Determining accurate measurements of the growth rate from the galaxy correlation function in simulations

Carlos Contreras,* Chris Blake, Gregory B. Poole and Felipe Marin
Centre for Astrophysics and Supercomputing, Swinburne University of Technology, PO Box 218, Hawthorn, VIC 3122, Australia

Accepted 2012 December 18. Received 2012 December 17; in original form 2011 November 18

ABSTRACT
We use high-resolution N-body simulations to develop a new, flexible empirical approach for measuring the growth rate from redshift-space distortions in the 2-point galaxy correlation function. We quantify the systematic error in measuring the growth rate in a 1 h⁻³ Gpc³ volume over a range of redshifts, from the dark matter particle distribution and a range of halo-mass catalogues with a number density comparable to the latest large-volume galaxy surveys such as the WiggleZ Dark Energy Survey and the Baryon Oscillation Spectroscopic Survey. Our simulations allow us to span halo masses with bias factors ranging from unity (probed by emission-line galaxies) to more massive haloes hosting luminous red galaxies. We show that the measured growth rate is sensitive to the model adopted for the small-scale real-space correlation function, and in particular that the ‘standard’ assumption of a power-law correlation function can result in a significant systematic error in the growth-rate determination. We introduce a new, empirical fitting function that permits the galaxy pairwise velocity distribution, the quantity which drives the non-linear growth of structure, to be measured as a non-parametric stepwise function. Our (model-independent) results agree well with an exponential pairwise velocity distribution, expected from theoretical considerations, and are consistent with direct measurements of halo velocity differences from the parent catalogues. In a companion paper, we present the application of our new methodology to the WiggleZ Survey data set.

Key words: cosmological parameters – cosmology: theory – large-scale structure of Universe.

1 INTRODUCTION
The growth rate of cosmic structure is a key parameter which quantifies the cosmological model. While distance–redshift probes such as Type Ia supernovae (e.g. Riess et al. 1998; Perlmutter et al. 1999; Kowalski et al. 2008; Hicken et al. 2009; Amanullah et al. 2010) or baryon acoustic oscillations (e.g. Blake & Glazebrook 2003; Seo & Eisenstein 2003; Eisenstein et al. 2005; Percival et al. 2010; Blake et al. 2011) give us information about the cosmic expansion history, and observations of the cosmic microwave background radiation allow the study of the composition and physics of the early Universe (e.g. Komatsu et al. 2011), the growth rate describes how the small density perturbations present in the early Universe evolve to form the large-scale structure which populates the Universe today. Hence, the growth rate provides a fundamental test of the laws of gravity which operate in the expanding cosmos.

Measurements of the growth rate have assumed a special importance as evidence has accumulated that the expansion of the Universe has entered a phase of acceleration. Two principle explanations have been put forward for accelerating expansion. The first is the presence of some unknown material constituent of the Universe with the exotic property of negative pressure, whose energy density has become dominant within the last half of the age of the Universe. This material is known as ‘dark energy’. The second explanation is that our current theory of gravity, General Relativity, must be modified on large scales to account for the observations without invoking exotic constituents. Various methods for modifying General Relativity have been explored (see Tsujikawa 2010). Measurements of the growth rate of structure over different cosmic epochs can help discriminate between these two physical interpretations of the observations (Linder & Jenkins 2003; Linder & Cahn 2007; Guzzo et al. 2008; Wang 2008).

The growth rate of structure within a given cosmological model can be derived by solving the differential equation for the linear density of matter perturbations δ at scale factor a in an expanding Universe (Peebles 1980). The growth rate at redshift z is defined by \( f(z) = \frac{\delta(z)}{\delta_0(z)} \), where \( \Omega_m(z) \) is the matter density relative to the critical density and \( \gamma \) is a phenomenological parameter which takes the value 0.55 for General Relativity (Linder & Cahn 2007).

* E-mail: ccontrer@astro.swin.edu.au
A powerful method for measuring the growth rate is to exploit the anisotropic signature it imprints in the clustering within galaxy redshift surveys, known as redshift-space distortions (RSD; see e.g. Hawkins et al. 2003; Tegmark et al. 2004, 2006; da Angela et al. 2008; Guzzo et al. 2008; Okumura et al. 2008; Cabré & Gaztañaga 2009; Blake et al. 2011). The growth of structure is driven by coherent flows of matter into clusters and superclusters, and the resulting coherent galaxy peculiar velocities produce a correlated signature in the pattern of 2-point galaxy clustering. The galaxy clustering anisotropy on large scales is parameterized by a measured distortion parameter $\beta$, which is related to the growth rate by $f = b\beta$, where $b$ is the bias parameter which relates galaxy overdensity to matter overdensity.

An alternative method to map out the clustering pattern and measure the growth rate is weak gravitational lensing, which involves modelling the observed correlated alignment of distant background galaxy shapes by the foreground large-scale structure. However, the current sensitivity of the method is limited (e.g. Bernstein 2009), despite extremely promising future prospects. Other methods that have been developed to constrain the growth rate include the luminosity function and gas-mass fraction of X-ray-selected clusters (Rapetti et al. 2009), and galaxy bulk flows measured in the local neighbourhood (Abate & Lahav 2008; Watkins, Feldman & Hudson 2009; Nusser & Davis 2011). However, at the present time RSD in galaxy clustering represent the most accurate method for measuring the growth rate of structure.

Utilizing RSD as a cosmological probe has two requirements. First, a large galaxy redshift survey must be performed in order to reduce the sources of error in the measurement, cosmic variance and galaxy shot noise. Secondly, reliable models must be fitted to the clustering measurements. The second issue, correlation-function modelling, is the subject of the current paper, whereas applications of these techniques to the WiggleZ Dark Energy Survey (Drinkwater et al. 2010) are presented in a companion paper.

