THE UNIVERSITY *of* EDINBURGH
School of Physics and Astronomy

Summer Student Report

Redshift Space Power Spectrum Analysis with Density Splits

Julian Wack

Lay Summary

The Universe is permeated by a field of dark matter which only reveals itself through the gravitational interaction with ordinary, baryonic matter. This field is largely responsible for the clustering of galaxies into larger structures and their evolution. The details in which such large scale structures are formed provide important tests for cosmological theories and our understanding of gravity. In this project I consider a simple model to describe the clustering of galaxies and assess its predictions when considering galaxies belonging to different density intervals. I find that the considered model does not provide the percent level accuracy desired for precision cosmology regardless of the considered scale and density interval. This assessment informs the choice of model which will be used in the analysis of DESI's BGS data regarding relativistic effects.

Acknowledgements

Supervisors:
Dr Florian Beutler
Dr Shadab Alam

Duration:
June 13 - August 5 2022

Contents

1 Introduction

Even though at the time of writing this report no dark matter has not been detected directly, the wealth of evidence for its existence allowed it to become a well established component of modern cosmology. A quantity commonly used to describe the dark matter field is the overdensity field, defined as the normalised difference between the dark matter density field at position $\boldsymbol{r}$ and its spatial average

$$\delta(\boldsymbol{r}) \equiv \frac{\rho(\boldsymbol{r}) - \langle\rho\rangle}{\langle\rho\rangle}.$$

Understanding how this field influences baryonic matter and causes the growth of large scale cosmological structures provides important tests for modified theories of gravity and competing cosmological theories. Since the mean of the overdensity field is by definition zero, the simplest, non-trivial statistic is its variance, also referred to as the two-point correlation function or two-point statistic

$$\xi(\boldsymbol{r_{12}}) \equiv \langle\delta(\boldsymbol{r_1})\delta(\boldsymbol{r_2})\rangle. \tag{1}$$

Assume galaxies are positioned uniformly through space and consider a pair of galaxies to be separated by $\boldsymbol{r_{12}}$ with occurs with some probability p. The two-point statistic gives the probability exceeding p in the case of arbitrarily positioned galaxies. The object that this report will focus on is the power spectrum, simply defined as the Fourier transform of the two-point statistic.

$$P(\boldsymbol{k}) \equiv \mathcal{FT}[\xi(\boldsymbol{r})]. \tag{2}$$

Since we cannot directly probe the dark matter density field, we must rely on tracers, such as galaxies, to infer properties of the field. However, on large scales galaxies appear point-like causing the underlying field to be probed at discrete points only, introducing a discrepancy between the field of galaxies and the dark matter field commonly known as the *galaxy bias*. The simplest mathematical approach to account for the problem is to introduce a proportionality constant between the galaxy and dark matter field

$$\delta(\boldsymbol{r}) = b_1 \delta_{gal}(\boldsymbol{r}).$$

Due to the proposed linear map between the two fields, b_1 is referred to as the linear bias parameter.

A second issue that arises when using galaxies as tracers is connected to the fact that their position is much easier measured in terms of redshifts rather than physical distances. Hubble's law: $cz = H_0 d$ provides a mapping between these two measures of radial distance. The component of a galaxy's peculiar velocity along the line of sight perturbs the cosmological redshift such that the galaxy appears at a distorted position in redshift space, leading to an incorrectly inferred position in real space. This phenomenon is commonly known as *redshift space distortion* (RSD) and quantified by the linear redshift distortion parameter β. On the other hand, RSD offer cosmological information which can be extracted through the measurement of β. Specifically, the non-negative growth rate of cosmological structure at redshift z is defined by
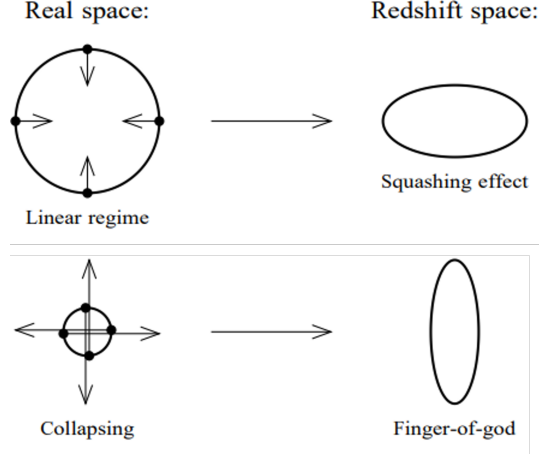
$$f(z) = b_1 \beta \tag{3}$$

Figure 1. Illustration of emergence of redshift space distortions on large and small scales. Adapted from [2].

and can equally be expressed as the logarithmic derivative of the growth factor with respect to the scale factor [1]. Inferring the value of $f(0.2)$ will be one of the key objectives of the main analysis. A metric well suited to discriminate between modified gravity models is the product of $f(z)$ and the amplitude of mass fluctuations in some region of space, conventionally chosen to be a sphere of radius $8\,\mathrm{Mpc/h}$ and thus defined by

$$\sigma_8^2(z) = \int_0^\infty \frac{k^3 P(k,z)}{2\pi^2} W_T^2(8k) \frac{dk}{k}, \quad W_T(x) = 3\frac{\sin x - x\cos x}{x^3} \tag{4}$$

where $W_T(8k)$ is the Fourier transform of a top-hat filter of width $8\,\mathrm{Mpc/h}$. The product of $f(z)$ and $\sigma_8(z)$ will be referred to as *fs8* hereafter.

Figure 1 illustrates the process of how RSD emerge. The observer is positioned at the bottom of the figure and looks upwards, such that the cosmological recession velocity is directed towards to the top of the figure. The arrows indicate the direction and scale of the peculiar velocities of galaxies which are symbolically represented by black dots. First, consider the top row, illustrating the process on large scales. The motion of the top galaxy towards to the observer reduces the cosmological redshift such that the galaxy appears less distant in the redshift space. The exact opposite effect occurs to the bottom galaxy, while peculiar velocities perpendicular to the line of sight have no effect of the measured radial redshift and thus do not introduce any distortions. The second row of the figure considers the behaviour on small scales, specifically regions close to the centre of dark matter halos where the peculiar velocities of galaxies are large. As in the previous case, no effect is observed from the perpendicular peculiar velocities. However due to the magnitude of the velocities, the observer concludes that the top galaxy is actually closer than the bottom one. The flipping of the apparent position in redshift space causes the cluster to appear elongated along the line of sight which is commonly referred to as the Finger of God (FoG) effect [2].

As the power spectrum is a measure of the galaxy clustering, we expect it to be damped by the FoG effect. A simple approach to account for this damping is based on the reasonable assumption that in a viralised cluster the velocities of galaxies are largely uncorrelated with position and typically of the order of a few hundred $\mathrm{km\,s^{-1}}$. A common choice for the distribution of velocities is a Gaussian, which I will adapt henceforth. By convolving

this with the distribution of galaxies, one obtains the corrected galaxy distribution [2, 3]. The convolution theorem then implies that the correction factor to the power spectrum in redshift space is simply by the Fourier transform of the velocity distribution, yielding

$$P^s(k, \mu) \xrightarrow{FoG} \exp\left(-\frac{1}{2}(\sigma k \mu)\right) P^s(k, \mu). \tag{5}$$

This immediately raises a further important implication of performing observations in redshift space: Statistics of the overdensity field have to be mapped from real to redshift space (denoted by the superscript in the Equation 5). Such mapping can be derived from the conservation of mass equation

$$[1 + \delta^s(\boldsymbol{s})]d^3\boldsymbol{s} = [1 + \delta(\boldsymbol{r})]d^3\boldsymbol{r}.$$

By the *Cosmological Principle* we expect homogeneous fields on large scales, implying $\delta \ll 1$ which motivates to rewrite the above relation as an expression of $\delta(\boldsymbol{r})$ in terms of powers of $\delta^s(\boldsymbol{s})$. Treating this to linear order and substituting into the definition of the power spectrum leads to the celebrated Kaiser model [3]

$$P^s(k, s) = (1 + \beta\mu^2)^2 b_1^2 P_{lin}(k). \tag{6}$$

Here μ is defined as the cosine between the wave vector $\boldsymbol{k}$ and the line of slight while P_{lin} denotes the real space matter power spectrum obtained when neglecting greater than linear powers of δ (hence the name linear theory). Combining the FoG prescription with the Kaiser model yields the model of the redshift space power spectrum I will be using throughout the remainder of the report. Namely,

$$P^s_{model}(k, \mu) = (1 + \beta\mu^2)^2 b_1^2 P_{lin}(k) \exp\left(-\frac{1}{2}(\sigma k \mu)^2\right) \tag{7}$$

to which I will refer to as the *KaiserFoG* model. In fact, rather than considering the whole power spectrum, I will focus on individual multipoles, defined as the projection of the power spectrum on the Legendre polynomials of order ℓ

$$P^s_\ell(k) = \int_{-1}^{1} \frac{2\ell + 1}{2} L_\ell(\mu) P^s(k, \mu) d\mu \tag{8}$$

As I will exclusively work with redshift space quantities, the superscript will be omitted from here on.

