Applying Neural Network Emulators to Void-Galaxy Clustering

by

Tristan Sohrab Fraser

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Physics

Waterloo, Ontario, Canada, 2026

© Tristan Sohrab Fraser 2026

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Kelly Douglass Professor, Dept. of Physics and Astronomy, University of Rochester
Supervisor:	Will Percival Professor, Dept. of Physics and Astronomy, University of Waterloo
Internal Member:	Michael Balogh Professor, Dept. of Physics and Astronomy, University of Waterloo
Internal Member:	Dustin Lang Professor, Dept. of Physics and Astronomy, University of Waterloo
Internal-External Member:	Eduardo Martin-Martinez Professor, Dept. of Applied Mathematics, University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

I am the sole author of Chapters 1, 5 and 6 which were written under the supervision of Prof. Will Percival and have not been published. For other chapters, the contributions from others are detailed below.

Chapter 2 The work described in this chapter, namely in Sections 2.3-2.5, encompasses work that was published in Fraser et al. 2024, alongside the results of Chapter 4. I was the primary investigator of this work, which was co-authored by Will Percival, Enrique Paillas, Seshadri Nadathur, Hans A. Winther and Sladana Radinovic. Sections 2.4.3 and 2.4.4 utilize results and analysis from a co-authored manuscript (Woodfinden et al. 2023). As a co-author of this work, I explored the limitations of the reconstruction code and produced void-catalogues as part of this. I collaborated with the principal investigator on the results and their interpretation.

Chapter 3 This chapter was adapted in part from the lead authored publication of Fraser et al. 2024. This chapter encompasses a review of the implementation of a neural network emulator that was discussed in the publication.

Chapter 4 This chapter was adapted from the first authored publication Fraser et al. 2024. As part of the work, I conducted all of the analysis shown, made all of the figures and wrote the majority of the text. My co-authors aided in the interpretation of the results and provided editorial feedback on the manuscript. The work done in this chapter was compared with the work done in a publication I co-authored, Radinovic et al. 2024. In that work, I was involved in modifying the voxel-void finder to correctly incorporate Alcock-Paczynski distortions, testing its resultant effect on the VG-CCF, which was compared to the methods explored by the authors.

Abstract

The discovery of the accelerating expansion of the Universe was a defining moment in the field of cosmology. The fallout of this has been three decades dedicated to the search for the cause of this accelerated expansion, known as dark energy. Cosmologists have developed many tests for dark energy, and have embarked on increasingly ambitious campaigns to map the sky, in order to precisely characterise the evolving Universe. While the standard model of cosmology has seen an enormous number of successes, in this present era, the added precision comes with a price, as our picture of the Universe has some tensions. The latest of these is that a long-supported model of dark energy, the cosmological constant, does not enjoy unanimous support from observations. The next generation of cosmological surveys are poised to tackle these tensions with greater precision, to potentially resolve them, or to underscore the need for a new standard model of cosmology.

Cosmic voids are vast, underdense regions of the Universe, that contain very few galaxies. They contain a wealth of cosmological information that has only recently been leveraged in large-scale structure cosmology. Cosmic voids have seen increasing adoption in the literature to obtain more information out of the same catalogue of galaxies than what is obtained with standard techniques. One key use is through modeling voids as a standard object, wherein a stack of voids is obtained from a survey, and the distortions of this stack can be used to test the Alcock-Paczynski (AP) effect. The AP effect arises from our choice of cosmological model, should that model be incongruent with the true cosmology of the Universe, the ratio between D_H, the Hubble distance parallel to the light of sight and D_M, the comoving distance transverse to the line of sight, would no longer be unity. Voids allow us to measure this effect to remarkable precision, so long as redshift-space distortions (RSD), which are distortions in distance due to the Doppler effect applied to a galaxy's peculiar velocity, are sufficiently modelled or removed from the sample. Extending this success to the next generation of galaxy redshift surveys, such as DESI, demands an increasing level of precision.

The demands for precision to match other observational probes of the Universe mean that in order to use voids, we need to understand the density and velocity of galaxies around them. One way of doing this is to utilise numerical N-body simulations, but they are expensive. N-body simulations cannot feasibly be run for every model that we would need to test. Instead, we created a model that utilises a neural network to build an emulator from a small set of N-body simulations. This model interpolated over different points in parameter space to model how a measurement changes with different cosmological parameters, an emulator. We trained a neural network emulator, that interpolated the measured void-galaxy clustering, specifically the two-point correlation function, across a

range of cosmological simulations with different cosmological parameters. This model allowed for the cosmological sensitivity of this void-galaxy cross-correlation function (VG-CCF) to be explored. This neural network emulator was applied to a series of N-body simulations, `AbacusSummit` constructed to match the requirements of the DESI survey.

We outlined how emulators are constructed, and which modeling decisions we made to construct one that is suitable for cosmological inference. Using the latest developments in the field, we constructed a neural network from our training sample of VG-CCF measurements, and train a model. This neural network model was first trained with `AbacusSummit` mock catalogues without the addition of any survey selection effects for the BOSS CMASS sample, and enclosed a single redshift window $0.45 < z < 0.60$. Various parameter recovery and consistency tests were run with this training sample, and later with the `Nseries` N-body simulations to ensure its robustness in recovering cosmological parameters from a measurement of the VG-CCF. Having confirmed the model's parameter recovery was unbiased, it was applied to the CMASS sample. This result was compared to its predecessor velocity template models, in particular the AP ratio D_M/D_H and the growth rate of large-scale structure $f\sigma_8$, and this comparison was also extended to measurements of $\Omega_{\rm m}$ and σ_8, where for CMASS it was measured to be $\Omega_{\rm m} = 0.330 \pm 0.020$, $\sigma_8 = 0.777^{+0.047}_{-0.062}$. This work demonstrated that a neural network emulator can be trained with the VG-CCF, and can extract cosmological information that is complementary to the BAO constraints obtained from the BOSS sample.

The emulator model was then adapted for application to a larger catalogue obtained from the first data release (DR1) of the Dark Energy Survey Instrument (DESI). Several adaptations needed to be implemented, among them was the need for a training sample that reflected the selection effects of the DESI survey. Varying completeness, number density and fibre assignment were all factors that limited the transferability of the BOSS iteration of the emulator to the DESI sample. A pipeline was constructed to impart the same selection effects on the training data, which in turn will be learned by the emulator. This step was resource intensive, but necessary, as failure to do so would exacerbate an existing bias in the emulator. The resulting training sample was a series of DR1 mock catalogues from the North Galactic Cap (NGC) of the `LRG-1` bin ($0.4 < z < 0.6$) of the DR1 sample. Incorporating this information allowed for the emulator to recover cosmological parameters, but these additional steps enlarged the errors. This resulted in σ_8 being prior restricted, and we obtained $\Omega_{\rm m} = 0.291^{+0.011}_{-0.027}$ for ΛCDM cosmologies. As a proof-of-concept, we explored how this methodology could be extended to evolving dark energy models; importance sampling was employed to allow for the constraining power of the emulator to be used in tandem with the BAO and BAO+SNIa datasets from DESI DR1. We demonstrated how constraints on $\Omega_{\rm m}, w_0, w_a$ could be obtained this method, as a proof-of-concept for future

work. This demonstrated the feasibility of the void emulator to be a possible source of complementary information, since most of the reduction in errors came from the VG-CCF emulator's sensitivity to $\Omega_{\rm m}$. Future work would need to be done to characterize how much more constraining power could arise if redshift evolution information about the VG-CCF were included in an emulator.

This thesis outlines the feasibility of void-galaxy emulation as a source of complementary constraining power to more established techniques of measuring the large-scale structure. We prescribe a series of potential improvements to this technique that range from expanding the training sample of the neural network to tackling the emergent bias in these neural network models. Even with a subset of the DESI DR1 sample, we illustrated the power of void-galaxy emulation to obtain standalone constraints on $\Omega_{\rm m}$, and joint constraints on w_0, w_a. Extending this work to the entirety of DR1's `LRG-1` bin makes obtaining standalone constraints on σ_8 a strong possibility, as this has been demonstrated with the BOSS sample. Moreover, extending the work to incorporate the DR2 sample, taking advantage of the higher number density and higher completeness of the sample would mitigate some of the confounding factors in this analysis. Future work will encompass an emulator that analyzes the VG-CCF from multiple redshift bins, yielding competitive constraints on w_0, w_a in addition to $\Omega_{\rm m}$ and σ_8, either as a standalone set of constraints, or through leveraging the complementary information of the VG-CCF and BAO.

Acknowledgements

I would like to thank my supervisor, Professor Will Percival, who taught me more things than I could list in this thesis. Will introduced me to the world of large-scale structure and all of the exciting physics it encompasses, and throughout this thesis he challenged me to be a better scientist, approach problems with rigour and to get things done. Will also introduced me to a network of researchers that have really been an invaluable support throughout this journey. I would also like to thank the postdocs I worked with, Professor Enrique Paillas, who showed an amount of patience and understanding for when the code broke more times than I can list, the model failed or when some other challenge emerged from the shadows. I would like to thank Dr. Simone Paradiso, whose knowledge of the cobaya likelihoods helped make the seemingly impossible possible. I would also like to thank the members of Will's cosmic void telecons - Professor Seshadri Nadathur, Professor Hans A. Winther, Dr. Alex Woodfinden, Dr. Elena Massara and Dr. Sladana Radinovic. They made taking the plunge into voids approachable, exciting and stimulating. My deepest thanks to my thesis committee members: Professor Dustin Lang, Professor Brian McNamara and Professor Michael Balogh, who helped keep me on pace for this challenging undertaking.

I would also like to acknowledge Professor Ashley Ross for patiently sharing their knowledge of the DESI LSS pipeline and answering even the most inane questions about every module within. I would also like to acknowledge Aurelio Carnero-Rosell and Cheng Zhao, their mastery of the DESI mock catalogue pipeline eased me into the challenges that awaited me with DR1.

I would like to extend a special thanks to Professor Rita Tojeiro, whose mentorship and guidance over my Master's gave me the confidence to pursue a PhD. This project would not be possible without the academic village that surrounded me.

It takes a village shape a person: Professor Yahya Nazer and Dr. Nassrin Nazer, who encouraged me and helped me appreciate the joys of scientific computing, my grandmothers, Mehri Mahboobi and Joyce Fraser, who showed me that being kind and resilient are the most important things in life. Professor Emeritus Yuri Sergeev, whose kindness and encouragement have been invaluable throughout this journey. This is to Mercer and Paul, and our late-night discussions over the anthropology of beige culture while playing Smash Bros and Mario Kart. To Jason and our Saturday strolls through Kensington Market, getting upsold on good cheese. To Adam, watching JoJo and commiserating over grad school together. To Frey and Matt, whose appearances from across the border brightened everyone's day. To Alyssa and Arin and Hemish, for reminding me how cool this all is. To all the feasts and merriment and warm summer days we have all shared together.

Dedication

This is dedicated to my amazing wife, Sofya, who kept me going and helped me see the wonder in everything, I am the luckiest person on Earth. This is dedicated to my parents, Peter and Ladan, for all of your love and support, and for encouraging me to never stop asking *why*, I am so lucky to know you.

Table of Contents

Examining Committee ii

Author's Declaration iii

Statement of Contributions iv

Abstract v

Acknowledgements viii

Dedication ix

List of Figures xv

List of Tables xxi

1 Introduction 1

1.1 The Homogeneous and Isotropic Universe 2

1.1.1 Isotropy and homogeneity 3

1.1.2 Obtaining equations of motion 4

1.1.3 Components of the Universe 5

1.1.4 Latest measurements of energy densities in the Standard Model of cosmology 7

1.2 A timeline of the Universe under ΛCDM 8
1.2.1 The Big Bang, inflation and the first minutes 8
1.2.2 Matter-radiation equality 9
1.2.3 Recombination and Decoupling 10
1.2.4 Growth of structure 10
1.2.5 Dark energy dominated era and the present 12
1.3 Cosmological Distances and Redshift 12
1.3.1 Defining distance 13
1.4 The anisotropic and inhomogeneous Universe 15
1.4.1 Adding perturbations 15
1.4.2 Peculiar Velocities and the Distance-to-Redshift relation 18
1.4.3 The Two-Point Correlation Function 20
1.4.4 Galaxy two-point correlation function and its multipoles 21
1.4.5 Distance-redshift relation and the distortion of geometry 24
1.5 Observational cosmology 24
1.5.1 Big Bang Nucleosynthesis (BBN) 24
1.5.2 Cosmic Microwave Background 25
1.5.3 Large-Scale Structure 28
1.5.4 Type Ia Supernovae (SNIa) 34
1.6 Statistical tensions 35
1.6.1 H_0 tension 35
1.6.2 S_8 tension 36
1.6.3 Dark Energy Equation of State 36
1.7 Obtaining more from large-scale structure 37
1.8 Thesis Outline 38

2 Voids and VG-CCF modelling **40**
2.1 Introduction . . . 40
2.2 Void-finders . . . 42
2.2.1 Spherical voids . . . 42
2.2.2 Watershed voids . . . 43
2.3 Measuring void-galaxy clustering . . . 48
2.3.1 A review of estimators of the two-point correlation function . . . 48
2.4 Cosmic voids in previous analyses using galaxy-redshift surveys . . . 51
2.4.1 Linear RSD and the linear bias approximation . . . 51
2.4.2 Template modelling . . . 54
2.4.3 Cosmological constraints from the VG-CCF . . . 55
2.4.4 Biases in the void-finder . . . 59
2.4.5 Template-free modelling of the void-galaxy cross-correlation function 61
2.4.6 Limitations of previous modelling . . . 63

3 Emulators **66**
3.1 Emulators . . . 66
3.1.1 Emulation of large-scale structure summary statistics . . . 66
3.1.2 Neural network emulators . . . 68
3.1.3 Modelling objectives in emulating the VG-CCF . . . 74
3.2 Ingredients for a neural network emulator of the VG-CCF . . . 74
3.2.1 Preparation of training data . . . 75
3.2.2 Covariance matrices . . . 78
3.2.3 Choice of model architecture . . . 79

4 Emulation of the VG-CCF in BOSS **82**
4.0.1 SDSS-III: BOSS CMASS . . . 82
4.0.2 Preparation of BOSS CMASS training data . . . 85

4.0.3 Void-galaxy cross-correlation . . . 88
4.0.4 Neural network model design . . . 89
4.0.5 Covariance estimation . . . 90
4.0.6 Inference with the emulator and MCMC . . . 95
4.0.7 Angular CMB scale prior . . . 95
4.1 Testing the pipeline using mock catalogues . . . 96
4.1.1 `AbacusSummit` based tests . . . 97
4.1.2 `Nseries` based tests . . . 97
4.2 CMASS data fits . . . 102
4.2.1 Baseline ΛCDM . . . 103
4.2.2 Extending to wCDM cosmologies . . . 108
4.2.3 Testing the parameters of the fits . . . 111
4.3 Discussion . . . 111
4.3.1 Impact of the θ_* prior . . . 113
4.3.2 The Alcock–Paczyński effect . . . 114
4.3.3 Comparison with other simulation-based clustering analyses . . . 116

5 Void Galaxy Cross-Correlation Function in DESI DR1 118
5.1 Boundary effects in DESI footprint on void-finder . . . 121
5.2 Preparation of training data . . . 125
5.2.1 Creating cubic box mocks for DESI DR1 LRGs . . . 125
5.2.2 Creating DR1 DESI mocks . . . 126
5.2.3 Adding fibre assignment information to cutsky mocks . . . 128
5.2.4 Downsampling . . . 129
5.2.5 Void-finding . . . 133
5.3 Data covariance . . . 133
5.4 Model . . . 134
5.4.1 Covariances . . . 135

5.5 Testing the pipeline with `AbacusSummit` DR1 mocks . . . 136
5.6 Fibre assigned cutsky mock tests . . . 141
5.7 DESI data fits . . . 145
5.7.1 Baseline ΛCDM . . . 145
5.7.2 wCDM results . . . 149
5.7.3 w_0w_aCDM results . . . 155
5.8 Discussion . . . 157
5.8.1 Effects of fibre-collision and DR1 mock window on emulator . . . 158
5.8.2 Emulator bias estimation . . . 159
5.8.3 w_0w_a constraining power . . . 159
5.8.4 Roadmap for future VG-CCF analyses . . . 160

6 Conclusions and future work **163**
6.0.1 Summary . . . 163
6.0.2 Roadmap for DESI DR2 . . . 167
6.0.3 Future work with VG-CCF emulators . . . 171