In order to model the 2D redshift-space galaxy correlation function transverse and parallel to the line of sight, $\xi(s, \pi)$, we must model the pattern of peculiar velocities acquired by galaxies as part of the growth of structure. There are (crudely speaking) two sources for these peculiar velocities: the random velocity a galaxy possesses with respect to its own group or cluster, and the velocity the galaxy acquires as part of bulk flows into bigger structures such as superclusters. These two effects distort the shape of the measured $\xi(s, \pi)$ in a characteristic way which can be modelled and split from the underlying isotropic real-space correlation function. The challenge is to measure the RSD parameter due to bulk flows, $\beta$, in a manner free from systematic error, and relate that to the growth rate of structure using parallel measurements of the galaxy bias factor.

The aim of this paper is to address the obstacles which must be overcome to obtain reliable measurements of the cosmic growth rate from modelling the correlation function, namely (1) provision of an accurate model for the underlying isotropic real-space correlation function, (2) modelling the non-linear and random effects of galaxy pairwise velocities and (3) determining an accurate measurement of the galaxy bias factor. We assume throughout that the background cosmological parameters are known, for example from cosmic microwave background observations, hence we neglect Alcock–Paczynski distortions in this study.

There currently exists no model which is able to describe accurately the matter distribution and its clustering properties over the complete range of scales of our interest: $0-50\,h^{-1}\,\text{Mpc}$. At least three different physical regimes can be identified which shape the galaxy correlation function and RSD in this range: the non-linear, quasi-linear and linear regimes. On large scales ($\gtrsim 20\,h^{-1}\,\text{Mpc}$), perturbation theory provides a good description of the growth of the clustering pattern (Bernardeau et al. 2002; Crocc & Scoccimarro 2006; Carlson, White & Padmanabhan 2009; Nishimichi et al. 2009; Taruya et al. 2009), a continuity equation holds between velocity and density, and the RSD pattern is well described by a simple anisotropy known as the ‘Kaiser limit’ (Kaiser 1987). At intermediate scales ($\lesssim 20\,h^{-1}\,\text{Mpc}$), in the quasi-linear regime, perturbation theory breaks down (Peebles 1980; Hatton & Cole 1998; Ratchiffe et al. 1998; Landy 2002; Scoccimarro 2004) and significant corrections are needed to the Kaiser limit formulation. At small scales ($\lesssim 3\,h^{-1}\,\text{Mpc}$), in the non-linear or ‘1-halo’ regime, the matter distribution is virialized in groups and clusters of galaxies and high velocity dispersions imprint the ‘fingers-of-god’ feature into the correlation function. All of these physical ingredients leave different signatures in the correlation function, currently making it difficult to construct a physically motivated model which is valid across the entire range of scales. In this study, we instead focus on empirical approaches for modelling the observations.

*N*-body dark matter simulations, and the halo catalogues that can be constructed from them, are a powerful method for modelling the full range of non-linear processes described above. Because the cosmological parameters and input growth rate in the simulations are known, we can study the systematic errors that arise from using particular algorithms to extract these observables. Simulations allow us to see in detail pieces of information that are typically hidden in the real data samples, such as the velocity distributions in virialized clusters or the shape of the real-space correlation function.

Other studies in the literature have used simulations to test and improve models fitted to RSD. In Cabré & Gaztañaga (2009), the pairwise velocity distribution in redshift space was studied in detail and found to be scale dependent, leading to the inclusion in the model of two independent velocity dispersion terms applying at scales smaller and larger than $2\,h^{-1}\,\text{Mpc}$ (see also Slosar, Seljak & Tasitsiomi 2006). Jennings, Baugh & Pascoli (2011) used *N*-body simulations to test models of varying complexity for recovering the true value of RSD for different cosmologies via the power spectrum. They demonstrate that linear models by themselves do not extract an unbiased growth rate (see also Matsubara 2008; Taruya et al. 2009; Kwan, Lewis & Linder 2011; Okumura & Jing 2011; Bianchi et al. 2012).

In our work, we use a new set of *N*-body simulations – the Gigaparsec WiggleZ (GiggleZ) Simulations (Poole et al., in preparation) – to develop a new, empirical method for extracting the growth rate of structure from a range of halo catalogues and redshifts. The distinguishing feature of these simulations is their low particle mass in comparison with most large-volume simulations, which is appropriate for modelling the relatively low-mass haloes probed by emission-line galaxies mapped by the WiggleZ Dark Energy Survey. In contrast to the studies cited above, which largely considered the galaxy power spectrum, we base our analysis on the galaxy correlation function. This is a commonly used statistic for quantifying galaxy clustering which has several merits including that (1) different physical processes (such as shot-noise) are confined to distinct sets of scales, and (2) it is less sensitive than the power spectrum to modelling the survey selection function. In the absence of a complete physical model, it has been standard in the literature to fit a power-law model as the real-space correlation function when modelling the data in redshift space, and include non-linearities via an exponential pairwise velocity distribution function. In this paper, we use the GiggleZ simulations to critically examine these assumptions, proposing a new, improved empirical fitting
function and new techniques for studying the non-linear velocity dispersion.

Our paper is structured as follows: in Section 2, we describe the N-body simulations we employ in more detail. In Section 3, we outline how the 2-point correlation function is measured from the simulation data, and in Section 4, we specify the models we fit to these measurements. In Sections 5 and 6, we describe the performance and results of these fits to the correlation functions measured from the dark matter distribution and halo-mass catalogues, respectively. In Section 6.2, we discuss the determination of the galaxy bias parameter from these data, and in Section 7, we summarize and discuss our findings.

2 THE GiggleZ SIMULATION

We conducted our analysis using the Gigaparsec WiggleZ Survey simulations (GiggleZ). The main simulation, which we utilize here, is a 2160^3 particle dark matter simulation run in a periodic box 1 h^{-1} Gpc on a side. The resulting particle mass of this simulation is 7.5 \times 10^9 h^{-1} M_\odot which permits us to resolve bound systems with masses \gtrsim 1.5 \times 10^{11} h^{-1} M_\odot, facilitating studies of haloes with clustering bias factors ranging from near unity (e.g. galaxies in the GiggleZ Dark Energy Survey) to in excess of 2 [e.g. luminous red galaxies (LRGs) in the Sloan Digital Sky Survey].

