshift. Therefore, even for a flux limited selection function, redshift distortions may be important. The size of the effect depends on the slope of $k_{corr}(z)$, and one can minimise the effect by carefully choosing the band used for

selection and the type(s) of galaxies included in the sample. (One can envision cases where slope of the average k-correction is zero, thus removing any effect even before applying k-corrections.)

4 ANALYSIS OF HUBBLE VOLUME SIMULATIONS

In order to test the effect of redshift-space distortions on the projected correlation function for a realistic non-linear distribution of galaxies, we have analysed results from the ΛCDM Hubble Volume (HV) simulations (Evrard et al. 2002). The ΛCDM HV simulation, covering a $(3000\,h^{-1}\,\mathrm{Mpc})^3$ box, assumes a cosmological model with $\Omega_m = 0.3$, $\Omega_{CDM} = 0.25$, $\Omega_b = 0.05$, $\Omega_\Lambda = 0.7$, $h = 70$, $\sigma_8 = 0.9$, & $n_s = 1$.

We make a number of simplifications in order to help with the calculation of projected real-space and redshift-space correlation functions. For each sample to be analysed, along the two non-projection axes, we use the periodic nature of the numerical simulation to eliminate boundaries. This means that we can confidently use the natural estimator $\xi + 1 = D/R$, where the expected number of galaxy pairs in the absence of clustering R can be calculated analytically. We also do not introduce a galaxy-bias model, and assume that galaxies Poisson sample the matter particles. The inclusion of such a model would not alter the conclusions of this work.

We start by applying a top-hat selection function to the galaxy positions, calculating projected correlation functions for window widths $50\,h^{-1}\,\mathrm{Mpc}$ and $100\,h^{-1}\,\mathrm{Mpc}$ in real and redshift space. Fig. 6 shows the correlation function after reducing noise by averaging over 30 samples ($100\,h^{-1}\,\mathrm{Mpc}$ bins) or 60 samples ($50\,h^{-1}\,\mathrm{Mpc}$) bins. In real-space the projected correlation function tends towards the 3D correlation function at large scales, as expected. In line with the analysis presented in Section 2.4, the scale at which ξ_p becomes $\sim \xi_{3D}$ is larger for the $100\,h^{-1}\,\mathrm{Mpc}$ bin. For each bin size, the inclusion of redshift-space distortions clearly has a strong effect and this effect grows dramatically as the scale gets larger. Notably, it is larger even than the effect of redshift-space distortions on the 3D spherically averaged correlation function (or power spectrum). The effect is enhanced in the narrower projection window. As well as increasing the amplitude of the projected correlation function, we see that redshift-space distortions also act to wash out the baryon acoustic oscillation signal.

Selecting galaxy pairs solely based on the position of their pair-centre removes the effect of redshift-space distortions. To see this, suppose we split along the projection axis into N slices, and average the DD counts over all slices. Then the average is independent of N as all pairs are counted however many bins are selected. In addition, the periodic nature of the simulation means that no pairs are gained or lost between real-space and redshift-space: we always count all pairs of galaxies, so there will be no change in the measured correlation function. As explained in Section 3.1, if we select based on an apparent magnitude limit, we can remove redshift distortions by applying a more stringent magnitude limit based on k-corrected luminosities. Here we have to cut the luminosity limit back to make sure that the new sample is complete, in that it contains all of the possible galaxies. However, there is a further practical problem in that including galaxy pairs with wide radial separation might complicate the modelling of cosmological evolution required to fit the correlation function. Consequently, it might be difficult to analyse the measured correlations function for a pair-centre binned sample in practice.

We therefore introduce a **constrained pair-centre** binning scheme that includes an upper limit on the pair separation along the projection axis, in addition to pair-centre binning. This is equivalent to locating *each* galaxy included in the analysis in the centre of a top-hat bin. We should expect that the effect of redshift-space distortions will be reduced compared with binning galaxy distributions in a top-hat with the same width, as boundaries will only affect galaxy pairs with the maximum radial separation, whereas for top-hat bins they affect galaxy pairs with a range of radial separations (see Fig. 5). Results calculated using this binning scheme are shown in Fig. 6. Here we see that the effect of redshift-space distortions is reduced, especially for the larger $|d_z|$ limit.

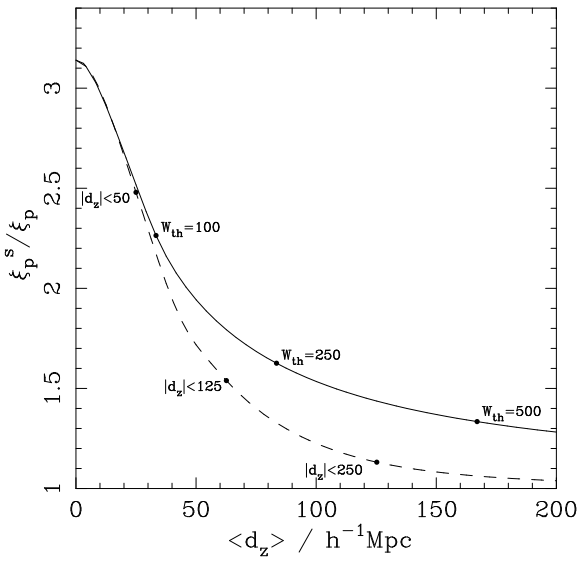

Figure 7. The expected ratio of the projected correlation functions in redshift-space and in real-space, averaged for "angular" separations between $40\,h^{-1}\,\mathrm{Mpc}$ and $80\,h^{-1}\,\mathrm{Mpc}$, as a function of bin width. The solid line show the difference as a function of the width of the top-hat window. The dashed line show the result for constrained pair-centre binning as a function of an additional constraint placed on the radial galaxy separation. We have plotted results (and therefore matched filters) as a function of the mean radial galaxy separation.