A key point of the project lies in considering the same volume of space through tracers with different linear bias. By separating the entire galaxy field by the density of galaxies (henceforth referred to as *density bins*), one obtains galaxy fields with the desired property of varying bias. Considering such density splits has many advantages, a few of which I want to highlight here. Firstly, McDonald (2009) argues that relativistic effects manifest themselves in the cross-correlation between two such tracer fields. Beutler and Dio (2020) have recently examined the potential to detect these effects in DESI's bright galaxy survey (BGS) [6, 7]. These relativistic effects offer new tests for gravity on large scales and their accurate modelling is important for detections on large scales, such as primordial non-Gaussianity which would allow to discriminate between Inflation and competing theories. This application forms the main motivation of this study. Secondly, to conduct precision

cosmology a large signal to noise ratio is required, which can be increased through the consideration of density splits. First note that the noise is composed of shot-noise and sample variance. While shot-noise is a measures for how well the dark matter density fields has been sampled by the point-like galaxies, sample variance originates from being able to only sample a finite volume of a single realisation of a random field which in this context is the density field (each Universe provides one realisation of the density field). Typically, sample variance is the dominating source of noise in large scale observations. Seljak (2009) argues that by correlating tracer fields with different biases, sample variance cancels out. The core logic behind this conclusion is that different tracers for the same volume of space sample the same realisation of the underlying density field, often encapsulated by the expression that the density field is not stochastic. It was further pointed out by Paillas, Cai, Padilla, and Sánchez (2021) that by partitioning the density field and cross-correlating the resulting galaxy fields with the whole galaxy field, one obtains more cosmological information than from the two-point correlation function which therefore offers an alternative to computing higher order statistics.

The aims of this project are to investigate down to what scales the KaiserFoG model offers an accurate description of the power spectrum monopole ($\ell = 0$) and quadrupole ($\ell = 2$) in redshift space. The novelty of this work lies in performing this analysis separately for tracers with different linear bias which are obtained by partitioning the analysed region of space into density bins. Whether or not the KaiserFoG proves to be accurate enough for precision cosmology on large scales will decide if a more sophisticated model must be employed when looking for relativistic effects in future data releases. A further, more personal, aim of the project is to provide me with an introduction into typical workflows in the field of observational cosmology.

2 Methods

My analysis is fully contained in the directory `fitmodel_densitysplit/` of the repository referenced in Wack (2022). The workflow as well as the files and folders performing specific tasks is listed below.

1. Compute and store the 2D redshift space power spectrum for each density bin: `getpower_densitybins/`
   A visualisation of the density partition is provided by `show_densityfield.ipynb`.

2. Estimate the covariance matrix of the monopole and quadrupole for each density bin: `bruteforce_covmat/`

3. Perform a fast estimation of the KaiserFoG model parameter values and store these: `pre-analysis_fitting.ipynb`

4. Fit the power spectrum model in each density bin using Markov Chain Monte Carlo, store the results, and make plots illustrating the quality of the fit: `do_fitting/`
   This fitting is performed for an incrementally enlarged range of k and the fitted model parameters are plotted in terms of k_{max}, the upper bound of the consider k range.

5. Compare the data and the predictions by the pure Kaiser and KaiserFoG models: `compare_data_fittedmodels.ipynb`

The computationally expensive steps 1, 2, and 4 are designed to be submitted as jobs to a computing cluster. To provide additional explanations, the functionality of step 4 has been implemented in the notebook `fitmodel_densitysplit/fitting.ipynb`. Finally, the file `helper_funcs.py` contains several helper functions to load stored data and perform frequent computations. These will be discussed in detail later in this section.

Other base directories in the referenced repository provide intermediate steps I took to arrive at the final analysis workflow. Particularly, they include general tutorials on using the toolkit `nbodykit` which is documented by Hand et al. (2018), computing power spectra of *eBOSS* simulation data, and exploring different methods to perform the parameter inference of the main analysis. The latter requires steps 1 and 2 of the main analysis to be performed previously.

2.1 The Data

The model fitting will be performed using BGS mock data. The simulations are part of the *AbacusSummit* suite and for this study have been selected to assume a ΛCDM cosmology with parameter values inferred by the *Planck Collaboration 2018* (Table 2, TT,TE,EE+lowE+lensing) [12, 13]. The N-body simulation runs on a set of dark matter particles, leading to the formation of dark matter halos. To obtain a catalogue of galaxies (also known as mock catalogues), these halos are then populated following the *Halo Occupation Distribution* technique which assigns each halo a mass-dependent probability $P(N|M)$ of containing N galaxies [14]. My supervisors provided me with such mock catalogues which are available on the Institute for Astronomy's computing cluster *cuillin*. Note that read permissions are required to access the data. For this project I use two kinds of mock catalogues, differing in the volume of space that they simulate. The larger of the two, with an edge length of $2\,\text{Gpc/h}$ and hereafter referred to as the analysis box, will be used to fit the model while a large number of smaller boxes, each of size $500\,\text{Mpc/h}$, is used to estimate the covariance matrix for the multipole measurements of the analysis box. This process is discussed in Section 2.3.

From Equations 1 and 2, one can derive an equivalent definition of the power spectrum in terms of the Fourier modes of the overdensity field $\hat{\delta}(\boldsymbol{k}) \equiv \mathcal{FT}[\delta(\boldsymbol{r})]$:

$$\langle \hat{\delta}(\boldsymbol{k_1}) \hat{\delta}(\boldsymbol{k_1}) \rangle = (2\pi)^3 \delta_D(\boldsymbol{k_1} - \boldsymbol{k_2}) P(\boldsymbol{k_1})$$

with δ_D denoting the Dirac delta [2]. From this, it becomes apparent that the computation of the power spectrum in some k bin requires that at least one Fourier mode of the overdensity field exists in that k bin. This fails to be the case for the first bin when choosing a, for redshift space analysis conventional, bin width of $dk = 0.01\,\text{h/Mpc}$. Hence the first k bin will be omitted from the analysis.

I am working the parallel line of slight approximation, meaning the line of slight to every galaxy in the catalogue is assumed to be the Cartesian z axis. This approximation is valid in the present case as the considered objects are sufficiently far away.

Bin index	Percentage range	Density range in units of $(\mathrm{Gpc/h})^{-3}$
0	$0 - 10\%$	$3.747 \times 10^{-4} - 1.417 \times 10^{-2}$
1	$10 - 20\%$	$1.418 \times 10^{-2} - 2.489 \times 10^{-2}$
2	$20 - 30\%$	$2.490 \times 10^{-2} - 3.876 \times 10^{-2}$
3	$30 - 40\%$	$3.877 \times 10^{-2} - 5.848 \times 10^{-2}$
4	$40 - 50\%$	$5.849 \times 10^{-2} - 8.951 \times 10^{-2}$
5	$50 - 60\%$	$8.952 \times 10^{-2} - 1.458 \times 10^{-1}$
6	$60 - 70\%$	$1.459 \times 10^{-1} - 2.754 \times 10^{-1}$
7	$70 - 80\%$	$2.755 \times 10^{-1} - 7.755 \times 10^{-1}$
8	$80 - 90\%$	$7.756 \times 10^{-1} - 8.390$
9	$90 - 100\%$	$8.391 - 9.229 \times 10^{1}$

Table 1. Percentile edges for the mock BGS box which will be used for the fitting.

2.2 Density Split

The galaxies of the mock catalogues are treated as point particles but in order to compute the density field and partition it subsequently, it is necessary to prescribe a volume to each galaxy. This was achieved through *Voronoi Tessellation*, which refers to partitioning a space containing N base points into N polygon cells. Each cell contains one base point and encloses the region of space which is closer to the associated base point than to any other [15]. Considering the galaxies of the catalogue as the base points, the density of a cell is given by the reciprocal of its volume. As more massive halos are populated by more galaxies, the cell sizes will be smaller, resulting in a high density regions around massive halos. Having computed the galaxy density field on this grid of cells, one can partition it into percentiles by finding the density threshold values that cause a certain percentage of cells to have a smaller density. A split into density bins of width 10% was applied to the analysis box, leading to the bin edges as detailed in Table 1.

Due to the size of the mock analysis catalogue (containing just over 2.95×10^8 galaxies), the computation of the density bin edges is very memory heavy and exceeds 10 GB. To avoid running out of memory, I request 100 GB of memory when submitting the script `getpower_densitybins.py` to cuillin. The purpose of this file is to perform the density split and to compute and store the $2D$ power spectrum for each density bin.

For the visualisation of these density splits, I keep records of the galaxies belonging to each percentile. Again due to the large number of galaxies in the mock catalogue, these records are rather larger and exceed 100 MB, avoiding them to be pushed to the GitHub repository. Instead of attempting to plot a $3D$ representation of the galaxy field, I generate figures showing a slice in the $x - y$ plane with an arbitrarily chosen line of sight depth of 10 Mpc/h along the z axis.