A Wilkinson Microwave Anisotropy Probe 5 (WMAP-5) cosmology with (\Omega_M, \Omega_k, h, \sigma_8, n) = (0.727, 0.273, 0.0456, 0.705, 0.812, 0.960) was assumed for this simulation, with the initial conditions constructed to yield a CAMB (Lewis, Challinor, & Lasenby 2000) power spectrum for a starting redshift of z = 49 using the Zel'dovich approximation (Zel'Dovich 1970; Buchert 1992).

Bound structures were identified using SUBFIND (Springel et al. 2001), which uses a friends-of-friends (FoF) scheme followed by a substructure analysis to identify bound overdensities within each FoF halo. We use the SUBFIND substructures for all our analysis in this paper and use the value of each halo's maximum circular velocity $V_{\text{max}}$ as a proxy for mass. This choice of mass proxy was made to increase the robustness and reproducibility of our results, since it avoids many numerical effects and biases associated with specific halo finding schemes and halo mass definitions. We use the centre of mass velocities of each halo when computing RSD.

In order to explore clustering systematics as a function of halo mass, we rank-ordered the GiggleZ substructures by their maximum circular velocities and selected contiguous groupings of 250 000 systems. This from, we chose a series of six halo groupings, ranging in halo mass from GiggleZ galaxies to Sloan LRGs (with median values of $V_{\text{max}} = 130, 160, 190, 230, 260, 300 \text{ km s}^{-1}$) with a number density $2.5 \times 10^{-4} h^3 \text{ Mpc}^{-3}$, which is well matched to that of the WiggleZ survey and ongoing Baryon Oscillation Spectroscopic Survey (BOSS). Fig. 1 illustrates the ranges of maximum circular velocities and halo masses contained in these six catalogues for the $z = 0$ snapshot. We note that the completeness limit of the halo catalogue is around 120 km s$^{-1}$ at $z = 0$.

3 MEASUREMENTS OF THE 2D CORRELATION FUNCTION IN THE GiggleZ SIMULATION

In this paper, we quantify clustering using the 2D 2-point correlation function. This statistic is determined by counting the number of unique galaxy pairs as a function of transverse and parallel separation to the line of sight ($\sigma$ and $\pi$), and comparing the result to a similar pair-count performed on a randomly distributed catalogue. We averaged over 30 random catalogues, each containing an equal number of particles as the data set. We used two estimators to measure the correlation function. The first is the (Landy & Szalay 1993) estimator, the standard minimum-variance procedure performed with real galaxy catalogues and corresponding random catalogs. In the second estimation, we exploit the fact that our simulation is a cube with periodic boundary conditions and only count data-pairs, estimating the random pair-count using analytic methods. The results agree very closely, and the second method allows a much more rapid computation. We computed the pair counts in square (\sigma, \pi) bins with side 2 h^{-1} Mpc, up to 40 h^{-1} Mpc in \sigma and 30 h^{-1} Mpc in \pi (300 data bins).

We obtained the data covariances using the jack-knife procedure, in which we divided the 1 h^{-1} Gpc cube into a number $N_{\text{JK}}$ of identical jack-knife regions. Singular value decomposition (SVD) analysis of the resulting covariance matrices is crucial for understanding their robustness in the $\chi^2$ fitting procedure. Fig. 2 illustrates the spectrum of SVD eigenvalues for a range of choices of the number of jack-knife regions. We found that for our data set, 7^3 jack-knife subvolumes produced noisy covariance matrices with both of the employed methods, and this effect is not completely ameliorated if the lowest eigenvalues are truncated in the re-constructed covariance matrix. SVD analysis showed that increasing the number of jack-knife subregions improves asymptotically the quality of the covariance matrix. In our case, the number in the range $10^3$--20^3 are good choices. For our default choice $N_{\text{JK}} = 10^3$, we note that the size of each jack-knife region is 100 h^{-1} Mpc, significantly exceeding the clustering scales of interest. We checked that the growth-rate measurements presented in this paper do not depend significantly on whether we use the full covariance matrix, a truncated matrix in which the lowest-amplitude eigenvalues of an SVD decomposition are excluded, or a diagonal error matrix.

We performed these measurements using simulation snapshots at redshifts $z = 0$ and $z = 0.6$, in both real space and redshift space, for the dark matter distribution and for the six different halo-mass catalogues. We selected $z = 0.6$ because this is the median redshift of the WiggleZ Dark Energy Survey, and $z = 0$ because here the...
Determining the growth rate from simulations

Figure 2. The curve of eigenvalues in an SVD of the covariance matrix which results as we vary the number of jack-knife subregions. The robustness of the covariance matrix improves asymptotically until we reach $N_{JK} = 20^3$, which is approximately the maximum number of subvolumes in which the original data cube can be divided whilst still retaining dimensions bigger than the relevant scales in our 2-point correlation function measurement. The shape of these curves is similar for both methods of calculating the covariance matrix described in Section 3. In the case of lower numbers of jack-knife subdivisions, we were forced to truncate the lowest eigenvalues of the covariance matrix to attain stable fits, whilst for $N_{JK} > 10^3$ full and truncated covariances gave consistent results.

non-linearity in the clustering pattern that we are modelling will be most significant. We randomly subsampled the dark matter catalogue for each snapshot to $10^6$ particles before performing the correlation function estimation. The expected values for the growth rate $f$ for these two snapshots are 0.49 and 0.76, respectively, corresponding to the $\Lambda$ cold dark matter cosmological parameters used to construct the simulations. When generating redshift-space positions, we used a plane-parallel approximation, shifting coordinates along one axis. In Fig. 3, we plot the 2D real-space and redshift-space correlation functions of the GiggleZ simulation dark matter distribution for the $z = 0.6$ snapshot. In Fig. 4, we plot the corresponding 2D redshift-space correlation functions of the six halo-mass catalogues.