In order to investigate the effect of different binning schemes further, Fig. 7 shows a comparison on the large-scale redshift-space and real-space correlation function amplitude. These are averaged for galaxy separations between $40\,h^{-1}\,\mathrm{Mpc}$ and $80\,h^{-1}\,\mathrm{Mpc}$. We have plotted these as a function of average radial galaxy separation, in order to compare filters in an unbiased way. We clearly see that, when binning radially using the constrained pair-centre binning scheme, the effect of redshift-space distortions is significantly reduced.

The relative importance of redshift-space distortions depends on the average galaxy bias of the populations being considered; there is a balance between the impacts of b and f in Eq. (10). In order to demonstrate this, Fig. 8 shows that the relative effect of redshift-space distortions decreases as the bias of the galaxy sample analysed increases. This explains why the effect of redshift-space distortions was reduced in the work of Baldauf et al. (2009).

In this section, we have considered the cases of a top-hat or pair-centre galaxy selection. We have argued that while, in principle, pair-centre binning removes the effects of redshift-space distortions provided k-corrections are included when magnitude limits are applied, there are good reasons to remove galaxies of wide separation if we are to measure the evolution in the correlation function. Therefore, in Section 6, we will test how these binning schemes

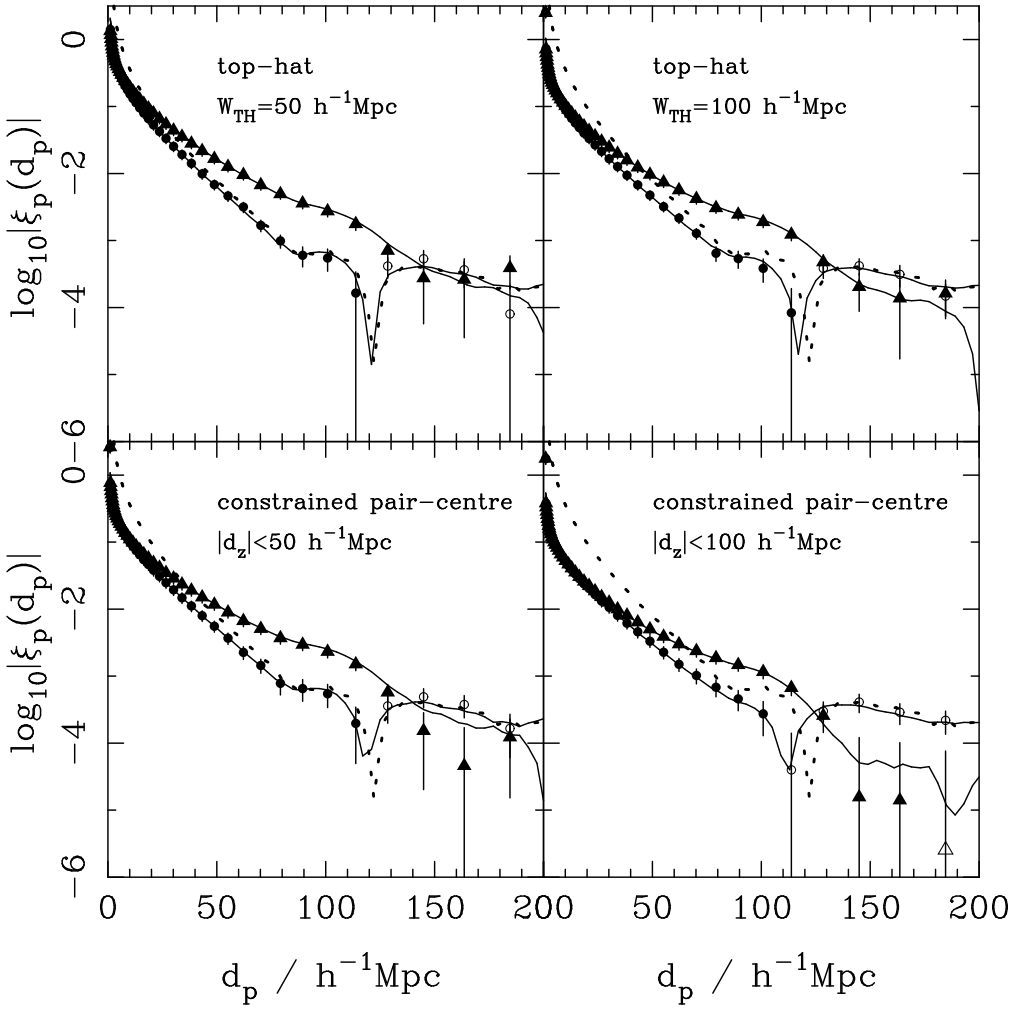

Figure 6. Top row: Correlation functions calculated from Hubble Volume data in radial bins of width $50\,h^{-1}\,\mathrm{Mpc}$ and $100\,h^{-1}\,\mathrm{Mpc}$. Solid symbols are plotted where the correlation function is positive, while open symbols show where the correlation function is negative. Bottom row: Correlation functions calculated from Hubble Volume data for galaxy pairs selected based on constrained pair-centre binning scheme with both their pair-centres and radial separation less than $50\,h^{-1}\,\mathrm{Mpc}$ or $100\,h^{-1}\,\mathrm{Mpc}$. 1σ error bars are plotted in both cases, assuming that the slices analysed draw correlation functions from a Gaussian distribution. The dotted line gives the 3D HV correlation function, ξ_{HV}, (plotted assuming $r = r_p$) as measured from the simulation. The solid lines denote the projected correlation function one expects in real-space with ξ_{HV} as the 3D correlation function (lower curves in each panel), and in redshift-space using Eqns. (9) & (10) to estimate the 3D redshift-space correlation function from ξ_{HV} (upper curves in each panel).

work when the background galaxy distribution has a more realistic radial distribution, similar to that expected for a survey like the Dark Energy Survey. Before we can do that, we need to consider the case where we have a boundary that consists of a mix of real- and redshift-space constraints.

