Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **457,** 3652–3665 (2016)
Advance Access publication 2016 January 27

doi:10.1093/mnras/stw223

Recovering dark-matter clustering from galaxies with Gaussianization

Nuala McCullagh,[1⋆] Mark Neyrinck,[2⋆] Peder Norberg[1,3] and Shaun Cole[1]

[1]*Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*
[2]*Henry A. Rowland Department of Physics and Astronomy, The Johns Hopkins University 3400 N Charles St., Baltimore, MD 21218, USA*
[3]*Centre for Extragalactic Astronomy, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*

Accepted 2016 January 23. Received 2016 January 22; in original form 2015 November 6

ABSTRACT

The Gaussianization transform has been proposed as a method to remove the issues of scale-dependent galaxy bias and non-linearity from galaxy clustering statistics, but these benefits have yet to be thoroughly tested for realistic galaxy samples. In this paper, we test the effectiveness of the Gaussianization transform for different galaxy types by applying it to realistic simulated blue and red galaxy samples. We show that in real space, the shapes of the Gaussianized power spectra of both red and blue galaxies agree with that of the underlying dark matter, with the initial power spectrum, and with each other to smaller scales than do the statistics of the usual (untransformed) density field. However, we find that the agreement in the Gaussianized statistics breaks down in redshift space. We attribute this to the fact that red and blue galaxies exhibit very different fingers of god in redshift space. After applying a finger-of-god compression, the agreement on small scales between the Gaussianized power spectra is restored. We also compare the Gaussianization transform to the clipped galaxy density field and find that while both methods are effective in real space, they have more complicated behaviour in redshift space. Overall, we find that Gaussianization can be useful in recovering the shape of the underlying dark-matter power spectrum to $k \sim 0.5\,h\,\mathrm{Mpc}^{-1}$ and of the initial power spectrum to $k \sim 0.4\,h\,\mathrm{Mpc}^{-1}$ in certain cases at $z = 0$.

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

1 INTRODUCTION

Local density transforms, such as the log and Gaussianization transform (Neyrinck, Szapudi & Szalay 2009; Neyrinck 2011; Neyrinck, Szapudi & Szalay 2011), the log transform modified for a Poisson-sampled field, known as the A^*-transform (Carron & Szapudi 2014; Wolk, Carron & Szapudi 2015a; Wolk et al. 2015b), clipping (Simpson et al. 2011, 2016), and the copula transformation (Scherrer et al. 2010) have been proposed recently as methods to efficiently extract cosmological information from galaxy clustering data. The Gaussianization transform has been shown to restore small-scale information in the two-point statistics of the matter density field by reducing the covariance on small scales and providing better fidelity to the linear-theory shape, and thus tightening constraints on cosmological parameters (Neyrinck et al. 2009; Neyrinck 2011). It has also been suggested that Gaussianization may have the ability to separate galaxy bias from underlying clustering statistics (Neyrinck et al. 2014; Neyrinck 2014). Under the assumption that the bias of different fields (red and blue galaxies, dark matter) is encoded in the different one-point probability density functions (PDFs) of the fields, and the underlying clustering of the fields is the same, the Gaussianized fields of the biased tracers and the dark matter will have the same two-point statistics.

While both the A^*-transform and clipping have been applied to real galaxy data (Wolk et al. 2015a; Simpson et al. 2016), previous work on the Gaussianization transform for the power spectrum has focused largely on theoretical motivation and idealized cases such as the dark-matter distribution in N-body simulations. It is not known to what extent the benefits of the transform can be achieved in realistic galaxy samples. The effects of number density, shot noise, clustering properties, redshift-space distortions, and size of the density grid on the effectiveness of the Gaussianization transform have not been thoroughly explored in the context of two-point statistics, though they have been studied more systematically in the related context of the topology of large-scale structure (Melott, Weinberg & Gott 1988; Park, Kim & Gott 2005; Kim et al. 2014). Gaussianization has also been applied to power spectrum recovery from the Lyman α forest, though in this case the natural 'clipping' that arises from saturation of absorption in high-density regions causes the non-linear flux power spectrum to track the linear matter power spectrum down to relatively small scales, even without Gaussianization (Croft et al. 1998, 1999).

In this paper, we apply the Gaussianization transform to realistic galaxy samples, produced from the semi-analytic galaxy formation

⋆ E-mail: nuala.mccullagh@durham.ac.uk (NMcC); mark.neyrinck@gmail.com (MN)

code GALFORM (Cole et al. 2000; Baugh et al. 2005; Bower et al. 2006; Gonzalez-Perez et al. 2014; Lacey et al. 2015), with the aim of testing the ability of the transform to remove the effect of non-linear (scale-dependent) galaxy bias to recover the small-scale shape of the matter power spectrum. We apply the transform to blue and red galaxies from GALFORM (Gonzalez-Perez et al. 2014) and compare the shape of the power spectra of the transformed fields to each other as well as to those of the underlying dark matter and initial density fields. We study how the transform performs in real and redshift space, and explore the effects of varying the sample selection and the size of the density grid.

In Section 2 we describe the simulations and galaxy samples used in this work. In Section 3 we discuss the Gaussianization transform in detail and present the methods used to estimate the Gaussianized power spectrum, including a prescription for shot noise correction. In Section 4 we present our results comparing the Gaussianized and usual density statistics of galaxies and discuss the effects of sample selection and varying cell size. In Section 5, we compare the Gaussianization transform to clipping of the galaxy density field as a method for removing non-linearity and non-linear bias. We conclude in Section 6.

2 SIMULATION AND GALAXY SAMPLES

We use simulated galaxy samples from the GALFORM semi-analytic galaxy formation model (Gonzalez-Perez et al. 2014), which was run on the Millennium dark-matter-only *N*-body simulation (Springel et al. 2005; Guo et al. 2013).[1] Semi-analytic galaxy formation models such as GALFORM model the formation and evolution of galaxies using simple, physically motivated equations to predict baryonic physics within dark-matter halo merger trees (Cole et al. 2000; Baugh et al. 2005; Bower et al. 2006; Lacey et al. 2015). The output is a galaxy catalogue with realistic clustering. Such a catalogue is sufficient for our purposes because it allows us to apply the Gaussianization transform to realistic galaxy samples with different clustering properties, and to test its effectiveness in recovering the known shape of the underlying dark-matter and linear power spectra.

The dark-matter only 7-year *Wilkinson Microwave Anisotropy Probe* (*WMAP*7) Millennium simulation (MR7) uses a periodic box with $L_{\rm box} = 500\,{\rm Mpc}\,h^{-1}$ on a side, with 2160^3 dark-matter particles, and *WMAP*7 cosmological parameters: $\Omega_{\rm m} = 0.272$, $\Omega_{\rm b} = 0.0455$, $\Omega_\Lambda = 0.728$, $h = 0.704$, $n_{\rm s} = 0.967$, and $\sigma_8 = 0.810$ (Komatsu et al. 2011; Guo et al. 2013). We use the galaxy and dark-matter distributions at $z = 0$.

Fig. 1 shows the total stellar-mass function of the GALFORM galaxies (black solid line), as well as that of the red and blue galaxies separately. We create four samples using stellar-mass cuts of $10^{8.5}$, $10^{8.9}$, $10^{9.6}$ and $10^{10.1}\,h^{-1}\,{\rm M}_\odot$. The number densities, fraction of red galaxies, and other properties of the four samples are summarized in Table 1. We use the colour–magnitude cut shown in Fig. 2 (solid black line) to split the galaxies into red and blue. The equation of the colour cut is: $(g-r)_0 = a(M_r^{0.0} + 20) + b$, with $a = -0.047$ and $b = 0.55$. The colour–magnitude distributions of each sample are shown in different shades of red and blue.

In general, red galaxies are much more strongly clustered than blue galaxies (e.g. Norberg et al. 2002; Zehavi et al. 2002, 2011), and this can also be seen in the GALFORM galaxies (Campbell et al. 2015; Farrow et al. 2015). The left panel of Fig. 3 shows the positions of galaxies in a slice of the simulation box in real space. In redshift space, where the positions of the galaxies are distorted along the line of sight by their peculiar velocities, the small-scale behaviour of the two fields is very different. Red galaxies have much stronger fingers of god (FoG) because they live in much more massive clusters than blue galaxies. This is shown in the middle panel of Fig. 3. As we will discuss in Section 4, we find it useful to also consider galaxy fields where the FoG have been collapsed. This is possible to do in real galaxy data with high enough number density and a reliable group catalogue (Tegmark et al. 2004; Berlind et al. 2006). We perform an idealized version of such a procedure here by assigning each galaxy the central subhalo velocity as opposed to its own peculiar velocity. This leads to the field in the right panel of Fig. 3, which includes the large-scale distortion effect, but the FoG are collapsed. We refer to this as 'collapsed FoG space' throughout the paper.

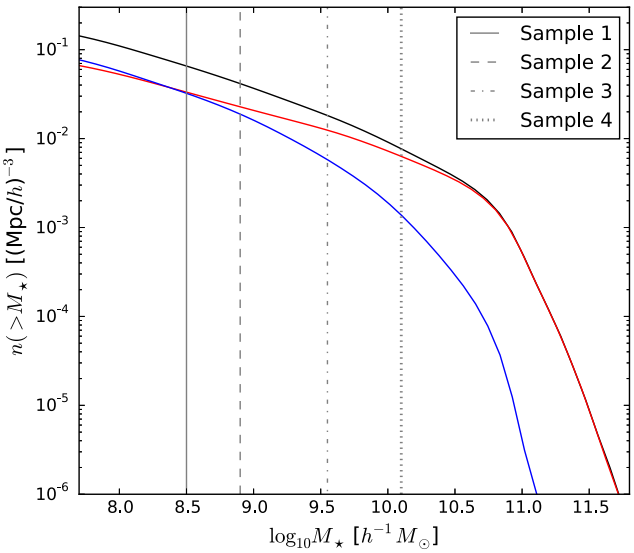

Figure 1. Cumulative stellar-mass function of the GALFORM galaxies. The red and blue lines refer to the stellar-mass functions of the red and blue galaxies, and the black solid line is the total. We make cuts in stellar mass to form four different samples, which are shown by the grey solid, dashed, dot–dashed, and dotted vertical lines.

3 GAUSSIANIZATION TRANSFORM AND POWER SPECTRUM ESTIMATION

3.1 Gaussianization transform

It has long been known that cosmological density fields are roughly lognormal (Coles & Jones 1991). The statistics of the log-transformed density field have been shown to have lower covariance on small scales and better fidelity to the linear-theory shape (Neyrinck et al. 2009). In principle, using the statistics of the log-density as opposed to the usual density variable can give much tighter constraints on cosmological parameters (Neyrinck 2011; Neyrinck et al. 2011). In the case of a perfectly lognormal field, analysis of the log-density correlation function accesses all of the Fisher information, whereas using arbitrarily high-point density correlation functions can give only a small fraction of that total information on small scales (Carron 2011; Carron & Neyrinck 2012).

Previous studies of the log transform have focused on the dark-matter density field from *N*-body simulations. The dark-matter

[1] Data from the Millennium simulation are available on a relational data base accessible from http://galaxy-catalogue.dur.ac.uk:8080/Millennium (Lemson & the Virgo Consortium, 2006).

Table 1. Summary of galaxy samples used. The columns give the minimum stellar mass, total number density, fraction of red galaxies ($f_{\rm R}$), fraction of empty cells in a (real space) 128^3 CIC grid for red and blue galaxy samples, and the average number of galaxies of a given type per halo with a galaxy of that type.