Recalling the definition of the two-point statistic, it becomes apparent that it is zero for a uniformly random field. This causes a trivial power spectrum, corresponding to a vanishing bias. On the contrary, this implies that large scale structures are predominately formed by members of high density bins. A clear view of halos seeding structure formation is given by Figure 2, showing all galaxies in the neighbourhood around the the most massive halo. This figure can be decomposed into contributions from all density bins leading to Figure 3. The point of structure emerging as the bias of the tracer field increases is becomes apparent in this figure.
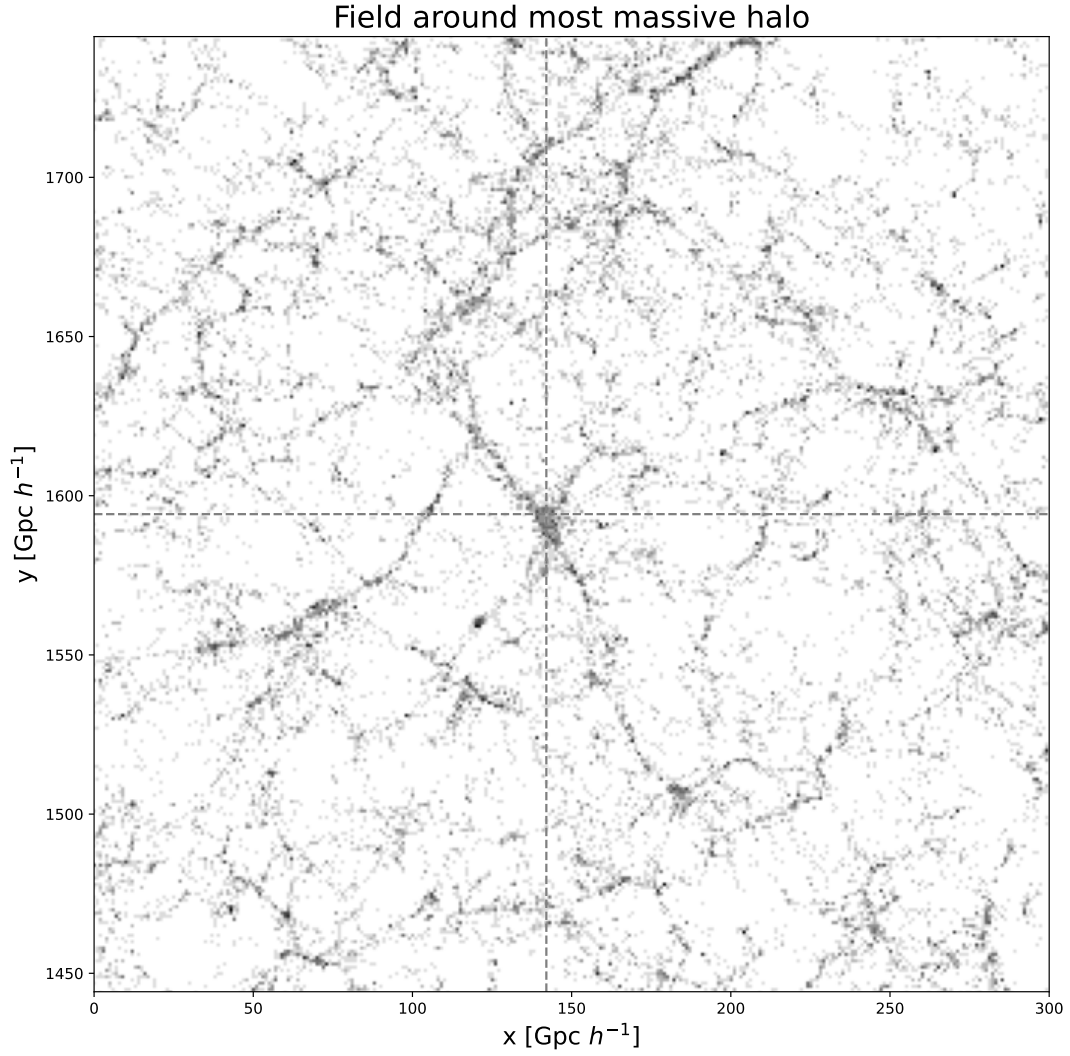
Figure 2. Full galaxy field close to the most massive halo of the simulation whose position is indicated by the crossing of the dashed lines. A large scale filament structure formed by galaxies is visible. The figure has a depth of $10\,\mathrm{Mpc/h}$ along the z axis.
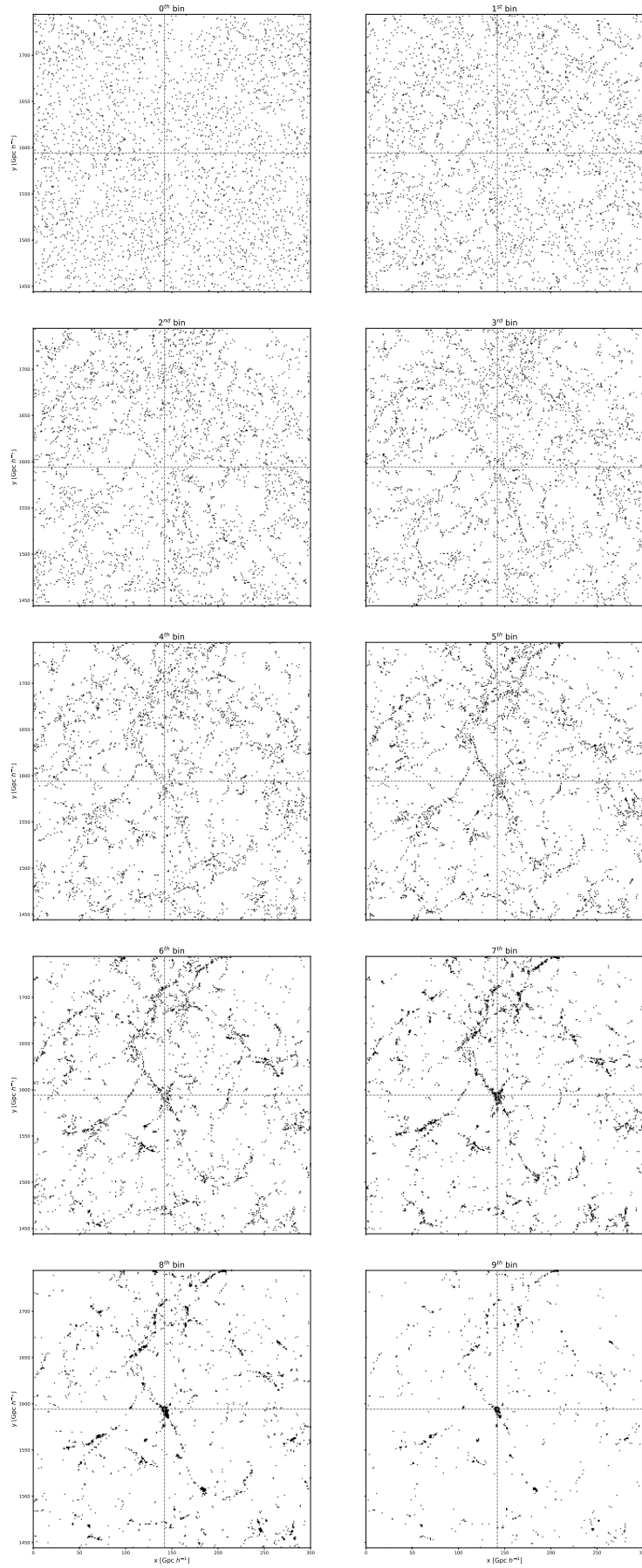
Figure 3. The same region of the galaxy field as in Figure 2 but split into contributions from different density bins with labels following Table 1. While low bias tracer fields appear random, high bias ones clearly show structure.

8

2.3 Covariance Matrix

The covariance between two random variables X, Y with respective expectation values $\langle X \rangle, \langle Y \rangle$ is defined as

$$Cov(X, Y) \equiv \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle = \langle XY \rangle - \langle X \rangle \langle Y \rangle = Cov(Y, X).$$

By considering an array of N random variables, one can compute the covariance between any two elements leading to the symmetric $N \times N$ covariance matrix.

In this study, each random variable will be a power spectrum multipole evaluated in some k bin and thus takes the form $P_\ell(k_i)$ for $\ell = 0, 2$ and $i \in [1, M]$ for a total of M considered k bins. The array of random variables is therefore

$$[P_0(k_1), P_0(k_2), \ldots, P_0(k_M), P_2(k_1), \ldots, P_2(k_M)]. \tag{9}$$

To reduce the noise in the expectation values and thus obtain a good estimate of the covariance matrix, many samples of the random variables are needed. I achieve this by computing the monopole and quadrupole for 1643 different realisation of the $500 \, \text{Mpc/h}$ cubic boxes. The full covariance matrix can be subdivided into 4 blocks:

$$C = \left[\begin{array}{c|c} C_{0,0} & C_{0,2} \\ \hline C_{2,0} & C_{2,2} \end{array} \right], \quad (C_{i,j})_{p,q} = Cov\left(P_i(k_p), P_j(k_q)\right). \tag{10}$$

While $C_{0,0}$ and $C_{2,2}$ represent the covariance between monopole and quadrupole bins amongst themselves, $C_{0,2}(= C_{2,0})$ gives the covariance between a monopole and a quadrupole bin. Loading the simulation data, splitting it into density bins, computing the power spectrum multipoles, and deducing the full covariance matrix for each density bin is all done by `bruteforce_covmat/get_covmats/power_covmat_densitybins.py`. I compute these covariance matrices up to $k \approx 1.4$ and in the later analysis slice them down to the desired range of k. As mentioned in Section 2.1, the first k bin will always be excluded. Note that each box used to compute the covariance matrices has only $\frac{1}{4^3}$ of the volume of the analysis box. A larger volume provides more data which in general reduces the noise in the measurements. Hence, the covariance matrices deduced from the $500 \, \text{Mpc/h}$ boxes must be divided by 4^3 to describe the errors for the $2 \, \text{Gpc/h}$ analysis box accurately.