References **174**

Appendix **225**

A Figures from the Joint Likelihood Analysis **226**
A.1 Additional Joint Analysis for w_0CDM . . . 226
A.2 Additional Joint Analysis for w_0w_aCDM . . . 233

List of Figures

1.1 Slice of galaxies from the first year of the DESI survey 11

1.2 The matter-power spectrum as measured through galaxy clustering and the CMB 19

1.3 The power spectrum of the CMB temperature anisotropy, as measured by the Planck collaboration. 29

1.4 Figure of redshift space distortions and their effect on measured galaxy positions 31

1.5 Diagram of the cumulative effects of AP stretching and RSD, shown in both real- and redshift-space. 32

1.6 Figure summarizing different measurements of H_0 through direct and indirect measurements. 39

2.1 Figure spherical voids obtained from the Millenium XXL simulations. . . . 44

2.2 Figure of a large void obtained from the ZOBOV void-finder. 47

2.3 Void+2PCF cosmological constraints for BOSS eBOSS. 60

2.4 The real-space VG-CCF with the Alcock-Paczynski effect included. 62

2.5 An implementation of the template-free method, which attempted to recover f from a series of cosmological simulations. 64

3.1 A fully connected neural network with two hidden layers, with an input dimension of d, producing a single dimensional output $\hat{\boldsymbol{y}}$. Each layer, ℓ takes the previous layer's activation function of $f_{\ell-1}(x)$ as its input, for each neuron. For a detailed example, see Equation 3.1. 72

4.1 An aluminum fibre plug plate from SDSS-III BOSS. 83
4.2 Completeness maps of the CMASS and LOWZ samples from BOSS program. Each patch of colour corresponds to a *plate* from the survey. BOSS utilized well over 2000 unique plates [67]. 84
4.3 Correlation matrices associated to the different components that contribute to the total error budget in our likelihood analysis. 91
4.4 Emulator predictions of the monopole and quadrupole, for each of our six test cosmologies and averaged across 100 HODs, with their residuals. . . . 93
4.5 68th and 95th percentiles of the residuals for the VG-CCF measured from the 600 test samples. 94
4.6 Full posterior of the `c000` test cosmology. 98
4.7 Posteriors obtained from the emulator, marginalized over $\omega_{\rm b}$, $n_{\rm s}$ and the HOD parameters. 99
4.8 Marginalized posteriors of the emulator's predictions of the `Nseries` cosmology. 100
4.9 Predicted monopole and quadrupole from the `Nseries` cutsky mocks. . . . 101
4.10 Emulator's fit of the monopole and quadrupole of the void-galaxy cross-correlation function from the CMASS sample with residuals. 103
4.11 Posterior from fitting CMASS data using the emulator for a flat ΛCDM model, where $\{\omega_{\rm b}, \omega_{\rm cdm}, \sigma_8, n_{\rm s}\}$ were our cosmological free parameters and where we had marginalized over the HOD parameters. H_0 was derived from fixing θ_*. Consequently, $\Omega_{\rm m}$ can be computed from our free parameters $\omega_{\rm cdm}$ and $\omega_{\rm b}$. Fixing θ_* will also influence the baryon fraction, and subsequently $\omega_{\rm b}$. 105
4.12 Marginalized constraints on $\rm D_M/D_H$ and $f\sigma_8$ obtained from our BOSS emulator of the VG-CCF and template models of the VG-CCF. 106
4.13 Posterior from fitting CMASS data using the emulator for a flat wCDM model, where $\{\omega_{\rm b}, \omega_{\rm cdm}, \sigma_8, n_{\rm s}, w_0\}$ were our cosmological free parameters and where we have marginalized over the HOD parameters. H_0 is derived from fixing θ_*. Consequently, $\Omega_{\rm m}$ can computed from our free parameters $\omega_{\rm cdm}$ and $\omega_{\rm b}$. Fixing θ_* will also influence the baryon fraction, and subsequently $\omega_{\rm b}$. 109

4.14 Marginalized constraints on D_M/D_H and $f\sigma_8$ obtained from our BOSS emulator of the VG-CCF and template models of the VG-CCF, this time for wCDM cosmologies. . . . 110
4.15 Whisker plot of marginalized constraints on Ω_m and σ_8 from the BOSS VG-CCF emulator. . . . 112
4.16 A comparison of the constraints obtained on $f\sigma_8$ and D_M/D_H through two sets of training data generated with two different assumptions for the Alcock–Paczyński (AP) effect applied to voids. . . . 115

5.1 A view of the 5000 fibres mounted onto the Mayall 4.2m telescope (Left). Backlit image of the fibres mounted onto the instrument, grouped into *petals* (Right). . . . 119
5.2 The DESI DR1 footprint, split across the north and south galactic caps (left, right), with each blue dot representing a `DARK` tile, the higher the opacity, the higher the completeness of a given tile. . . . 120
5.3 Void size function, scaled by the fractional number of voids per size bin, for `AbacusSummit` cubic mock catalogues and `AbacusSummit` fibre assigned (`altMTL`, `FFA`) cutsky catalogues, and `AbacusSummit` cutsky complete catalogues. Each of the VSF computed here were averaged across the 25 realizations available. . . . 122
5.4 VG-CCF with weights computed from the VSF and applied to the voids when the correlation function is computed. . . . 124
5.5 Different stages of the DESI pipeline to convert cubic periodic box mocks into DR1 mocks. . . . 130
5.6 The distribution of galaxies in the `complete` and DR1 mocks from the downsampling stage of the pipeline. . . . 132
5.7 The two components of the DESI VG-CCF emulator's covariance matrix. . . 136
5.8 Full posterior obtained from the emulator for the `c000` cosmology and HOD parameters. While the cosmological parameters such as σ_8 and ω_{cdm} are well constrained, the same does not apply for the HOD parameters, continuing a trend seen in Chapter 4, where the VG-CCF is largely agnostic to galaxy-halo connection models. . . . 137
5.9 Building off of the results of Figure 5.8, we explore the constraining power of the VG-CCF exclusively for the HOD parameters in our model. We find the VG-CCF is agnostic to galaxy-halo connection models. . . . 138

5.10 The emulator's predictions of the concatenated monopole and quadrupole of the void-galaxy cross-correlation function from the `AbacusSummit` test cosmology `c000`, averaged across 100 HODs, with their residuals in the panel below. 139

5.11 Posteriors obtained from the emulator, marginalized over ω_b and the HOD parameters, using the priors described in Table 4.1, with the test cosmologies and their true values for σ_8, n_s and $\omega_{\rm cdm}$. As discussed in the previous chapter, θ_* is fixed in the `AbacusSummit` simulations, allowing for $\Omega_{\rm m}$ to be obtained from $\omega_{\rm cdm} = (\Omega_c h^2), \omega_{\rm b} = (\Omega_b h^2)$. 140

5.12 The marginalized posteriors of the emulator's predictions for the `altMTL` and `FFA` cutsky mocks prepared from the `AbacusSummit` cubic boxes using the `lss` pipeline. We averaged the 25 realizations and scaled the errors to the DESI NGC DR1 volume. We find that in both cases, the emulator can successfully predict $\Omega_{\rm m}, \sigma_8$ to within $1-\sigma$, across two sets of mocks that had two different fibre assignment treatments. The fiducial cosmology used to prepare these mocks is overplotted as dashed lines in this corner plot. . 142

5.13 The emulator's predictions of the monopole and quadrupole of the VG-CCF from the mean of the VG-CCF for 25 `AbacusSummit` FFA mocks, with their errors scaled to match a single DESI LRG-1 volume. The residuals are shown in the lower panel. 143

5.14 The emulator's predictions of the monopole and quadrupole of the VG-CCF from the mean of the VG-CCF for 25 `AbacusSummit altMTL` mocks, with their errors scaled to match a single DESI LRG-1 volume. The residuals are shown in the lower panel. 144

5.15 ΛCDM best-fit for the VG-CCF obtained from the emulator when the DESI DR1 sample is analyzed with it. The monopole and quadrupole predictions lie within 2.5σ of each other, save for the first 6 bins of the monopole, where the small-scale bias of the emulator is apparent. 146

5.16 The marginalized posterior distribution of ΛCDM cosmological parameters inferred from fitting the emulator to the DESI DR1 data. We left ω_{cdm}, ω_b, n_s,σ_8 as free parameters and kept all other cosmological and HOD parameters fixed before marginalizing to obtain σ_8 and $\Omega_{\rm m}$ parameter constraints. We overplot the survey's fiducial cosmological parameter values, `c000`. . . . 148

5.17 Marginalized posterior distribution of wCDM cosmological parameter inferred from fitting the emulator to DESI DR1 data. 150

5.18 Joint likelihood analysis utilizing the DESI DR1 BAO dataset (blue) and the void emulator (red) to obtain joint constraints on Ω_m and w_0 (grey). . 152

5.19 Joint likelihood analysis utilizing the DESI DR1 BAO+SNIa (Pantheon Plus) dataset (blue) and the void emulator (red) to obtain joint constraints on Ω_m and w_0 (grey). 153

5.20 Cosmological parameter constraints from our joint likelihood analysis. Importance sampling was used to construct a joint likelihood for the BAO and void emulator chains. The combined information from the BAO (blue) and void emulator (red) does significantly reduce the marginalize constraints (grey). 155

5.21 Joint likelihood analysis combining the BAO+Pantheon result (blue) with the void emulator (red). The grey dashes demarcate the prior walls for the void emulator, and the grey contour denotes the joint likelihood. 156

6.1 A diagram of the void catalogue pipeline, starting from dark matter simulations and culminating in voxel void catalogues. 169

A.1 Joint likelihood analysis combining the BAO+DESY5 result (blue) with the void emulator (red). The grey dashes demarcate the prior walls for the void emulator, and the grey contour denotes the joint likelihood. The joint likelihoods help produce dark energy constraints that are competitive with the BAO+SN+CMB results. 227

A.2 Joint likelihood analysis combining the BAO+Union3 result (blue) with the void emulator (red). The grey dashes demarcate the prior walls for the void emulator, and the grey contour denotes the joint likelihood. The joint likelihoods help produce dark energy constraints that are competitive with the BAO+SN+CMB results. 228

A.3 Joint likelihood analysis combining the BAO+DESY5 result (blue) with the void emulator (red). The grey dashes demarcate the prior walls for the void emulator, and the grey contour denotes the joint likelihood. The joint likelihoods help produce dark energy constraints that are competitive with the BAO+SN+CMB results. 229

A.4 Joint likelihood analysis combining the BAO+Pantheon plus result (blue) but unlike Figure A.1, we explored the constraints for ω_{cdm} instead of Ω_m. 230

A.5 Joint likelihood analysis combining the BAO+Pantheon plus result (blue) but unlike Figure A.2, we explored the constraints for ω_{cdm} instead of Ω_m. 231

A.6 Joint likelihood analysis combining the BAO+Pantheon plus result (blue) but unlike Figure A.3, we explored the constraints for $\omega_{\rm cdm}$ instead of $\Omega_{\rm m}$. 232

A.7 Joint likelihood analysis combining the BAO+DESY5 result (blue) with the void emulator (red). The grey dashes demarcate the prior walls for the void emulator, and the grey contour denotes the joint likelihood. The joint likelihoods help produce dark energy constraints that are competitive with the BAO+SN+CMB results. 234

A.8 Joint likelihood analysis combining the BAO+Union3 result (blue) with the void emulator (red). The grey dashes demarcate the prior walls for the void emulator, and the grey contour denotes the joint likelihood. 235

A.9 Joint likelihood analysis combining the BAO+PantheonPlus result (blue) with the void emulator (red). The grey dashes demarcate the prior walls for the void emulator, and the grey contour denotes the joint likelihood. 237

List of Tables

3.1 Parameter space spanned by the `AbacusSummit` cosmologies used in our training data. Starred parameters are not considered when developing the training set for the BOSS CMASS emulator, and are instead set to their default values ($\alpha_c = 1, \alpha_s = 0$). For all of our parameter inference, our references to α_s will purely focus on the HOD parameter, not the cosmological parameter with the same notation, which is fixed to the fiducial value of $\alpha_s = 0$. 77

4.1 Prior distributions used during the parameter inference stage. Parameters marked with a star are fixed to their default base-ΛCDM values when not considering these model extensions. 90

4.2 Table of cosmologies used in the test set. Bold values highlight changes from the fiducial `c000` case. The full description of these simulations, the cosmologies used and the way these simulations are obtained can be found in [92]. 96

4.3 Table of credible intervals for cosmological parameters from various fits. "CMASS template – 2 bins" describes the constraints obtained with a template using the two redshift bins of [234] covering most of the CMASS galaxies: $0.4 < z < 0.5$ and $0.5 < z < 0.6$. "CMASS template – 1 bin" indicates the results from applying the template method using the CMASS sample to a single redshift bin $0.4 < z < 0.73$ from [158], while "SDSS template" results include all 5 redshift bins. Note that 2D posteriors for both fits hit the prior walls, defined by the emulator, for ω_b and σ_8. 107

4.4 Table of derived parameters and σ_8, with their means and standard deviations shown. The density-split clustering and the galaxy two-point correlation function (2PCF) emulator were fitted to the same data, the BOSS CMASS NGC catalogue with similar redshift binning: $0.45 \leq z \leq 0.60$ for Density Split and $0.46 < z < 0.60$ for the 2PCF emulator. The wavelet scattering transform (WST) analysis was run across a redshift range of $0.46 < z < 0.57$, with $z_{\rm eff} = 0.515$ used for the $f(z)\sigma_8(z)$ calculation. . . . 108

5.1 HOD Parameter space spanned by the DESI OnePercent survey. For all of our parameter inference, our references to α_s will purely focus on the HOD parameter, not the cosmological parameter with the same notation, which is fixed to the fiducial value of $\alpha_c = 0, \alpha_s = 1$. This model was introduced by [237], where α_c, α_s modulate the peculiar velocities of central and satellite galaxies by a Gaussian-distributed line-of-sight component. 126

5.2 Table of marginalized constraints of $\Omega_{\rm m}$, σ_8,w_0 and w_a obtained from emulation of the VG-CCF using DESI DR1 LRG-1 voids, and from the DESI BAO and Full-Shape analysis of the DR1 sample. Where possible, we select results from the LRG-1 redshift bin, which corresponds to the redshift bin used in our analysis ($0.4 < z < 0.6$). The dashed rows contain the results of the joint analysis with DESI DR1 BAO and BAO+SNIa. Starred quantities denote prior-limited constraints. 154

5.3 Table of marginalized constraints for $\omega_{\rm cdm}$, w_0 and w_a utilizing a joint-likelihood analysis of the void emulator with BAO and SNIa, exploring wCDM and w_0w_aCDM cosmological models. 157

Chapter 1

Introduction

Modern cosmology arose from the publication of the theory of General Relativity in 1916, which characterized gravity in terms of mass, energy and the curvature of spacetime [79]. It provided a mathematical framework to describe our Universe in terms of the Einstein field equations. Assuming that the Universe is homogenous and isotropic, these differential equations reduce to the Friedmann equations (also known as the Friedmann-Lemaitre equations) which govern the evolution of the Universe and provide a framework for modern cosmology. The Friedmann-Lemaitre-Roberston-Walker (FLRW) solution to these field equations [87, 88, 146, 195, 196, 197], accurately models the Universe at large scales, and allows for the expansion history of the Universe to be described by different energy densities.

This approach has helped cosmologists determine the composition and age of the Universe, notably one that is finite in observable size and age. This characterization of our finite universe was controversial for decades. Hubble's discovery of receding "nebulae" in 1929 fueled the nascent field of cosmology [125]. Crucially, Hubble's discovery meant that the Universe was no longer defined by the scale of the Milky Way, but something far more massive, and it was expanding. In his 1929 paper, Hubble described the recessional velocity of galaxies v_r in terms of their physical distance, D and a proportionality constant, H_0, now called the Hubble constant in Equation (1.1), which is now known to be valid *locally*.

$$v_r = H_0 D \tag{1.1}$$

In deriving the Friedmann equations, we will illustrate how the concordance model of cosmology was assembled, the timeline of the Universe and outline how cosmology has entered an era of precision measurements. Through these precise measurements the main

components of the Universe, namely ordinary (baryonic) matter, radiation, dark matter and dark energy can be inferred. The Universe is statistically homogeneous and isotropic at the largest scales $\geq 100h^{-1}$Mpc; this is not the case at the scale of individual galaxies and clusters. Making use of the Friedmann equations, perturbations in the density field can be used with them to chart the development of our universe from a hot, dense, form of high energy into one that has hierarchical structure and complicated networks of galaxies and galaxy clusters. This set of perturbations to the density field is crucial for understanding the evolution of the cosmos at late-times (the present day). Compressing the information obtained from measuring the large-scale structure of the Universe, and using it to place constraints on the parameters of the standard model of cosmology is the goal of this thesis.