5 DEALING WITH HYBRID SELECTION FUNCTIONS

In practice, the radial selection function will be dependent on both observational constraints such as the limiting apparent magnitude of the survey, and additional binning. One expects that the boundary based on observational constraints can be treated as a real-space boundary (though, this even, is not so simple; see Section 3.1). Thus, when one applies a top-hat selection in redshift to an observed sample of galaxies, the resulting boundaries of the selection function will include both real-space and redshift-space components.

Fig. 9 shows a schematic representation of a top-hat selection in redshift made at positions s_{z1} and s_{z2} along a non-uniform real-space radial selection function. It shows that we can split galaxies within this bin into three sub-samples, with different boundaries:

- A_s (redshift-redshift): Selected with both boundaries in redshift-space.
- B_h (redshift-real): Selected with one boundary in real-space and one boundary in redshift-space (*hybrid*-space).
- C_r: Selected with both boundaries in real-space.

The real-space and redshift-space boundaries of Fig. 9 are represented by solid and dotted lines respectively. Any auto-correlation of galaxies with this selection function will essentially be a weighted sum (based on the amplitude of the selection function) of the auto-correlations of galaxies within the individual subsamples and the cross-correlations of galaxies in different subsamples.

In order to investigate the projected clustering of these different subsamples, we have drawn samples of particles from the HV simulation (see Section 4), created in top-hat bins of width

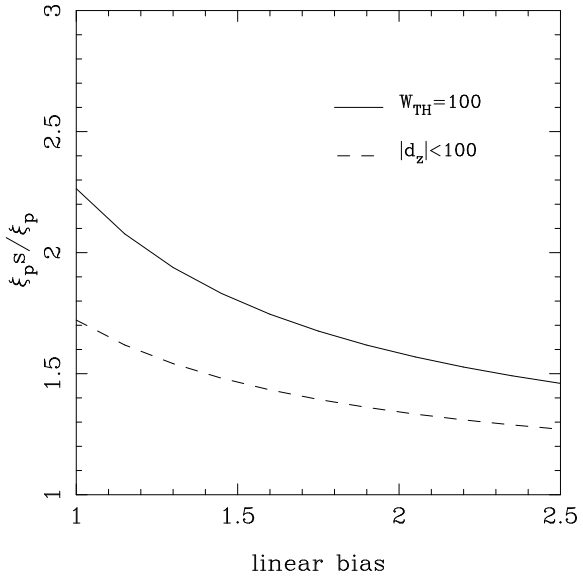

Figure 8. The expected ratio of the projected correlation functions in redshift-space and in real-space, averaged for "angular" separations between $40\,h^{-1}\,\mathrm{Mpc}$ and $80\,h^{-1}\,\mathrm{Mpc}$, as a function of galaxy bias, assuming a ΛCDM cosmology with $\Omega_m = 0.25$. The solid line show the difference as a function of the width of the top-hat window. The dashed line show the result for constrained pair-centre binning as a function of the additional constraint placed on the radial galaxy separation. As in Fig. 7, we have plotted results (and therefore matched filters) as a function of the mean radial galaxy separation.

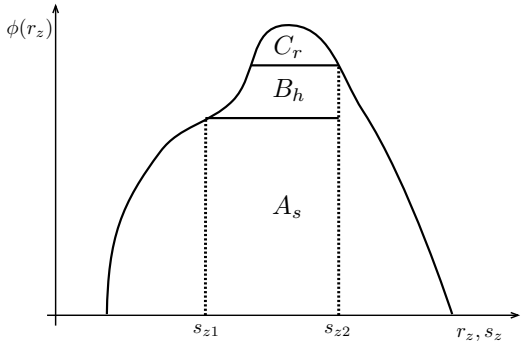

Figure 9. Schematic representation of an evolving real-space radial selection function with populations A_s, B_h and C_r defined according to where a top-hat bin with redshift-space boundaries at s_{z1} and s_{z2} intersect the radial selection. Populations have boundaries in: A_s redshift-space, B_h hybrid-space and C_r real-space.

$100\,h\,\mathrm{Mpc}^{-1}$. Sample A_s has top-hat selection boundaries in redshift-space, sample B_h has one real-space and one redshift-space boundary, while sample C_r has both boundaries in real-space. These samples cover the same region of the simulation.