	$M_\star^{\rm min}$	$\bar{n}$	$f_{\rm R}$	$f_{\rm empty}^{128}$		$\langle N_{\rm gal}\rangle$		
	$[{\rm M}_\odot\, h^{-1}]$	$[({\rm Mpc}\, h^{-1})^{-3}]$		Red	Blue	Red	Blue	All
Sample 1	$10^{8.5}$	0.061	0.51	0.24	0.05	3.89	1.08	1.74
Sample 2	$10^{8.9}$	0.039	0.55	0.27	0.10	3.07	1.04	1.69
Sample 3	$10^{9.6}$	0.017	0.69	0.33	0.30	2.07	1.02	1.61
Sample 4	$10^{10.1}$	0.007	0.83	0.44	0.66	1.69	1.01	1.52

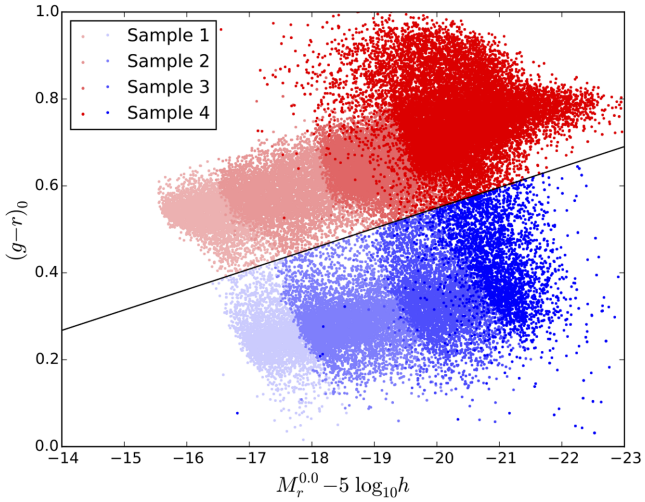

Figure 2. Colour–magnitude diagram for our four samples. The solid line shows the colour cut used to separate red (above the line) and blue (below the line) galaxies.

density (ρ) is interpolated to a grid using either a nearest-grid-point (NGP) or a cloud-in-cell (CIC) interpolation scheme (Hockney & Eastwood 1981), and the statistics of $A = \ln(1+\delta)$ (where $\delta \equiv \rho/\bar{\rho} - 1$ is the overdensity) are analysed as opposed to those of δ itself. In reality, we cannot observe the full dark-matter distribution, and instead we infer it through the presence of galaxies, which act as biased tracers of the dark-matter distribution. Depending on the number density and grid size, the log transform of the galaxy density field may not be well defined, as there may be many grid cells with zero density. For these cases, one option is the A^*-transform, providing an optimal local estimate of the log-density for a Poisson-lognormal field (Carron & Szapudi 2014). Another is Gaussianization, which is agnostic regarding an assumed underlying continuous field and how it is point-sampled. Gaussianization maps the one-point PDF of the density field on to an exactly Gaussian PDF with a specified variance by rank-ordering the density cells and transforming them using the inverse error function (Neyrinck 2011; Neyrinck et al. 2011):

$$\mathrm{Gauss}(\delta) = \sqrt{2}\sigma\ \mathrm{erf}^{-1}\left(2 f_{<\delta} - 1 + \frac{1}{N}\right), \quad (1)$$

where $f_{<\delta}$ is the fraction of cells below δ, σ is the standard deviation of the Gaussian, and N is the total number of cells. The $1/N$ term in equation (1) ensures that the argument in the inverse error function is between -1 and 1, and thus returns a finite value for all δ. For an exactly lognormal distribution, Gaussianization is equivalent to a log transform. In the case of poor sampling, where there are empty cells, all zero cells are mapped to a single value, which is set by requiring the Gaussian to have mean zero and variance σ^2. The transform thus can increase the effect of shot noise by accentuating the difference between cells with zero and non-zero density. The large number of cells with zero density is inconvenient from the point of view of Gaussianization. One way to avoid this problem would be to use a non-compact smoothing kernel extending over multiple cells, such as a Gaussian. However, this necessarily introduces an additional (smoothing) scale beyond the grid scale. Also, we found that Gaussianized power spectra respond in a complicated way to the additional smoothing scale of the Gaussian window on top of a (smaller-scale) CIC grid. Thus we choose to study the transform in the context of a CIC grid. As we will test and confirm in Section 4, for a CIC grid, choosing a grid size corresponding to an average of about 1 particle per cell ensures that the contribution from shot noise is manageable and the transform delivers the greatest benefits.

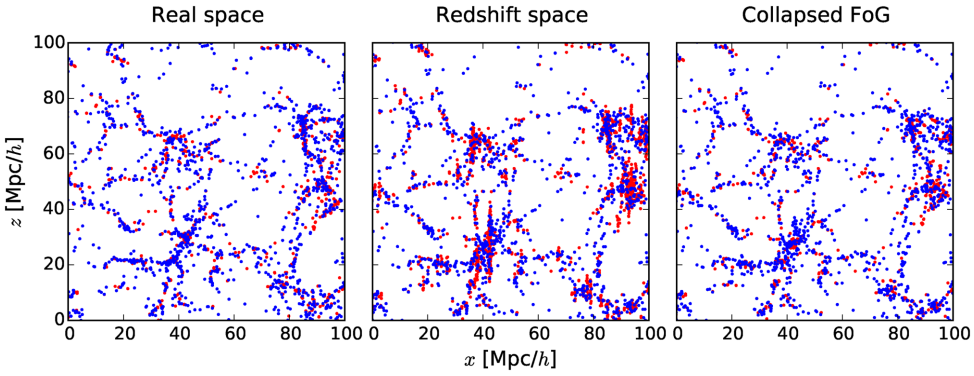

Figure 3. Spatial distribution of red and blue galaxies (with corresponding colour dots) in a 7 Mpc h^{-1} slice from Sample 2 in real space (left), redshift space (centre), and with collapsed FoG (right). The z-axis is the line-of-sight direction and we use the distant observer approximation for the redshift-space distribution.

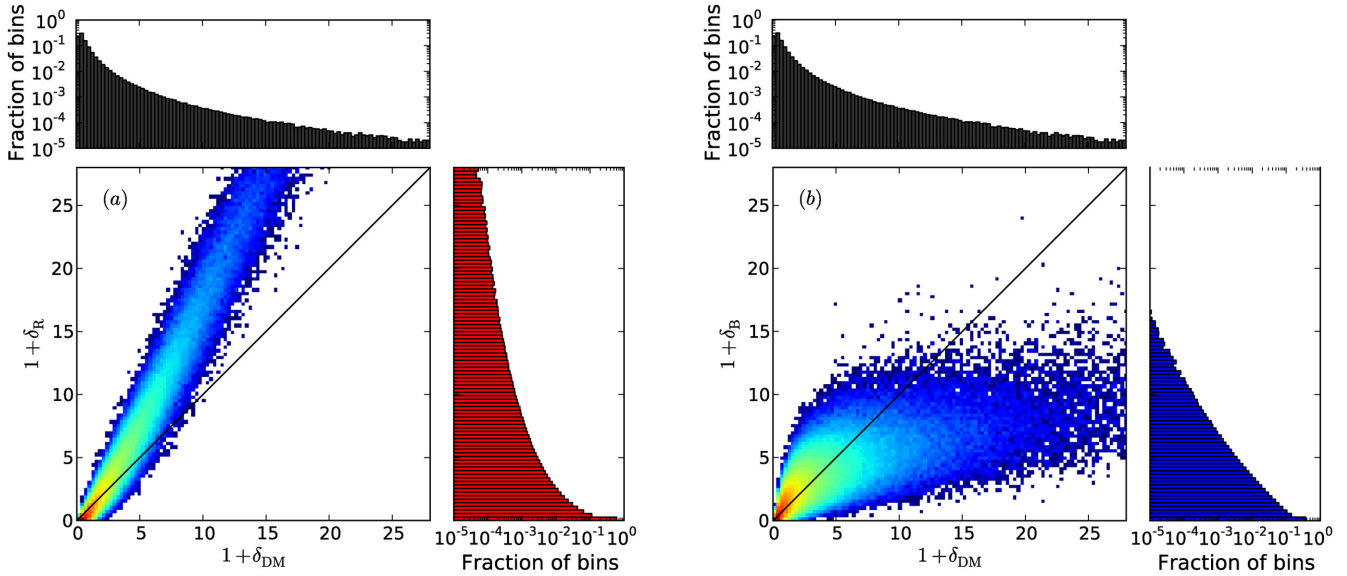

Figure 4. Distribution of red (left panels) and blue (right panels) galaxy densities versus dark-matter density on a 128^3 grid for Sample 2 in real space. Side panels show the one-point PDFs of the dark matter (top) and galaxies (right) and the centre panels show the two-dimensional distribution in each case. Colours in centre panels show number of cells using a logarithmic colour mapping.

We explore the Gaussianization transform first through studying the relationship between the usual and transformed galaxy and dark-matter density fields. We use the galaxies from Sample 2 in real space to illustrate these relationships. Fig. 4 shows two-dimensional histograms of the red (a) and blue (b) galaxy densities from Sample 2 versus the dark-matter density on 128^3 CIC grids, as well as the one-point PDFs of each field. From these figures it is clear that red galaxies are highly biased with respect to the dark matter, whereas blue galaxies have bias less than 1. The relationship between red galaxy density and dark-matter density is also quite tight compared to the blue galaxies over this range of densities.

Next, we consider the relationship between the Gaussianized galaxy fields and the Gaussianized dark-matter field. Fig. 5 shows the two-dimensional histograms of the Gaussianized red (a) and blue (b) fields versus the Gaussianized dark-matter field from Sample 2. Note that we only show the non-zero density cells in these figures. For the Gaussianization transform, we have used equation (1) with $\sigma = 1.0$ in all cases. The value of σ used is arbitrary, and only affects the amplitude of the Gaussianized field, and thus affects the amplitude of the Gaussianized statistics but not the scale-dependence. All the zero-density cells in the galaxy fields map to a single value in the Gaussianized field. The relationship between the Gaussianized red galaxy field and dark matter is remarkably tight, especially in high-density regions. The Gaussianized fields of blue galaxies and dark matter are clearly correlated, but with more scatter in high-density regions than in the case of the red galaxies. For example, if we consider cells in both the red and blue galaxy fields that have Gaussianized density values close to 3.0 (± 0.05), the standard deviation of the corresponding Gaussianized dark-matter density cells is much smaller for the red (0.08) than the blue (0.41).