4 MODELLING THE REDSHIFT-SPACE CORRELATION FUNCTION

4.1 Constructing the model

The fact that large-scale galaxy surveys observe redshifts, not distances, implies that their clustering pattern is distorted by galaxy peculiar velocities. These peculiar velocities are generally modelled in the correlation function by a combination of the two effects which dominate in the large-scale and small-scale limit: the large-scale coherent flow of galaxies into clusters and superclusters, and the small-scale random motions of galaxies within virialized structures. The large-scale effects of coherent flows on the power spectrum and correlation function can be described by the standard treatment of Kaiser (1987) and Hamilton (1992) and the small-scale random velocity distribution can be introduced by convolving with a function $f(v)$, as summarized for example by Hawkins et al. (2003).

We note some potential systematic errors in this approach, which we discuss in turn in the remainder of this section. First, the isotropic real-space correlation function $\xi_r(s)$ must be modelled reliably in order to extract the anisotropic signature. Historically, a power law has been employed (Hawkins et al. 2003; Madgwick et al. 2003; Cabrè & Gaztañaga 2009). However, with increasing quality of data and simulations, a power law has become a bad approximation to the true non-linear clustering pattern. We discuss some improvements below. Secondly, the non-linear behaviour of small-scale peculiar velocities is entirely modelled by the single function $f(v)$. However, in reality this function is describing a scale-dependent combination of at least two physical effects: the virialized motion of galaxies within haloes, and the non-linearity in the coherent flows of galaxies which damp the velocity power spectrum on quasi-linear scales (Sheth 1996; Slosar et al. 2006). In detail, a single function is unlikely to provide a good match in both regimes, and indeed more complex models have been considered for matching small-scale data (Cabrè & Gaztañaga 2009). Another possibility is to reduce the impact of systematic modelling errors by excluding data at small scales from the fit (Hawkins et al. 2003), although this comes

Figure 3. Measurements of the 2D real-space (left-hand panel) and redshift-space (right-hand panel) correlation functions of dark matter particles in the $z = 0.6$ snapshot of the main GiggleZ simulation.
4.2 Models for the real-space correlation function

4.2.1 Fitting formulae from \textit{CAMB} and \textit{HALOFIT}

For a given set of cosmological parameters, the matter power spectrum (hence correlation function) at recombination can be numerically calculated by solving the Boltzmann transport equation; a popular publicly available code which provides this solution is \textit{CAMB} (Lewis et al. 2000). The effect of the non-linear growth of structure at an arbitrary redshift can be incorporated using the ‘\textit{HALOFIT}’ recipe calibrated by N-body simulations (Smith et al. 2003). In this model, we can determine the galaxy correlation function by combining this with a linear galaxy bias parameter $b$ as a simple normalization. We refer to the non-linear real-space galaxy correlation function generated in this manner as the ‘\textit{CAMB} model’, and by combining this real-space correlation function with the RSD parameters described above we can determine the corresponding redshift-space correlation function:

$$\xi_r(\sigma, \pi) = \text{CAMBmodel}[b, \beta, f(v)].$$  \hfill (1)

We explore below the dependence of the results on the fitted range of scales. We obtained the input \textit{CAMB} power spectrum using the WMAP-7 best-fitting cosmological parameters (Larson et al. 2011) which are consistent with the input parameters for the GiggleZ simulation.

4.2.2 Power-law correlation function

In previous studies, the real-space galaxy correlation function at small scales has often been modelled with a power-law form $\xi_r = (\frac{r}{r_0})^{-\gamma}$ and the full model for the redshift-space correlation function can be written as

$$\xi_r(\sigma, \pi) = \text{PowerLawModel}[^{\gamma, r_0, \beta, f(v)}].$$  \hfill (2)

4.2.3 Quadratic correlation function ($\xi_r, \text{QCF}$)

Improving data from both galaxy surveys and numerical dark matter simulations has demonstrated that the real-space galaxy correlation function deviates from a power law at scales beyond $\sim 15 h^{-1} \text{ Mpc}$ (Hawkins et al. 2003). Thus, in the case of high signal-to-noise data, a power-law model produces a poor fit to the correlation function. This motivates our definition of a new empirical fitting formula with greater flexibility than the \textit{CAMB} and power-law models. We introduce here the \textit{quadratic correlation function (QCF) model},

$$\xi_r(r) = \left(\frac{r}{r_0}\right)^{-\gamma + q \log_{10}(\frac{r}{r_0})},$$  \hfill (3)

where $q$ is the additional quadratic parameter. This is just a simple quadratic equation in logarithmic space:

$$y = a + bx + qx^2; \quad x = \log_{10}(r); \quad y = \log_{10}(\xi_r).$$  \hfill (4)

Our full model in this scenario is then

$$\xi_r(\sigma, \pi) = \text{QCFModel}[^{\gamma, r_0, q_0, \beta, f(v)}].$$  \hfill (5)

Although this model is not physically motivated, it produces an impressive fit over the wide range of scales 1–50 $h^{-1}$ Mpc to both a suite of non-linear matter correlation functions generated by \textit{CAMB}, and (as we show below) to the real-space correlation function of a range of halo-mass catalogues from the GiggleZ simulation. The flexibility of the QCF model enables us to achieve fits to the redshift-space correlation function with lower systematic errors. We note that the QCF model, when applied to a halo correlation function, can also accommodate a scale-dependent bias factor.
In this study, we consider two variations. First, we explore both possibilities by fitting to our data the weighted combination,

\[ f_i(v) = x f_i(v) + (1 - x) f_i(v), \]

thus adding a final parameter, \( x \), to our correlation-function fits, which we required to lie in the range \( 0 \leq x \leq 1 \).