Fig. 10 shows the projected auto-correlation functions for these subsamples. The measured ξ_p for the C_r and A_s samples are essentially the same as those shown in the top-right panel of Fig. 6, and just as before they return the expected real and redshift-space correlation functions calculated via Eqns. (9) & (10). However, the hybrid-space correlation function, ξ_p^h of sub-sample B_h has an amplitude that lies in-between those of the pure real and redshift-space

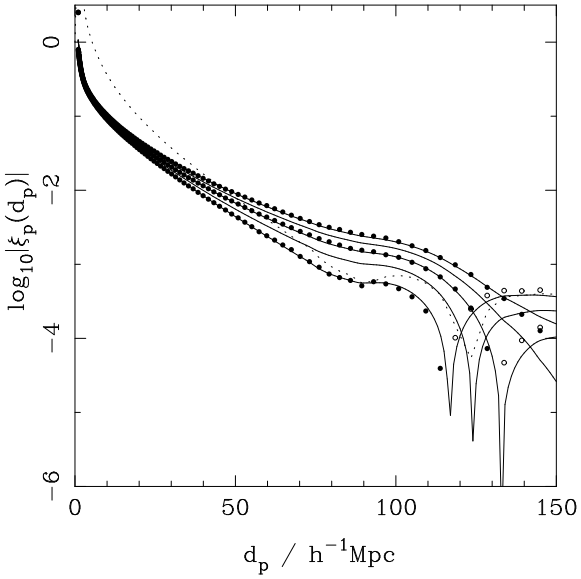

Figure 10. The average recovered auto-correlation function (solid circles) for galaxies from 90 samples drawn from the Hubble Volume simulation using three different radial selections, each with top-hat width 100 h^{-1}Mpc. These are compared against model correlation functions calculated for different galaxy samples Eq. (36). The three radial selections are: 1) two real-space boundaries (lowest points), which best matches the model calculated using the real-space correlation function, 2) two redshift space boundaries (highest points), which best matches the model calculated using the redshift-space correlation-function and, 3) a real-space boundary on one side and a redshift space boundary on the other side (points in the middle), which best matches the model calculated using the geometric mean of the real- and redshift-space correlation functions.

correlation functions. We find that we can effectively model ξ_p^h by assuming the underlying 3D overdensity field has a correlation function ξ^h given by

$$\xi^h + 1 = \sqrt{(1+\xi^r)(1+\xi^s)}. \tag{35}$$

Note that we are using ξ^r to represent the real-space 3-dimensional correlation function. As can be seen in Fig. 10, this model is well-matched to the measured ξ_p. The justification for this model is that the multiplicative boost to the projected density fluctuations (R if we consider that $\xi = D/R - 1$) can be decomposed into multiplicative contributions from each boundary. Following this model, we should find that the relative effect of redshift-space distortions on each population, and their cross-correlations, are simply proportional to the number of redshift-space boundaries present. If we choose galaxies from a sample with $n \in \{0,1,2\}$ redshift-space boundaries, and another from a sample (possibly the same one) with $m \in \{0,1,2\}$ redshift-space boundaries, then expected correlation function is given by

$$\xi^h + 1 = (1+\xi^r)^{1-l/4}(1+\xi^s)^{l/4}, \tag{36}$$

where $l = m + n$.

Fig. 11 displays the cross-correlations between the our three HV subsamples. As expected, the model calculated using the appropriate ξ^h from Eq. (36) is the closest match to the measured cross-correlation in every case. All of the models do over-predict all three measurements at large scales, but we believe this is reflective of the error associated with our measurements (one would expect it to be covariant between each sample as they all sample the same density field). It is possible that we are seeing effects caused

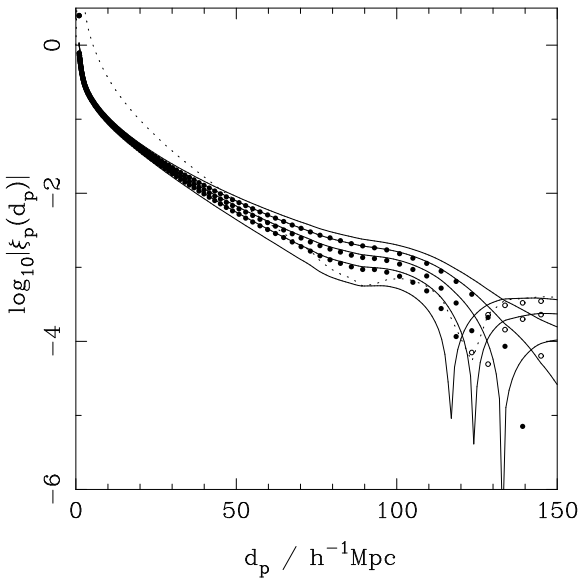

Figure 11. The average measured cross-correlation functions from 90 radial slices of width 100 h^{-1}Mpc in real-space, redshift-space or a hybrid with one real-space and one redshift-space boundary, each containing 10^6 galaxies (solid circles). These are compared against the model ξ_p^h of Eq. (36) (solid lines), for different total numbers of redshift boundaries. The amplitude of both model and data correlation functions increase with increasing dependence on the redshift-space correlation function.

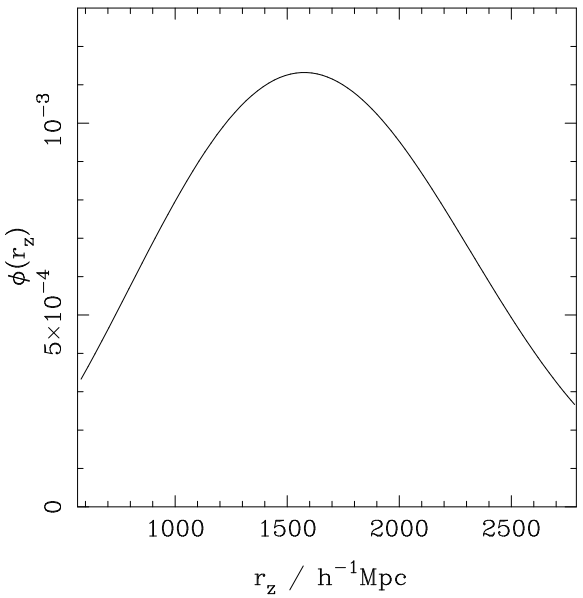

Figure 12. Approximate redshift distribution similar to that expected from the Dark Energy Survey. In order to use this distribution of galaxies to easily measure cosmological acceleration using projected clustering measurements, this population will have to be subdivided or binned in redshift.

by the coherence of the boundaries with each other that would be removed for wider bins, such as those we consider in Section 6.