Fig. 6 shows the Gaussianized dark-matter PDF, which is by definition a Gaussian with mean zero and $\sigma = 1.0$. The red and blue PDFs show the distributions of dark-matter cells that have zero density ($\delta = -1$) in the red and blue galaxy fields, respectively. For the CIC density, zero-density cells correspond to cells with no particles and whose neighbouring cells also have no particles. Note

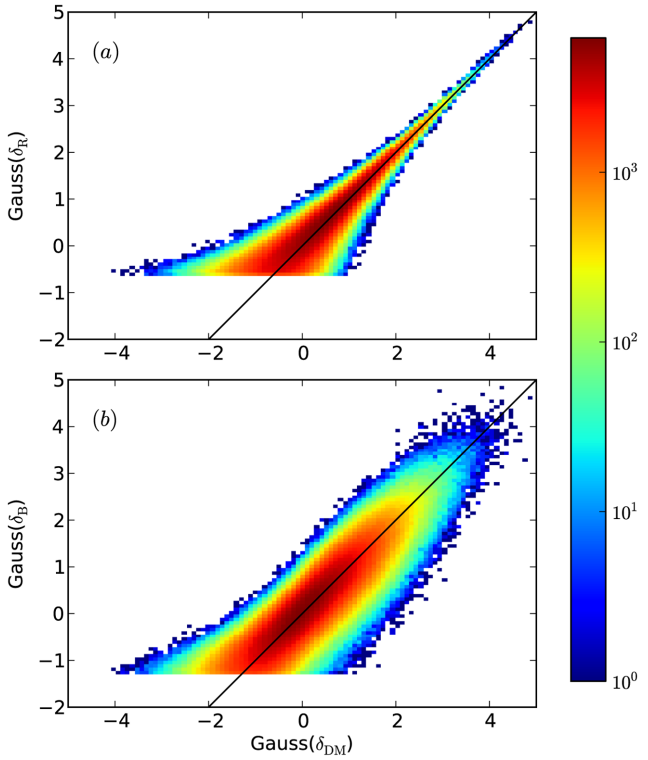

Figure 5. Distribution of Gaussianized density of red (a) and blue (b) galaxies versus Gaussianized dark-matter density on a 128^3 CIC grid for Sample 2. In both panels, we show only cells where the galaxy density is non-zero and there are more empty cells for the red galaxy sample. Colour bar gives the number of cells with a logarithmic colour mapping.

that although the galaxy fields have similar number densities in this sample ($\bar{n}_{\rm R} = 0.021$ and $\bar{n}_B = 0.018$ (Mpc $h^{-1})^{-3}$), there are more empty cells in the red galaxy field because it is more highly clustered than the blue.

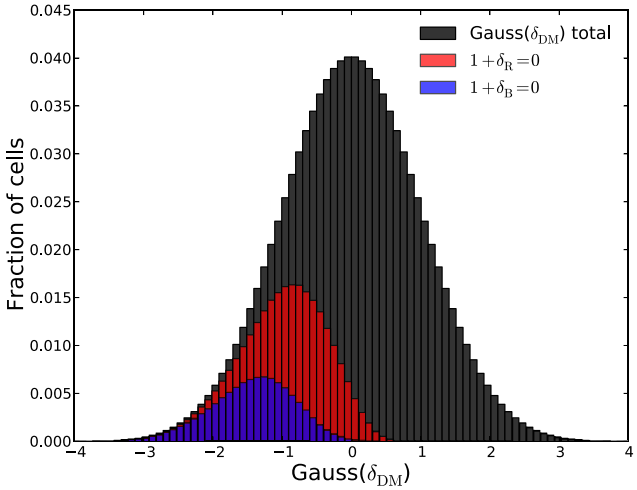

Figure 6. Distribution of Gaussianized dark-matter density cells. In black, all cells; in red (blue), cells that are empty in the red (blue) galaxy density field for Sample 2.

3.2 Gaussianized statistics

Next, we discuss methods used to estimate the power spectra of the Gaussianized fields, normalize the large-scale amplitude of the power spectra, and correct for shot noise.

Raw power spectra are computed through an FFT of the field under consideration (the δ field or the transformed field). We square the Fourier amplitudes to obtain the three-dimensional power spectrum of the field, and then average over shells in $|\boldsymbol{k}|$ to find $P(k)$. In redshift space, this computes the angular-average or monopole of the full two-dimensional redshift-space power spectrum.

The large-scale amplitude of a Gaussianized power spectrum depends on various quantities, including the intrinsic amplitude of fluctuations, the large-scale bias, the variance of the Gaussian, and the grid size used. While this amplitude does contain information, in this work we focus only on the shape of the power spectrum, and thus we choose to normalize all galaxy power spectra to the large-scale amplitude of the real-space dark-matter power spectrum at $z = 0$. We define the large-scale amplitude of a given power spectrum as:

$$B = \frac{\sum_i P(k_i) N_i}{\sum_i N_i}, \tag{2}$$

where N_i is the number of modes in bin i, and the sum is taken for $k < 0.07\ h\ \mathrm{Mpc}^{-1}$, to include only linear modes. For example, to normalize a Gaussianized power spectrum $P_{\rm Gauss}$ to the dark-matter power spectrum $P_{\rm DM}$, we compute B for each and multiply the Gaussianized power spectrum by $B_{\rm DM}/B_{\rm Gauss}$.

We must also correct for the effects of shot noise in our power spectra. For the usual δ power spectrum, we correct for Poisson shot noise by subtracting $1/\bar{n}$, where $\bar{n}$ is the tracer number density. For the power spectra of the Gaussianized fields, we apply a shot-noise correction procedure, detailed below.

In general, the contribution of shot noise in the Gaussianized power spectrum is more complicated than in the usual density power spectrum, and may depend on scale, number density, grid size, and intrinsic clustering properties of the sample. However, Neyrinck (2011) found that the contribution from shot noise in the Gaussianized power spectrum does not diverge greatly from a constant form on quasi-linear scales. Therefore, we make the ansatz that on these scales, the shot noise will contribute in a scale-independent way:

$$P^{\bar{n}}_{\rm meas}(k) = P_{\rm true}(k) + \frac{A}{\bar{n}}, \tag{3}$$

where A is an unknown amplitude that we must fit for in a given galaxy sample. In general, A may depend on the clustering of the given sample, so its value will be different for red and blue galaxies, and will depend on the sample definition.

We test this ansatz by randomly down-sampling a given galaxy sample to several different values of $\bar{n}$ and computing the resulting Gaussianized power spectra. We look at the differences between these power spectra, which arise from the difference in shot noise and so should depend only on the difference between the reciprocals of the number densities, if our ansatz is valid:

$$P^{\bar{n}_1}_{\rm meas}(k) - P^{\bar{n}_2}_{\rm meas}(k) = A\left(\frac{1}{\bar{n}_1} - \frac{1}{\bar{n}_2}\right). \tag{4}$$

We estimate A for each pair of $\bar{n}_1, \bar{n}_2$ using the above expression and weight by the number of modes in each k-bin, N_i:

$$\hat{A}(\bar{n}_1, \bar{n}_2) = \left(\frac{\bar{n}_2 - \bar{n}_1}{\bar{n}_1 \bar{n}_2}\right) \frac{\sum_i \left(P^{\bar{n}_1}_{\rm meas}(k_i) - P^{\bar{n}_2}_{\rm meas}(k_i)\right) N_i}{\sum_i N_i}, \tag{5}$$

where the sum is over bins $0.1 \leq k \leq 0.3\ h\ \mathrm{Mpc}^{-1}$, as we expect our ansatz to apply over this range of scales. We use four different sample sizes for each sample, giving six combinations of pairs of $(\bar{n}_1, \bar{n}_2)$, and we weight all pairs equally to find the best-fitting value of A to correct for the shot noise in equation (3).

Fig. 7 shows the Gaussianized power spectra of the red and blue galaxies from Sample 1, down-sampled to different values of $\bar{n}$. The red lines show the Gaussianized power spectra of red galaxies at different number densities, and the blue lines show the power spectra of blue galaxies at the same number densities. The left panels show the power spectra before shot noise correction, and the

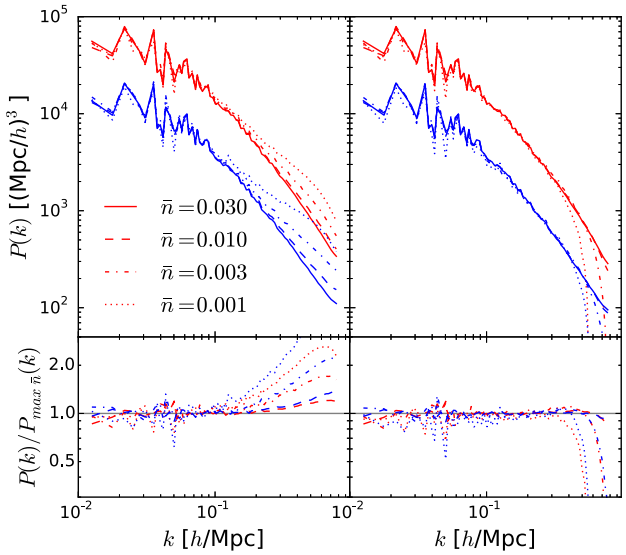

Figure 7. Gaussianized power spectra with fixed large-scale amplitudes, before (left) and after (right) shot-noise correction. The lower panels show the ratios of the down-sampled power spectra to that of the largest sample ($\bar{n} = 0.030$). The deviation from unity in the lower-left panel is the shot-noise term. The ratios in the lower right panel agree much better with each other, which indicates that our model for shot noise is reasonable on these scales.

right panels show the shot-noise-corrected power spectra. In these plots, the large-scale amplitudes of each have been normalized to that of the usual density power spectra of the galaxies of each colour using equation (2). In the left panel the effect of shot noise can be seen in the high-k tail of the power spectra as the sample becomes smaller. The lower left panel shows the ratio of the power spectra of the sampled galaxy population to that of the densest sample [$\bar{n} = 0.030$ (Mpc h^{-1})$^{-3}$]. The deviation grows with decreasing $\bar{n}$. The differences between these power spectra are used to fit for the value of A as in equation (5). The upper right panel shows the power spectra after the shot noise correction has been applied, and the lower right panel shows the corresponding ratios. Note that the range of k that this shot noise correction works well on depends on the number density $\bar{n}$.

From this figure, we see that after shot-noise correction, the power spectra converge as $\bar{n}$ increases. We conclude that for $\bar{n} \geq 0.010$ (Mpc h^{-1})$^{-3}$, the procedure results in the correct power spectrum [recovers '$P_{\rm true}(k)$' to within 10 per cent] on scales larger than the Nyquist frequency $k \leq 0.8\,h$ Mpc^{-1}. For $\bar{n} = 0.003$ (Mpc h^{-1})$^{-3}$ the correction is accurate up to $k \sim 0.6\,h$ Mpc^{-1}. For the lowest number density shown here, $\bar{n} = 0.001$ (Mpc h^{-1})$^{-3}$, the shot noise correction is especially poor even on relatively large scales. In general, we find that the accuracy of the shot noise procedure depends on several variables, including number density and fraction of empty cells. For the 128^3 grid size, we limit our analysis to samples with number densities $\bar{n} > 0.001$ (Mpc h^{-1})$^{-3}$. This excludes the blue galaxy sample from Sample 4 from our analysis, as it has a number density of $\bar{n} = 0.001$ (Mpc h^{-1})$^{-3}$. For each sample and cell size, we test the shot noise correction procedure to determine on what scales it can be trusted.

4 RESULTS

We first present the measured statistics from a single sample (Sample 2 in Table 1) on a 128^3 CIC grid. We choose this grid size because it corresponds to roughly one particle per cell for this sample. In Section 4.1, we show the comparison between the galaxy and dark-matter statistics for both the usual and transformed galaxy density fields. In Section 4.2, we compare the usual and transformed statistics to that of the linear power spectrum, and in Section 4.3 we compare the statistics of the red galaxy field to those of the blue. In Section 4.4, we discuss the effects of varying both the sample selection and the grid size on the results.