Secondly, we considered expressing the pairwise velocity distribution as a general stepwise function:

\[ f(v) = a_i \quad \text{for} \quad v_{i-1} \leq |v| < v_i \quad i = 1, 2, \ldots, N \]

for a number of intervals \( N \) from \( v_0 = 0 \) up to some maximum pairwise velocity \( v_N = v_{\text{max}} \). The distribution is normalized such that \( \int_{-\infty}^{\infty} f(v) dv = 1 \), which implies that \( 2 \sum_{i=1}^{N} a_i (v_i - v_{i-1}) = 1 \). The quantities \( N \) and \( v_{\text{max}} \) are set by hand (depending on the quality of the data) by inspecting the solutions and requiring that \( f(v) \) should be generally positive and smoothly decreasing from \( v = 0 \) to \( v_{\text{max}} \).

For our simulation data set, we typically obtain robust solutions using \( N \) in the range \( 6-9 \) and \( v_{\text{max}} \) between 1500 and 2500 km s\(^{-1}\).

For each combination of \( \beta \) and the parameters describing the real-space correlation function, we obtained the set of coefficients \( a_i \) which minimized the chi-squared statistic between model and data by solving an \( N \times N \) linear system of equations. As explored in more detail below, we obtain stable, physically sensible solutions provided that we include correlation function measurements at small transverse separations \( \sigma \) in our fitted range (which have most sensitivity to virialized motions). Using this technique we can determine the real underlying shape of the pairwise velocity distribution, and test whether or not the exponential or Gaussian models indeed provide a good description. We give algebraic details of this calculation in Appendix.

5 MODEL FITS TO THE DARK MATTER CORRELATION FUNCTION

5.1 Growth rate of the dark matter

First, we explored how accurately the three real-space correlation function models we have defined (\( \text{CAMB}, \text{QCF}, \text{power law} \)) described the real-space dark matter correlation function of the simulation, before the addition of RSD. One example of these results (for \( z = 0 \)) is displayed in Fig. 5. The minimum values of chi-squared, which are listed in Table 1, demonstrate that the \( \text{CAMB} \) model provides the

Table 1. This table explores the effect of excluding the small-scale data bins from the fitting of the real-space dark matter correlation function. \( r_{\text{min}} \) is the minimum value of the total separation \( r = \sqrt{x^2 + y^2} \) for the data bins included in the fit. We note that the \( \text{CAMB} \) model provides a good fit to the real-space dark matter correlation function, even in the small-scale regime. The discrepancy in normalization between the fitted \( \text{CAMB} \) correlation function and the simulation is less than 1 per cent.

<table>
<thead>
<tr>
<th>( \text{CAMB} ) bias ((h^{-1} \text{ Mpc}))</th>
<th>( r_{\text{min}} )</th>
<th>( \chi^2/\text{dof} ) ( \text{CAMB} )</th>
<th>( \chi^2/\text{dof} ) ( \text{QCF} )</th>
<th>( \chi^2/\text{dof} ) ( \text{PLAW} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.006</td>
<td>0</td>
<td>1.246</td>
<td>6.909</td>
<td>9.470</td>
</tr>
<tr>
<td>1.006</td>
<td>2</td>
<td>1.234</td>
<td>1.687</td>
<td>9.425</td>
</tr>
<tr>
<td>1.006</td>
<td>4</td>
<td>1.192</td>
<td>1.161</td>
<td>8.491</td>
</tr>
<tr>
<td>1.004</td>
<td>6</td>
<td>1.019</td>
<td>0.931</td>
<td>6.438</td>
</tr>
<tr>
<td>1.000</td>
<td>8</td>
<td>0.674</td>
<td>0.704</td>
<td>4.102</td>
</tr>
<tr>
<td>0.999</td>
<td>10</td>
<td>0.633</td>
<td>0.669</td>
<td>2.559</td>
</tr>
</tbody>
</table>
C. Contreras et al.

Table 2. This table compares models and measurements in real space, for the six halo-mass catalogues from the GiggleZ simulation considered in this study. The power-law model consistently fails to match the measured correlation function. The CAMB model produces a good description of the clustering of low-mass haloes, but breaks down for higher mass catalogues due to scale-dependent halo bias. The QCF model is flexible enough to produce a good fit to the real-space correlation function of all the halo catalogues.

<table>
<thead>
<tr>
<th>Halo group</th>
<th>CAMB</th>
<th>QCF</th>
<th>PLAW</th>
</tr>
</thead>
<tbody>
<tr>
<td>bias</td>
<td>$\chi^2$/dof</td>
<td>$\chi^2$/dof</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.069 (005)</td>
<td>1.403</td>
<td>1.370</td>
</tr>
<tr>
<td>2</td>
<td>1.260 (003)</td>
<td>1.717</td>
<td>1.472</td>
</tr>
<tr>
<td>3</td>
<td>1.417 (003)</td>
<td>1.494</td>
<td>1.475</td>
</tr>
<tr>
<td>4</td>
<td>1.572 (003)</td>
<td>2.893</td>
<td>1.696</td>
</tr>
<tr>
<td>5</td>
<td>1.707 (003)</td>
<td>1.815</td>
<td>1.368</td>
</tr>
<tr>
<td>6</td>
<td>1.856 (002)</td>
<td>2.532</td>
<td>1.729</td>
</tr>
</tbody>
</table>

best-fitting and a stable value of the galaxy bias $\sim 1$ for a variety of fitting ranges, whilst the power-law model provides a poor fit to the data. The QCF model also yields a good fit for all scales $> 4 \ h^{-1} \text{Mpc}$. 

Next, we included the effects of RSD in the dark matter correlation functions of the $z = 0.0$ and $z = 0.6$ snapshots, and fitted the clustering models described in Section 4. Fig. 6 displays how the measured parameters for the two snapshots depend on the minimum transverse separation fitted, $\sigma_{\text{min}}$, for the three different real-space correlation function models we are considering. We fix the maximum transverse separation fitted at $\sigma_{\text{max}} = 40 \ h^{-1} \text{Mpc}$. We performed the fits using a Monte Carlo Markov Chain procedure, exploring the multidimensional space of variables of the models and obtaining their joint and individual probability distributions. The horizontal lines indicate the input values of the simulation, determined from its fiducial cosmological parameters. We find that the CAMB and QCF models are both able to recover the input growth rate with low systematic error and reduced $\chi^2 \approx 1$, for $\sigma_{\text{min}} > 2 \ h^{-1} \text{Mpc}$. On the other hand, the assumption of a power-law model produces a significant systematic error and bad fit.