We will first explore how cosmology emerged from general relativity, what assumptions stipulated the Friedmann equations, and how these laid the foundation for a model of the Universe that can be described through the measured abundances of different forms of matter and energy.

1.1 The Homogeneous and Isotropic Universe

The foundation of the FLRW solution to the Einstein field equations rests on the cosmological principle which assumes that the Universe is both isotropic and homogeneous [87, 146]. *Statistical* isotropy and homogeneity is observationally supported at large scales ($\geq 100\ h^{-1}$Mpc), and is a useful approximation to analyze the dynamics of the Universe. The composition of the Universe is statistically similar, and there is no preferred observer or direction. Galaxies across disparate parts of the sky inhabit the same universe, and obey the same field equations. This is crucial, as it simplifies several tensor components of the Einstein field equations

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \tag{1.2}$$

where the Greek letters are tensor indices, μ, $\nu = (0...3)$, representing the four spacetime components, $T_{\mu\nu}$ is the stress-energy tensor, which describes momentum and energy in terms of spacetime, Λ is the cosmological constant, $g_{\mu\nu}$ is the metric, which describes distances in spacetime, and $G_{\mu\nu}$ is the Einstein tensor, which describes how the metric changes in terms of curvature.

1.1.1 Isotropy and homogeneity

Applying the isotropic and homogeneous approximation means the above equation has no preferred angle or position - this simplifies $T_{\mu\nu}$, the stress-energy tensor to be dependent on the energy density of non-interacting components $\rho(t)$, and their pressure $P(t)$, which can be regarded respectively as the energy density and the pressure of an ideal fluid. The metric is simplified into the equation

$$ds^2 = c^2dt^2 - a^2(t)\left(d\chi^2 + S_k(\chi)d\Omega^2\right) \tag{1.3}$$

where χ is the comoving radial coordinate and Ω is the solid angle.

$$S_k(\chi) = \begin{cases} S_1(\chi) = \sin(\chi) \\ S_0(\chi) = \chi \\ S_{-1}(\chi) = \sinh(\chi) \end{cases} \tag{1.4}$$

Equation (1.3) can be rewritten in terms of the curvature parameter, k as follows:

$$ds^2 = c^2dt^2 - a^2(t)\left(\frac{d\chi^2}{(1-k\chi^2)} + \chi^2(d\theta^2 + \sin^2\theta d\phi^2)\right) \tag{1.5}$$

Comoving coordinates were used for the entirety of this thesis, unless otherwise specified.

Equation (1.3) describes the *comoving* distance between two points in a homogeneous, isotropic spacetime as being dependent on the global curvature of the Universe, $S_k(\chi)$, and a unitless scale factor, $a(t)$, which describes the expansion in distances between objects as a function of time. This latter point was motivated by Hubble's discovery of an expanding Universe.

From the metric Equation (1.5), the proper distance between two events A and B, in a flat universe ($k = 0$), which occur simultaneously to the observer ($dt = 0$) is computed as the integral of ds over a photon trajectory

$$s(t) = \int ds' = a(t)\int \frac{d\chi}{(1-k\chi^2)^{\frac{1}{2}}} = a(t)S_k(\chi) \tag{1.6}$$

Further motivated by this distinction of *comoving* distances and *proper* distances, this metric can be used by inserting Equation (1.5) into the Einstein field equations (Equation (1.2)), to obtain an equation that describes the time evolution of comoving and proper distances in terms of the time-dependent scale factor $a(t)$, the curvature of the Universe, k, and the stress-energy properties, ρ, energy density, and P.

1.1.2 Obtaining equations of motion

Using this distance element obtained from applying the cosmological principle allows us to obtain a solution to the second-order differential equations that follow from Equation (1.2). Applying the FLRW metric, these solutions yield the equations of motion

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho}{3} - \frac{kc^2}{a^2} + \frac{\Lambda}{3} \tag{1.7}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G(\rho + 3P)}{3c^2} + \frac{\Lambda}{3} \tag{1.8}$$

Motivated by the parameterization of the comoving distance between objects, $d\chi$, the physical distance $r_{AB}(t)$ between objects A and B can be written in terms of the scale factor, $a(t)$:

$$r_{AB}(t) = a(t)\chi_{AB} \tag{1.9}$$

Comoving distances are time-independent. To find the physical velocity between objects A and B, taking the time derivative of the physical distance gives the following equations:

$$\frac{dr_{AB}(t)}{dt} = \frac{d}{dt}(a(t)\chi_{AB}) = \chi_{AB}\frac{da}{dt} \tag{1.10}$$

$$\dot{r}_{AB}(t) = \frac{\dot{a}(t)}{a(t)} r_{AB} \tag{1.11}$$

It then follows that Equation (1.11) is identical to Equation (1.1), where $H(a) = \frac{\dot{a}}{a}$ is the Hubble parameter at a scale factor of $a(t)$ at a given time of t. Substituting this quantity and its time derivative back into Equations (1.7) and (1.8) leads to the following equations for $H(a)$:

$$H(a)^2 = \frac{8\pi G\rho}{3} - \frac{kc^2}{a^2} + \frac{\Lambda}{3} \tag{1.12}$$

$$\dot{H} + H^2(a) = -\frac{4\pi G(\rho + 3P)}{3c^2} + \frac{\Lambda}{3} \tag{1.13}$$

where k is the spatial curvature of the Universe. k determines the shape of the Universe, where $k = 1$ corresponds to a closed universe (spherical geometry), $k = -1$ to an open universe (hyperbolic geometry), and $k = 0$ corresponds to a flat universe.

Additionally, cosmologists often report measurements of H_0, the Hubble constant measured at the present-day. Cosmological quantities are often written in terms of $h = \frac{H_0}{100 \text{kms}^{-1}\text{Mpc}^{-1}}$, which rescales them in terms of H_0.

These equations also admit a term with a cosmological constant, Λ. This solution was put forth by Einstein, as it allows for a solution to Equation (1.8), where for a particular value of Λ the solution is static, allowing for a non-expanding static universe. Over time, this Λ term once lost the support of cosmologists; where Einstein later believed that introducing Λ was his greatest blunder. With the discovery of the accelerating expansion of the Universe, commonly referred to as *dark energy*, this term once again found purchase, but this time to allow for late-time acceleration in the Universe's expansion, where Λ could be chosen to align with the observed acceleration, sometimes referred to as Λ *dark energy*. The expansion history of the Universe can be described by the co-evolution of different energy densities, culminating in the present-day dominance of the energy density of dark-energy.

1.1.3 Components of the Universe

Recalling how the assumptions of isotropy and homogeneity restrict the form of $T_{\mu\nu}$ to consist of self-interacting energy densities, ρ and pressures P, this can be taken further. Suppose the Universe consists of N different energy densities, such that the ith component has an energy density of ρ_i, and its equation of state is parameterized as $P_i = w_i \rho_i$. Assuming that these components only interact with each other, this reduces the fluid equations to the form:

$$\dot{\rho}_i + 3\frac{\dot{a}}{a}\left(\rho_i + P_i\right) = 0 \tag{1.14}$$

This decomposition of energy densities can be used to rewrite the Friedmann equations Equations (1.12) and (1.13) in terms of the different energy densities of the Universe. The combination of the equation of state, the fluid Equation (1.14), and the Friedmann Equation (1.12) gives the tools needed to describe the evolution of the scale factor $a(t)$ and dynamics of the Universe in terms of its component energy densities. This is the foundation of precision cosmology, as precise measurements of these different energy densities allow for the measurement of the expansion history of the Universe, as well as for a characterization of the growth of its structures. Solving Equation (1.14) for a constant w_i (equation of state) gives the solution for the time evolution of energy densities:

$$\rho(a) = \rho_0 a^{-3(1+w)} \tag{1.15}$$

where

$$\begin{cases} w = 0, \text{ for matter} \\ w = 1/3, \text{ for radiation} \\ w = -1/3, \text{ for curvature} \\ w = -1, \text{ for cosmological constant dark energy} \end{cases}$$

The form of Equation (1.15) changes if w is non-constant. If this assumption is relaxed, as is the case for *dynamic dark energy*, and the equation of state assumes the form: $w(a) = w_0 + (1 + a)w_a$, where w_0,w_a are the scalar parameters, the integral of Equation (1.14) changes:

$$\rho = \rho_0 a^{-3(1+w_0+w_a)} e^{-3w_a(1-a)} \tag{1.16}$$

where $w_a = 0, w_0 = -1$ simplifies into Λ dark energy. Λ implies dark energy is a negative pressure fluid that accelerates the expansion of the Universe, and has a constant energy density.

We know the Universe is composed of at least non-relativistic matter ($w = 0$) and radiation ($w = 1/3$), allowing us to decompose the energy density ρ into ρ_m and ρ_r. For a spatially flat universe where $\Lambda = 0$, we can rewrite Equation (1.12) in terms of a critical density in Equation (1.17):

$$\rho_c(t) = \frac{3H^2(t)}{8\pi G} \tag{1.17}$$

Dividing the energy densities of all components by the critical density at t_0, we can then write:

$$\Omega_i = \rho_{i,0}/\rho_{c,0} \tag{1.18}$$

where Ω_i denotes the present day value, unless otherwise specified. We can also write:

$$\omega_i = \Omega_i h^2 \tag{1.19}$$

This means the Friedmann equation can be rewritten in terms of Ω_i. This links the evolution of the scale factor $a(t)$ with present-day energy densities Ω_i. Typically, $\Omega_k = 0$ for flat universes. For dynamical dark energy, at late times ($a \approx 1$), with $w_0 \approx -1$,$w_a \approx 0$, dark energy will evolve similarly to its Λ counterpart (i.e. Ω_Λ is fixed, and will dominate at late-times). But at earlier times, its behaviour will be different. For dynamical dark energy, Equation (1.12) can be rewritten as

$$H(a)^2 = H_0^2 \left(\Omega_{m,0} a^{-3} + \Omega_{r,0} a^{-4} + \Omega_{k,0} a^{-2} + \Omega_{\Lambda,0} a^{-3(1+w_0+w_a)} e^{-3w_a(1-a)}\right) \tag{1.20}$$

For Λ dark energy, it takes the form

$$H(a)^2 = H_0^2 \left(\Omega_{m,0} a^{-3} + \Omega_{r,0} a^{-4} + \Omega_{k,0} a^{-2} + \Omega_{\Lambda,0}\right) \tag{1.21}$$

This latter equation describes the standard model of cosmology - ΛCDM. For the aforementioned curvature, Ω_k is defined as:

$$\begin{cases} \Omega_k = 0 & \text{flat, } k = 0 \\ \Omega_k > 0 & \text{open, } k = -1 \\ \Omega_k < 0 & \text{closed, } k = +1 \end{cases}$$

1.1.4 Latest measurements of energy densities in the Standard Model of cosmology

The current scientific consensus has formed around the following measurements of these components of the Universe. We will later discuss how these measurements are obtained, and how this information relates to the chronology of the Universe.

Since the late 1990s, the concordance model of cosmology has centered around a model with cosmological constant dark energy Λ, with a flat ($k = 0$) spatial curvature. This is known as the ΛCDM model of cosmology, which holds that cold dark matter comprises the majority of the Universe's matter, and that the acceleration in its expansion is driven by cosmological constant dark energy Λ. The evolution of the Universe, outlined in Section 1.2 was identified through measuring the energy density of different components of Equation (1.18) as they fit into Equation (1.21). The latest measurements for the energy density of baryons obtained $\Omega_{\rm b}h^2 = 0.02166 \pm 0.00015$ [56] using Big-Bang Nucleosynthesis measurements, and $\Omega_{\rm b}h^2 = 0.02237 \pm 0.00015$ [182] from measurements of the cosmic microwave background (CMB) radiation measured by the Planck satellite. The CMB is rich in information, and has given us the strongest constraints on $\Omega_{\rm m}$, which was measured to be: $\Omega_{\rm m} = 0.315 \pm 0.007$ with the Planck satellite [182], and $\Omega_{\rm r} = 9.2 \times 10^{-5}$. In a flat ΛCDM universe, baryons amount to a small percentage of the energy density. Measurements of the Universe's large-scale clustering also yield a constraint on $\Omega_{\rm m}$, where the latest BAO analysis of the second data release (DR2) from DESI obtained $\Omega_{\rm m} = 0.297 \pm 0.0086$ [51]. Under flat ΛCDM, the energy density of dark energy can be inferred through $\Omega_{\rm m,0} + \Omega_\Lambda = 1$.

1.2 A timeline of the Universe under ΛCDM

Now that we can describe the ΛCDM model and its components, we can assemble a timeline of the Universe as it relates to different components and cosmic history. Using the Friedmann equations, we can trace the history of the Universe back to when it was infinitely small and infinitely dense.

1.2.1 The Big Bang, inflation and the first minutes

The Friedmann equations imply that if our universe has a finite age and is expanding, then it was once much smaller, denser and hotter. Observational evidence supports this picture of a finite age for our universe (Section 1.5). This encompasses a rapid cascade of events that shapes the initial density fluctuations we will discuss in this chapter, the formation of baryons, and the subsequent formation of the first atoms through nucleosynthesis in this hot and dense state of matter.

Measurements of the cosmic microwave background temperature fluctuations support the conclusion that the nascent (approximately 380,000 years old) Universe's temperature was statistically isotropic to the order of 1 part in 10^5. However, this temperature map contains cold-spots, temperature fluctuations that are large enough and far enough apart that when the Universe was the size it was 380,000 years after the Big Bang, these cold spots could not have been in causal contact with each other. This is called the *Horizon* problem. Given the scale of the Universe in the first instants following the Big Bang, the lack of spatial curvature is another notable feature, even though small quantum fluctuations could have altered it drastically in the early Universe. This is the *Flatness* problem.

The leading solution to both of these problems is to introduce a process called inflation [103]. After about 10^{-36} seconds into the life of the Universe, inflation causes it to rapidly expand until the Universe is about 10^{-32} seconds old. Small-scale quantum fluctuations are stretched to macroscopic scales during this period of inflation. Any curvature that would have been present was rapidly flattened in this process, and it allowed for enough causal contact that ensured the first observable radiation of the universe would have small anisotropies. This results in an early Universe that is flat and statistically homogeneous and isotropic. Following the end of inflation, the Universe continues to rapidly expand, and around this point in time, matter and antimatter pairs annihilate, but a small excess of matter is left over. This is the process of baryogenesis. Soon, at $t \sim 10^{-5}s$ quarks are confined into hadrons (including protons and neutrons). Up until the first 0.01 seconds, the thermal conditions of the Universe are dominated by high energy photons and neutrinos.

At around 1 second after the Big Bang, as the temperature continues to drop, high energy neutrinos decouple from the background, freezing in the quantity of light neutrino species. From 1 second into the Universe's life until about 3 minutes, protons and neutrons rapidly form isotopes of Deuterium, Helium-3, Helium-4 and Lithium-7 [91, 17]. This process is called Big Bang Nucleosynthesis (BBN), and is influenced by the energy density of baryons in the Universe.

1.2.2 Matter-radiation equality

The early universe was dominated by high energy particles. In the period following Big Bang Nucleosynthesis, the Universe rapidly cooled and expanded. The small density fluctuations began to grow under gravity. However, this growth was suppressed at sub-horizon scales for a significant period, arising from two factors: i) the rapid expansion of the Universe and ii) the photon pressure, as the growth in density fluctuations under gravity would be significantly suppressed during this period of radiation energy density dominance. The characteristic scale at which a gravitational instability would balance with this gas pressure is called the physical Jeans scale [168], λ_J. Overdense regions larger than this length collapsed under gravity, while those smaller than this scale underwent fluctuations in their density, the pressure gradient balancing against the gravitational collapse.

$$\lambda_J = \frac{2\pi a}{k_J} = c_s \sqrt{\frac{\pi}{G\bar{\rho}}} \tag{1.23}$$

As the Universe continued to expand, photons and neutrinos (radiation) began to lose energy with this expansion, leading to the energy density of matter reaching equilibrium with radiation at $a_{eq} = \frac{\Omega_{\mathrm{r},0}}{\Omega_{\mathrm{m},0}}$. Following this matter-radiation equality in the energy density, dark matter was able to gravitate and form structures. The growth of density perturbations was no longer suppressed. This growth was still suppressed at smaller scales by the radiation pressure from photons coupled with baryons, consequently the Jeans length was still large. This scale leaves a feature that can be observed in the present-day large-scale structure of the Universe, freezing-in the super-horizon and sub-horizon scales at k_{eq}. As we will show later on, these density perturbations grew logarithmically during the radiation-dominated period, and linearly after matter-radiation equality.