Computing the covariance matrix by brute force as described here is computationally expensive, making estimating this matrix by other means an interesting problem. Grieb, Sánchez, Salazar-Albornoz, and Dalla Vecchia (2016) have proposed an analytic form for the covariance matrix based on the simplifying assumption that the overdensity field follows a multi-variate Gaussian distribution which causes the sub-matrices of Equation 10 to be diagonal. An explicit form of this analytic covariance matrix can be found in equations 15 and 16 of [16] and is implemented in `helper_funcs.py` to be optionally used as an alternative to the brute force covariance. The Gaussian assumption of the overdensity field is a good one at high redshift when little structure formation has occurred. As the considered mock galaxies sit at a redshift of $z = 0.2$, this does not apply. As mentioned above and illustrated in Figure 3, low bias tracer fields tend to be much more random that high bias ones. Hence, we expect the two methods of obtaining the covariance matrix to give similar results in the low density regime but deviate significantly for high density bins. This is illustrated in Figure 4 which, to allow an easier comparison, shows the correlation rather than the covariance matrix. The matrices are computed on the largest k range considered in the main analysis. Throughout my analysis, I employ the covariance matrix determined by brute force.
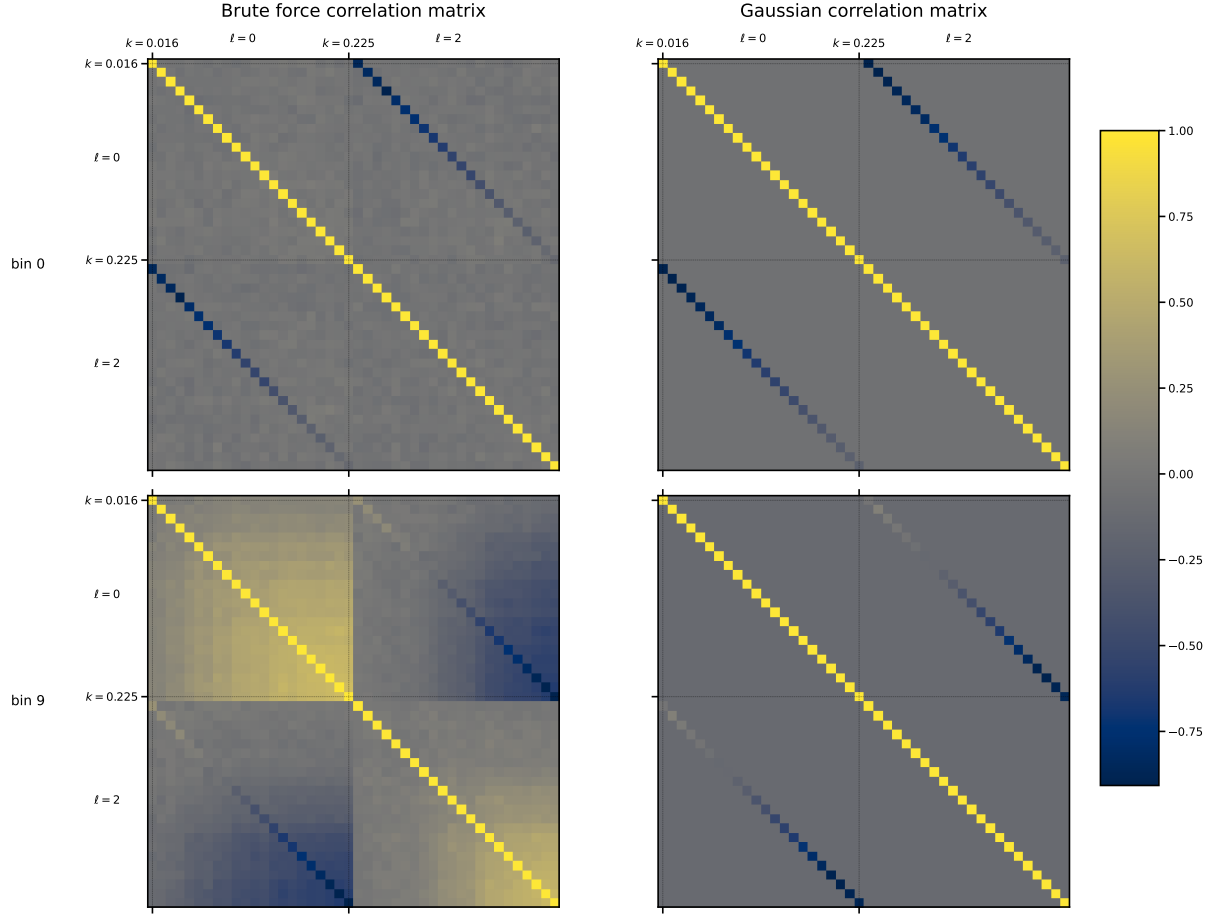
Figure 4. Correlation matrix of the lowest and highest density bins deduced by brute force and in a Gaussian approximation of the overdensity field following equations 15, 16, and 24 of Grieb, Sánchez, Salazar-Albornoz, and Dalla Vecchia (2016). In the low density case, the Gaussian approximation works well but fails in the high density bin. Sub-matrices, as explained in the main text, are indicated by the annotated values of ℓ.

2.4 Fitting the Model

The redshift space power spectrum model that I will be fitting is the KaiserFoG model containing 3 free parameters to be determined: $\boldsymbol{\theta} = [b_1, \beta, \sigma]$. As mentioned above, only the monopole and quadrupole will be fitted. Since the model is quite simple, these multipoles have a closed form given by

$$
P_0^{model}(k) = \frac{1}{2k^5\sigma^5} \left[\sqrt{2\pi}\mathrm{erf}\left(\frac{k\sigma}{\sqrt{2}}\right) \left(3\beta^2 + k^4\sigma^4 + 2\beta k^2\sigma^2\right) \right.
$$
$$
\left. + e^{-\frac{1}{2}k^2\sigma^2} \left(-2\beta(\beta+2)k^3\sigma^3 - 6\beta^2 k\sigma\right) \right] b_1^2 P_{lin}(k),
$$
$$
P_2^{model}(k) = -\frac{5}{4k^7\sigma^7} \left[\sqrt{2\pi}\mathrm{erf}\left(\frac{k\sigma}{\sqrt{2}}\right) \left(-45\beta^2 + k^6\sigma^6 + (2\beta-3)k^4\sigma^4 + 3(\beta-6)\beta k^2\sigma^2\right) \right.
$$
$$
\left. + e^{-\frac{1}{2}k^2\sigma^2} \left(2(2\beta(\beta+2)+3)k^5\sigma^5 + 12\beta(2\beta+3)k^3\sigma^3 + 90\beta^2 k\sigma\right) \right]
$$
$$
b_1^2 P_{lin}(k)
$$

where $\mathrm{erf}(z)$ refers to the error function. These results were obtained using *Mathematica* following Equation 8. One can also perform this integration numerical, as implemented in `helper_funcs.py`. However, as the analytic approach is more precise and faster, I will use it henceforth. For each density bin I fit these multipole models to the data, first only across two k bins up to $k_{max} \approx 0.025\,\mathrm{h/Mpc}$, then across 3 bins and so on until the maximum k range of the analysis with $k_{max} \approx 0.225\,\mathrm{h/Mpc}$ is reached. The inferred model parameters are the displayed as a function of the upper bound of the k ranges. For each considered k range the covariance matrix is sliced down accordingly by `fitmodel_densitysplit/helper_funcs.slice_covmat`.

2.4.1 Bayesian Inference

I consider a Bayesian approach towards the model fitting. This means that the model parameters are treated as random variables and thus follow some distribution. The aim of a Bayesian parameter inference is to update these parameter distributions through the information contained in the data. Bayes' theorem encapsulates the process of how data informs the parameter distribution, stating

$$
p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{M}) = \frac{p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{M})p(\boldsymbol{\theta}|\mathcal{M})}{p(\mathcal{D}|\mathcal{M})} \tag{11}
$$

for a data vector $\mathcal{D}$, a model prediction vector $\mathcal{M}$, and a model parameter vector $\boldsymbol{\theta}$. The quantity $p(\boldsymbol{\theta}|\mathcal{M})$ describes the parameter distributions subject to constraints imposed by the model but crucially before analysing the data and is thus referred to as the prior distribution. On the contrary, the posterior distribution $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{M})$ quantifies the knowledge about the model parameters including the information contained in the data. One of many benefits of Bayesian inference is that one may use the posterior distribution

of a previous analysis as the prior distribution for a new one, thus incrementally improving the understanding of the true parameter values. The likelihood-to-evidence ratio

$$\frac{p(\boldsymbol{\mathcal{D}}|\boldsymbol{\theta}, \boldsymbol{\mathcal{M}})}{p(\boldsymbol{\mathcal{D}}|\boldsymbol{\mathcal{M}})}$$

describes how the prior should be updated in light of the data presented. The likelihood $\mathcal{L}(\boldsymbol{\theta}) = p(\boldsymbol{\mathcal{D}}|\boldsymbol{\theta}, \boldsymbol{\mathcal{M}})$ gives the probability that the chosen model can reproduce the data as a function of the parameters $\boldsymbol{\theta}$. The parameters that produce a model prediction that agrees best with the data have the largest likelihood associated to them. The model evidence $p(\boldsymbol{\mathcal{D}}|\boldsymbol{\mathcal{M}})$ is the likelihood marginalised over the prior parameter distribution, meaning

$$p(\boldsymbol{\mathcal{D}}|\boldsymbol{\mathcal{M}}) \equiv \int p(\boldsymbol{\mathcal{D}}|\boldsymbol{\theta}, \boldsymbol{\mathcal{M}})p(\boldsymbol{\theta}|\boldsymbol{\mathcal{M}})d\boldsymbol{\theta}$$

which can be viewed as a kind of expectation value for the likelihood. The likelihood-to-evidence ratio in Equation 11 causes the prior to be enhanced to greater values in regions of the parameter space which produce a greater-than-average likelihood and diminishes it in parameter regions leading to a less-than-average likelihood. Through this process it is ensured that the posterior provides an updated understanding of the parameter distribution compared to the prior [17].