5.2 Pairwise velocity distribution of the dark matter

In Fig. 6, we show the fitted values of the pairwise velocity dispersion parameters $x$ and $a$, which are fitted for each model jointly with the other parameters. The results strongly favour an exponential rather than a Gaussian velocity distribution ($x \approx 1$). The systematic discrepancy between the values of $a$ fitted in the QCF and CAMB models indicate that there is some cross-talk between $a$ and the shape of the real-space correlation function on small scales.

We now compare these measurements with a direct determination of the shape of the pairwise velocity distribution $f(v)$, which is possible using our stepwise fitting method described in Section 4.3. This constitutes a further check for systematic errors in the models which describe this distribution. We fitted the stepwise distribution in seven velocity bins up to a maximum velocity of 2000 km s$^{-1}$ (our results are not sensitive to these choices). We assumed a CAMB real-space matter correlation function in these fits.

The results are displayed in Fig. 7, comparing the best-fitting stepwise distributions to both the best-fitting exponential and Gaussian models, and to a direct measurement of the pairwise velocity distribution from the catalogues. In the top row, the stepwise velocity distribution is fitted to the correlation function of the dark matter subsample, for three different values of the minimum value of $\sigma$ of the data bins included in the fit. In the bottom row, the pairwise velocity distribution is directly measured from the dark matter catalogues, this time varying the maximum total separation ($R = \sqrt{\sigma^2 + \pi^2}$) of the pairs considered for this measurement. We set a maximum value of $R$ for this calculation because we only want to estimate the dispersion within a single bulk flow, not between different bulk flows. The distribution obtained from the stepwise fitting process is in excellent agreement with its direct measurement from the catalogue, demonstrating the capacity of the stepwise approach to recover the pairwise velocity distribution from the measured

Figure 6. Fits for the growth rate, bias, the pairwise velocity dispersion, the parameter $x$ and the resulting reduced $\chi^2$, from the GiggleZ redshift-space 2D dark matter correlation function, for different values of the minimum transverse scale included in the fit, $\sigma_{\text{min}}$. Top: results from the $z = 0.6$ snapshot of the simulation. Bottom: results from the $z = 0.0$ snapshot of the simulation. Measurements are shown for three different real-space correlation function models. The CAMB and QCF models (red circles and green triangles) show good agreement with the expected theoretical simulation growth rate (represented by the horizontal line) and rest of parameters, while the power-law model fit (blue squares) is strongly affected by systematics. The parameter $x$ controls whether the pairwise velocity distribution is modelled as an exponential or a Gaussian, with the data clearly favouring an exponential form ($x \approx 1$).
Determining the growth rate from simulations

6 MODEL FITS TO THE HALO CORRELATION FUNCTIONS

We now turn our attention to fitting models to the halo correlation functions, which represent galaxies in the simulation. We explore how accurately our empirical models can describe the halo clustering in redshift space.

6.1 RSD parameters of the halo catalogues

Table 2 compares models and measurements in real space, for the six halo-mass catalogues from the GiggleZ simulation considered in this study. The power-law model consistently fails to match the measured correlation function. The CAMB model produces a good description of the clustering of low-mass haloes, but breaks down for higher mass catalogues due to scale-dependent halo bias. The QCF model is flexible enough to produce a good fit to the real-space correlation function of all the halo catalogues.

We then fitted the RSD models to the 2D redshift-space correlation function measurements for these six halo-mass catalogues. In Fig. 8, we show the reduced $\chi^2$ statistic corresponding to fitting our three RSD models to the range $\sigma_{\min} < \sigma < 40 h^{-1}$ Mpc, $0 < \pi < 30 h^{-1}$ Mpc, as a function of the choice of the minimum value $\sigma_{\min}$. Our motivation for this analysis is that systematic errors in the recovery of the RSD parameter are likely to be most serious at the lowest values of $\sigma$, where non-linear velocity effects such as the ‘fingers-of-god’ are most significant [note however Bianchi et al. (2012), who show in a similar analysis that increasing $\sigma_{\min}$ does not make systematics completely vanish]. We find that the power-law real-space model is a bad fit to the data, whereas the CAMB and QCF models show mutual agreement and better $\chi^2$ values. The QCF model provides the best fit to the clustering pattern, particularly for high-mass haloes.

The best-fitting values for the pairwise velocity dispersion $a$ depend on the real-space correlation function model in a similar way to the fits to the dark matter catalogues. For these halo catalogues the $x$ parameter, which was introduced to determine the shape of the pairwise velocity distribution, does not favour the exponential function over the Gaussian function as strongly as in the case of dark matter. We speculate that a halo catalogue with a narrow range of masses will also possess a more uniform dispersion of pairwise correlation function. The distributions we obtain are consistent with an exponential and not a Gaussian shape for $f(v)$, and we recover the standard deviation of the pairwise velocity distribution $a \sim 300$ km s$^{-1}$, independently of specifying a model for $f(v)$. This agreement between the model-independent stepwise measurement and the assumed functional form increases our confidence in the reliability of the exponential small-scale RSD model.
velocities, producing a closer match to a Gaussian function than before. In order to convert these results into growth-rate measurements, we require an estimate of the galaxy bias, which we consider in the next section.

6.2 Bias factor of the halo catalogues

Modelling the RSD in the 2-point correlation function gives us one of the two quantities, $\beta$, that are necessary to deduce the growth rate $f = \beta b$. We now discuss and test possible methods to measure the bias factor $b$, which describes the clustering of galaxies (or dark matter haloes) relative to the underlying matter distribution. For example more massive haloes, which are sampled preferentially from more clustered regions of the Universe, will possess a higher bias factor $b$ and a correspondingly lower value of $\beta$ (resulting in a less flattened large-scale redshift-space 2D correlation function). These more massive haloes are sites of early galaxy formation which today can be observed as LRGs (Eisenstein et al. 2001; Cabré & Gaztañaga 2009); less massive haloes preferentially host blue, star-forming galaxies (Madgwick et al. 2003; Blake et al. 2010).