1.2.3 Recombination and Decoupling

When the Universe had cooled to a temperature of $T = 3000K$ and $z \approx 1100$, the average energy of photons had dropped to such a point that the photoionization of Hydrogen nuclei becomes less frequent. Neutral atoms form by radiative recombination which leads the ionization fraction of the Universe to rapidly drop. This rapid process proceeds over tens of thousands of years, and culminates when the Universe is about 380,000 years old.

Photon-electron decoupling rapidly followed. Decoupling was characterized by the scattering rates of photons compared to electrons diverging. The sudden drop in free electrons in this electrically neutral universe, and their rapid dilution by the expansion of the Universe meant that the number density of electrons is no longer closely tied to that of photons. This caused the populations to decouple from each other, causing the Universe to be optically thin to photons. Photons would freely travel with an exceptionally low probability of ionizing hydrogen, helium or other nuclei. This marked change in behavior occurred at around $z = 1090$.

These photons freely streamed through the Universe, retaining most of their energy. Prior to this, the small scale fluctuations in baryon density had been oscillating with the previously coupled photons as a fluid, forming standing waves, with the photon pressure stalling against gravitational collapse. These baryon acoustic oscillations (BAOs) are a key feature of the universe, which we will discuss in Section 1.5. These baryons over the course of recombination and decoupling have become increasingly neutral, and the disappearance of the photon pressure allows these overdense baryon regions to gravitate and collapse into the gravitational wells formed by dark matter. It is from these seeds that the first stars will be born and the first galaxies will assemble. This epoch of the photons decoupling from the hot baryons is the "surface of last scattering" - the first photons from a universe that has become optically thin. We will discuss the CMB as an observable in Section 1.5.2, and how the baryon-acoustic oscillations (BAO) from this period have a dramatic impact on the large-scale structure of the Universe.

1.2.4 Growth of structure

As we will discuss in more detail in Section 1.4, small fluctuations in the density field have caused dark matter and later baryons to infall. Prior to decoupling, the baryon-photon fluid experienced oscillations (standing sound waves). The decoupling of photons allowed gravity to become dominant. The baryons and cold dark matter gravitated and rapidly grew into the first populations of stars and galaxies. It is during this epoch that the density

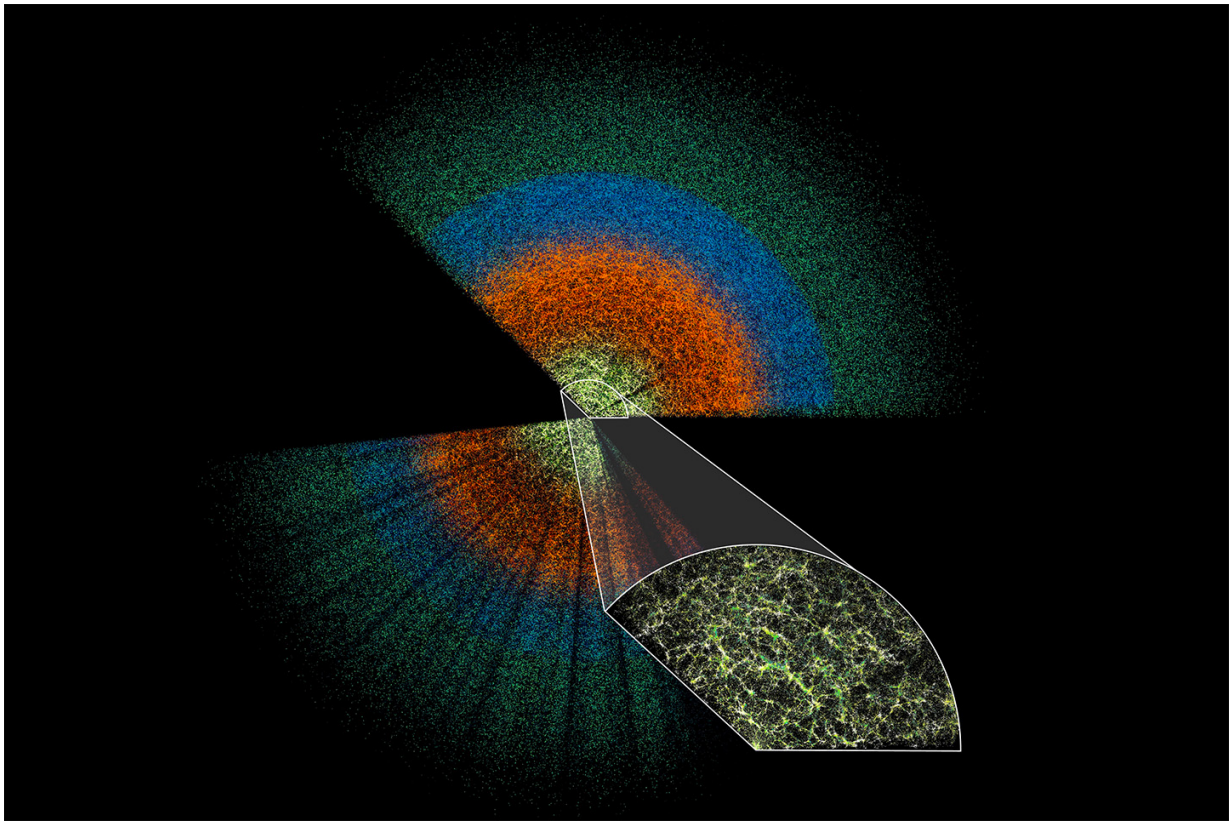

Figure 1.1: A slice of the first year of the DESI survey, image credit to C.Lamman/DESI survey. Each point is a galaxy, with different colours corresponding to the different galaxy samples.

fluctuations $\delta(t)$ grew (in a purely matter-dominated universe, $\delta(t)$ grew proportional to $a(t)$) which allowed large scale structures to form. This in turn led to the hierarchical clustering that is observed in the present (Figure 1.1) where galaxies inhabit high density clusters, separated by lower density regions called voids. The large-scale structure of the Universe is a relic of early-time acoustic oscillations. These BAO will be discussed as an observable clue for structure formation in Section 1.5.3.

1.2.5 Dark energy dominated era and the present

In the ΛCDM model, at $z \approx 0.5$, or 10 Gyr after the Big Bang, the energy density of dark energy becomes dominant as the matter energy density continues to dilute with the expanding volume of the universe $\rho_{\mathrm{m},0} = \rho_{\mathrm{m}} a^{-3}$. This leads to the accelerated expansion of the Universe that was observed at late times, which motivated the inclusion of Λ into the Friedmann equations as discussed above. The dominance of dark energy also suppresses the formation and growth of linear perturbations, leaving them frozen-in. Galaxy clusters during this period are still growing, but the space between these clusters is rapidly increasing.

1.3 Cosmological Distances and Redshift

Cosmological distances are not measured directly. Given the notation introduced in Section 1.1.1, we note that there are *physical* and *comoving* distances, and these are related through Equation (1.9). As the Universe expands, physical separations and photon wavelengths are stretched. This stretching is proportional to the Doppler shift of a photon's wavelength upon emission λ_e to the wavelength it is observed at λ_o. This shift is defined by Equation (1.24). It describes the relative shift in wavelength.

$$1 + z = \frac{\lambda_o}{\lambda_e} \tag{1.24}$$

This Doppler shift of light can be related to the expansion of the Universe - the increase in space between objects as a function of the scale factor $a(t)$ can be related to the redshift:

$$1 + z = \frac{a(t_o)}{a(t_e)} \tag{1.25}$$

By defining $cz = v_r$, where v_r is the recessional velocity of the source emitting light, with respect to the observer, then the Doppler shift of this light can be linked to the Hubble constant at low z:

$$cz_{\mathrm{obs}} = H_0 d_p \tag{1.26}$$

This equation assumes that the only significant velocity with respect to the observer comes from the expansion of the Universe. Section 1.5.3 will outline how this expression is modified by the velocity fields of galaxy clusters.

1.3.1 Defining distance

In prior sections, physical and comoving distances were distinguished in terms of their relationship to the redshift of an object (see: Equation (1.9)). In cosmology, there are various definitions for distance. Here we will review most crucial ones, adopting the notation of [121].

We established that Hubble's law can be linked to the observed Doppler shift of astronomical sources in Equation (1.26). For photons travelling at the speed of light, Equation (1.1) can be rewritten as the Hubble distance, Equation (1.27), which is the Hubble time: $1/H_0$ multiplied by the speed of light. Hubble time, and the Hubble distance

$$D_H = c/H_0 \tag{1.27}$$

both set the scale of the Universe, and can be used to define other distances with respect to this scale.

Comoving distance

The line-of-sight comoving distance is the distance between objects scaled by $a(t)$. If a pair of objects are moving solely due to the expansion of the Universe (Hubble flow), then their comoving distance should be identical, as their proper distance would increase in direct proportion with the scale factor.

$$E^2(z) = H_0^2 \left(\Omega_M(1+z)^3 + \Omega_R(1+z)^4 + \Omega_k(1+z)^2 + \Omega_\Lambda\right) \tag{1.28}$$

Equation (1.28) is related to the time derivative of the scale factor, as $H(z) = E(z)H_0$.

$$D_C = \chi = D_H \int \frac{dz'}{E(z')} \tag{1.29}$$

Integrating $c/H(z)$ in terms of the redshift yields (Equation (1.29)), the line-of-sight comoving distance. This expression can be later utilized for other distances discussed in this work.

The transverse component of the comoving distance, separating a pair of objects at identical redshifts but varying solid angles, is determined by the spatial curvature of the Universe, k.

$$D_M = \begin{cases} D_H \frac{1}{\sqrt{\Omega_\mathrm{k}}} \sinh\left[\Omega_\mathrm{k} D_C / D_H\right] \text{ for } \Omega_\mathrm{k} > 0 \\ D_C \text{ for } \Omega_\mathrm{k} = 0 \\ D_H \frac{1}{\sqrt{\Omega_\mathrm{k}}} \sin\left[\Omega_\mathrm{k} D_C / D_H\right] \text{ for } \Omega_\mathrm{k} < 0 \end{cases} \tag{1.30}$$

Equation (1.30) summarizes this for the three cases of spatial curvature.

Angular diameter distance

The angular diameter distance is the physical distance of the object to the observer scaled by its angular size. At small distances, this is typically defined as $d_A = \frac{x}{\theta}$. With a curved spacetime and large distances, we write it in terms of the curvature of the Universe, the comoving distance and the redshift in Equations 1.31, 1.32. Where the curvature affects the numerator as described by Equation 1.4.

$$D_A = \frac{S_k(\chi)}{(1+z)} \tag{1.31}$$

Equation 1.31 can be further simplified by using the transverse component of the comoving distance, D_M.

$$D_A = \frac{D_\mathrm{M}}{(1+z)} \tag{1.32}$$

Luminosity distance

If the flux F of an object is measured, then a corresponding distance can be obtained through the flux-luminosity relationship. Often, this distance is particularly useful when measuring the distance to an object with a well known luminosity profile, such as a standard candle.

$$D_L = \sqrt{\frac{L}{4\pi F}} \tag{1.33}$$

For cosmological scales, the luminosity distance D_L is related to the transverse comoving distance D_M by a factor of $1+z$ in Equation (1.34), and the angular diameter distance D_A by a factor of $(1+z)^2$.

$$D_L = (1+z) D_\mathrm{M} \tag{1.34}$$

$$D_L = (1+z)^2 D_\mathrm{A} \tag{1.35}$$

1.4 The anisotropic and inhomogeneous Universe

The Friedmann equations we described earlier come from assuming homogeneity and isotropy. Observational evidence supports the assumption of *statistical* homogeneity and isotropy at large scales. We will explore how the small-scale fluctuations in the statistically homogenous and isotropic density field of the early Universe will grow into the large-scale overdensities that are observed in the present, such as the ones seen in Figure 1.1.

1.4.1 Adding perturbations

First consider the evolution in the matter density where $\bar{\rho}(\boldsymbol{x}, t)$ is defined as the mean matter density field at spacetime coordinates $\boldsymbol{x}, t$, and $\rho(\boldsymbol{x}, t)$ is the local matter density. The local density field can be related to the mean density field in terms of a time-dependent perturbation $\delta(t)$(as $\rho_m \sim a^{-3}$). The local density field can be defined in terms of the mean matter density field as

$$\rho(\boldsymbol{x}, t) = \bar{\rho}(\boldsymbol{x}, t)\,(1 + \delta(t)) \tag{1.36}$$

.

Equation (1.36) can be rearranged to define the overdensity field, i.e. the relative magnitude of the density field with respect to the mean density of the Universe.

$$\delta(\boldsymbol{x}, t) = \frac{\rho(\boldsymbol{x}, t) - \bar{\rho}(\boldsymbol{x}, t)}{\bar{\rho}(\boldsymbol{x}, t)} \tag{1.37}$$

The time evolution of these small $|\delta| \ll 1$ perturbations in the density field can be modeled by treating it as a fluid, and in combining Equation (1.37) with the continuity, Euler and Poisson equations written in terms of comoving coordinates, resulting in a differential equation describing the evolution of these overdensities:

$$\frac{\partial^2 \delta}{\partial t^2} + 2H\frac{\partial \delta}{\partial t} = 4\pi G \bar{\rho}_m \delta + \frac{c_s^2}{a^2}\nabla^2\delta + \frac{2}{3}\frac{\bar{T}}{a^2}\nabla^2 S \tag{1.38}$$

This equation can be further simplified. At this stage, it is useful to consider the Fourier modes of δ, δ_k. If the Universe is statistically homogeneous and isotropic, then the Fourier modes of the density field, δ_k are Gaussian distributed. Consequently, perturbations at different k will be independent of each other.

$$\delta_k = \frac{1}{V}\int \delta(\mathbf{r}) \exp(-i\mathbf{k}\cdot\mathbf{r}) d^3\mathbf{r} \tag{1.39}$$

This simplifies the partial differential equation in Equation (1.38) to an ordinary differential equation in k-space.

$$\ddot{\delta}_k + 2H\dot{\delta}_k = 4\pi G\bar{\rho}_m\delta_k - \frac{k^2c_s^2}{a^2}\delta_k \tag{1.40}$$

This equation describes the time evolution of Fourier modes δ_k in terms of the sound speed, and the background growth in terms of matter density. Solving for a static background growth case $H = 0$ leads to a solution that defines the Jeans length λ_J in configuration space and Fourier space $k_J = \frac{2\pi a}{\lambda_J}$. This solution indicates that physical scales above (below) k_J experience a suppression (permission) in growth. It should be noted that k_J changes over time. At recombination, k_J is at such a scale that the growth of galaxy clusters is suppressed, while following decoupling this limitation is relaxed.