4.1 Galaxy versus dark-matter statistics

Fig. 8 shows the normalized, shot-noise corrected power spectra in real space (top), redshift space (middle), and with collapsed FoG (bottom). In redshift space and with collapsed FoG, we measure the angular-averaged (monopole) power spectrum. The solid black, red, and blue lines show the Gaussianized dark matter, red galaxy, and blue galaxy power spectra, respectively. The dashed lines show the corresponding power spectra of the usual density fields, and the green line shows the linear (input) power spectrum computed using CAMB (Lewis, Challinor & Lasenby 2000). In real space, the usual density power spectra (dashed lines) of the red and blue galaxies deviate significantly from the dark-matter power spectrum for $k \gtrsim 0.15\,h$ Mpc^{-1}. This is due to non-linear galaxy bias. The Gaussianized power spectra (solid lines) lie more or less on top of one another to the smallest scales shown here (to the Nyquist frequency of the 128^3 grid, $k_{\rm Ny} = 0.8\,h$ Mpc^{-1}), which supports the hypothesis that the transform removes non-linear bias. The Gaussianized

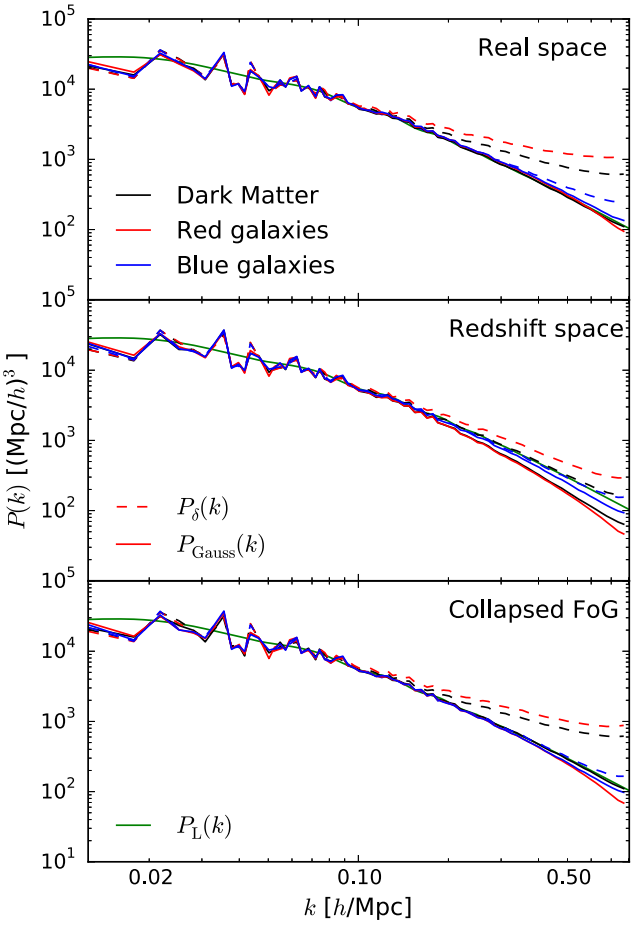

Figure 8. Power spectra of dark matter (black lines), red galaxies (red lines), and blue galaxies (blue lines) in real space (top panel), redshift space (middle panel), and collapsed FoG space (bottom panel). Solid lines show Gaussianized density power spectra and the dashed lines show the usual density power spectra. The green line (identical in all panels) shows the linear (input) power spectrum at $z = 0$. Large-scale amplitudes of all $P(k)$ are normalized to the real-space dark-matter power spectrum, as described in Section 3.2.

power spectra are also much closer to the linear power spectrum on small scales in real space.

The agreement with dark matter can be seen more clearly in Fig. 9, which shows the ratio of the galaxy power spectra with the corresponding dark-matter power spectrum. The dashed lines show the ratio of the usual density power spectra of the galaxy fields to that of the dark-matter field. The solid lines show the ratio of the Gaussianized power spectra of galaxies to the Gaussianized dark-matter power spectrum. The shaded regions show 10 per cent deviation from the dark-matter spectra. In the top and bottom panels, we compare galaxy statistics to real-space dark-matter statistics. In the middle panel, we compare redshift-space galaxy statistics to redshift-space dark-matter statistics.

In real space, the Gaussianized galaxy statistics have the same shape as that of the dark matter to much smaller scales ($k \sim 0.4\,h$ Mpc^{-1} at 10 per cent level) than the usual density statistics ($k \sim 0.15\,h$ Mpc^{-1} at 10 per cent level). However, the situation is more complicated in redshift space, where the Gaussianized red galaxy power spectrum agrees with the underlying dark-matter statistics to even smaller scales than in real space ($k \sim 0.6\,h$ Mpc^{-1}) but the Gaussianized blue power spectrum deviates at larger scales

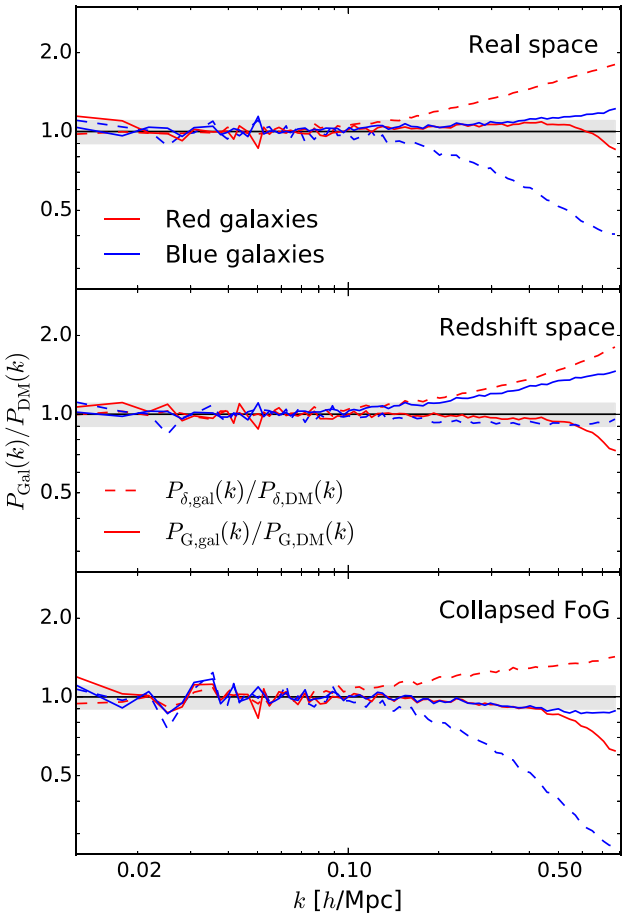

Figure 9. Ratio of red and blue galaxy power spectra with dark-matter power spectra in real space (top panel), redshift space (middle panel), and collapsed FoG space (bottom panel). For real space and collapsed FoG space, we compare to the real-space dark-matter statistics. For redshift space, we compare to the redshift-space dark-matter statistics. As in Fig. 8, solid lines show Gaussianized power spectra and dashed lines show the usual density power spectra. Red and blue lines corresponding to red and blue galaxy statistics. The shaded regions show 10 per cent deviation from the dark-matter spectra.

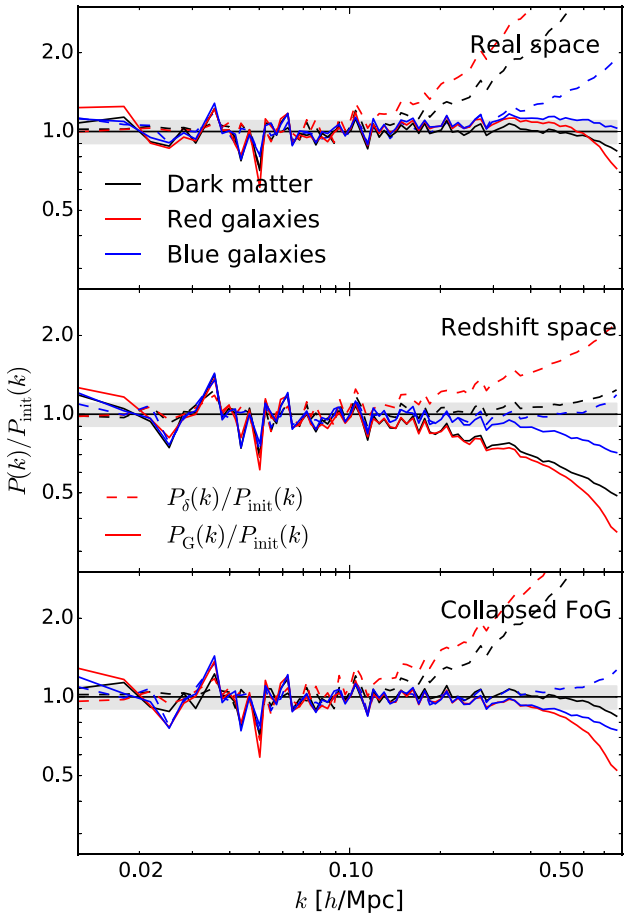

Figure 10. Ratio of red and blue galaxy power spectra with the initial (linear) power spectrum in real space, redshift space, and collapsed FoG space. The denominator in all three panels is the same. Line colours and styles are the same as in Fig. 8. Power spectrum of the Gaussianized galaxy density field recovers well the initial power spectrum in real and collapsed FoG space.

($k \sim 0.2\ h$ Mpc^{-1}). Collapsing the FoG brings the agreement back to similar levels as in real space.

This comparison suggests that the hypothesis that Gaussianization removes non-linear galaxy bias holds only in real space on quasi-linear scales ($k \lesssim 0.4\ h$ Mpc^{-1}), and that it breaks down in redshift space for the blue galaxies. With traditional density statistics, the agreement with the underlying dark matter is reversed between red and blue, with blue galaxies likely to be better described in redshift space by dark matter.

4.2 Galaxy versus initial statistics

One of the original proposed goals of the log transform and Gaussianization was to reconstruct the initial power spectrum from the final density field (Weinberg 1992; Neyrinck et al. 2009). We test this by comparing the usual and transformed density statistics to the initial power spectrum. We can see already in Fig. 8 that the real-space shapes of the Gaussianized spectra are much closer to the linear power spectrum than the usual density statistics. Fig. 10 shows the ratio of the galaxy and dark-matter power spectra with the dark-matter power spectrum measured from an early snapshot of the Millennium simulation in real space. We use the measured initial power spectrum because it has the same shape as the linear (input) power spectrum on small scales, but includes the random modes from the simulation on large scales. The line styles and colours are the same as in Fig. 8.

In real space, the Gaussianized statistics retain the shape of the initial power spectrum better than the usual density power spectrum. The shape of the Gaussianized dark-matter power spectrum agrees with the initial power spectrum to $k \sim 0.7\ h$ Mpc^{-1}, as compared to $k \sim 0.13\ h$ Mpc^{-1} for the usual density power spectrum. The shape of the Gaussianized galaxy power spectra agrees with the initial power spectrum to about $k \sim 0.4\ h$ Mpc^{-1}, whereas the usual density power spectra agree to $k \sim 0.1\ h$ Mpc^{-1} (red) and $k \sim 0.3\ h$ Mpc^{-1} (blue).

However, the agreement again breaks down in redshift space. Interestingly, the usual density power spectrum of dark-matter and blue galaxies agree with the initial power spectrum to smaller scales ($k \sim 0.5\ h$ Mpc^{-1}) than they do in real space. The Gaussianized power spectra only agree with the initial power spectrum to $k \sim 0.15\ h$ Mpc^{-1} (dark matter and red galaxies) and $k \sim 0.3\ h$ Mpc^{-1}

(blue galaxies). In this case, the Gaussianization transform seems not to be beneficial in recovering the small-scale shape of the initial power spectrum. However, collapsing the FoG restores the real-space agreement to a large extent.