The bias factor cannot be deduced directly from the measured 2-point correlation function without some other assumption (such as the underlying amplitude of the real-space matter correlation function), although we note that such a determination may be possible using the 3-point correlation function (Verde et al. 2002; Gaztañaga & Scoccimarro 2005; Marin 2011). However, the use of N-body simulation catalogues allows us to compare the bias measurements resulting from fits of a real-space correlation function model with those directly determined from the data via

$$b^2(r) = \frac{\xi_G(r)}{\xi_{DM}(r)},$$

where $\xi_G(r)$ and $\xi_{DM}(r)$ are the real-space correlation functions of the halo (galaxy) catalogue and dark matter distribution, respectively. We assume hereafter that bias is a constant in our fits, which is a good approximation on large scales.

Table 3 shows the consistency of the bias factors resulting from fitting a CAMB model to the halo-mass catalogues, with the direct measurements of the bias obtained by dividing the different halo-mass correlation functions by the dark matter correlation function and using equation 9. We divided the two correlation functions in each different separation bin, and averaged over each measurement as an independent estimate of the bias. We note that the two independent measurements are broadly consistent, with the main discrepancy appearing for small scales due to systematic non-linear effects. This discrepancy can be ameliorated by restricting the fitting range to exclude small scales, which are polluted by the highest amplitude systematic errors.

The broad agreement between the bias factors fitted by the CAMB model to the 2D redshift-space correlation function, and those measured directly from the 1D real-space correlation function of the simulation, gives us confidence to use the CAMB bias factors to measure the growth rate of our halo catalogues.

6.3 Growth rate of the halo catalogues

The resulting growth-rate measurements for the $z = 0.6$ simulation snapshot, combining the separate determinations of the RSD parameter $\beta$ and the galaxy bias, are plotted in Fig. 9. This figure illustrates the amplitude of the systematic error in measuring the growth rate
Determining the growth rate from simulations

Figure 9. The best-fitting value of the growth rate as a function of the minimum value of $\sigma$ included in the fit. Results are shown for each of the six halo-mass catalogues for the $z = 0.6$ snapshot, for the three different RSD models. This figure shows that we can recover the growth rate with minimal systematics when using QCF and CAMB models for the real-space correlation function.

as a function of halo mass, minimum transverse scale fitted, and model adopted for the real-space correlation function. All models contain a systematic error $\Delta f \approx 0.05$ for $\sigma_{\text{min}} < 6 \, h^{-1} \, \text{Mpc}$. However, our fitting procedure, and the adoption of the QCF model, allows us to recover accurate measurements when excluding the small-scale region ($\sigma_{\text{min}} > 6 \, h^{-1} \, \text{Mpc}$). Similar results are obtained for the $z = 0.0$ snapshot. We note in general that the model fits to halo catalogues described in this section contain somewhat higher systematic errors than the fits to the underlying dark matter distribution.

7 CONCLUSIONS

We have measured the 2D correlation function of the dark matter and halo-mass catalogues for two snapshots of the GiggleZ simulation at redshifts $z = 0.0$ and $z = 0.6$, and fitted different clustering models to recover the growth rate. We carefully study the effect on the results of the range of scales fitted and the halo mass, spanning halo bias factors between unity (hosting emission-line galaxies) and high-mass haloes traced by LRGs. We list our conclusions as follows.

(i) The commonly used power-law model for the real-space correlation function produces a poor fit to the clustering pattern and a systematic error in the resulting growth rate. A real-space correlation function based on a non-linear CAMB model does better, but breaks down for high-mass haloes. We introduce a new empirical model, the quadratic correlation function (QCF model, which has one more degree of freedom than a power law), which provides a better description of the real-space correlation function (particularly for high-mass haloes which possess significant scale-dependent bias) and produces a measurement of the input growth rate with a lower (5–10) amplitude of systematic error.

(ii) We introduce a new technique which permits the measurement of the pairwise velocity distribution as a stepwise function from the redshift-space correlation function. The pairwise velocity distribution measured directly from the simulation catalogues is consistent with this model, and matches closely to the exponential function expected from theoretical considerations.

(iii) We have quantified the amplitude of systematic error in the measured growth rate from our $1 \, \text{h}^{-3} \, \text{Gpc}^3$ simulation as a function of halo mass, the model employed and the minimum transverse scale $\sigma_{\text{min}}$ fitted. We find that for $\sigma_{\text{min}} < 6 \, h^{-1} \, \text{Mpc}$, the systematic measurement error using our procedure is $\Delta f \approx 0.05$. The adoption of the QCF model allows us to recover accurate measurements of the growth rate from halo catalogues when excluding the small-scale region ($\sigma_{\text{min}} > 6 \, h^{-1} \, \text{Mpc}$). This is consistent with the recent analysis of Bianchi et al. (2012).

(iv) We note that the halo correlation function contains a higher level of systematic modelling errors than the dark matter correlation function, due to scale-dependent galaxy bias. Our modelling allows us to recover the growth rate from the dark matter particle catalogue with no detectable systematic error.

We conclude that $N$-body simulations are an essential tool for testing and developing methods to measure the cosmic growth rate, and that our empirical techniques should be useful for modelling correlation functions measured in the latest large-volume galaxy redshift surveys. A companion paper will represent the application of these methods to the WiggleZ Dark Energy Survey data set.

ACKNOWLEDGMENTS

We acknowledge financial support from the Australian Research Council through Discovery Project grants funding the positions of GP and FM. This research was supported by CAASTRO: http://caastro.org. CC thanks Karl Glazebrook who suggested the stepwise approach, Eyal Kazin for his useful comments on the previous draft of this work, and Mercedes LópezMorales (IEEC), Mercedes Mollá (CIEMAT), Ana María Martín (UCLM) and
C. Contreras et al.