Given a hybrid selection function such as that shown in Fig. 9, we must split the sample into populations where we can assume simple boundary conditions for each. In fact, we can consider solving the projection equation (e.g. Eqns. 2 & 9 in real-space and redshift-space) by Monte-Carlo integration over pairs of radial galaxy locations. For each pair of locations we can determine the relative contributions from galaxies in each of the subsamples, and therefore construct a full model for the correlation function.

6 IMPLICATIONS FOR FUTURE PHOTOMETRIC REDSHIFT SURVEYS

A number of extremely wide angle imaging surveys are planned over the next few years: the Dark Energy Survey (DES), the Panoramic Survey Telescope & Rapid Response System (Pan-Starrs) and the Large Synoptic Survey Telescope (LSST). One goal of these surveys is to constrain the current acceleration of the Universe. In general, one can hope to use such surveys to make four measurements of dark energy using complimentary techniques: cluster counting, BAO, weak lensing and supernovae. In this paper we consider BAO measurements. For these experiments, radial distances to galaxies will be estimated from photometric redshifts, so there will be little information in the radial direction on the scale of BAO. Consequently, analyses will tend to rely on making projected galaxy clustering measurements in redshift slices that are sufficiently narrow to be able to reveal cosmological acceleration.

In order to assess the effect of redshift-space distortions on such measurements, we now consider one of these surveys, DES, in more detail. The DES will use a 500 Mega-pixel camera on the Blanco 4-metre telescope in Chile to conduct a galaxy survey over a sky area of 5000 deg^2. Multi-band observations using *g, r, i* and *z*

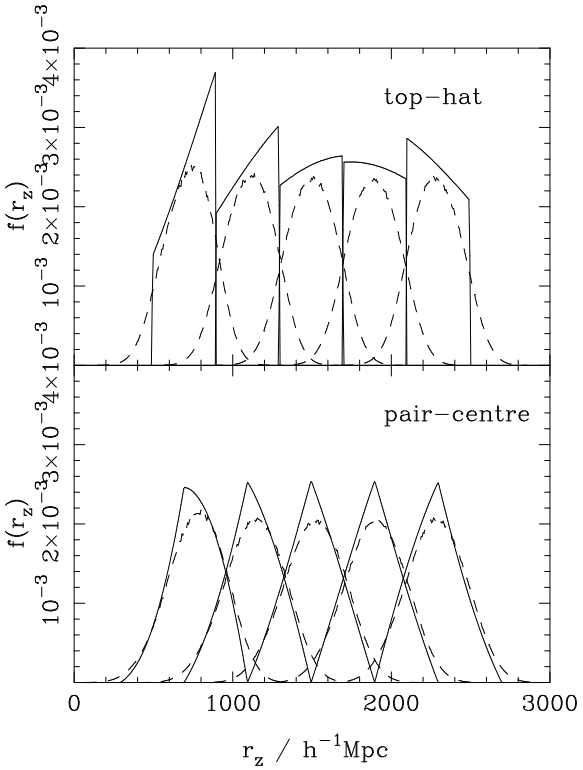

Figure 13. Top panel: normalised radial selection functions for top-hat slices of width $400\,h^{-1}\,\mathrm{Mpc}$ created from a DES-like distribution. Bottom panel: We also consider bins in radial galaxy pair-centre of the same width $400\,h^{-1}\,\mathrm{Mpc}$. While we bin in distances derived from photometric redshifts (solid lines), the true distribution of radial galaxy distances is shown by the dashed lines.

filters will allow photometric redshifts to be obtained over a range $0.2 < z < 1.4$. The expected redshift distribution of the galaxies will be approximately [2]

$$\phi_{DES}(z) \propto \left(\frac{z}{0.5}\right)^2 \exp\left(-\frac{z}{0.5}\right)^{1.5}, \quad (37)$$

after applying approximate survey depths to basic luminosity functions. This function is plotted in Fig. 12. This distribution of galaxies will then be sub-divided into bins in order to assess the evolution of the BAO scale across the survey. As discussed above, measurements of the projected correlation function will be affected by redshift-space distortions, which will increase the signal strength and decrease the importance of BAO features. We now consider how the choice of binning methodology affects the impact of redshift-space distortions.

We consider splitting this galaxy distribution into five redshift slices each of width $400\,h^{-1}\,\mathrm{Mpc}$ for distances estimated from photometric redshifts, assumed to be Gaussian with $\sigma_z = 0.03(1+z)$. These bins cover radial distances of $500 \rightarrow 2500\,h^{-1}\,\mathrm{Mpc}$, related to redshifts $z = 0.15$ to $z = 1.06$ (assuming a flat ΛCDM cosmology with $\Omega_m = 0.25$. The upper panel of Fig. 13 shows the distributions of galaxies in these slices. The lower panel of Fig. 13 shows the redshift distributions when we bin the galaxies based on the centre of the radial separation, calculated from the photometric redshifts. Because we are using photometric redshifts, there is no way to bin without leaving overlap in the true radial distributions. Consequently, the top-hat binning scheme does not provide an obvious advantage over other schemes in terms of analysing disjoint regions.