Note that on large scales ($k < 0.1\ h\ \mathrm{Mpc}^{-1}$), the Gaussianized power spectra seem to deviate more from the initial power spectrum than the usual density power spectra. Gaussianization generally does not preserve fluctuations (arising from cosmic variance) in mode amplitudes away from an ensemble-mean power spectrum. For linear-scale modes of the matter field, the validity of linear theory assures that the pattern of fluctuations away from an ensemble-mean (e.g. from CAMB) is exactly preserved from the initial conditions. A transform applied to the matter field will generally change this pattern. Already, a galaxy biasing function will change the fluctuation pattern, but an additional Gaussianization step is likely to change it more, since it amplifies discrete-sampling changes to the field. Another effect that could cause changes in the pattern of large-scale fluctuations is the migration of power from small to large scales (Neyrinck & Yang 2013), in addition to the migration from large to small scales that is present in the usual power spectrum. However, we find that the deviation introduced by Gaussianization on large scales is below the level of cosmic variance.

This comparison is a test of how well the transform removes the effects of both non-linear bias and non-linear gravitational evolution from the power spectra of galaxies. Fig. 10 shows that the transform is effective in achieving both goals to $k \sim 0.4\ h\ \mathrm{Mpc}^{-1}$ in real space but not in redshift space. We attribute this to the FoG, which alter the small-scale behaviour of the galaxy and dark-matter fields in different ways. Gaussianization is almost as effective at removing non-linear bias and non-linearity after FoG have been collapsed as it is in real space.

4.3 Red versus blue statistics

Because we want to test the ability of the Gaussianization transform to remove non-linear bias, we also consider the ratio between the red and blue galaxy power spectra of both the usual and transformed fields. Fig. 11 shows this ratio in real (top), redshift (middle), and collapsed FoG (bottom) space. The dashed lines show the ratio for the usual density power spectra and the solid lines show the ratio for the Gaussianized density power spectra.

In real space, the transformed galaxy density statistics agree with each other to $k \sim 0.5\ h\ \mathrm{Mpc}^{-1}$, whereas the usual density power spectra deviate around $k \sim 0.1\ h\ \mathrm{Mpc}^{-1}$. This further supports what we observed in Fig. 9, that the statistics of Gaussianized fields of differently biased tracers agree to smaller scales than those of the usual density fields. Interestingly, it appears that the agreement between the shapes of the Gaussianized red and blue power spectra extends to smaller scales than the agreement between the galaxies and the Gaussianized dark matter (Fig. 9, top panel).

However, the agreement is lost in redshift space, presumably due to the FoG. In this case, the agreement of the Gaussianized statistics is only marginally better than the usual density statistics. By collapsing the FoG, we again see the agreement in the transformed galaxy spectra, to similar scales as in real space.

4.4 Dependence on sample selection and cell size

So far, we have only presented results from a single sample (Sample 2) using a cell size of 4 Mpc h^{-1}, which corresponds to an average of roughly one particle per cell in both the red and blue samples.

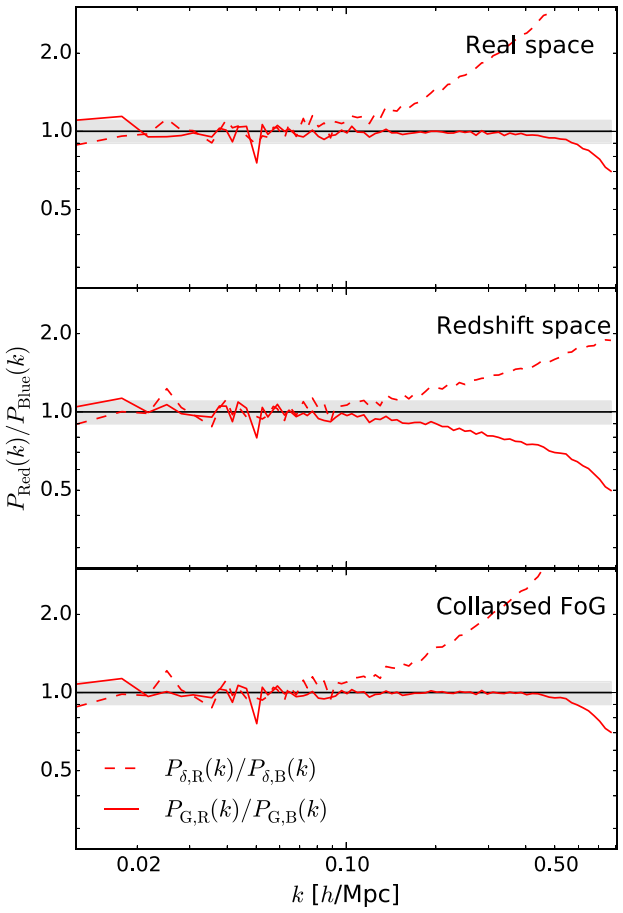

Figure 11. Ratio of red to blue galaxy power spectra in real space, redshift space, and collapsed FoG space. Solid lines show the ratio of Gaussianized statistics of red and blue galaxies, and dashed lines show the ratio of the usual density statistics. The agreement between Gaussianized power spectra for the two galaxy types, once FoG are accounted for, demonstrates the potential power of Gaussianization.

However, it is important to quantify the effects of varying sample selection and cell size on the results quoted above. First, we study the effect of varying the grid size in Sample 2 on the ratios considered in the previous section. Next, we look at these ratios for the four different samples, which have different clustering due to the stellar-mass cuts, as well as different number densities. Finally, we comment on how to choose the optimal grid size for a given sample. In this section, we only show power spectrum ratios in real space, but the variations with grid size and sample are similar in redshift space and with collapsed FoG.

The left panels of Fig. 12 show real-space ratios of the galaxy power spectra of the usual density and transformed density fields to the underlying dark matter (top), initial power spectrum (middle), and the ratio between red and blue power spectra (bottom). Gaussianized power spectra for red and blue galaxies are shown with the red and blue lines, and the usual density power spectra are shown with magenta and cyan lines, respectively. The different line styles in each colour show three different cell sizes used in the CIC density assignment. The solid line shows a cell size of 2 Mpc h^{-1}, which corresponds to a 256^3 CIC grid in our simulation box. The dashed line is the 4 Mpc h^{-1} grid size that was used in the previous section, corresponding to a 128^3 grid. The largest cell size shown here is

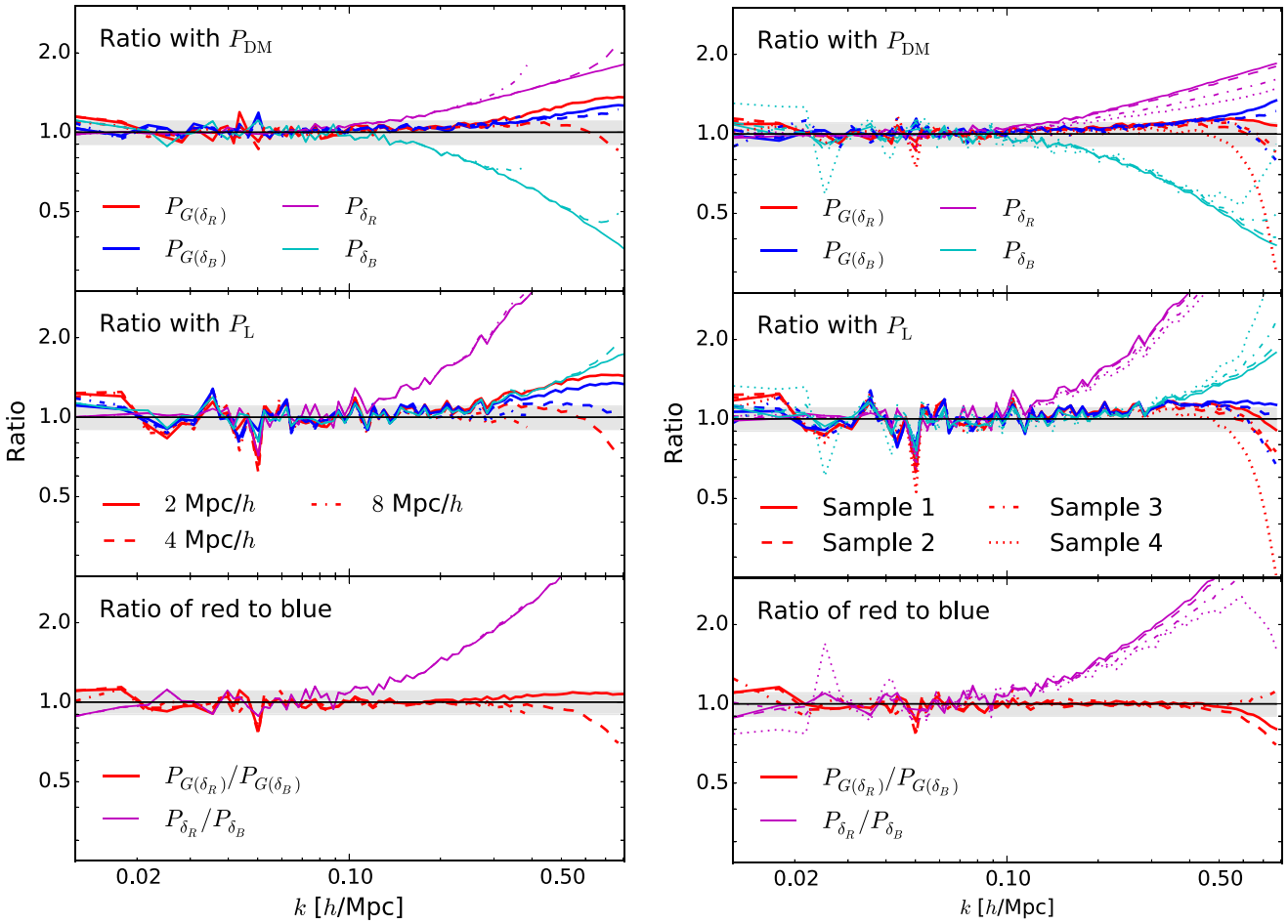

Figure 12. Variation of results with respect to grid size (left panels) and sample definitions (right panels). Ratios of the galaxy power spectra of the usual density and transformed density fields to the underlying dark matter (top), linear power spectrum (middle), and the ratio between red and blue power spectra (bottom). In both panels: red and magenta lines show the Gaussianized and usual density power spectra of red galaxies. Blue and cyan lines show those of blue galaxies. Left panels: the different line styles show different grid sizes. Right panels: the different line styles show the different samples, summarized in Table 1. The transformed density power spectrum for the blue galaxies is not shown for Sample 4 (see text).

8 Mpc h^{-1}, shown in the dot–dashed lines, which corresponds to a 64^3 grid.

In the usual density power spectra (magenta and cyan lines in Fig. 12), the cell size has almost no effect on these ratios, except near the Nyquist frequency of each grid, as expected. The shape of the Gaussianized power spectra, on the other hand, is in general dependent on the grid size chosen. For example, with very large cells, there are few empty cells, and the cells in the Gaussianized field populate the full Gaussian PDF. However, with very small cells, there are many empty cells, and after the transformation, the non-zero cells only populate a small fraction of the Gaussian PDF. Therefore, there is a tradeoff between the reducing number of empty cells, which make the PDF of the Gaussianized field somewhat non-Gaussian, and reaching smaller scales with smaller cells.