For $k \ll k_J$, i.e. large-scales, this above equation is further simplified. Relating $\bar{\rho}_m$ to $\Omega_{\rm m}$ allows the equation to be solved using the present day matter energy density.

$$\ddot{\delta} + 2H\dot{\delta} - \frac{3}{2}\Omega_{\rm M}\delta H^2 = 0 \tag{1.41}$$

While this differential equation can be solved for combinations of radiation, matter and dark energy density, it is informative to consider the solutions for flat universes where each of these components dominate. These solutions to Equation (1.41) are summarized below:

$$\begin{cases} \delta(t) = A_1 + A_2 \ln t & (\text{radiation} - \text{dominated}) \\ \delta(t) = D_1 t^{2/3} + D_2 t^{-1} & (\text{matter} - \text{dominated}) \\ \delta(t) = B_1 + B_2 e^{-2Ht} & (\Lambda - \text{dominated}) \end{cases} \tag{1.42}$$

Solving this differential equation for an Einstein-deSitter universe (i.e. $\Omega_{\rm m} = 1$) helps illustrate how structures grow in the matter-dominated era, the EdS solution is then: $\delta(t) = D_1 t^{2/3} + D_2 t^{-1}$, which consists of a growth and decay mode respectively. Over time, the former solution dominates over the latter, so for this matter-dominated universe, $\delta(t)$ evolves as described in Equation 1.42. This means that small density fluctuations in the linear regime will ultimately grow proportionally to the scale factor. Solving this equation for late times (i.e. $\Omega_\Lambda \gg \Omega_{\rm M}$) means growth is frozen-in. For radiation domination, the growth is logarithmic, i.e. suppressed.

$$D(a) = \frac{\delta(a)}{\delta(1)} \tag{1.43}$$

$$D(z) \sim g(z)/(1+z) \tag{1.44}$$

From these solutions, the growth rate of matter $D(a)$ can be defined in Equation (1.43), where $g(z)$ for a matter-dominated solution is defined:

$$g(z) \approx \frac{5}{2}\Omega_m(z)\left[\Omega_{\rm m}^{4/7}(z) - \Omega_\Lambda(z) + \right]^{-1} \tag{1.45}$$

Once $\delta \sim 1$, this approximation is no longer sufficient. One extension to the model of the linear growth of structure is to treat overdensities as spherically symmetric perturbations that will collapse under gravity. If these structures have a linear overdensity of $\delta_c > 1.7$, then the resultant structure will form a virialized dark matter halo at late times. N-body simulations have been used to model this non-linear growth period (e.g. [210, 132, 167]). We will instead focus on the larger-scale behaviour of $\delta(x)$.

$$\delta(\boldsymbol{x}) = \frac{V}{(2\pi)^3}\int \delta_k e^{-i\vec{k}\cdot\boldsymbol{x}} d^3k \tag{1.46}$$

It is useful to consider the Fourier modes δ_k. The power spectrum is defined as follows:

$$P(k) = \left\langle |\delta_k|^2 \right\rangle \tag{1.47}$$

If all Fourier modes δ_k are uncorrelated, then the density field described by this decomposition is a Gaussian field. Theories of inflation predict that the density field should be primarily Gaussian [103]. In addition, if the density field $\rho(\boldsymbol{x})$ is homogeneous and isotropic, then this Gaussian field is statistically complete. This means that overdensities are distributed in a homogeneous and isotropic probability distribution function (PDF), described by the equation:

$$p(\delta) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-\delta^2}{2\sigma^2}\right) \tag{1.48}$$

The variance of this field can also be defined:

$$\sigma^2 = \frac{V}{(2\pi)^3}\int P(k)d^3k = \frac{V}{2\pi}\int P(k)k^2dk \tag{1.49}$$

The variance of the overdensities in a sphere of comoving radius, R can be defined by first defining a *window* function, $W_R(x)$. One common window function is the spherical

top-hat defined as follows:

$$W_R(x) = \begin{cases} \frac{3}{4\pi R^3} & , \ x < R \\ 0 & , \ x \geq R \end{cases}$$

Defining the observed matter overdensity $\delta_{m,R}(x)$ within a comoving radius, R as:

$$\delta_{m,R}(x) = \int d^3x \delta_m(x') W_R(|x - x'|) \tag{1.51}$$

This allows the variance within this comoving radius to be calculated:

$$\sigma_R^2 = \left\langle \delta_{m,R}(x)^2 \right\rangle \tag{1.52}$$

Often, this variance is scaled to match the observed scale of galaxy clusters, around $8h^{-1}$Mpc. This scaled quantity of Equation (1.49) is called σ_8, or the amplitude of matter fluctuations.

Most models of inflation describe the power spectrum $P(k)$ as following a power-law relation, i.e. $P(k) \sim A_s k^{n_s}$, where measurements of the temperature anisotropies from CMB limit us to $n_s \sim 1$. This means that the initial density field is nearly Gaussian [103]. This primordial power spectrum significantly evolves as the Universe reaches the epoch of matter and dark-energy domination, as shown in Figure 1.2. In addition to the shape of the power spectrum changing at the characteristic scale of k_{eq} (from matter-radiation equality), the power spectrum will become increasingly nonlinear as a result of the formation of large scale structure at late times.

1.4.2 Peculiar Velocities and the Distance-to-Redshift relation

Cosmological distances are obtained by converting the Doppler shift of an object's light (redshift, z) into a distance. This conversion depends on the cosmological model adopted. For instance, in Equation (1.29), the comoving distance depends on whether the Universe's curvature is flat, open or closed and the energy densities of $\Omega_{\rm m}, \Omega_\Lambda, \Omega_r$. Furthermore, if there were any external factors in our observation of the redshift, then this could frustrate measurements of the distance.

Peculiar velocities are one such example. The local environment of galaxy clusters produces strong gravitational forces, imparting velocities whose vector may deviate from the direction of Hubble flow. Any component of the peculiar velocity that falls along

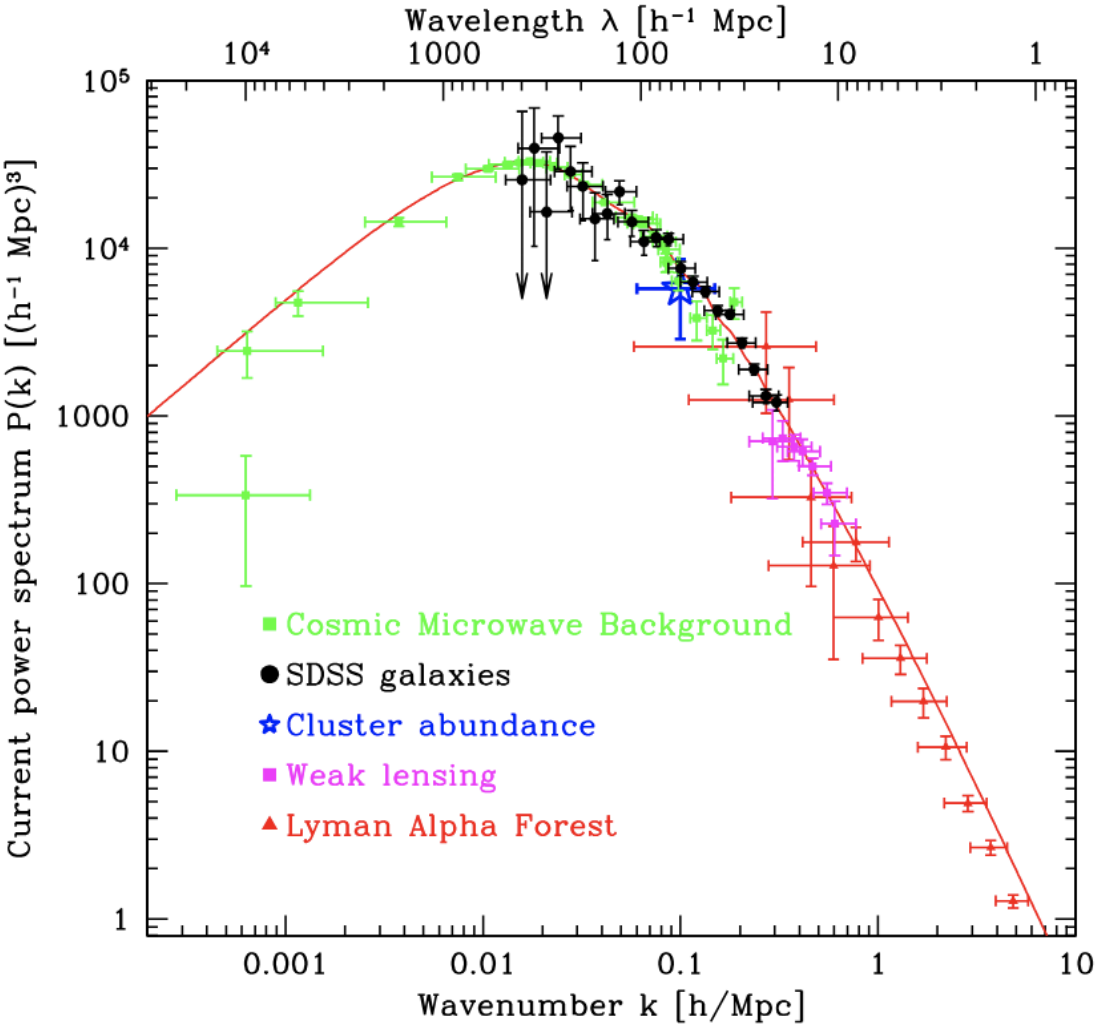

Figure 1.2: The matter-power spectrum as measured through galaxy clustering and the CMB, image from [223]. The peak in this power spectrum corresponds to wavenumber (scale) of the Universe' comoving Hubble radius when matter-radiation equality occurred, k_{eq}. For larger scales $k < k_{eq}$, the relation is a power law, as it corresponds to structures that were unable to interact with each other (super-horizon modes), while for smaller scales $k > k_{eq}$, this corresponds to modes that entered the comoving Hubble radius prior to matter-radiation equality, and thus had their growth suppressed by radiation pressure (sub-horizon modes).

the line-of-sight will be added to the recessional velocity component when the redshift is measured. This peculiar velocity component makes the object move slower or (faster) than the velocity given in Equation (1.26), and thus the inferred redshift is lower (greater) than an object with no peculiar velocity. This modulates the inferred position of an object, as the peculiar velocity component along the line-of-sight will skew the measured redshift. Consequently, while Equation (1.26) is the assumed relation between observed distances to redshifts, it neglects the contribution of peculiar velocities, hence the additional term in Equation (1.53).

$$cz_{obs} = H_0 D + v_{pec} \tag{1.53}$$

Redshift-space distortions (RSD) greatly alter the observed density field, thus we distinguish between *real*-space positions and *redshift*-space positions. An object's position in *redshift*-space $\boldsymbol{s}$ is related to its *real*-space position $\boldsymbol{r}$ through the equation:

$$\boldsymbol{s} = \boldsymbol{r} + (\boldsymbol{r} \cdot \boldsymbol{v_{pec}})\hat{\boldsymbol{r}} \tag{1.54}$$

Because the positions of galaxies are preferentially distorted along the line-of-sight, this renders any observation of galaxy positions anisotropic. Modelling the effect of RSD allows for the linear growth rate to be obtained, which describes how rapidly density fluctuations grow over time. It is useful to define the linear growth rate, $f(z)$:

$$f(z) = \frac{\mathrm{d}\ln\delta}{\mathrm{d}\ln\mathrm{a}} \approx \Omega_{\mathrm{m}}(z)^{\gamma} \tag{1.55}$$

This is a quantity that can be derived from measuring RSD in galaxy-redshift surveys. The approximation in Equation (1.55) holds for FLRW spacetimes, and is typically set as $\gamma \approx 0.55$. The latest measurements of RSD will be discussed in Section 1.5.3.

1.4.3 The Two-Point Correlation Function

The Gaussian nature of these primordial density fields means that the matter field can be described entirely by its auto correlation $\xi(\vec{r})$.

$$\xi(\vec{r}) = \left\langle \delta(\vec{x_1})\delta(\vec{x_2}) \right\rangle \tag{1.56}$$

Defining $\vec{r} = \vec{x_1} - \vec{x_2}$, then the above equation simplifies into an expression for the two-point correlation function. This expression has a Fourier transform, $P(\vec{k})$.

$$\xi(r) = \frac{V}{(2\pi)^3} \int P(\vec{k}) e^{-i\vec{k}\cdot\vec{r}} d^3k \tag{1.57}$$

The power spectrum $P(k)$ depends on the scalar k and its Fourier counterpart, $\xi(r)$ only depends on the scalar r, rendering Equation 1.58 as the standard.

$$\xi(r) = \frac{V}{(2\pi)^3} \int P(k) e^{-ikr} d^3k \tag{1.58}$$

This formalism can be applied to galaxies obtained from galaxy-redshift surveys, where Equation 1.59 describes the excess probability of finding a pair of galaxies g_1, g_2 at some separation of r_{12}, modulated by the correlation function $\xi(r_{12})$, enclosing volumes dV_1, dV_2 and the average galaxy number density $\bar{n}_g$.

$$dP(g_1, g_2) = \bar{n}_g \left[1 + \xi(r_{12})\right] dV_1 dV_2 \tag{1.59}$$

1.4.4 Galaxy two-point correlation function and its multipoles

The matter power spectrum $P(k)$ and the two-point correlation function $\xi(r)$ describe the large-scale clustering of the matter density field in Fourier space and configuration space. At times, it is more illustrative to utilize the Fourier space function than its configuration space counterpart. For this thesis, $P(k)$ and $\xi(r)$ are discussed to give the reader the background knowledge on measurements of the galaxy two-point correlation function and the galaxy power spectrum. For later chapters, the void-galaxy two-point correlation function will be exclusively discussed in configuration space. The two-dimensional galaxy correlation function $\xi(s, \mu)$ is written in terms of the cosine of the angle, μ and the comoving distance of the object s.

$$\xi_\ell(s) = \frac{2\ell + 1}{2} \int_{-1}^{1} \mathrm{d}\mu \, \xi(\mathrm{s}, \mu) \mathrm{P}_\ell(\mu), \tag{1.60}$$

The first three moments are written in Equation (1.61). This decomposition has a clear physical interpretation where $\ell = 0$ (the monopole) is equivalent to integrating over all angles, taking a spherical average of the stacked galaxies around other galaxies.

$$P_\ell(\mu) = \begin{cases} 1 & \text{for } \ell = 0 \\ \mu & \text{for } \ell = 1 \\ \frac{3}{2}\mu^2 - \frac{1}{2} & \text{for } \ell = 2 \\ \dots \end{cases}$$

The quadrupole moment, $\ell = 2$, is an integral that will measure the anisotropy along the line-of-sight compared to the transverse direction. This is used to quantify RSD and AP distortions, which are discussed in Section 1.5.3. Galaxy catalogues, obtained from galaxy-redshift surveys will thus consist of *observed* galaxy positions that are neither statistically isotropic nor homogeneous due to AP and RSD.

Cold dark matter began to cluster before baryonic matter, and while most dark matter halos contain galaxies, the distribution varies greatly. The galaxy overdensity field traces the distribution of baryons and cold dark matter in an biased manner. Consequently, the two-point galaxy correlation function also retains this property: $\xi(r) \neq \xi_{\text{baryons}}(r)$. This is further complicated by the evolutionary history of these galaxy clusters - gravitational infall, ram pressure stripping and mergers all alter the local ratio between cold dark matter and galaxies.

Large-scale bias and RSD

Galaxies are biased tracers of the matter density field - different observed samples of the galaxy population will have differences in their clustering. The relationship between the observed galaxies to the matter density field can be described by

$$\delta_g = \sum_i^\infty \frac{b_i}{i!} \delta_m^i \tag{1.62}$$

where b_i are the bias factors for each population [50]. At large scales, it can be shown that the galaxy and matter fields are correlated, and with the appropriate assumptions, Equation (1.62) can be simplified into a linear bias relation:

$$\delta_g = b\delta_m \tag{1.63}$$

The measurement of this large-scale bias can allow for a constraint on the growth of structure. First consider the matter overdensity, $\delta_m(r)$. Mass conservation can be applied

to a volume element in real-space d^3r and redshift-space d^3s, relating the overdensity fields in real-space and redshift-space:

$$(1 + \delta_m(s))d^3s = (1 + \delta_m(r))d^3r \tag{1.64}$$

Using the peculiar velocity Equation (1.53), and taking the Jacobian for the volume elements in Equation (1.64) leads to:

$$(1 + \delta_m(s)) = (1 + \delta_m(r)) \left(1 - \frac{1}{H(z)} \frac{dv_\parallel}{dr_\parallel}\right) \tag{1.65}$$

Taking the linear regime allows this to be simplified, with the z-axis chosen as the line-of-sight direction, and assuming a plane-parallel approximation [131] for the Fourier transform of the above to be written in terms of f, the linear growth rate, μ, the angle between the LOS and the wavemode k.

$$\delta_m(k)^{(s)} = \delta_m(k) \left(1 + f\mu^2\right) \tag{1.66}$$

Since the anisotropic power spectrum can be written as $P(k, \mu) \sim \langle \delta(k)^2 \rangle$, then the density perturbations can be rewritten with respect to a power spectrum. Computing the power spectrum and linearizing the terms from Equation (1.66) leads to the power spectrum given below, relating the power spectrum for matter in real-space P^r to its redshift-space counterpart P^s .

$$P^s_{mm}(k, \mu) = (1 + f\mu^2)^2 P^r_{mm}(k) \tag{1.67}$$

This can be further extended, as the matter power spectrum can be linked to a galaxy power spectrum through Equation (1.63).

$$P^s_{gg}(k, \mu) = P^r_{mm}(k) \left[b + f\mu^2\right] \tag{1.68}$$

Defining $\beta = f/b$, then the matter-power spectrum can be linked to the galaxy power spectrum in redshift-space as:

$$P^s_{gg}(k, \mu) = (1 + \beta\mu^2)^2 P^r_{gg}(k) \tag{1.69}$$

Measuring the redshift-space power spectrum or its correlation function, and modelling the contribution of RSD is an important test for modified gravity models (e.g. [104, 216]), as well as constraining the growth rate of structure $f(z)$, and indirectly, $\Omega_{\rm m}$. These measurements are often normalized by the variance in density fluctuations within a sphere of radius $8h^{-1}$Mpc, i.e. $\sigma_8(z)$, using Equation (1.49). A thorough review of linear RSD can be found in Hamilton [118].