In light of the discussion in Section 3.1, we consider both the case in which ϕ_{DES} is treated as a real-space boundary (results presented in Fig. 14), and the case in which it is treated as a redshift-space boundary (as may be the case when the slope of $k_{corr}(z)$ is especially large; results presented in Fig. 15). For the hybrid boundary, we employ the techniques described in Section 5 to determine the full form of the projection. When we treat ϕ_{DES} as a redshift-space boundary, we can simply use Eq. (2) & (9) to determine ξ_p in real and redshift-space.

The left-hand panels of Figs. 14 and 15 show the expected projected correlation functions when a top-hat binning scheme is applied with width $400\,h^{-1}\,\mathrm{Mpc}$. Even for this large bin width, in every radial bin there is a significant difference between the result obtained using the redshift-space correlation function and the real-space correlation function. The difference is made clear by observing the ratios between the two, displayed in the bottom panels. The ratios are slightly higher in the case where we treat ϕ_{DES} as a redshift-space boundary (Fig. 15), and the difference between the two treatments is largest for the lowest redshift bin (which has its selection most affected by the overall DES selection). In every case, the ratio is significant (~ 1.5) around $100\,h^{-1}\,\mathrm{Mpc}$ and the shape of the predicted ξ_p and ξ_p^s measurements differ substantially.

The effects of redshift distortions are completely removed when a pair-centre binning scheme is employed and the ϕ_{DES} boundary is assumed to be real-space, as made clear in the middle panel of Fig. 14. Based on the discussion in Section 3, we can simply use Eq. (2) for both and thus their ratio is identically 1. Even when the ϕ_{DES} boundary is assumed to be in redshift-space, as displayed in the middle panel of Fig. 15, the difference between the redshift-space and real-space model is considerably smaller than for the top-hat binning. The right-hand panels of Figs. 14 and 15 show that even if one applies the constraint that the separation between pairs be less than $400\,h^{-1}\,\mathrm{Mpc}$ to be included in a constrained pair-centre bin, redshift space distortions introduce a much smaller effect than for a top-hat binning scheme.

[2] We thank the DES LSS working group for providing this approximation

7 DISCUSSION

Redshift distortions produce a strong effect on projected clustering measurements — one that is far stronger than the redshift-space distortion effect on the 3D clustering signal for galaxy samples with low bias and a narrow radial window. It is clear that redshift distortion effects must be included when modelling the projected galaxy clustering in redshift slices.

If we consider the apparent motion of galaxies as we move from real- to redshift-space, then redshift-space distortions cause an apparent coherent motion of galaxies into and out of samples. This is true whether samples have sharp boundaries, or if the selection function changes more gradually with distance. In fact, we have argued that such motion does not in itself alter the projected correlation function — we would recover the real-space projected correlation function if we could correct for the movement of the boundary (i.e. allow for the depth of the survey to change with the distortions). However, this is not easy to do, although it is theoretically possible and is an interesting alternative approach. The effect of redshift-space distortions is due to the redshift-space boundaries themselves having an angular clustering signal, and their correlation with the overdensity field. We can alternatively view the effect from a Lagrangian standpoint, where we have to consider that the projection does not remove redshift-space effects from the anisotropic correlation function.

We have used Hubble Volume simulations to show that the projected correlation function can be modelled most easily by integrating the redshift-space correlation function over the radial selection function. Galaxy selection will often be a mix of real and redshift-space constraints, and we have shown that this can be modelled by splitting the population into samples that can be considered to have top-hat windows in either real-space, redshift-space or a hybrid of the two. In the hybrid situation, the projected correlation function can be modelled using both the real-space and redshift-space correlation function over the radial selection function, and that more complicated selection functions can be effectively modelled in a similar manner. Prior to this publication, no-one has considered how these hybrid selection functions affect the recovered projected clustering signal.

7.1 Pair-Centre Binning

We have presented a new measurement technique, *pair-centre* binning, and shown that it minimises the effects of redshift space distortions. In this new scheme, we only include galaxies where their apparent *pair-centres* lie within a given radial bin, whereas traditional methods select pairs where both galaxies lie within the bin. The new scheme includes individual galaxies that lie *outside* the traditionally applied top-hat boundaries. This simple modification acts to reduce the effect of the coherent movement of galaxies between slice boundaries on projected correlation function clustering analyses. It is important to note that this new technique does not *prevent* the movement of galaxies between slices; redshift-space distortions due to peculiar velocities will always exist in the radial