For the ratio with dark matter (Fig. 12, top left panel), the grid size does affect the small-scale shape of the ratio between the Gaussianized spectra, but does not significantly change the scale to which there is agreement between the Gaussianized galaxy power spectra and the underlying dark-matter spectrum. In all cases, the Gaussianized power spectra deviate from the dark-matter power spectrum around $k \sim 0.3$–$0.4\ h$ Mpc^{-1}.

For the ratio with the initial power spectrum (middle left panel), there is slightly more variation with cell size, with the 8 Mpc h^{-1} and 4 Mpc h^{-1} grids showing the best agreement with the initial power spectrum. For smaller cells, the Gaussianized power spectra begin to deviate at larger scales. This indicates that a larger grid size (more particles per cell) is preferable for this ratio.

For the ratio between the red and blue power spectra (bottom left panel), the cell size does not have a very large effect on the scale to which the power spectra agree. For the larger cells, the agreement extends to $k \sim 0.4$–$0.5\ h$ Mpc^{-1}, and for the larger cells it extends to the Nyquist frequency of the grid, $k = 0.4\ h$ Mpc^{-1}.

The right panels of Fig. 12 show how these ratios vary with sample selection. The right panels correspond to the same ratios as the left, with the same colour scheme, but the line styles now show the four samples given in Table 1. We use a 4 Mpc h^{-1} cell size for all power spectra in this figure. Note we have not included a measurement for the blue galaxies from Sample 4. As mentioned in Section 3.2, the shot noise correction scheme works well for a range of number densities and grid sizes. For Sample 4, the number density of the blue galaxies is [$\bar{n} = 0.001$ (Mpc $h^{-1})^{-3}$] and with a 128^3 grid, the shot noise correction procedure is not accurate on the scales of interest.

For the usual density statistics in the right-hand panels of Fig. 12 (cyan and magenta lines), the sample selection has an effect on the ratios shown. This is because the different stellar-mass cuts give samples with different clustering properties, including non-linear bias.

For the Gaussianized statistics, the sample selection has remarkably little effect on these ratios. The red galaxies in Sample 4 (red dotted line) show a deviation at larger scales than the other samples. We attribute this to the shot noise correction procedure, which is only accurate on scales below $k \sim 0.5\ h\ \mathrm{Mpc}^{-1}$ for the number density of this sample [$\bar{n} = 0.006\ (\mathrm{Mpc}\ h^{-1})^{-3}$]. Therefore, we do not expect agreement on scales smaller than $k \sim 0.5\ h\ \mathrm{Mpc}^{-1}$ for this sample.

To quantify the agreement between the galaxy power spectra and the underlying dark matter (or initial) power spectrum, we use a reduced χ^2 to test the goodness-of-fit of the dark-matter spectrum to that of the galaxies as a function of k-bin:

$$\chi^2_n \equiv \chi^2(k_n) = \frac{1}{\nu}\sum_{i=0}^{n}\frac{\left(\ln P_{\rm gal}(k_i) - \ln P_{\rm DM}(k_i)\right)^2}{\sigma^2}, \tag{6}$$

where ν is the number of degrees of freedom, and σ^2 is the variance of $\ln(P_{\rm G}(k_i))$. For bin n in the reduced χ^2, the number of degrees of freedom is $n + 1$. For simplicity, we assume that the underlying density field is Gaussian, implying that the variance in bin i is $2/N_i$, where N_i is the number of modes in the bin. While this assumption is certainly not true on small scales, we use this quantity only to estimate the scale of deviation of the power spectra, so it does not need to be extremely accurate.

With these assumptions, we arrive at the following expression for the reduced χ^2:

$$\chi^2_n = \frac{1}{2(n+1)}\sum_{i=0}^{n}\left(\ln\frac{P_{\rm gal}(k_i)}{P_{\rm DM}(k_i)}\right)^2 N_i. \tag{7}$$

We define the quantity $k_{\rm max}$ to be the highest k where the reduced $\chi^2_n < 1$. This corresponds to the scale at which the galaxy power spectrum $P_{\rm gal}(k)$ no longer provides a statistically adequate description of the underlying $P_{\rm DM}(k)$.

In order to test whether there is an optimal cell size for a given sample, we compute the $k_{\rm max}$ values for the ratios of the usual and transformed power spectra with the underlying dark-matter power spectrum in all four samples over a range of cell sizes. We check that $k_{\rm max}$ is never larger than the shot noise correction scale for a given grid size and sample. Fig. 13 shows $k_{\rm max}$ for the usual density (dashed lines) and Gaussianized density (solid lines) cases, as a function of the average number of particles per cell for red galaxies (left panels) and blue galaxies (right panels). The different colours of solid and dashed lines represent different cell sizes used for each of the four samples given in Table 1. The horizontal dotted lines mark the Nyquist frequencies of the corresponding grids. The three vertical panels in Fig. 13 show $k_{\rm max}$ values in real space (top), redshift space (middle), and collapsed FoG space (bottom).

These figures show that in the ratio of Gaussianized galaxy spectra to the Gaussianized dark-matter spectrum (solid lines), there is not a strong dependence on number of particles per cell in most cases, but that a higher number of particles per cell is preferred in some cases. Apart from red galaxies in redshift space, there is no need to use cells smaller than 8 Mpc h^{-1}, as $k_{\rm max}$ never exceeds the Nyquist frequency of that grid. In the case of the red galaxies in redshift space, $k_{\rm max}$ can be as high as $0.6\ h\ \mathrm{Mpc}^{-1}$ if a smaller grid size is used, corresponding to an average of one particle per cell.

We also computed the $k_{\rm max}$ values for the ratio with the initial power spectrum and for the ratio of red to blue power spectra in all cases. We found that for the ratio with the initial power spectrum, larger cells are optimal for both red and blue galaxies. In both cases, the agreement in real space extends to $k \sim 0.4\ h\ \mathrm{Mpc}^{-1}$ for the 64^3 grid. For the ratio between red and blue power spectra, the agreement can be extended to $k \sim 0.6\ h\ \mathrm{Mpc}^{-1}$ using a grid size corresponding to $\sim$1 particle per cell in both real space and with collapsed FoG.

Overall, we find that for Gaussianized galaxy power spectra, a minimum cell size of about 4 Mpc h^{-1} is sufficient, as the agreement in shape between the galaxy spectra and dark matter never exceeds $k \sim 0.6\ h\ \mathrm{Mpc}^{-1}$. Using smaller cells tends to make the agreement worse. While the optimal cell size has a complicated dependence on number density and clustering, and may also depend on which ratio one is interested in, we find that a cell size that roughly corresponds to an average of one or more particles per cell is a good rule of thumb.

5 COMPARISON TO CLIPPING

In this section we compare the Gaussianization transform to the clipping transform (Simpson et al. 2011), which has also been shown to reduce the effect of both non-linearity and non-linear bias in the galaxy power spectrum (Simpson et al. 2016). First, we describe the clipping procedure in general and how we apply it to the red and blue galaxy samples in Sample 1 in both real and redshift space. We then discuss how the method and results compare to Gaussianization.

The basic idea of the clipping transform is to impose an upper limit on the density in the field under consideration. For a given density threshold, δ_0, all cells with $\delta_i > \delta_0$ are set to the threshold value, δ_0. Formally, the clipped density field ($\delta_{\rm c}(\boldsymbol{x})$) is related to the usual density field ($\delta(\boldsymbol{x})$) through the following transformation:

$$\delta_{\rm c}(\boldsymbol{x}) = \begin{cases} \delta_0, & \delta(\boldsymbol{x}) > \delta_0 \\ \delta(\boldsymbol{x}), & \delta(\boldsymbol{x}) \leq \delta_0. \end{cases}$$

This suppresses the high-density peaks which contribute most to the non-linear behaviour of the density field. As shown in Simpson et al. (2011, 2016), the statistics of the clipped density field can be modelled accurately to higher wavenumbers ($k \sim 0.5\ h\ \mathrm{Mpc}^{-1}$) than those of the usual density field.

Simpson et al. (2016) applied clipping to the full galaxy sample from the GAMA survey to constrain $f\sigma_8$. The various threshold values used were chosen such that the large-scale amplitude of the galaxy power spectrum was reduced by 30 per cent to 60 per cent, corresponding to removal of roughly 10 per cent to 20 per cent of the objects in the field. This balances between reducing the effects of non-linearity and maintaining the signal-to-noise ratio.

Here, we are interested in testing clipping on galaxy fields separated by colour. As the red and blue galaxy fields we are considering have very different large-scale linear biases ($b_{\rm blue} = 0.80$ and $b_{\rm red} = 1.55$ in Sample 1, as computed with equation 2), it is unclear what the optimal clipping threshold is, and whether it should be the same for both galaxy fields. Thus, we test a number of different threshold values in the two fields. See Table 2 for details of the various thresholds used for Sample 1.

In order to model the resulting clipped galaxy power spectrum, we follow equations (27)–(29) in Simpson et al. (2016), which relate the clipped galaxy power spectrum to the clipped dark-matter power spectrum. Here, instead of using an analytic prediction for the non-linear dark-matter power spectrum in real or redshift space, we use the known underlying dark-matter field from the simulation. To