1.4.5 Distance-redshift relation and the distortion of geometry

We have already established the distance-redshift relation is model dependent. Different cosmological model parameterizations will alter D_M and D_C significantly. This gives rise to the Alcock-Paczynski effect (AP), which is caused by the chosen fiducial cosmology being inconsistent with the true cosmology of the Universe [15]. Comoving distances parallel and transverse to the line-of-sight are distorted, with stretching governed by the AP parameters:

$$q_\perp = \frac{D_{\rm M}(z)}{D_{\rm M}^{\rm fid}(z)}, \quad q_\parallel = \frac{D_{\rm H}(z)}{D_{\rm H}^{\rm fid}({\rm z})} \tag{1.70}$$

The AP parameters in Equation (1.70) are a ratio of the comoving distance parallel and transverse to the line-of-sight in terms of the fiducial and true cosmologies. Given that galaxy clustering is isotropic in real-space, the AP effect is yet another source of anisotropy in survey catalogues. Section 1.5.3 describes how these distortions can be measured if some standard shape can be inferred. Often, these measurements are quoted in terms of $\frac{D_M}{D_H}$.

1.5 Observational cosmology

Modern cosmology, and the abridged timeline presented in Section 1.2 is anchored by a series of measurements that correspond to different epochs of the Universe, but whose physics contains information about the expansion of the Universe and the co-evolution of its energy densities.

1.5.1 Big Bang Nucleosynthesis (BBN)

Immediately following the Big Bang and up until $1s$ after the Big Bang, the Universe was filled with a sea of hot particles. Baryogensis led to the primacy of matter over anti-matter. The Universe's expansion helped cool this primordial soup of particles, until only neutrons, protons, electrons, neutrinos and their antiparticle counterparts remained in thermal equilibrium. At $t = 1s$, the temperature of the Universe was around $k_B T \approx 1$ MeV, leading to neutrinos and antineutrinos decoupling [91, 17, 168]. The Universe is populated with neutrinos, photons, protons and neutrons. It is under these conditions that nuclear reactions between the latter two particles occur. Around $k_B T = 0.06$ MeV, deuterium can synthesize from a proton-neutron reaction. This began a runaway process where Tritium, He^3, He^4 and Li^7 are all rapidly formed. This was a short lived process.

Once the temperature dropped below $T \approx 30$ keV, the production of these nuclides could not be sustained. Ultimately, BBN predicts the relative baryon abundances to be 75% H, 24% He, 0.01% D and trace quantities of Li. The exact abundances are predicated on the energy density of baryons prior to nucleosynthesis $\Omega_{\rm b}h^2$.

BBN observations rely on measurements of the abundances of the aforementioned nuclides (He^3, He^4, Deuterium, Li^7). He^4 is measured through studying the emission lines of highly ionized gas in low metallicity dwarf galaxies, i.e. extragalactic HII regions. He^3 is harder to measure, but is detected in local HII regions. Li^7 is another isotope that is utilized in calculating the primordial photon to baryon ratio, and its isotope can be observed in the atmospheres of population II (low metallicity) stars in the Milky Way. The most crucial abundance to detect is the Deuterium/Hydrogen ratio, as the Deuterium formed in stars is immediately fused into Helium, while primordial Deuterium comes from BBN. Primordial Deuterium is detected by measuring the absorption lines of high redshift quasar systems, where the quasar's light is absorbed by gas clouds in the foreground. Measurements by [56] have yielded the most precise constraints, where they measure $\Omega_{\rm b} = 0.02156 \pm 0.0002$ [56].

1.5.2 Cosmic Microwave Background

Recombination and decoupling resulted in the Universe becoming optically thin to photons, minimizing their probability of interacting or ionizing [172, 168]. These photons are the cosmic microwave background radiation (CMB), and contain a wealth of information about the early universe, the photon-baryon fluid they had decoupled from and the evolving state of the Universe as they traversed it. The field's spectral distribution is characterized by a blackbody function, centered at a temperature of $T_{CMB} = 2.7255 \pm 0.0006K$ [85]. This measurement is foundational in Big Bang cosmology. Measurements of the angular correlations in the temperature field as well as its polarization allow us to link the observed anisotropies in the CMB to cosmological parameters. These measurements contain information about the baryon energy density, the curvature of the Universe, the energy density of matter, and, crucially, can be linked to theories of inflation. Measurements of the secondary anisotropies have allowed us to characterize the equation of state for dark energy.

CMB temperature fluctuations

The cosmic microwave background consists of a field of photons from the surface of last scattering. These photons have been measured to come from all directions across the sky,

but it is their temperature fluctuations that hold a wealth of information. The temperature map is binned into angles $T(\theta, \phi)$, and its anisotropy is decomposed using spherical harmonics:

$$\delta T(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} a_{lm}^{T} Y_{lm}(\theta, \phi) \tag{1.71}$$

In particular, the modes of $l = 0$ (monopole) and $l = 1$ (dipole) are of less interest, as the monopole corresponds to the temperature map across the sky from blackbody radiation with $T = 2.7255K$, and the dipole introduces anisotropy from the Doppler shift of our frame of reference with respect to the cosmological rest frame. Removing these two features from the temperature map and masking out foreground sources of microwave radiation (i.e. the galactic plane and other nearby galaxies) produces the CMB temperature map. The observed fluctuations in this map are of the order 10^{-5}, corresponding to a remarkably small anisotropy. The physics of these anisotropies lie in the early quantum fluctuations in the instants after the Big Bang, as discussed in Section 1.2.1.

With the assumptions of statistical isotropy and homogeneity, the temperature-temperature power spectrum can be computed in terms of the spherical harmonic modes l, m. Specifically the sample variance for the multipole l, C_l^{TT} is computed:

$$C_l^{TT} = \left\langle a_{lm}^{T} a_{lm}^{*} \right\rangle = \frac{1}{2l+1} \sum_{m=-l}^{+l} |a_{lm}|^2 \tag{1.72}$$

And the power spectrum is defined as:

$$D_l^{TT} = \frac{l(l+1)}{2\pi} C_l \tag{1.73}$$

This latter equation describes the angular contribution per mode l to the temperature perturbations δT. In Figure 1.3, the upper portion of the figure compares D_l^{TT} with respect to the angular modes l. At lower l (larger angular scales), the power spectrum is flat, with an upturn around $l > 10$. At scales $100 < l < 1000$, the power spectrum is dominated by oscillations, which at $l \geq 1000$ are progressively damped. Each of these features reflects the presence of either primary anisotropies in the CMB, anisotropies that were extant on the surface of the last scattering, or secondary anisotropies, which affected the CMB photons after decoupling and prior to their observation.

Primary anisotropies

The peaks seen from $100 < l < 1000$ are relics of the baryon acoustic oscillation waves, with different peaks corresponding to different modes of compression (odd peaks) and extension (even peaks). The positions of these peaks are determined by the curvature of the universe, which informs the angular scale θ_s of the sound horizon. The main (first) peak can be measured to obtain constraints on $\Omega_{m,0}, \Omega_\Lambda$ and to constrain H_0 through forward modelling. The first of these peaks is also referred to as *the* BAO peak. The amplitude of subsequent peaks and their locations is determined by the ratio of baryons to photons, in turn permitting constraints on $\Omega_b h^2 = \omega_b$ and $\Omega_c h^2 = \omega_{\rm cdm}$ to be obtained. At higher l, thermal diffusion in the photon-baryon fluid suppresses the anisotropies at smaller angles, damping the oscillations in Figure 1.3.

Furthermore, at scales of $l < 100$, the temperature-temperature power spectrum is defined by a plateau in Figure 1.3. This is the result of a primary anisotropy called the Sachs-Wolfe effect where CMB photons, during decoupling, are gravitationally redshifted by the primordial density fluctuations [201].

The power spectrum of the polarization modes of the CMB also contain a wealth of information. Because the CMB contains anisotropies, the outgoing radiation field is linearly polarized. For instance, in the matter power spectrum, the location of the peak in Figure 1.2 corresponds to the Fourier mode k_{eq}, which refers to a characteristic scale of the Universe during matter-radiation equality. This turnover corresponds to which scales entered the cosmic horizon prior to this epoch, where radiation pressure suppressed the growth of structure, and which scales entered following. The coupling of matter and photons prior to decoupling allows for A_s and n_s to be estimated using the polarization E-modes of the CMB [19, 135].

In terms of measurements, the most precise measurements of the CMB power spectrum come from the Planck satellite, reporting the following cosmological constraints: $\Omega_{\rm b} h^2 = 0.0224 \pm 0.0001$, $\Omega_{\rm cdm} h^2 = 0.120 \pm 0.001$ [11]. For the primordial power spectrum, they report $\ln(10^{10} A_s) = 3.044 \pm 0.014$ and $n_s = 0.9649 \pm 0.0042$ [19]. If a ΛCDM model is adopted, then it also obtains: $\Omega_{\rm m} = 0.315 \pm 0.007$, $\sigma_8 = 0.811 \pm 0.006$, $H_0 = 67.4 \pm 0.5$ km/s/Mpc. Finally, if these CMB measurements are combined with SNIa and BAO measurements, then $w_0 = -0.957 \pm 0.086$, $w_a = -0.295^{+0.32}_{-0.26}$ for the dark energy equation of state parameters [11].

Secondary Anisotropies

Secondary anisotropies of the CMB arise from the photon interactions with intervening matter after last scattering. One such secondary anisotropy to the CMB arises from how CMB photons travelling through large gravitational potentials that are varying with time, namely through dark energy accelerating the expansion of the Universe [201]. This is the late-time or integrated Sachs-Wolfe effect (ISW), as it describes the contributions to the shift in the spectra of the CMB photons due to the gravitational redshift of evolving gravitational potentials. This observable links the information from high redshifts (CMB) to low redshifts (large-scale structure). The power of this observable lies in its direct sensitivity to dark energy and models of modified gravity.

Photons travelling from the surface of last scattering through large gravitational wells are gravitationally blue-shifted as they enter a potential well, and gravitationally red-shifted as they exit. The integral calculates the accumulated redshift and the associated change in the photon's temperature. Crucially, the cosmic gravitational potential does not alter with time if the Universe is spatially flat and has no dark energy. Conversely, it gains a time-dependence if either of the above are present. Consequently, if there is a significant ISW, then this introduces a significant secondary anisotropy into the CMB. Its effect is apparent in Figure 1.3 at very small $l < 10$, tilting the temperature-temperature power spectrum, however the uncertainties are large. This intersection of large-scale structure and CMB cosmology is crucial for establishing the ΛCDM model. Refer to: [94, 122, 227, 120, 93] for a more detailed exploration and application of this curious observable.

This is not the only interaction of CMB photons and the large scale structure, as hot intracluster gas can also scatter the CMB photons. Scattering caused by the gas interacting with the CMB photons can alter the peak of the CMB blackbody. This is the kinetic Sunyaev-Zeldovich effect (kSZ) [218]. Another secondary anisotropy comes from a similar scattering process, where high temperature X-ray gas filled with charged particles can scatter the CMB photons, distorting the CMB blackbody. This is the temperature Sunyaev-Zeldovich effect (tSZ) [219, 26]. All of these secondary anisotropies can be combined with the primary ones to obtain more information about the large-scale structure of the Universe.

1.5.3 Large-Scale Structure

Following matter-radiation equality, the small quantum fluctuations in the matter field can begin to grow, first as dark matter gravitates to overdense regions, and later with baryons,

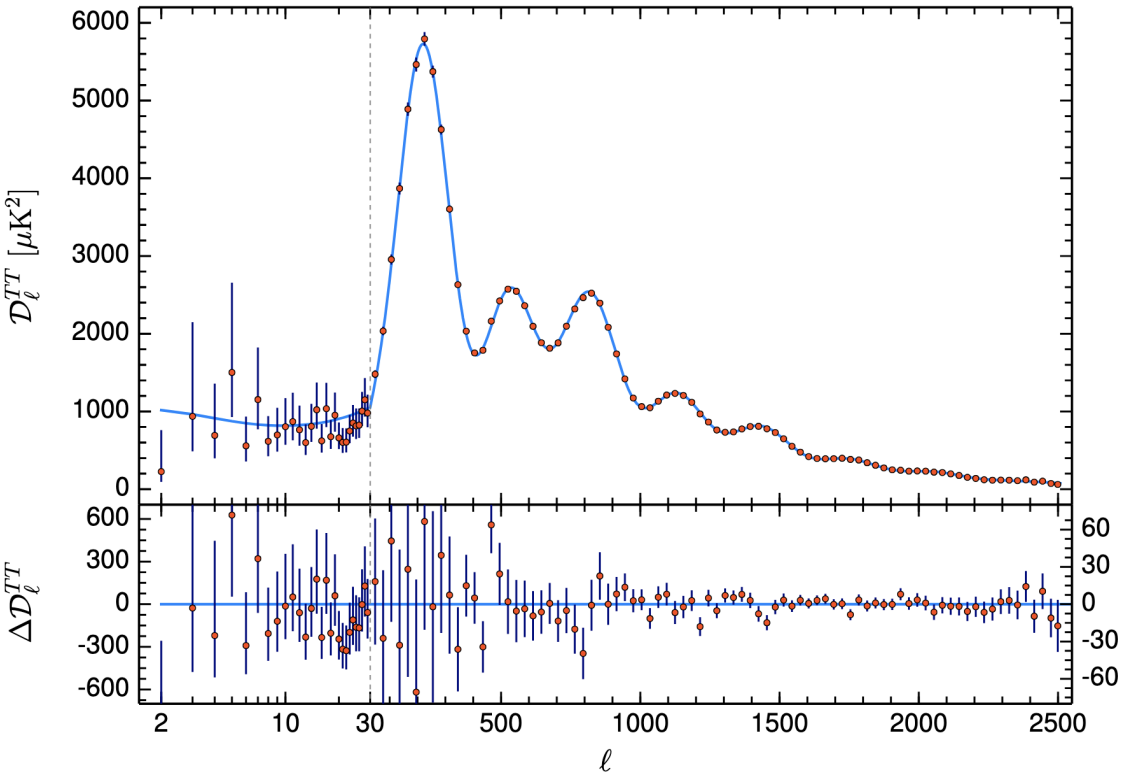

Figure 1.3: The power spectrum of the CMB temperature anisotropy, as measured by the Planck collaboration [11], with the line of best fit representing a ΛCDM cosmology.

following their decoupling from photons. The growth of these structures is traced primarily through measuring the density field via galaxies. Through measuring their position (Right Ascension - RA, Declination - Dec) on the sky, and their redshift, and upon selecting a fiducial cosmology, these galaxy redshifts can be used to construct a catalogue of their positions. We have established (Section 1.4.4) galaxies are not precise tracers of the underlying matter density field. While most dark matter halos that grew from these small fluctuations do accumulate clusters of galaxies, their distribution is inhomogeneous. This means that galaxies are a biased tracer of the matter field. This can be resolved through modelling of the galaxy-halo connection, which aims to bridge the distribution of directly observable baryons to the inferred distribution of dark matter.

With a suitable galaxy-halo connection model and fiducial cosmology, then any cata-

logue of angular coordinates (RA, Dec) and redshifts (z) can be converted into comoving distances for galaxies across the sky. This enables us to use the primary tool for measuring large-scale structure, the two-point galaxy correlation function discussed in Section 1.4.4. Through measuring the two-point galaxy correlation function, the physics of large-scale structure and the environments of galaxies can be linked together.

Measurements of Redshift-Space Distortions

As previously discussed, one significant observational effect is that of the local velocity fields in galaxy clusters. Galaxy clusters have large gravitational wells and can greatly modulate the spectroscopic measurement of a galaxy's redshift relative to us, impacting the inferred distance used to build these galaxy catalogues. This scale-dependent effect, known as redshift-space distortions (RSD) greatly alters the observed density field. See Figure 1.4 for an illustration of the influence of RSD on observing large objects in redshift-space .

These redshift-space distortions can be seen at large scales (e.g. first two rows of Figure 1.4) and smaller scales (bottom row of Figure 1.4). In the linear (Kaiser) [131] regime, redshift-space distortions can be described to the first order (assuming a plane-parallel approximation) as described in Section 2.4.1. In measuring the clustering of galaxies, strong constraints on the linear growth rate of structure, $f(z)$, and the growth rate combined with the scale of fluctuations in the matter field, σ_8 are obtained [175]. $f\sigma_8$ is a crucial observable that measures the redshift-space distortions found in galaxy clustering.