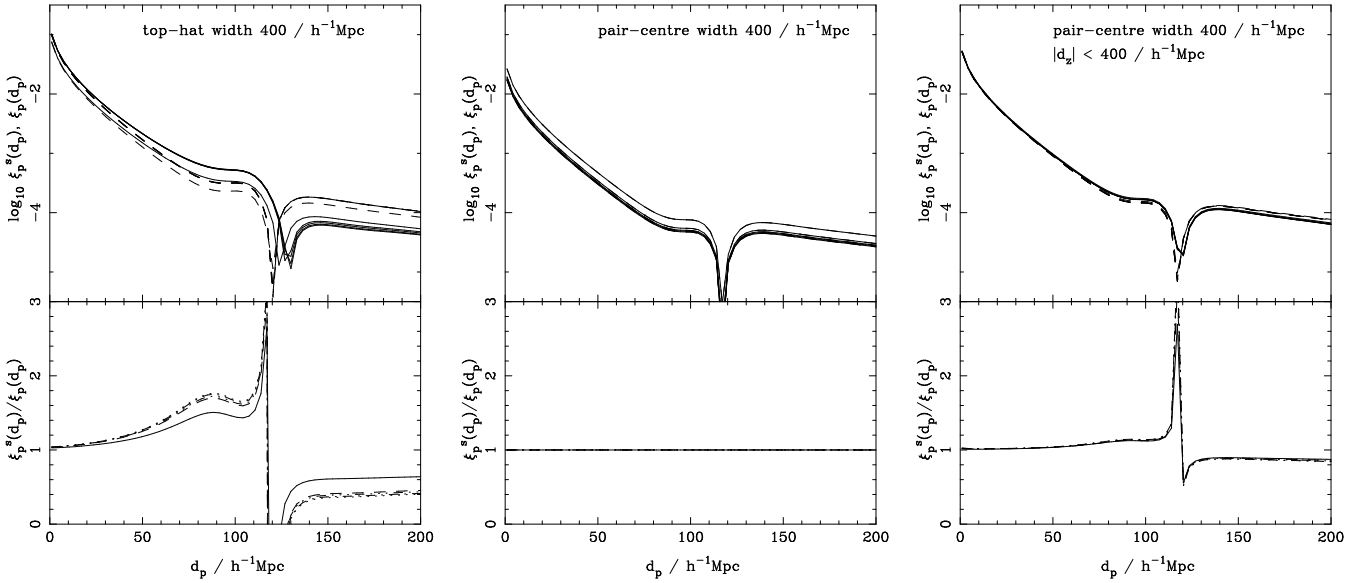

Figure 14. Top panels: Real-space (dashed lines) and redshift-space (solid lines) correlation functions predicted for the 5 radial bins drawn from the DES-like selection function, assuming it can be treated as a real-space boundary. Bottom panels: The ratio between the redshift-space and real-space projected correlation function. Here different line styles correspond to different bins: in the order of increasing redshift, they are solid, dashed, dot-dash, dotted, dot-dot-dash. From left to right: Top-hat bins of width $400\,h^{-1}\,\mathrm{Mpc}$ in the radial direction, pair-centre bins of width $400\,h^{-1}\,\mathrm{Mpc}$, and constrained pair-centre bins of width $400\,h^{-1}\,\mathrm{Mpc}$, including an additional constraint on the radial separation of $|d_z| < 400\,h^{-1}\,\mathrm{Mpc}$.

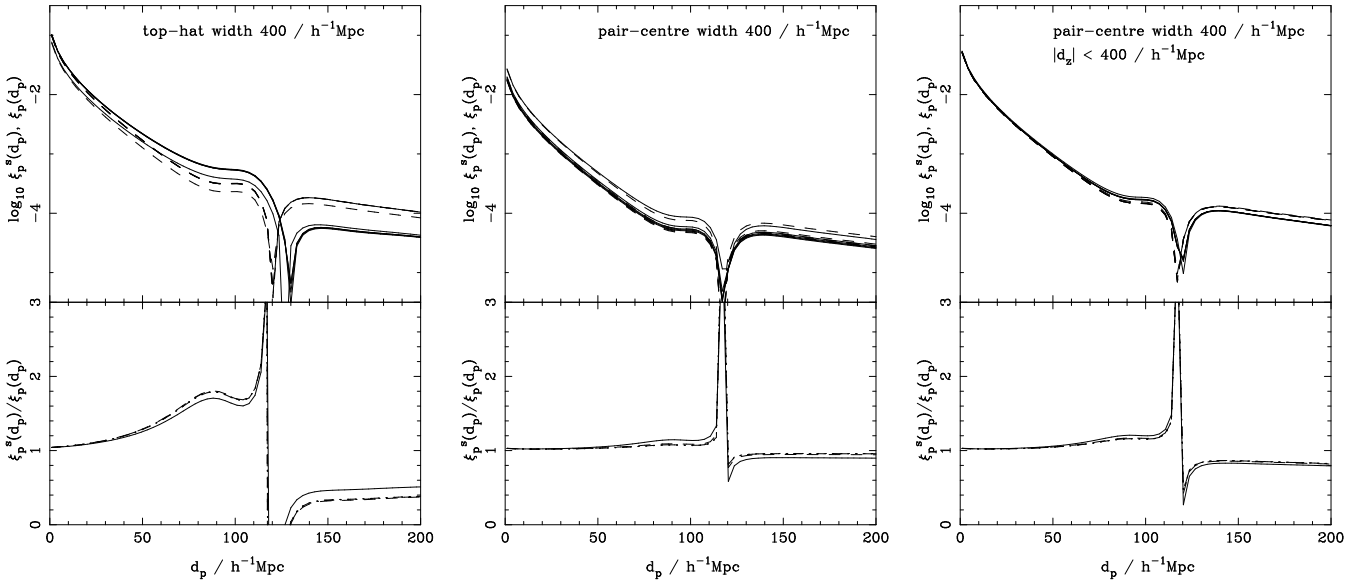

Figure 15. As Fig. 14, only in this case we treat the DES selection function as a redshift-space boundary.

direction. It simply makes sure that they do not produce a coherent effect on the measurements.