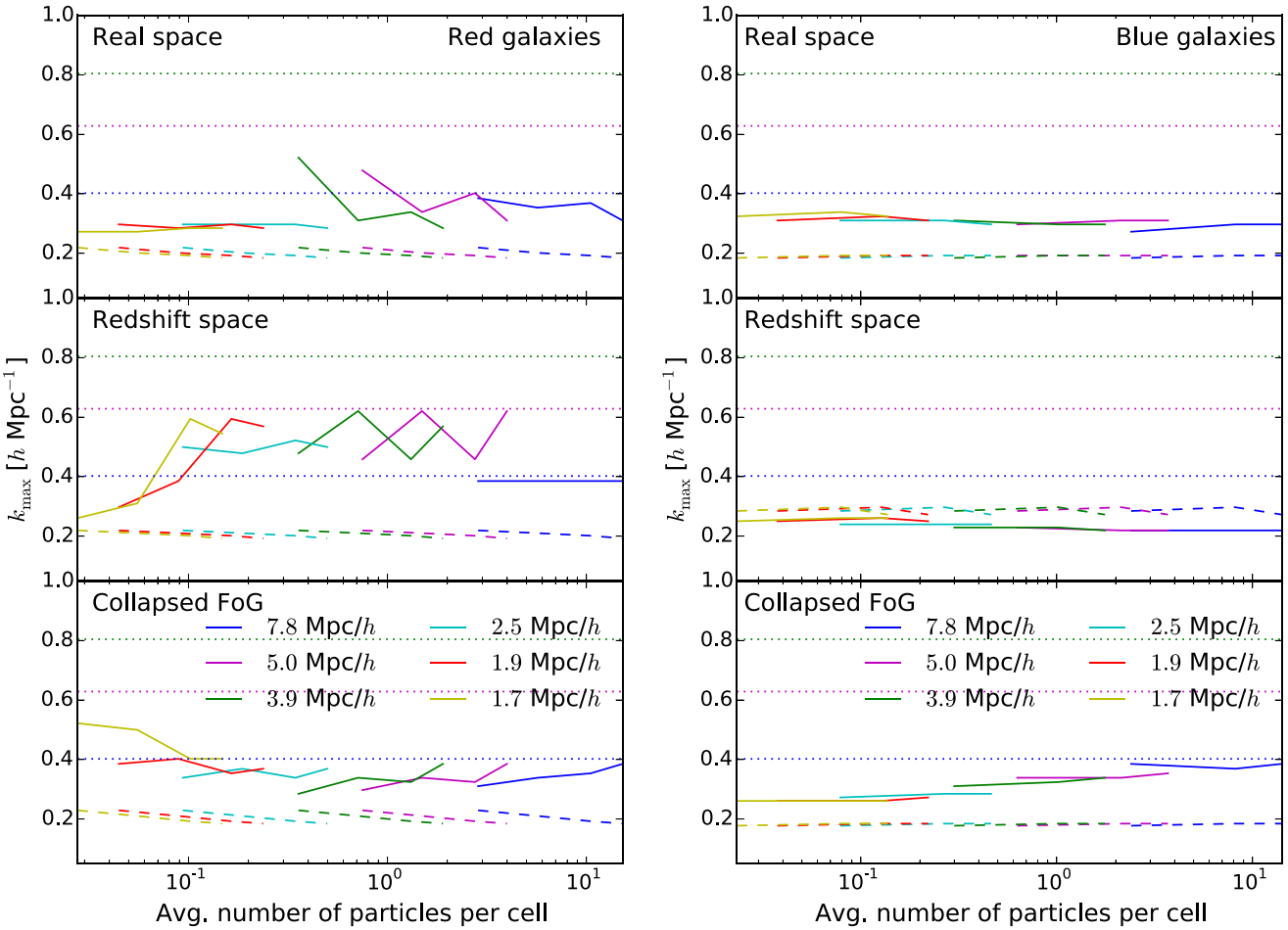

Figure 13. Maximum wavenumber, $k_{\rm max}$, for which the galaxy power spectrum $P_{\rm gal}(k)$ provides a statistically adequate description of the underlying $P_{\rm DM}(k)$, as qualified by equation (7). Left (right) panel shows $k_{\rm max}$ as a function of the average number of particles per cell for red (blue) galaxies and top to bottom shows real space, redshift space, and collapsed FoG space, respectively. Dashed lines show $k_{\rm max}$ for the usual density power spectrum and solid lines show $k_{\rm max}$ for the Gaussianized power spectrum. Nyquist frequencies for the largest three cell sizes are indicated as horizontal dotted lines.

Table 2. Summary of different clipping thresholds used for Sample 1 in real space. The first two columns give the large-scale amplitude of the clipped power spectra relative to the unclipped dark-matter power spectrum and unclipped blue power spectrum, respectively [see equation (2) for definition of the large-scale amplitude]. The middle four columns give the real-space density thresholds (and corresponding percentage of objects removed within parentheses) for the red galaxy field, the dark-matter field scaled by the bias of the red galaxies, the blue galaxy field, and the dark-matter field scaled by the bias of the blue galaxies. The final column gives notes about the chosen thresholds.

$B_{\rm clip}/B_{\rm DM}$	$B_{\rm clip}/B_{\rm blue}$	$\delta_0^{\rm red}$	$(b_{\rm red}\delta^{\rm DM})_0$	$\delta_0^{\rm blue}$	$(b_{\rm blue}\delta^{\rm DM})_0$	Notes
1.0	1.56	9.8 (17.8 per cent)	5.4 (19.6 per cent)	–	–	Red galaxies clipped to match the large-scale amplitude of DM power spectrum
0.51	0.8	3.95 (31.9 per cent)	2.05 (35.8 per cent)	3.85 (3.1 per cent)	10.5 (2.2 per cent)	Red and blue galaxies clipped so large-scale amplitudes are 80 per cent of unclipped blue power spectrum
0.26	0.4	1.7 (44.8 per cent)	0.7 (51.1 per cent)	1.32 (15.8 per cent)	1.6 (10.7 per cent)	Red and blue galaxies clipped so large-scale amplitudes are 40 per cent of unclipped blue power spectrum

model the power spectrum of a galaxy field clipped to a given large-scale amplitude, we scale the dark-matter field by the measured linear bias of the galaxy field and apply clipping at a level that results in a clipped biased dark-matter power spectrum with the same large-scale amplitude as the clipped galaxy power spectrum. In other words, our model for the clipped galaxy power spectrum is:

$$\mathrm{Cl}[P_{\rm gal}(k)] = \mathrm{Cl}[b_{\rm gal}\,P_{\rm DM}(k)], \quad (8)$$

where the strength of the clipping operation, represented by Cl[···], is defined by the reduction in large-scale amplitude of the resulting clipped power spectrum. In general, the density thresholds used (and percentage of mass removed) for the galaxy and linearly biased dark-matter fields may be different for each galaxy sample to achieve the same large-scale amplitude (see Table 2). Note that the change is not necessarily huge in terms of mass fractions: a few per cent, at most.

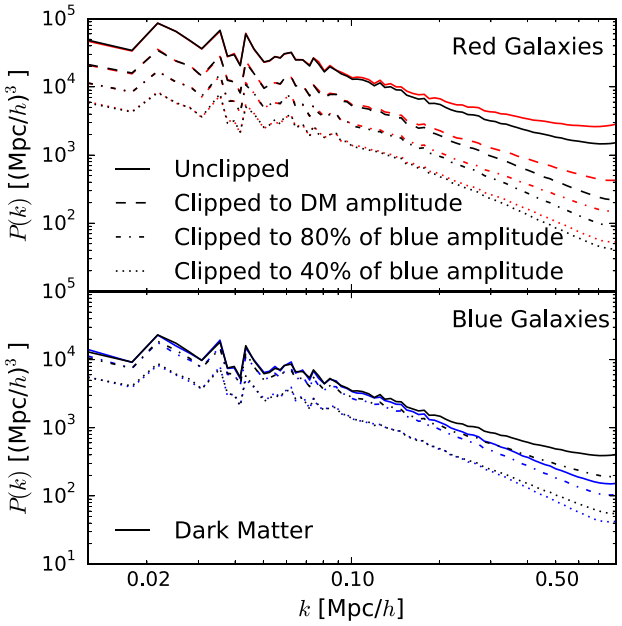

Figure 14. The effect of clipping on red and blue galaxy power spectra in real space. Line styles indicate clipping level (see key), and top and bottom panels show red and blue galaxies, respectively. Black lines in both panels show power spectra of dark matter, linearly biased by the appropriate $b_{\rm gal}$, and clipped to the same large-scale amplitude as the corresponding galaxy power spectrum. Stronger clipping results in agreement on smaller scales between galaxy and dark-matter power spectra.

We estimate the clipped power spectrum as described in Section 3.2 using an FFT of a 128^3 CIC grid, after subtracting the mean of the clipped overdensity field to ensure that $\langle\delta_{\rm c}\rangle = 0$. As described in Simpson et al. (2016), the shot noise contribution to the clipped power spectrum can be modelled for clipped Gaussian fields as $P_{\rm s} \simeq f_V^2/\bar{n}$, where f_V is the fraction of the volume of the field lying below the clipping threshold. For the thresholds used here, f_V is higher than about 85 per cent in all cases.

Fig. 14 shows the clipped and unclipped power spectra in real space. The top panel shows the unclipped (solid) and clipped (dashed, dot–dashed, and dotted) red galaxy power spectra (red lines) and corresponding clipped linearly biased dark-matter power spectra (black lines). The bottom panel shows the same for the blue galaxies. By construction, the large-scale amplitudes of the clipped galaxy and clipped (biased) dark-matter power spectra are the same. It is clear from this figure that clipping reduces the small-scale power of both the galaxy and dark-matter power spectra, bringing them closer to each other on small scales, as expected.

This can be seen more clearly in Fig. 15, which shows the ratios of the galaxy and dark-matter power spectra in the top panels, and the ratios of the galaxy and initial power spectra in the bottom panels. The left panels of Fig. 15 show these ratios in real space and the right panels show the ratios in redshift space (monopole). Red and blue lines correspond to red and blue galaxy samples in all panels, and the various clipping thresholds are shown in different line styles. From this figure, we see that increasing the clipping strength improves the agreement between the clipped galaxy fields and both the (clipped biased) non-linear dark-matter power spectrum and the linear power spectrum. However, even at the strongest clipping level in the red galaxies, which removes nearly 45 per cent of the galaxies, the agreement with both the dark-matter and linear power spectra only extends to $k \sim 0.2\ h\ {\rm Mpc}^{-1}$. The agreement is better in the case of blue galaxies at this clipping level, extending to $k \sim 0.3$–$0.4\ h\ {\rm Mpc}^{-1}$.

Note that when we repeat the clipping procedure in redshift space, we find our clipping thresholds relative to the redshift-space large-scale amplitudes. This means that for a given field, the density threshold used in real and redshift space may be different to achieve the same relative amplitude reduction. For example, to clip the red field to the dark-matter large-scale amplitude, we use a density threshold of $\delta_0^{\rm red} = 9.8$ in real space, and a threshold of $\delta_0^{\rm red} = 6.4$ in redshift space. As the right panels of Fig. 15 show, increasing the clipping strength improves the agreement on small scales between the galaxy power spectra and (clipped) dark matter, although the strongest clipping strength (dotted lines) does not seem to improve the agreement for blue galaxies. In the ratio with the initial power spectrum, the strongest clipping threshold overly suppresses both galaxy power spectra on small scales relative to the initial power spectrum.

From this analysis we can make several general comparisons with the Gaussianization transform, in both the applicability and effectiveness at removing non-linearity and bias from the galaxy power spectra. There is a degree of ambiguity when applying clipping to galaxy fields that is not present in Gaussianization. For example, it is not clear what the optimal clipping threshold is for a given field, whether that threshold depends on the clustering of the field, and whether it is the same in real and redshift space. It is also not obvious whether the red and blue galaxy fields, clipped to a given large-scale amplitude, should agree with each other on small scales, whereas an explicit goal of the Gaussianization transform is to separate bias from intrinsic clustering resulting in fields with the same statistics. However, the shot noise contribution is much better behaved in the case of clipping, where it can even be reduced compared with the usual density field. With Gaussianization, the shot noise contribution is increased in a complicated way that depends on grid size and number density.

Overall, it is clear that there are subtleties and complications in both methods that must be well understood in order to achieve any benefits from their application. Both transforms appear to work reasonably well at removing non-linearity and galaxy bias in real space. Clipping is more well behaved than Gaussianization in redshift space, provided the clipping strength is moderate.

6 CONCLUSION

The Gaussianization transform has previously been shown to be a promising method for restoring information to the two-point statistics of the matter density field on quasi-linear scales. This is due to two effects: reduced covariance on small scales and the shape of the small-scale power spectrum. In this paper, we have studied the effectiveness of the transform to remove non-linear galaxy bias and thus restore the small-scale shape of the underlying dark-matter power spectrum when applied to differently biased galaxy samples from the semi-analytic model GALFORM.

Overall, our results raise a few key points. The approximation that Gaussianization removes bias, which would imply that the Gaussianized statistics of differently biased fields agree, seems to hold in real space on quasi-linear scales ($k \sim 0.4$–$0.6\ h\ {\rm Mpc}^{-1}$). Gaussianization in real space also recovers the shape of the linear power spectrum on these scales ($k \sim 0.4\ h\ {\rm Mpc}^{-1}$).