The latest measurements come from the DESI peculiar velocity survey, which for DR1 reports $f\sigma_8 = 0.483^{+0.080}_{-0.043}$ through measuring over 76,000 peculiar velocities [140]. Measuring RSD through $f\sigma_8$ also serves as a test for modified theories of gravity [185].

Measuring the Alcock-Paczynski Effect

The AP effect is caused by a mismatch between the fiducial cosmology of a survey and the true cosmology of the Universe Equation (1.70). This effect has been detected by identifying and measuring *standard shapes* in cosmology, objects that are expected to be spherical, where a distinct measurement of the transverse and line-of-sight comoving distances can be obtained.

Distinguishing between RSD and AP stretching can be a non-trivial task, in particular for some void-galaxy models and BAO measurements, as Figure 1.5 illustrates. The dimensionless parameters in Equation (1.70) can be obtained from using an object, such

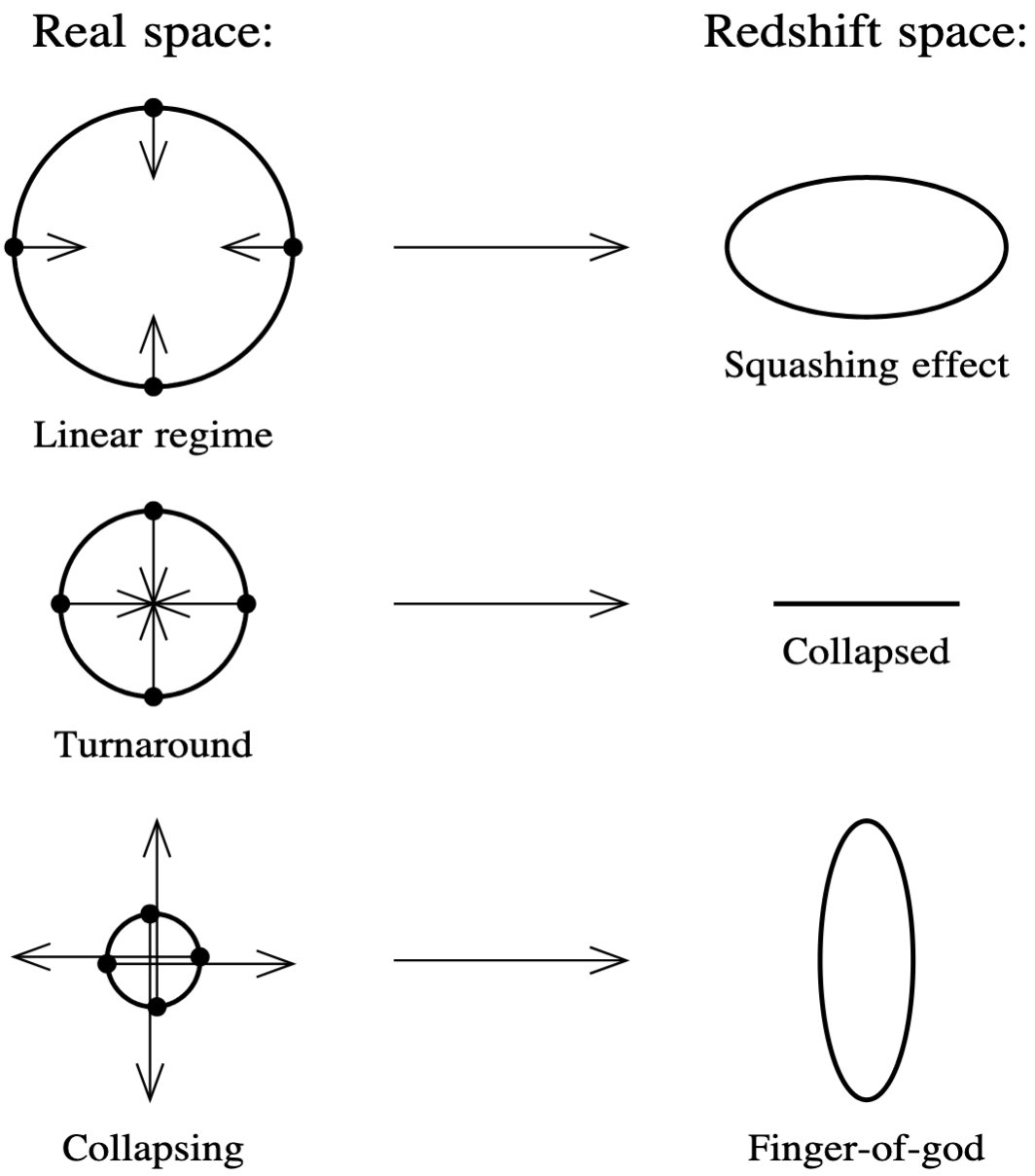

Figure 1.4: Image courtesy of [118]. The left column illustrates the positions of galaxies in real-space, the right column in redshift-space, illustrating how the effect of RSD is preferential along the line-of-sight (along the vertical axis).

as the BAO peak or the void-galaxy monopole, both of which are *standard shapes*, objects that are predicted to be physically spherical, but distorted by the AP parameters. In BAO analyses, these distances can be disentangled using r_d as a standard ruler [12, 4, 6].

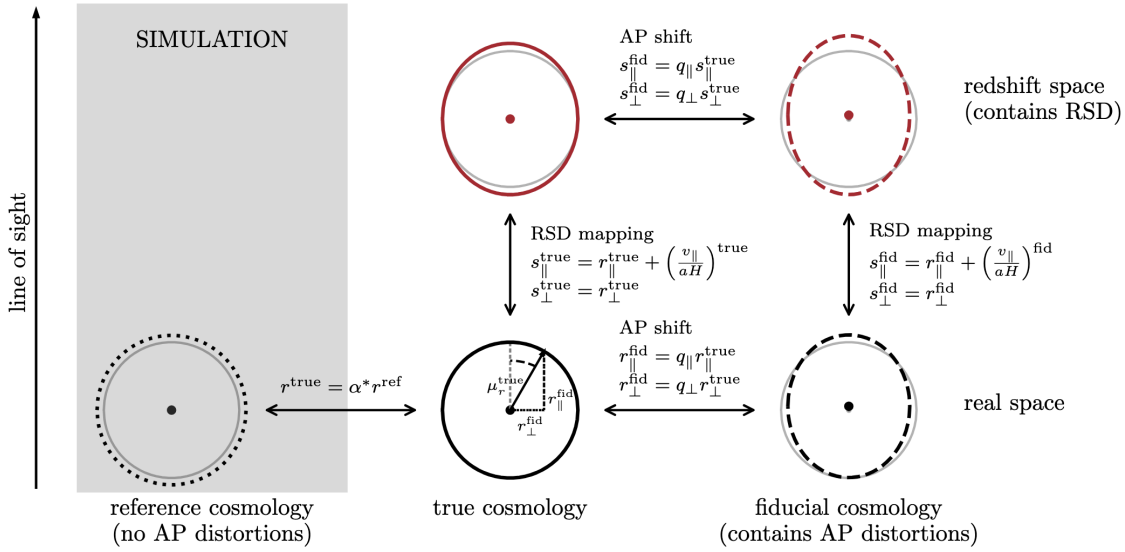

Figure 1.5: Figure courtesy of [187]. The cumulative effects of AP distortions and RSD, shown in *real* space and *redshift* space are outlined here. On the left column, a spherical object is inscribed in *real*-space, with the lower row illustrating how the shape of the object is altered along and transverse to the line-of-sight through choosing the true and fiducial cosmologies. Along the top row, the effect of RSD is applied to the positions in this spherical object, creating a line-of-sight distortion in *redshift* space.

Baryon Acoustic Oscillations

Recalling the conditions of the Universe prior to recombination and decoupling, the Universe was composed of a hot baryon-photon plasma, rendering it optically thick. Small quantum fluctuations in density had seeded the growth of anisotropies in higher density (colder) regions. This fluid was balanced by gravitational forces and photon radiation pressure, which for a time led to the Baryon Acoustic Oscillations (BAOs) discussed previously. Between recombination and decoupling, dark matter had accumulated in these overdense

perturbations, but their growth was suppressed. The photon-baryon fluid accumulated in density peaks, separated by the scale of the sound horizon of these waves. The scale of this horizon is determined by the balance of photons to baryons. When the photons became optically thin, the photon-radiation pressure vanishes, and the baryons were deposited at these acoustic peaks, *freezing in* the distribution. Baryons and cold-dark matter gravitated, allowing the density perturbations to grow into galaxies and later galaxy clusters. As the Universe expanded, the physical separation between these peaks expanded, and at late times, the comoving scale of the sound horizon is estimated to be $r_d \approx 150h^{-1}\text{Mpc}$.

$$r_d = \int_z^\infty c_s \frac{\text{dz}'}{H(z)} \tag{1.74}$$

The scale of this sound horizon, and thus the scale of the BAO peak and its amplitude can be inferred. The large-scale structure of the Universe is seeded by a field of perturbations upon which these acoustic waves are imprinted.

The BAO is obtained through measuring the clustering of galaxies. The BAO is detected because there is an excess probability of detecting pairs of galaxies at a separation scale that directly corresponds to the scale of the BAO sound horizon. Detecting the separation scale, s of this BAO peak in the two-point galaxy correlation function, and computing $D_M(z)/r_d$, $D_H(z)/r_d$ allows for us to obtain a precise distance-redshift relation. The Alcock-Paczynski effect can be exploited, as even though the BAO feature is measured in redshift-space, with only a fiducial cosmological model, one can obtain a precise constraint on the scale of the BAO feature, and how much the AP distortions alter it. Doing so then means these precise distance measurements, taken from multiple redshift bins, can be connected to a cosmological model - consequently allowing for the estimation of best fitting parameters for $\Omega_{\text{cdm}}, w_0, w_a$. This method of converting a galaxy catalogue to a correlation function to cosmological parameter constraints is the blueprint for the project encompassed by this thesis.

The latest measurement of galaxy clustering comes from the BAO analysis of the second data release (DR2) of the DESI survey, where they obtained for ΛCDM, $\Omega_{\text{m}} = 0.297 \pm 0.0086$, $H_0 = 67.14 \pm 0.47$ km/s/Mpc [6]. If the dark energy equation of state is allowed to vary, then incorporating CMB constraints from Planck narrows parameter estimates for w_0 and w_a, going from $w_0 = -0.48^{+0.35}_{-0.17}, w_a < -1.34$ to $w_0 = -0.47 \pm 0.21$ and $w_a = -1.75 \pm 0.58$ [6].

Gravitational Lensing

General relativity linked matter with the curvature of spacetime. Photons passing through the gravitational wells of large galaxy clusters will have their trajectories altered. Strong lensing refers to the case where the gravitational influence on the photon trajectories is so large that the light is significantly bent, and multiple images of the same object can observed. This magnifies the background source, distorting its shape to the observer.

Weak lensing is far more ubiquitous, and refers to the case where the deflection is far less pronounced. It describes the distortion of light emitted from background galaxies, and the matter field in the foreground causes the observed distortion. It lacks the hallmarks of multiple images of magnification that strong lensing has, but it can be used to ascertain the underlying matter distribution much like strong lensing. Through calculating the cosmic shear in a catalogue of galaxies, or shear power spectrum, the matter energy density $\Omega_{\rm m}$ can be obtained, as well as the amplitude of matter density fluctuations at scales of 8 h^{-1}Mpc which is called σ_8 and the dark energy equation of state parameters, w_0, w_a. Often, lensing surveys report σ_8 and $\Omega_{\rm m}$ as the degenerate parameter S_8.

$$S_8 = \left(\frac{\Omega_{\rm m}}{0.3}\right)^{0.5} \sigma_8 \tag{1.75}$$

The best constraints from weak lensing, using the Dark Energy Survey's year 6 measurement of 3x2 galaxy weak lensing measure $\Omega_{\rm m} = 0.333^{+0.023}_{-0.028}$, $S_8 = 0.789 \pm 0.012$. An analysis of the year 5 data yields cosmological constraints of $\Omega_{\rm m} = 0.263^{+0.020}_{-0.025}$ when combining the constraints the DESI BAO, Planck sound horizon measurements and BBN. In combining the DES data with the DESI BAO, the Planck CMB, and the Pantheon supernovae yields a constraint on the dark energy parameters of $w_0 = -0.673^{+0.098}_{-0.097}$ and $w_a = -1.37^{+0.51}_{-0.80}$.

1.5.4 Type Ia Supernovae (SNIa)

A precise calibration of the distance-redshift relation can provide cosmological information. Standard candles, objects with a well understood luminosity profile are used to obtain this relation for increasingly higher redshifts. Type Ia supernovae occur (SNIa) when a white dwarf accretes enough matter to exceed the Chandrasekhar limit of $1.4M_\odot$, causing the gravity to overpower the white dwarf's electron degeneracy pressure, triggering the supernova. This explosion is bright enough to be seen in distant galaxies. Measuring

the peak luminosity profile, and the object's redshift allows us to obtain a corresponding luminosity distance for these type Ia supernovae.

The distance ladder is used to calibrate cosmological distance measurements. The upper rungs of the distance ladder are constructed by using *standard candles*, objects with well-defined luminosity profiles to precisely constrain distances. Obtaining the parallaxes of local Cepheids to calibrate the Cepheid period-luminosity relation. Subsequently, SNIa measurements can be anchored to the distance ladder by finding Cepheid variables in the same galaxies hosting SNIa to calibrate SNIa distances. Measurements of the SNIa were one of the first pieces of observational evidence for the accelerating expansion of the Universe, also known as dark energy. The best measurements come from Pantheon Plus, where they obtained: $\Omega_{\rm m} = 0.334 \pm 0.018$, $w_0 = -0.90 \pm 0.14$, $w_a = -0.1^{+0.9}_{-2.0}$ [32, 208].

A precise and direct measurement of the expansion rate at present times, H_0 is then obtained by using the SNIa distance measurements. This direct measurement of H_0 is currently in tension with the other probes of the Hubble constant, a fact we shall now discuss.

1.6 Statistical tensions

Improvements in galaxy-redshift surveys, the distance ladder and cosmic microwave background measurements have opened a new frontier into large-scale structure cosmology. The marked improvements in the precision and quality of measurements have led to increasingly precise determinations of the age and composition of the Universe. Increasingly precise measurements of these cosmological observables across the last 3 decades have helped establish ΛCDM as the standard model of cosmology. These improvements have come with a price. With the arrival of subpercent measurements, statistical tensions have emerged between early-time measurements of cosmological parameters (as derived from the CMB or using CMB datasets) and late-time constraints on cosmological parameters (as measured from lensing and SNIa). Over the past two decades these discrepancies, known as *statistical tensions*, have grown from a minor curiosity to one that merits serious attention. These tensions will be outlined below.

1.6.1 H_0 tension

Measurements of the anisotropies in the CMB gives us an estimate of the ΛCDM model's parameters. This early time observation contains a wealth of information, and the con-

straints obtained from the CMB in terms of $\Omega_{\rm m}, \Omega_{\rm b}, N_{\rm eff}, N_{ur}, n_s$ can be forward modelled into a present day estimate of late-time cosmological parameters, such as the Hubble constant, H_0. The latest SNIa distance ladder measurements (SH0ES) measured $H_0 = 73.04 \pm 1.04$ km/s/Mpc [194], while CMB measurements from the Planck satellite gave an estimate of $H_0 = 67.4 \pm 0.5$ km/s/Mpc [11]. This amounts to a $> 4.5\sigma$ tension between the distance ladder measurement (SH0ES) and the CMB projection for H_0. Improvements in galaxy-redshift surveys have meant large-scale structure measurements of the BAO have provided another avenue to explore this tension. The latest constraints from BAO in DR2 of the DESI survey, combined with the CMB gives a measurement of $H_0 = 68.17 \pm 0.28$ km/s/Mpc [3]. Many modifications to cosmological models have been proposed. See [72] and Figure 1.6 for an overview.

1.6.2 S_8 tension

The S_8 parameter measures the homogeneity of overdensities in the matter field. S_8 describes the variance in density fluctuations in a sphere with a radius of 8 h^{-1}Mpc, the scale of galaxy clusters, scaled by the present day density of matter. This parameter can be characterized by measurements in the CMB (early times) $S_{8Planck} = 0.834 \pm 0.016$ [11], and through a slew of measurements at late times, namely weak lensing $S_{8DESY6} = 0.789 \pm 0.012$ [2], modelling of RSD and measurements of local peculiar velocity fields $S_8 = 0.776 \pm 0.033$ [31]. For CMB and weak-lensing measurements of S_8, the current tension amounts to over 3σ.