Having obtained the posterior, one can extract information about the model parameters in the form of expectation values

$$\mathbb{E}(f) = \int f(\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{\mathcal{D}}, \boldsymbol{\mathcal{M}})d\boldsymbol{\theta}. \tag{12}$$

The mean and standard deviation are of particular interest since they serve as the best fit parameter value and its error. In general, these integrals are high-dimensional and the posterior is some complex function making it impossible to evaluate the integral analytically and expensive to approximate numerically. Instead, one aims to collect samples $\boldsymbol{\theta}_i$ from the posterior distribution to compute the discrete analogue of Equation 12. The key observation is that the product of integrand and measure in Equation 12 is only non-zero is a specific region of the parameter space which is referred to as the *typical set* [17]. To approximate expectation values efficiently, most of the samples should be drawn from the typical set.

2.4.2   Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) methods form a class of algorithms which can efficiently sample from a high-dimensional target distribution. The core idea is that rather than drawing samples entirely at random, a dependency on the previous sample is introduced. This is done according to a Markov chain, assigning a probability $T(\boldsymbol{\theta}'|\boldsymbol{\theta})$ to each point in parameter space of becoming the next sample $\boldsymbol{\theta}'$ depending on the previous one $\boldsymbol{\theta}$. The collection of so generated samples are referred to as a chain which will eventually follow the target distribution if the transition probabilities satisfy a set of properties as detailed in [18, Chapter 1].

The initial element of the chain is some arbitrary position in parameter space. As the chain progresses, the samples start moving towards the typical set. Once the set is reached

and subsequently explored, the chain is said to have converged, meaning its elements now represents the target distribution accurately. The samples collected during the approach of the typical set are discarded for the analysis [17]. In my case I declare the first half of my chain to be part of this burn-in phase and thus reject it.

There exist a range of options for how the transition probabilities are determined. In my analysis I rely on the python package *Zeus* which makes use of ensemble slice sampling [19, 20]. This MCMC implementation is particularly fast and easy to use for model parameter inferences. I employ 8 walkers each producing an independent chain running for 2500 steps. After rejecting the burn-in, a total of $10,000$ samples for each parameter distribution remain. *Zeus* requires the user to define the logarithm of the likelihood and prior. I chose a standard Gaussian likelihood, leading to log-likelihood of $-\frac{1}{2}\chi^2$ where

$$\chi^2 = (\mathcal{D} - \mathcal{M})^T C^{-1} (\mathcal{D} - \mathcal{M}). \tag{13}$$

Here $\mathcal{D}$ refers to the data vector consisting of the power spectrum monopole and quadrupole, while $\mathcal{M}$ contains the respective model predictions. The inverse of the covariance matrix C is often referred to as the precision matrix. To construct the prior distribution of the model parameters, I performed a fast grid minimisation of χ^2 to obtain a set of initial guesses of the parameter values for each density bin and k range. This processes is implemented in `pre-analysis_fit.ipynb`. It is reasonable to assume the so obtained parameter values to be the mean values of the prior distributions. The probability distribution of a non-negative random variable that does not make any assumptions about the variable other than reproducing a known mean is the exponential distribution. In general, distributions which conform to a set of constraints but contain no other assumptions are called maximum entropy distributions and find extensive use in statistical physics. However due to run time improvements, for the presented analysis I employ uniform priors with ranges deduced by eye from this χ^2 minimisation. Specifically, I use the following priors

$$b_1 \in \text{Uniform}(0,3), \quad \beta \in \begin{cases} \text{Uniform}(-3,3) & 0^{th} \text{ density bin} \\ \text{Uniform}(0,3) & \text{otherwise,} \end{cases}$$

$$\sigma \in \begin{cases} \text{Uniform}(1,5) & k < 0.075 \, \text{h/Mpc} \\ \text{Uniform}(1,30) & \text{otherwise.} \end{cases}$$

The special behaviour of the 0th density bin regarding the β prior will be discussed below. As the damping term only deviates from 1 if k is large enough, I found that the data cannot inform the parameter value over the first few k ranges. Choosing a initially narrow prior produces a more readable plot without effecting the results of the fit. The Bernstein–von Mises theorem, stating that for a sufficient amount of data the posterior distribution converges to a Gaussian regardless of the prior, assures that these choices have no implications on the analysis results [21]. This theorem is well illustrated by Figure 6b, showing the posterior distribution of the arbitrary chosen 6th density bin for $k_{max} \approx 0.12 \, \text{h/Mpc}$.

3 Results and Discussion

The top 3 panels of Figure 5 show the mean value and standard deviation of the model parameters deduced from the posterior distributions and coloured according to the cor-

responding density bin following the notation of Table 1. Starting with the results for the linear bias and redshift space distortion parameter, it is expected that high density regions have a larger bias associated to them than low density ones. In fact, that the bias of low density tracers is close to zero agrees with the random nature of the tracer field shown in Figure 3. As the growth rate defined in Equation 3 is constant for all densities and values of k_{max}, the behaviour of β across different densities and k ranges is expected to be roughly the inverse of the described behaviour of b_1. It is important to note that the KaiserFoG model only depends on b_1^2 and is therefore insensitive to the sign of the bias. However, as the growth rate must be non-negative, one can fix the sign of b_1 depending on the sign of β. This was applied to the 0th bin, indicating the first of many aspects in which this low density bin behaves differently to the remaining ones. While the data is able to inform the parameter values well across most densities and k ranges, causing the uncertainty bands not to be visible to the naked eye, it fails to do so for β in the lowest density bin. The rather jagged trend of the parameter over the different k ranges with a significant standard deviation for both β and σ is caused by the posterior distribution being split into two peaks as shown in Figure 6a. This means that two regions of the parameter space are partially degenerate, suggesting that the Kaiser-FoG model is in inadequate for this density bin which will be discussed further later in this section. The 0th density bin also behaves like an outlier for the inferred value of σ by exploring the full range of the prior while all other density bins remain in the region $\sigma \in [2, 7]\,\mathrm{Mpc/h}$. Considering that theoretically value of σ should be about $\sqrt{2}$ times the galaxy velocity dispersion (see [2, Section 7.2]), which is usually a few hundred $\mathrm{km\,s^{-1}}$ and as $1\,\mathrm{Mpc/h} = 100\,\mathrm{km\,s^{-1}}$, the inferred value for the low density bins seems much too large. However, when forcing σ to lie within the theoretically expected range, the agreement between the data and the model prediction for the low density bin is worsened. As mentioned above, the data can only inform the value of σ if $k \gtrsim \frac{1}{\sigma}$. For k_{max} below this threshold, the posterior will be largely identical to the prior, explaining the clustering of σ for all density bins around $3\,\mathrm{Mpc/h}$ up to $k_{max} \approx 0.075\,\mathrm{h/Mpc}$ as well as the size of the standard deviation band.
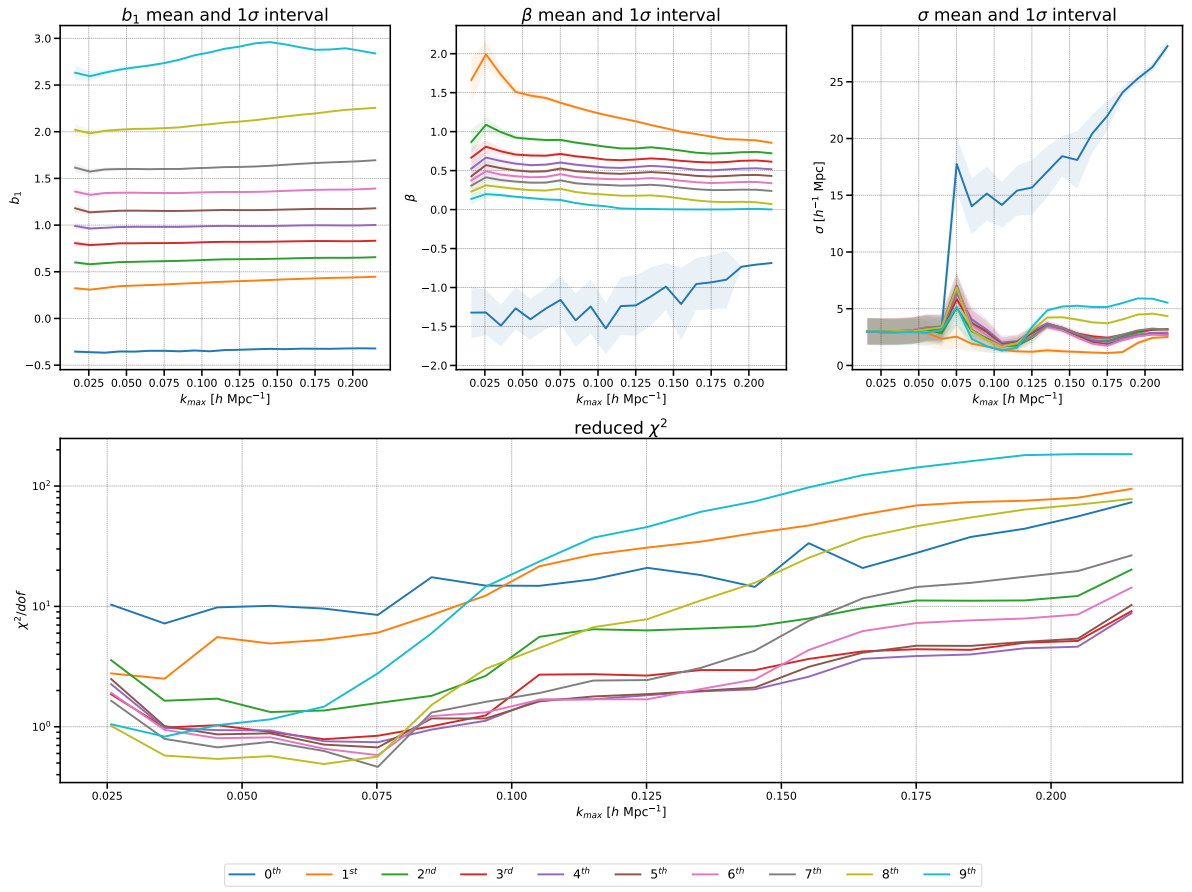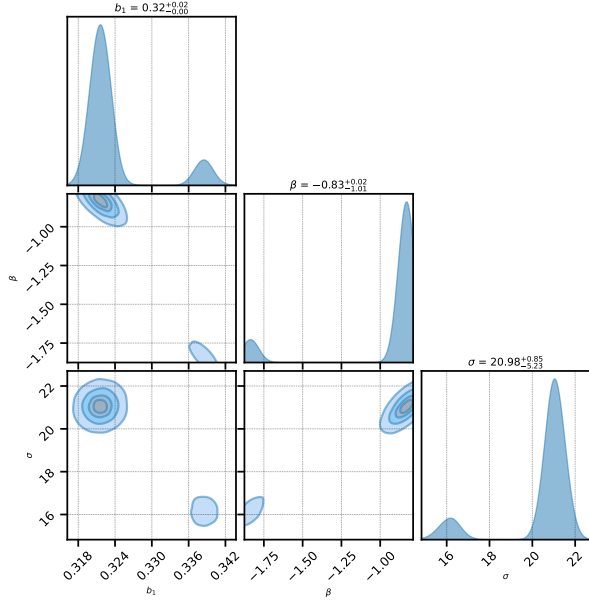
Figure 5. The top row shows the mean value and the standard deviation of the fitted model parameters for all density bins and for all consider k ranges. The bottom row shows the reduced χ^2 value corresponding to the mean parameter values.