Thus, the transform would seem to be an effective method for removing non-linear bias and non-linear gravitational evolution from

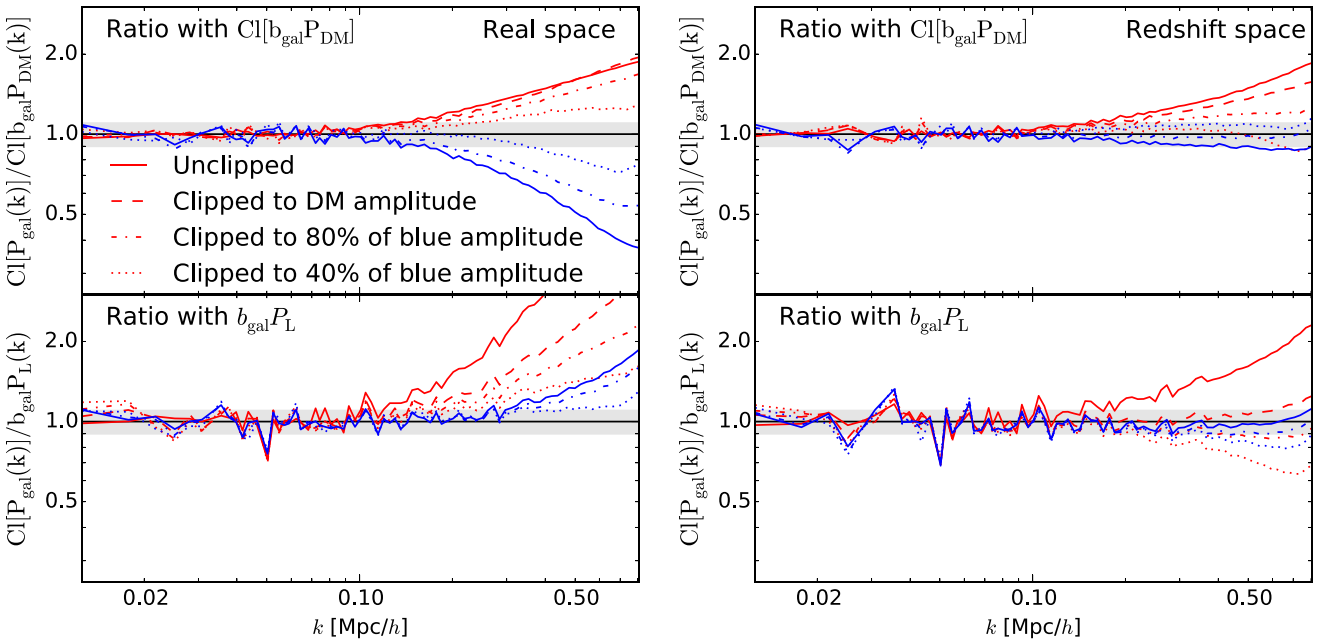

Figure 15. Ratios between unclipped (solid) and clipped (dashed, dot–dashed, and dotted) galaxy power spectra and corresponding linearly biased dark-matter power spectra (top) or initial power spectrum (bottom). Left panels show the ratios in real space and right panels show ratios in redshift space. Shaded grey regions show the 10 per cent range around 1. See equation (8) and text for discussion of the clipping procedure and modelling. Stronger clipping reduces the effects of both non-linearity and non-linear galaxy bias. In redshift space the effect of clipping is more complicated than in real space.

the two-point statistics of galaxies, and could simplify the modelling of modes up to $k \sim 0.4\,h\,\mathrm{Mpc}^{-1}$, likely improving constraints on cosmological parameters.

However, the galaxy-bias benefits observed in real space unfortunately do not extend to redshift space, for the most part. The Gaussianized statistics do not offer an improvement over the usual density statistics in estimating the linear-theory power spectrum in redshift space. While the shape of the Gaussianized blue galaxy power spectrum does not match that of the underlying dark matter in redshift space, we find that the shape of the Gaussianized red galaxy power spectrum in redshift space agrees very well ($k \sim 0.6\,h\,\mathrm{Mpc}^{-1}$) with that of the underlying dark matter. This suggests that Gaussianization of highly clustered galaxy samples, such as Luminous Red Galaxies (LRGs), may prove to be a useful method for recovering the non-linear dark-matter density field on quasi-linear scales. Further work is needed to understand why the Gaussianized statistics agree so well in this case. Another necessity is to investigate Gaussianization's effect on redshift-space clustering as a function of angle. It may be that Gaussianization is particularly adept at removing galaxy bias at angles where FoG do not dominate the signal, e.g. away from the line of sight.

We attribute the change in shape between the redshift-space Gaussianized matter and galaxy power spectra to the different behaviour of the galaxy samples on small scales. We test this by considering the redshift-space distribution with collapsed FoG. We find that in this case, the agreement is largely restored between the Gaussianized galaxy statistics and both Gaussianized dark-matter and initial statistics. This provides a possible avenue for recovering the small-scale shape of the underlying dark-matter and initial power spectra using Gaussianization.

We analyse the effect of sample selection and cell size on our results, and find that the agreement between the Gaussianized galaxy power spectra and underlying dark-matter (and initial) power spectra in general depends on cell size and number density, as well as other factors, in a complicated way. However, in all cases using a cell size that corresponds to 1 or more galaxies per cell on average gives the best results and ensures that the shot noise is manageable.

We also compare the Gaussianization transform to the clipping transform, which has been shown to reduce the effects of non-linearity and scale-dependent galaxy bias. We find that while clipping reduces the issue of shot noise and is less sensitive to resolution, there is ambiguity in the optimal clipping thresholds to use for each field. It is not clear how the optimal clipping threshold depends on the clustering of the field, and whether this threshold should be the same in real space and redshift space. In general, we favour Gaussianization as a method of overcoming galaxy bias for conceptual reasons: in the limit of no discreteness, and a monotonic biasing function, Gaussianization entirely eliminates galaxy bias. However, we have found that Gaussianization behaves subtly in the presence of discreteness, and with some stochasticity in the matter-to-galaxy-density biasing function. In practice, it seems that both clipping and Gaussianization can be effective at removing non-linearity and non-linear bias, at least in real space, if the subtleties of each procedure are carefully understood.

When applied appropriately, and if FoG compression is available, the Gaussianization transform may be able to provide better fidelity to the dark-matter and initial density power spectra, allowing for more modes to be included in analysis. Here we found that the agreement between the Gaussianized galaxy spectra and dark matter can extend to $k \sim 0.4$–$0.5\ h\ \mathrm{Mpc}^{-1}$, and similar agreement can be achieved between the Gaussianized galaxy spectra and the linear power spectrum at $z = 0$. We expect the agreement to extend to even smaller scales for $z > 0$, where non-linearities are smaller.

Further benefits are likely to be seen in the reduced covariance, which we have not studied here. Neyrinck et al. (2011) used galaxy samples somewhat different from those used here, particularly not

separated by colour, from the Millennium simulation to show that Gaussianization generally greatly reduces error bars in galaxy power spectra on small scales. The smaller error bars it provides can lead to smaller errors on any cosmological or galaxy-formation parameter to which galaxy power spectra may be sensitive, such as the power-spectrum tilt and neutrino masses.

In this work we have limited our analysis to relatively high number densities [$\bar{n} > 0.001$ (Mpc h^{-1})$^{-3}$] due to the limitations of our shot noise correction procedure. To use Gaussianization as a tool for recovering dark-matter clustering to higher wave numbers in future galaxy surveys such as DESI and Euclid, which may have lower number densities, a more sophisticated correction procedure will be necessary. Going forward, the full power and applicability of the Gaussianization transform must be tested via a suite of simulations with realistic galaxies, enabling us to quantify how well Gaussianization improves constraints on parameters due to the combination of the removal of non-local bias and the reduced covariance.

ACKNOWLEDGEMENTS

We thank Fergus Simpson and David Weinberg for useful discussions and comments. This work was supported by the Science and Technology Facilities Council (ST/L00075X/1). MN is grateful for financial support from a grant in Data-Intensive Science from the Gordon and Betty Moore and Alfred P. Sloan Foundations. PN acknowledges the support of the Royal Society through the award of a University Research Fellowship, and the European Research Council, through receipt of a Starting Grant (DEGAS-259586).

This work used the DiRAC Data Centric system at Durham University, operated by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). This equipment was funded by BIS National E-infrastructure capital grant ST/K00042X/1, STFC capital grant ST/H008519/1, and STFC DiRAC Operations grant ST/K003267/1 and Durham University. DiRAC is part of the National E-Infrastructure.

REFERENCES

Baugh C. M., Lacey C. G., Frenk C. S., Granato G. L., Silva L., Bressan A., Benson A. J., Cole S., 2005, MNRAS, 356, 1191
Berlind A. A. et al., 2006, ApJS, 167, 1
Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 645
Campbell D. J. R. et al., 2015, MNRAS, 452, 852
Carron J., 2011, ApJ, 738, 86
Carron J., Neyrinck M. C., 2012, ApJ, 750, 28
Carron J., Szapudi I., 2014, MNRAS, 439, L11
Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
Coles P., Jones B., 1991, MNRAS, 248, 1
Croft R. A. C., Weinberg D. H., Katz N., Hernquist L., 1998, ApJ, 495, 44
Croft R. A. C., Weinberg D. H., Pettini M., Hernquist L., Katz N., 1999, ApJ, 520, 1
Farrow D. J. et al., 2015, MNRAS, 454, 2120
Gonzalez-Perez V., Lacey C. G., Baugh C. M., Lagos C. D. P., Helly J., Campbell D. J. R., Mitchell P. D., 2014, MNRAS, 439, 264
Guo Q., White S., Angulo R. E., Henriques B., Lemson G., Boylan-Kolchin M., Thomas P., Short C., 2013, MNRAS, 428, 1351
Hockney R. W., Eastwood J. W., 1981, Computer Simulation Using Particles. Taylor & Francis Group, Madison Avenue, New York
Kim Y.-R., Choi Y.-Y., Kim S. S., Kim K.-S., Lee J.-E., Shin J., Kim M., 2014, ApJS, 212, 22
Komatsu E. et al., 2011, ApJS, 192, 18
Lacey C. G. et al., 2015, preprint (arXiv:1509.08473)
Lemson G., the Virgo Consortium, 2006, preprint (astro-ph/0608019)
Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473
Melott A. L., Weinberg D. H., Gott J. R., III, 1988, ApJ, 328, 50
Neyrinck M. C., 2011, ApJ, 742, 91
Neyrinck M. C., 2014, in Proc. IAU Symp. 306, Statistical Challenges in 21st Century Cosmology. Cambridge Univ. Press, Cambridge, p. 25
Neyrinck M. C., Yang L. F., 2013, MNRAS, 433, 1628
Neyrinck M. C., Szapudi I., Szalay A. S., 2009, ApJ, 698, L90
Neyrinck M. C., Szapudi I., Szalay A. S., 2011, ApJ, 731, 116
Neyrinck M. C., Aragón-Calvo M. A., Jeong D., Wang X., 2014, MNRAS, 441, 646
Norberg P. et al., 2002, MNRAS, 332, 827
Park C., Kim J., Gott J. R., III, 2005, ApJ, 633, 1
Scherrer R. J., Berlind A. A., Mao Q., McBride C. K., 2010, ApJ, 708, L9
Simpson F., James J. B., Heavens A. F., Heymans C., 2011, Phys. Rev. Lett., 107, 271301
Simpson F. et al., 2016, Phys. Rev. D, 93, 3525
Springel V. et al., 2005, Nature, 435, 629
Tegmark M. et al., 2004, ApJ, 606, 702
Weinberg D. H., 1992, MNRAS, 254, 315
Wolk M., Carron J., Szapudi I., 2015a, MNRAS, 454, 560
Wolk M., Szapudi I., Bel J., Carbone C., Carron J., 2015b, preprint (arXiv:1504.00069)
Zehavi I. et al., 2002, ApJ, 571, 172
Zehavi I. et al., 2011, ApJ, 736, 59

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