Luis David Moya for their kind hospitality, support and useful feedback.

REFERENCES

Guzzo L. et al., 2008, Nat, 451, 541
Komatsu E. et al., 2011, ApJS, 192, 18
Nishimichi T. et al., 2009, PASJ, 61, 321
Tegmark M. et al., 2006, Phys. Rev. D, 74, 123507

APPENDIX A: STEPWISE VELOCITY DISTRIBUTION METHOD

As an alternative to assuming a Gaussian or exponential pairwise velocity distribution, we explored fitting a stepwise function \( f(v) = a_i \) in a range of velocities to the data, together with the other model parameters. In order to avoid introducing a large number of additional free parameters, we have developed a method which allows these fits to be performed in a relatively simple and fast way.

The standard expressions for the correlation function multipoles in the large-scale limit (Kaiser 1987; Hamilton 1992) are

\[
\xi_0(r) = \left( 1 + \frac{2a_2}{3} + \frac{a_2^2}{5} \right) \xi_s(r) \tag{A1}
\]

\[
\xi_2(r) = \left( \frac{4a_2}{3} + \frac{2a_2^2}{7} \right) \left[ \xi_s(r) - \xi_s' \right] \tag{A2}
\]

\[
\xi_4(r) = \frac{8a_2^2}{35} \left[ \xi_s(r) + \frac{5}{2} \xi_s'(r) - \frac{7}{2} \xi_s''(r) \right] \tag{A3}
\]

where

\[
\xi_s(r) = \frac{3}{r^4} \int_0^r \xi_s(s) s^2 ds \tag{A4}
\]

\[
\xi_s'(r) = \frac{5}{r^5} \int_0^r \xi_s(s) s^4 ds \tag{A5}
\]

Starting from the real-space correlation function, such as that obtained from a CAMB matter power spectrum, we re-write equation A1 in the form:

\[
\xi_0(r) = C_0(\beta) J_0(\sigma, \pi) \tag{A6}
\]

where the form of the functions \( C_0 \) and \( J_0 \) can be deduced by analogy with equation A1, and we solve the integrals numerically. Now replacing the velocity distribution \( f(v) \) by a stepwise function \( a_i \) we obtain

\[
\xi(\sigma, \pi) = \int_{-\infty}^{\infty} \xi'(\sigma, \pi') f(v) dv = \sum_{i=1}^{N} a_i \left[ \int_{-\infty}^{0} \xi'(\sigma, \pi') dv + \int_{0}^{\infty} \xi'(\sigma, \pi') dv \right] \tag{A7}
\]

where we separate the negative and positive parts of the integral over \( v \) because of the loss of symmetry implied by the relation,

\[
\pi' = \pi + \frac{v}{H(z)} \tag{A8}
\]
Determining the growth rate from simulations

for the convolution with the pairwise velocity distribution. We now numerically calculate the following terms:

\[ Q_0(\sigma, \tau, k) = \int_{v_{1k-1}}^{v_{1k}} J_0(\sigma, \tau) P_0(\mu) dv \]

\[ Q_2(\sigma, \tau, k) = \int_{v_{1k-1}}^{v_{1k}} J_2(\sigma, \tau) P_2(\mu) dv \]

\[ Q_4(\sigma, \tau, k) = \int_{v_{1k-1}}^{v_{1k}} J_4(\sigma, \tau) P_4(\mu) dv, \] \hspace{1cm} (A9)

where \( P_i(\mu) \) are the Legendre polynomials in terms of \( \mu \), the cosine of the angle between the bin position and the line-of-sight axis \( \pi \).

We also define analogous functions \( Q'_i \) integrating over the negative part of the \( v \) axis. Now, we combine these functions for a given value of \( \beta \) to create a series of stepwise models \( \xi(\sigma, \pi) = \sum a_k F(\beta)_{\sigma,\pi,k} \), where the final \( F_{\sigma,\pi,k} \) terms correspond to

\[ F(\beta)_{\sigma,\pi,k} = C_0(\beta) \left[ Q_0(\sigma, \pi, k) + Q'_0(\sigma, \pi, k) \right] \]

\[ + C_2(\beta) \left[ Q_2(\sigma, \pi, k) + Q'_2(\sigma, \pi, k) \right] \]

\[ + C_4(\beta) \left[ Q_4(\sigma, \pi, k) + Q'_4(\sigma, \pi, k) \right]. \] \hspace{1cm} (A10)

Now, we have a model which is linearly dependent on \( a_k \), which constitutes an \( N \times N \) linear system of equations. Our \( \chi^2 \) equation is

\[ \chi^2(\beta) = \sum_{i,j} \left[ \left( \sum_k a_k F(\beta)_{i,k} - D_i \right) C^{-1}_{i,j} \left( \sum_k a_k F(\beta)_{j,k} - D_j \right) \right], \] \hspace{1cm} (A11)

where \( D \) is the data array, \( C \) is the covariance matrix, \( i \) and \( j \) represent bins in the correlation function data, and \( k \) is the index for the stepwise velocity distribution.

The \( n \)th equation of the \( N \times N \) linear system is

\[ \sum_{i,j} C^{-1}_{i,j} \left[ F(\beta)_{i,k} F(\beta)_{j,n} + F(\beta)_{j,k} F(\beta)_{i,n} \right] = \sum_{i,j} C^{-1}_{i,j} \left[ D_i F(\beta)_{j,n} + D_j F(\beta)_{i,n} \right]. \] \hspace{1cm} (A12)

The linear system may be solved by conventional methods and the parameter space in \( \beta \) can be quickly explored in search of the minimum \( \chi^2 \). The normalization of the stepwise function gives in this case the factor between the fitted galaxy correlation function and the \textsc{camb} matter correlation function, i.e. the bias \( b^2 \). If the real-space correlation function is a power law, the first set of numerical integrations may be replaced by analytical expressions, with the clustering length \( r_0 \) of the power law absorbed into the normalization factor.

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