(a) Unexpected behaviour of finding posteriors with two peaks, causing a large standard deviation in the inferred parameters. Since the separation of the peaks in b_1 is small, the error band is not be visible in Figure 5.
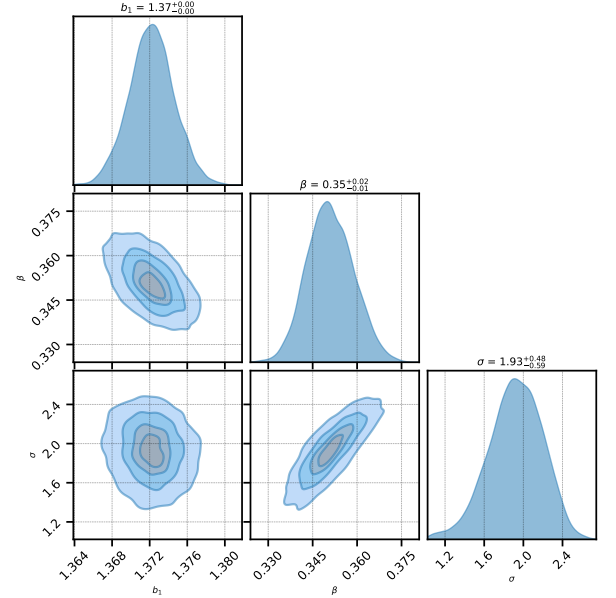
(b) Gaussian posteriors according to the Bernstein–von Mises theorem.

Figure 6. Corner plot of the posterior distributions for the 0th (**a**) and 6th (**b**) density bin when fitted up to $k_{max} \approx 0.165\,\mathrm{h/Mpc}$. Different contour colours in the $2D$ distributions indicate the $0.5, 1, 1.5,$ and 2 standard deviation levels.

To assess the quality of the fit, I use several metrics. The first and most fundamental is the value of χ^2 per degree of freedom, also known as reduced χ^2. The number of degrees of freedom is simply the number of random variables minus the number of parameters fitted. According to Equation 9, this leads to $DoF = 2\,(\#\ k\ \text{bins}) - 3$ where the number of k bins varies between 2 and 21. For a good quality fit, the value of the reduced χ^2 should be of order unity. The bottom panel of Figure 5 plots this metric for all density bins and all k ranges. Note that for all bins a trend of increasing reduced χ^2 with growing k_{max} is visible. This is expected as the fitted model is mainly based on linear theory and therefore only valid on large scales. For density bins 3 through 8, I deduce via a by-eye inspection that for $k_{max} \gtrsim 0.08\,\text{h/Mpc}$ the reduced χ^2 deviates significantly from 1, indicating the scale at which the KaiserFoG model breaks down. On the other hand, the large reduced χ^2 of extreme density bins (both low and high) signals poor agreement between the data and the model predictions. After the above discussion about the fitting of the 0th density bin, it should be noted that the large associated reduced χ^2 value expresses little confidence in found parameter values. The main finding of this figure is therefore that the KaiserFoG model can reproduce the first two even multipoles of the data only on large scales for medium density regions but fails on all scales for bins of extremely large or extremely low densities.

This observation is further illustrated in Figure 8, which provides a direct comparison between the data multipoles and the prediction by the KaiserFoG model when fitted up to $k_{max} \approx 0.225\,\text{h/Mpc}$. This and Figure 9 form my second metric to gauge the quality of the fit. The latter figure shows the residual between the data and the model, normalised to the error in the data. Hence, values of ± 1 form the threshold between the model being able to at least partially reproduce the data (the prediction lies within the error range of the data) and missing it completely. Note that a positive residual indicates the model under-predicts the data while a negative residual corresponds to over-prediction. To illustrate the effect of including the FoG term, I also ran the same fitting process for the pure Kaiser model, leading to the multipole predictions shown in blue. Just by assessing which model prediction yields a residual closer to zero, one is able to see that for most density bins in most k regions, including the FoG term does offer an improvement over the pure Kaiser model. Due to the large fitted value of σ in the 0th density bin, this is most clearly seen in the first row of Figures 8 and 9. Overall, these two figures support the finding from the discussion about the reduced χ^2 that the fit is best for medium density bins for small k_{max}. However, this can now be extended by noting that the monopole prediction is generally more accurate than the quadrupole prediction, suggesting to consider an improved model for the quadrupole before going to a more complex model of the redshift space power spectrum all together.

Further two interesting observations can be made from the plot of residuals for the medium density bins. Firstly, for these densities, both multipole residuals show dips at around $k \approx 0.075, 0.13, 0.2$ (monopole only) Mpc/h. Their origin must be investigated in future works. Secondly, the quadrupole residual (excluding the 0th bin) indicates some k^* for each density bin after which the models over-predict the data strongly. The inclusion of the FoG term reduces the intensity of this somewhat but is unable to increase the size of k^* significantly. Future work should investigate if introducing an additional damping term only for the quadrupole offers an improvement in that regard. This could be see as a first step towards a more complex quadrupole model as called for above. Towards the special behaviour of the 0th density bin noted in previous metrics, one can now
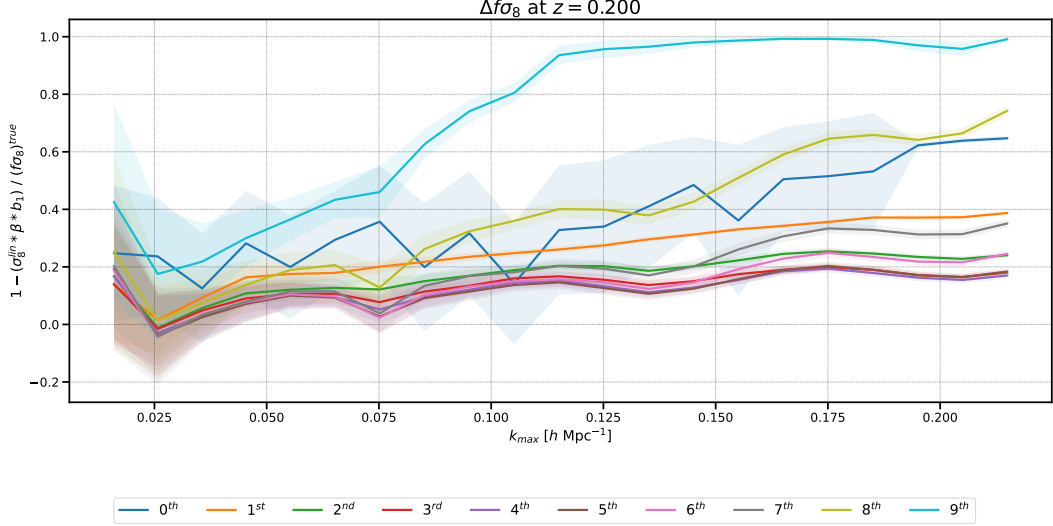
Figure 7. The relative difference of the product of growth rate and amplitude of matter fluctuations over all densities and k ranges. To be of practical use in modern cosmology, the error must be at the percent level which the KaiserFoG model fails to achieve even on large scales.

also add that the data quadrupole is entirely negative which is unexpected theoretically and must be investigated further.