There are two potential disadvantages of the pair-centre binning scheme. One is the fact that the same galaxy may be included in multiple radial bins — thus introducing a correlation between radial bins. Another is the fact that such a scheme results in necessarily wider radial bins, which causes the clustering signal to be diluted. We do not feel that either is a large problem. Applying the more traditional top-hat binning scheme to photometric surveys necessarily results in overlapping radial bins (due to photometric redshift errors) and there will always be considerable covariance between radial bins selected with photometric redshifts — we do not think that pair-centre binning will make this problem considerably worse. The dilution effect can be mitigated by imposing a maximum separation between the pairs included in a pair-centre bin: we call this constrained pair-centre binning. As can be seen by comparing the middle and right-hand panels of Figs. 14 and 15, imposing such a constraint increases the expected signal while not causing a significant change in the effects of redshift-space distortions. More detailed studies of these effects are warranted, but we are confident that the reduction in the redshift distortion effect we observe when utilising pair-centre binning will make this scheme considerably preferable to a top-hat binning scheme.

Pair-centre binning completely removes the effect of redshift

distortions when given a uniform galaxy distribution. Such perfect distributions do not exist — most galaxy samples selections are based on an apparent magnitude limit — and thus realistic radial distributions of galaxies are more complicated. However, we have argued that if galaxy samples selected based on an apparent magnitude limit are cut back so that no galaxies k-corrected galaxies are missing from the sample, then this does not matter: the boundaries of the bins are either in real-space, or based on pair-centres, neither of which introduces redshift distortion effects.

We have argued, and it is clear from previous work, that any interpretation of projected clustering measurements must account for redshift space distortions. In fact, comparing correlation functions calculated using different binning schemes might actually prove to provide a mechanism for measuring the amplitude of the redshift-space distortions. This is beyond the scope of our current draft, and we leave this for subsequent work.

7.2 Future Surveys

To quantify the effect of redshift-space distortions for future surveys, we have used the expected radial selection function and photometric redshift distribution for the Dark Energy Survey to predict the effect of redshift-space distortions on projected clustering measurements. This analysis is also relevant to other planned surveys such as PanStarrs and the LSST, which will have similar radial selection functions. We have contrasted two different types of binning: top-hat — in which we only allow galaxies between a given radial bound to enter our sample— and pair-centre — in which we only count galaxy pairs with an average radial position that lies within our bounds. For typical bin widths that will be applied to these surveys, we find that top-hat binning in the radial direction leaves a strong signal from redshift-space distortions. Using a pair-centre binning scheme reduces the redshift-space distortion signal, by as much as 80$\%$ in realistic situations (see Fig. 14) and should therefore allow the measurements to be more sensitive to the cosmological parameters one wishes to constrain.

In this analysis, we have only considered the simplified situation where the redshift-space distortions act along one axis of a Cartesian basis. However, the arguments we have put forward in favour of pair-centre binning do not rely on this assumption, and will remain valid even when wide-angle effects are included in any analysis.

8 ACKNOWLEDGEMENTS

The authors thank the UK Science and Technology Facilities Research Council for financial support. WJP is also grateful for support from the Leverhulme Trust and the European Research Council. Simulated data was calculated and analysed using the COSMOS Altix 3700 supercomputer, a UK-CCC facility supported by HEFCE and STFC in cooperation with CGI/Intel.

We thank the DES Large-Scale Structure working group members, and especially Enrique Gaztanaga, for many helpful discussions. We also thank the referee for carefully reading our manuscript and providing excellent suggestions for improvements.

REFERENCES

Baldauf T., Smith R.E., Seljak U., Mandelbaum R., 2009, [[arXiv:0911.4973]]
Blake C., Collister A., Bridle S., Lahav O., 2007, MNRAS, 374, 1527
Cole S., et al., 2005, MNRAS, 362, 505
Colless M., et al., 2003, [[astro-ph/0306581]]
Eisenstein D.J., et al., 2005, ApJ, 633, 560
Evrard A.E., et al., 2002 ApJ, 573, 7
Fisher K.B., Scharf C.A., Lahav O., 1993, MNRAS, 266, 219
Gaztanaga E., Cabre A., Hui L., 2008, [[arXiv:0807.3551]]
Hamilton A.J.S., 1992, ApJ, 385, L5
Hamilton A.J.S., "Linear redshift distortions: A review", in "The Evolving Universe", ed. D. Hamilton, pp. 185-275 (Kluwer Academic, 1998) [[astro-ph/9708102]]
Kaiser N., 1987, MNRAS, 227, 1
Padmanabhan N., et al., 2007, MNRAS, 378, 852
Peebles P.J.E., 1973, ApJ, 185, 413
Peebles P.J.E., 1980, *The Large Scale Structure of the Universe*, Princeton University Press
Percival W.J., et al., 2001, MNRAS, 327, 1297
Percival W.J., et al., 2007a, ApJ, 657, 51
Percival W.J., Cole S., Eisenstein D., Nichol R., Peacock J.A., Pope A., Szalay A., 2007b, MNRAS, 381, 1053
Percival W.J., Schafer B., 2008, MNRAS, 385, L78
Percival W.J., et al., 2009, MNRAS submitted, [[arXiv:0907:1660]]
Perlmutter S., et al., 1999, ApJ, 517, 565
Regos E., Szalay A. S., 1995, MNRAS, 272, 447
Riess A.G., et al., 1998, AJ, 116, 1009
Ross, A. J. & Brunner, R. J., 2009, MNRAS, 399, 878
Sanchez A.G., Crocce M., Cabre A., Baugh C.M., Gaztanaga E., 2009, MNRAS submitted, [[arXiv:0901.2570]]
Simpson F., Peacock J.A., Simon P., 2009, [[arXiv:0901.3085]]
York D.G., et al., 2000, AJ, 120, 1579