The final metric I analysed is fs8. For all density bins and k ranges I plot the difference between the true and my fitted fs8, normalised to the true fs8. The value of $\sigma_8(0.2)$ I use for my inferred fs8 is obtained by inserting the linear power spectrum in Equation 4 and denoted by σ_8^{lin} in Figure 7. Note that the true value of $\sigma_8(0.2)$ and the one inferred from the linear power spectrum agree up to 2.5%, allowing to conclude that the main contribution to the reduced difference of fs8 comes from the fitted values for b_1 and β. Figure 7 gives another perspective on the statements derived from the preceding metrics. Specifically, it shows that the error of the KaiserFoG model regrading cosmological parameter estimation lies on the order of 10% and is thus not suitable for precision cosmology, requiring percent level accuracy. Note again that extreme density bins produce the poorest agreement. For the 9th bin the error saturates at close to 100% for $k_{max} \gtrsim 0.125\,\mathrm{h/Mpc}$ as from there onward the fitted value of β is close to zero and b_1 is not large enough to produce a growth rate significantly different from 0. The error bands are the result of propagating the errors of b_1 and β, explaining the size of the band for the 0th bin. Even though the fitted value of fs8 is close to the true one on very large scales, the wide error bands for these k ranges (caused by the smaller number of data points used in the fit) express that even for small k_{max} the reliability of the model prediction is limited. Further, the consistent under-prediction of fs8 across all densities and k ranges indicates the systematic short coming of the KaiserFoG model.
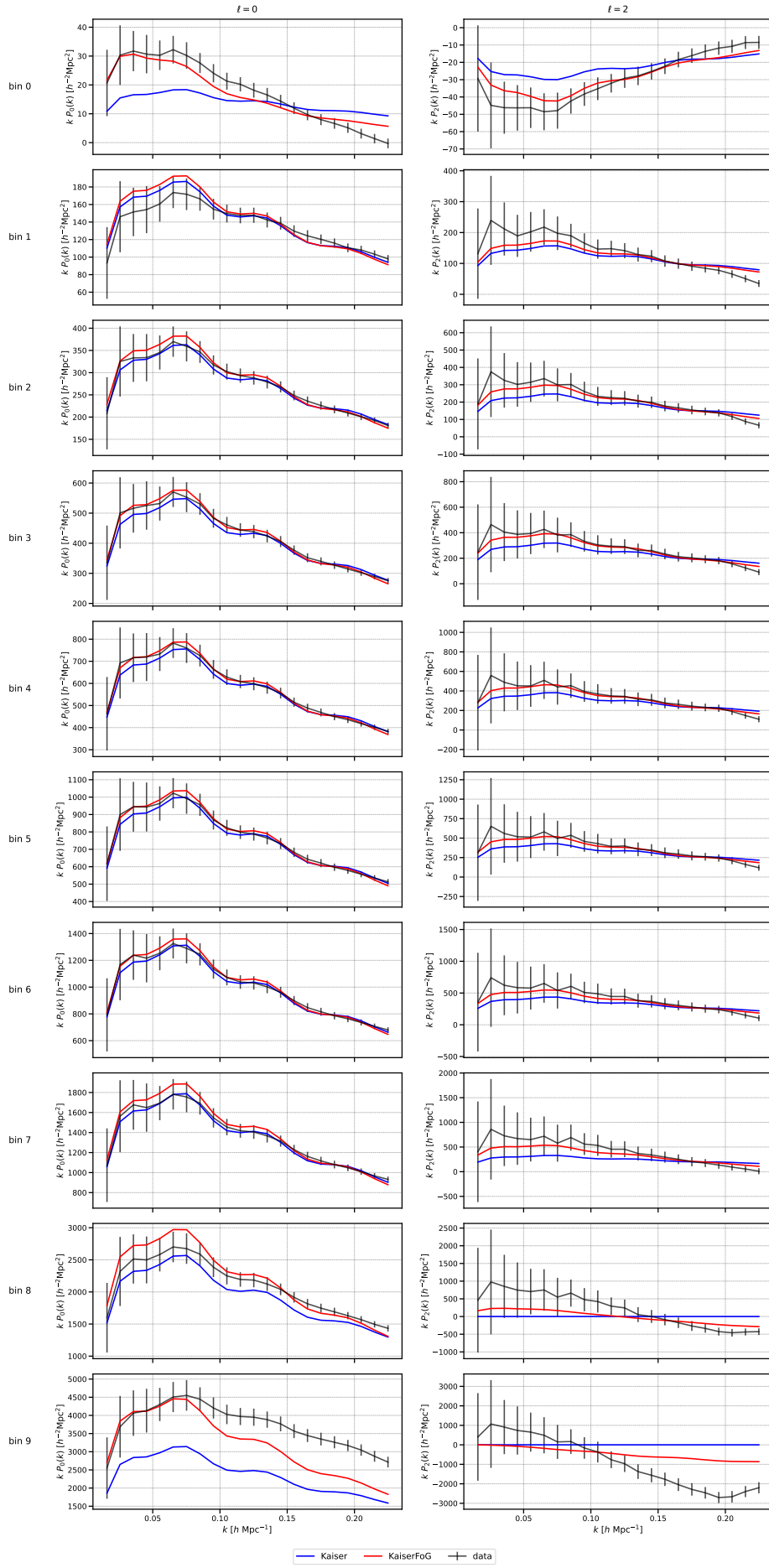
Figure 8. Plotting the data monopole and quadrupole and their errors deduced from the covariance matrix against the predictions by the pure Kaiser (blue) and KaiserFoG (red) model when fitted up to $k_{max} \approx 0.225 \, \mathrm{h/Mpc}$.
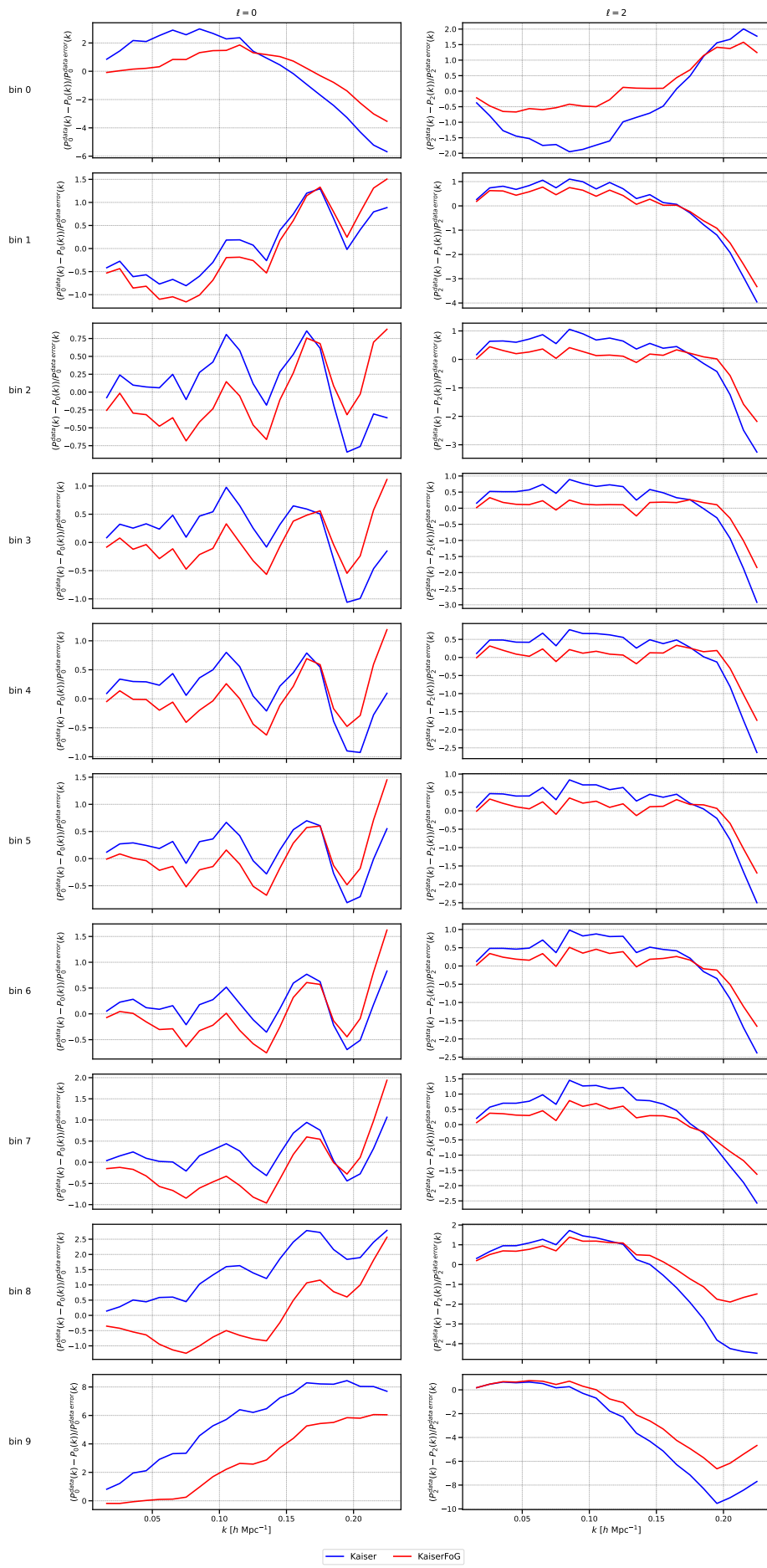
18

Figure 9. Plotting the normalised residual between the data monopole and quadrupole and the predictions by the pure Kaiser (blue) and KaiserFoG (red) model when fitted up to $k_{max} \approx 0.225\,\mathrm{h/Mpc}$. Values within $[-1, 1]$ indicate that the model predictions lie inside the error band of the data.

19

3.1 Conclusion

Due to its simplicity it is expected that the Kaiser model with a phenomenological Finger of God term fails to produce an accurate prediction of the redshift space power spectrum at fairly small k. I have shown that for medium density bins of BGS mock data, the breakdown occurs at $k \approx 0.08 \, \mathrm{h/Mpc}$ while the model is unable to reproduce the power spectrum of extremely high and extremely low density tracer fields across all scales. Moreover, the model prediction of fs8 fails to meet the desired percent level accuracy across all density splits and k ranges.

Throughout the presented analysis, the lowest density bin has proven itself to be an outlier offering several opportunities for future work including the investigation of the entirely negative quadrupole and negative linear redshift distortion parameter. Considering a finer split of the low density region might be a possible approach to these issues. Identifying the simplest redshift space power spectrum model that leads to accurate predictions on large scales will be a key next step following on from this study and hopefully of use when analysing real BGS data regarding relativistic effects. As the quadrupole predictions of the KaiserFoG model are particularly poor, and to keep the computational demand as low as possible, a first step towards identifying a suitable model would be to update the quadrupole prediction before turning to one of the many more sophisticated redshift space power spectrum models all together. I suggest an easy way of doing so by damping the KaiserFoG quadrupole prediction. The effectiveness of this remains to be investigated.

3.2 Personal Statement

This summer project allowed me to gain further insight into the workflow of academic research and thus reinforced my plans to pursue a PhD after completing my current degree. The project also offered me the opportunity to network with fellow summer students as well as academic staff, allowing me to benefit from their experiences. I believe that the open and independent nature of the project helped me to improve my organisation and time management skills. Through writing this project report, presenting my findings in a presentation, and by discussing encountered problems with my supervisors I was able to improve my communication skills. Designing an effective scientific poster, which I consider a key skill for my future career, further added to this. I learned a tremendous amount in the field of observational cosmology and was also able to build on my coding and problem-solving skills.

References

[1] Will J. Percival et al. "Redshift-space distortions". In: *Philosophical Transactions: Mathematical, Physical and Engineering Sciences* 369 (2011), pp. 5058–5067. URL: https://www.jstor.org/stable/pdf/23057251.pdf.

[2] A. J. S. Hamilton. "Linear Redshift Distortions: A Review". In: *Astrophysics and Space Science Library*. Springer Netherlands, 1998, pp. 185–275. DOI: 10.1007/978-94-011-4960-0_17. URL: https://arxiv.org/abs/astro-ph/9708102.

[3] Nick Kaiser. "Clustering in real space and in redshift space". In: *Monthly Notices of the Royal Astronomical Society* 227 (July 1987), pp. 1–21. DOI: 10.1093/mnras/227.1.1.

[4] Patrick McDonald. "Gravitational redshift and other redshift-space distortions of the imaginary part of the power spectrum". In: *Journal of Cosmology and Astroparticle Physics* 2009.11 (Nov. 2009), pp. 026–026. DOI: 10.1088/1475-7516/2009/11/026. URL: https://doi.org/10.1088%2F1475-7516%2F2009%2F11%2F026.

[5] Florian Beutler and Enea Di Dio. "Modeling relativistic contributions to the halo power spectrum dipole". In: *Journal of Cosmology and Astroparticle Physics* 2020.07 (July 2020), pp. 048–048. DOI: 10.1088/1475-7516/2020/07/048. URL: https://doi.org/10.1088%2F1475-7516%2F2020%2F07%2F048.

[6] DESI Collaboration et al. *The DESI Experiment Part I: Science,Targeting, and Survey Design.* 2016. DOI: 10.48550/ARXIV.1611.00036. URL: https://arxiv.org/abs/1611.00036.

[7] Omar Ruiz-Macias et al. "Preliminary Target Selection for the DESI Bright Galaxy Survey (BGS)". In: *Research Notes of the AAS* 4.10 (Oct. 2020), p. 187. DOI: 10.3847/2515-5172/abc25a. URL: https://doi.org/10.3847%2F2515-5172%2Fabc25a.

[8] Uroš Seljak. "Extracting Primordial Non-Gaussianity without Cosmic Variance". In: *Physical Review Letters* 102 (Jan. 2009). DOI: 10.1103/physrevlett.102.021302. URL: https://arxiv.org/pdf/0807.1770.pdf.

[9] Enrique Paillas, Yan-Chuan Cai, Nelson Padilla, and Ariel G Sánchez. "Redshift-space distortions with split densities". In: *Monthly Notices of the Royal Astronomical Society* 505.4 (June 2021), pp. 5731–5752. DOI: 10.1093/mnras/stab1654. URL: https://doi.org/10.1093%2Fmnras%2Fstab1654.

[10] Julian Wack. *JulianWack/IfA-SummerStudent-2022: Complete project.* Version v1.0.0. Aug. 2022. DOI: 10.5281/zenodo.7016884. URL: https://doi.org/10.5281/zenodo.7016884.

[11] Nick Hand et al. "nbodykit: An Open-source, Massively Parallel Toolkit for Large-scale Structure". In: *The Astronomical Journal* 156.4 (July 2018), p. 160. DOI: 10.3847/1538-3881/aadae0. URL: https://doi.org/10.3847%2F1538-3881%2Faadae0.

[12] Nina A Maksimova et al. "AbacusSummit: a massive set of high-accuracy, high-resolution N-body simulations". In: *Monthly Notices of the Royal Astronomical Society* 508.3 (Sept. 2021), pp. 4017–4037. DOI: 10.1093/mnras/stab2484. URL: https://doi.org/10.1093/mnras/stab2484.

[13] Planck Collaboration et al. "Planck 2018 results. VI. Cosmological parameters". In: *Astronomy & Astrophysics* 641 (Sept. 2020), A6. DOI: 10.1051/0004-6361/201833910. URL: https://doi.org/10.1051%2F0004-6361%2F201833910.

[14] Andreas A. Berlind et al. "The Halo Occupation Distribution and the Physics of Galaxy Formation". In: *The Astrophysical Journal* 593.1 (Aug. 2003), pp. 1–25. DOI: 10.1086/376517. URL: https://doi.org/10.1086%2F376517.

[15] Eric W Weisstein. "Voronoi diagram". In: *MathWorld–A Wolfram Web Resource.* (2000). URL: https://mathworld.wolfram.com/VoronoiDiagram.html (visited on 08/12/2022).

[16] Jan Niklas Grieb, Ariel G. Sánchez, Salvador Salazar-Albornoz, and Claudio Dalla Vecchia. "Gaussian covariance matrices for anisotropic galaxy clustering measurements". In: *Monthly Notices of the Royal Astronomical Society* 457 (Feb. 2016), pp. 1577–1592. DOI: 10.1093/mnras/stw065. URL: https://arxiv.org/pdf/1509.04293.pdf.

[17] Minas Karamanis. *"Bayesian Computation in Astronomy Novel methods for parallel and gradient–free inference"*. Unpublished doctoral dissertation, The University of Edinburgh. July 2022.

[18] Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. *Handbook of Markov Chain Monte Carlo.* Chapman & Hall/CRC handbooks of modern statistical methods. London, 2011. ISBN: 1420079417.

[19] Minas Karamanis and Florian Beutler. "Ensemble slice sampling: Parallel, black-box and gradient-free inference for correlated & multimodal distributions". In: *Statistics and Computing* 31.5 (Aug. 2021). DOI: 10.1007/s11222-021-10038-2. URL: https://doi.org/10.1007%2Fs11222-021-10038-2.

[20] Minas Karamanis, Florian Beutler, and John A Peacock. "zeus: A Python implementation of Ensemble Slice Sampling for efficient Bayesian parameter inference". In: *Monthly Notices of the Royal Astronomical Society* 508.3 (Oct. 2021), pp. 3589–3603. DOI: 10.1093/mnras/stab2867. URL: https://doi.org/10.1093%2Fmnras%2Fstab2867.

[21] A. W. van der Vaart. *Asymptotic Statistics.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998. DOI: 10.1017/CBO9780511802256.