This tension is present alongside the aforementioned H_0 tension, and presents a significant challenge to any competitor to ΛCDM, as many proposed changes that would resolve the tension in H_0 often increase the tension in S_8 or vice-versa, (e.g.[18, 126]). Both of these tensions point to a significant disagreement in between late- and early-time measurements of the Universe's growth and the variance in its overdensity field.

1.6.3 Dark Energy Equation of State

The growing precision of galaxy-redshift surveys has been made clear through its increasing contribution to existing statistical tensions, as well as identifying new ones. Following the first data release from DESI, measurements of the BAO, especially when coupled with information from the CMB or even distance ladder measurements have pointed to a new tension - the dark energy equation of state parameters. Recall that for Λ dark energy, $w_0 = -1$, $w_a = 0$ applied to Equation 1.20. DR1 from the DESI BAO gives a constraint

on $w_0 = -0.55^{+0.39}_{-0.21}$ and $w_a < -1.32$. Adding the Pantheon Plus SNIa dataset to DESI DR1 BAO gives tighter constraints, $w_0 = -0.831^{+0.066}_{-0.066}$ and $w_a = -0.73^{+0.32}_{-0.28}$. This gives rise to a mild discrepancy above 2σ. Adding in constraints from DES Y5 and the Planck CMB increases the tension, up to 3.9σ. The final analysis of DES Y5 data obtains a constraint for the dark energy parameters to be $w_0 = -0.673^{+0.098}_{-0.097}$ and $w_a = -1.37^{+0.51}_{-0.80}$, amounting to a 3.2σ tension with the fiducial Λ dark energy model and illustrates the growing tension that is seen between late and early time estimates of cosmological parameters. The equation of state of dark energy may not be constant as prescribed by the ΛCDM model, but may instead evolve as the Universe evolves. Adopting a model of evolving dark energy is not a solution to the H_0 and S_8 tensions. Rather, it stresses the value of high-precision measurements of the large-scale structure of the Universe, and the role it can play in either uncovering a looming systematics error, or in uncovering a crucial detail in cosmic history. This is a frontier with enormous potential, and one that is being actively explored.

1.7 Obtaining more from large-scale structure

This statistical tension between early- and late-time measurements of the Universe paints an exciting picture that the ΛCDM model might be missing a few elements from the Universe's chronology, leaving room for new physics. In order to diagnose this tension, it is crucial to make full use of the available probes for large-scale structure. Galaxy-redshift surveys are a strong avenue to pursue in order to characterise these growing tensions - the ongoing DESI, Euclid and LSST surveys and the upcoming Roman survey promise to bring increasingly precise measurements of the two-point galaxy-correlation function and subsequently, the BAO through drastically increasing the volume of the survey, obtaining millions of redshifts. As discussed in Section 1.4, the two-point galaxy-correlation function contains a wealth of information, but at late-times, more information can be accessed by exploiting the pronounced inhomogeneities of galaxy clusters. Cosmologists have reached for a variety of techniques to extract more information from the cosmic web. These include higher-order statistics, which encompass the galaxy bispectrum (three-point galaxy correlation function), which has been used to explore problems relating to primordial non-Gaussianity and the four-point galaxy correlation function, which has been used to explore parity violation. A wealth of techniques can also be utilized to extract more information from the existing galaxy catalogue, exploring cross-correlations between galaxies and other components of the density field - i.e. using as much information as possible from an existing catalogue. This thesis will outline and explore one of these techniques - using low density regions (voids) in tandem with the galaxies in extant catalogues to obtain more

information from existing galaxy-redshift survey data.

1.8 Thesis Outline

This thesis will first discuss the theoretical background behind cosmic voids, their identification and the information imparted by measuring the void-galaxy cross-correlation, offering a curated review of cosmic voids and their role in the large-scale structure cosmology (Chapter 2). In particular, we will then discuss how cosmic voids have been shown to offer complementary constraining power through AP and RSD measurements. Particular emphasis will be placed on the techniques used to measure the void-galaxy cross-correlation function, and how an existing set of models is challenged by the needs of next generation surveys. We will then, in the same chapter, talk about how these challenges can be addressed, and in the next chapter, how a machine learning model had to be developed to overcome the challenges presented by applying a void analysis to galaxy catalogues (Chapter 3).

In Chapter 4 we will discuss how this neural network was trained on measurements of the void-galaxy cross-correlation function, our tests on a previous generation galaxy-redshift survey catalogue (BOSS) and comparing it with precedent techniques, putting special emphasis on how this neural network emulator has drastically altered the considerations for a void-galaxy clustering analysis, and allowed the parameter dependence of this quantity to be explored. This chapter encapsulates the work of the paper [86].

In Chapter 5, we will describe how this emulator framework was extended to the DESI survey, and incorporated several key adjustments that arose from the development and testing of this emulator, namely through matching the number density and footprint of the survey for a large span of cosmological and galaxy-halo parameters.

In Chapter 6 we will summarize the conclusions of this work, the future avenues for voids as a tracer for the large-scale structure, and future considerations for the projects discussed in Chapter 4 and Chapter 5.

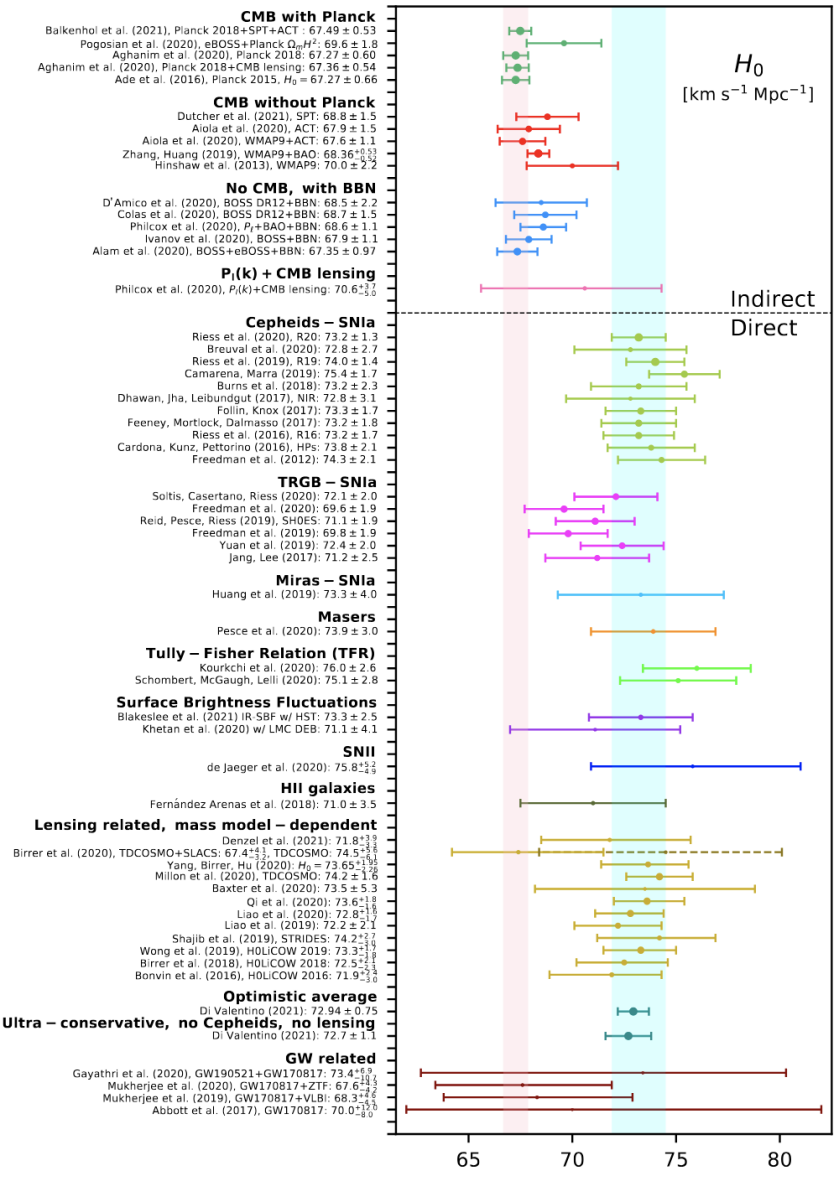

Figure 1.6: Figure courtesy of [72], which summarized the latest measurements of the Hubble constant, using the methods described in this section. The shaded blue region corresponds to the Cepheid+SNIa distance ladder measurements of H_0 in [193], the shaded red region corresponds to the Planck CMB estimate for H_0 under a flat ΛCDM cosmology in [11].

Chapter 2

Voids and VG-CCF modelling

2.1 Introduction

At late-times, the initial perturbations we discussed in Chapter 1 have grown, forming a complicated structure populated with clusters of galaxies, sheets of gas and baryonic matter, and empty spaces known as cosmic voids. This collection of clusters, sheets, voids and filaments is the cosmic web, a vast and complex structure which has been successfully measured with galaxy-redshift surveys [61, 130, 12, 4]. With successive generations of galaxy-redshift surveys, more galaxies and larger volumes of the sky have been observed. As alluded to in Chapter 1, the cosmological information obtained from measuring the clustering of galaxies is very informative. Measuring this galaxy clustering has led to strong constraints on cosmological parameters.

However, at late-times, the early-time density perturbations have made the Universe anisotropic and inhomogenous. There is a coherence in modes, k, such that the information in the amplitude of different modes is moved into the structure of mode correlations. Gravity has had billions of years to collapse the small fluctuations into structures, and the growth of large-scale structure has also influenced the subsequent collapse of smaller structures. This limits the constraining power of the galaxy two-point correlation function, and has led to the adoption of other summary statistics (i.e. higher-order correlation functions [89, 90, 169, 178]) or two-point statistics to trace different parts of the cosmic web as a way to extract more information. Additionally, measuring the geometry of the projected distance-to-redshift relationship instead of the fluctuations can be another source of cosmological information (e.g. [163, 217, 214, 233]).

Cosmic voids are regions of space that are underdense, and are $20-100\ h^{-1}$Mpc in scale. These regions were first observed in early galaxy-redshift surveys [226, 98, 129]. Cosmic voids and galaxy clusters were both shaped by the linear growth rate of structure, $f(z)$ during the matter-dominated era of the universe. Compared to galaxy clusters, the growth of voids can be easily modelled because of a lack of complex dynamics at late-times. This fact has made them a suitable laboratory for a variety cosmological topics, concerning: the tidal dynamics of cosmic voids (e.g. [184]); their lensing of the CMB [36, 42, 43, 35]; their weak lensing with respect to galaxy clusters [151, 21, 65, 30] modelling and observations of the velocity fields around voids and the application of these velocity fields to test GR and modified gravity (e.g. [54, 246, 24, 176]); the modelling of a void density profile to distinguishing between GR and competing theories of gravity (e.g. [114, 156, 83]); their measured ellipticity as a probe for dark energy [165, 145]; using the void size function in order to constrain the dark energy equation of state (e.g. [27, 47, 57, 54, 55]); as probes of the integrated Sachs-Wolfe effect [29, 34, 136, 74]; calculating their abundances to constrain dark energy [181, 37] and massive neutrinos [137]; and their impact on weak-lensing [48, 203, 100].

Of added interest is that cosmic voids can be used to perform an AP test, which was put forth by Lavaux & Wandelt [144]. Identifying voids in a galaxy redshift survey catalogue and computing the void-galaxy cross-correlation function (VG-CCF) is equivalent to stacking voids and measuring the profile. If the voids were found in real-space, then this stack of voids would be isotropic and spherical. Distortions of this sphere would arise from RSD and the AP effect. With a careful treatment of RSD, the AP test can be performed to obtain cosmological constraints on $\Omega_{\rm m}$, which we discussed in Section 1.5.3. This stack of voids, the VG-CCF, has been used to extract extra information out of the large-scale structure, and has been demonstrated to provide cosmological information that is complementary to the BAO measurement. In Section 2.2, we will discuss how voids are identified in galaxy-redshift survey catalogues. In Section 2.3, we will discuss how the VG-CCF is computed for surveys. In Section 2.4, we will outline how the VG-CCF has been modelled to obtain measurements of the AP distortions and RSD, first using a linear model and later using velocity templates. We then summarize the key results from this body of literature, and what limitations arose in modelling AP and RSD in void-finding, necessitating a new model of the VG-CCF. Chapter 3 will then outline the model we use to meet the challenges identified at the end of this chapter.

2.2 Void-finders

Voids are identifiable as low-density regions in galaxy catalogues, making them relatively straightforward to obtain. There are many different ways of finding voids with various levels of complexity. Here, we will outline two classes of void-finding algorithms that are relevant to this thesis. Our thesis used the `Revolver` void-finding package, which implements the `voxel` and `zobov` void-finding algorithms, which are both implementations of the watershed void-finder.

The literature is brimming with void-finders. Some groups use `voxel`, `zobov` or another watershed void-finder and some others use the spherical void-finding algorithm. Some examples include: `Photo-z` [203], Banerjee & Dalal [22]'s spherical void-finder in `Pylians`; `voidfinder` [82, 124]; `versus` [84]; `Popcorn` [166]; [161]'s spherical void-finder; Void IDentification and Examination toolkit (`VIDE`) [220]; VoronoiVoids (`V2`) [76]; and `delfin` [16]. An overview of some of these void-finders can be found in [49, 41, 162, 75].

We will discuss two major classes of void-finders: spherical void-finders, and watershed void-finders. The latter discussion will focus on the `zobov` and `voxel` algorithm.

2.2.1 Spherical voids

One algorithm lies in characterizing voids as underdense spheres in the cosmic web. The algorithm used in spherical void-finders takes the following generalized steps. First, the local overdensity $\delta = (\rho - \bar{\rho})/\bar{\rho}$ field is calculated, by counting galaxies or dark-matter halos, and estimated by either binning the tracers into volume pixels (voxels) or through two-dimensional angular pixels of the sky. This field is smoothed, using a spline or Gaussian kernel, usually prescribed by some smoothing scale, R_s. Within this smoothed overdensity field, a core-particle/voxel is identified, corresponding to the local minimum overdensity contrast. Within this minimum a sphere is grown around it, until the enclosing spherical volume either exceeds some overdensity threshold (e.g. $\delta > -0.7$) and/or is bounded by a certain number of galaxies on the threshold of a given void. This is repeated until all minima have been enclosed by a spherical volume. Finally, the void catalogue is then finalized by trimming voids based on size R_V or overlap with other spheres, and often sorted by the integrated overdensity of each void.

In following these broad steps, different implementations of spherical void-finders have varying implementations. In each of them, adjustments are made in correspondence to the type of catalogue used, whether photometric or redshift data is utilized (the latter is preferred, as it minimizes projection effects), trimming criterion for voids (e.g. overlap

threshold of different spheres, number of permitted field galaxies, location of minima), and the selection criteria for voids (e.g. number of bounding galaxies, overdensity threshold, size limitations, distance from survey boundaries). These reveal different nuances that arise not only from the spherical void-finding algorithm, but also across implementations.

Taking this further, in some implementations, spherical voids are used as "candidate" voids, whose central particle or voxel may be used as a starting point in assembling non-spherical voids through merging (e.g. [166]), or through iterating over these minima until the void-center does not move across iterations.

2.2.2 Watershed voids

The spherical void-finder has limitations because it is biased towards finding voids in the galaxy catalogue that tend to be more spherical. This might affect the resultant distribution, as identifying voids using a spherical smoothing algorithm, may not optimally find voids of different shapes.

An alternative approach lies in the watershed algorithm, which does not specify a shape. Watershed void-finding algorithms identify voids through low density catchment zones in the large-scale structure [183]. The implementations of this algorithm vary [159], but they operate with the following principles. The galaxy or halo catalogue is used to divide the estimated over-density field into different regions, in a process called tessellation. Each of these regions then has a corresponding mean overdensity. The minima in this overdensity field are identified, and the volumes pertaining to the voids are computed by sequentially moving away from the overdensity minima until it reaches a boundary, defined by the density contrast of the cells in the watershed. These watersheds, therefore, define the underlying catalogue of voids. The number of voids in the resultant catalogue is influenced by whether these watersheds are combined, and how steep the density contrast boundary is.

Each implementation of the watershed algorithm defines the void through a *void-centre*, but even at this stage of post-processing the voids, differences in the merging of watershed regions and the density contrast boundaries will delineate the resultant void catalogue, and their void-centres. Different watershed void-finders obtain different void-catalogues, although they all capture the low-density structure of the universe. These different void-finders have their own strengths for different applications. There are subtle differences between different watershed void-finders, especially between the `zobov` and `voxel` void-finders, but the overall principle is the same - voids are assumed as non-spherical, irregular features of the overdensity field, with a *void-centre* that varies across implementation. See

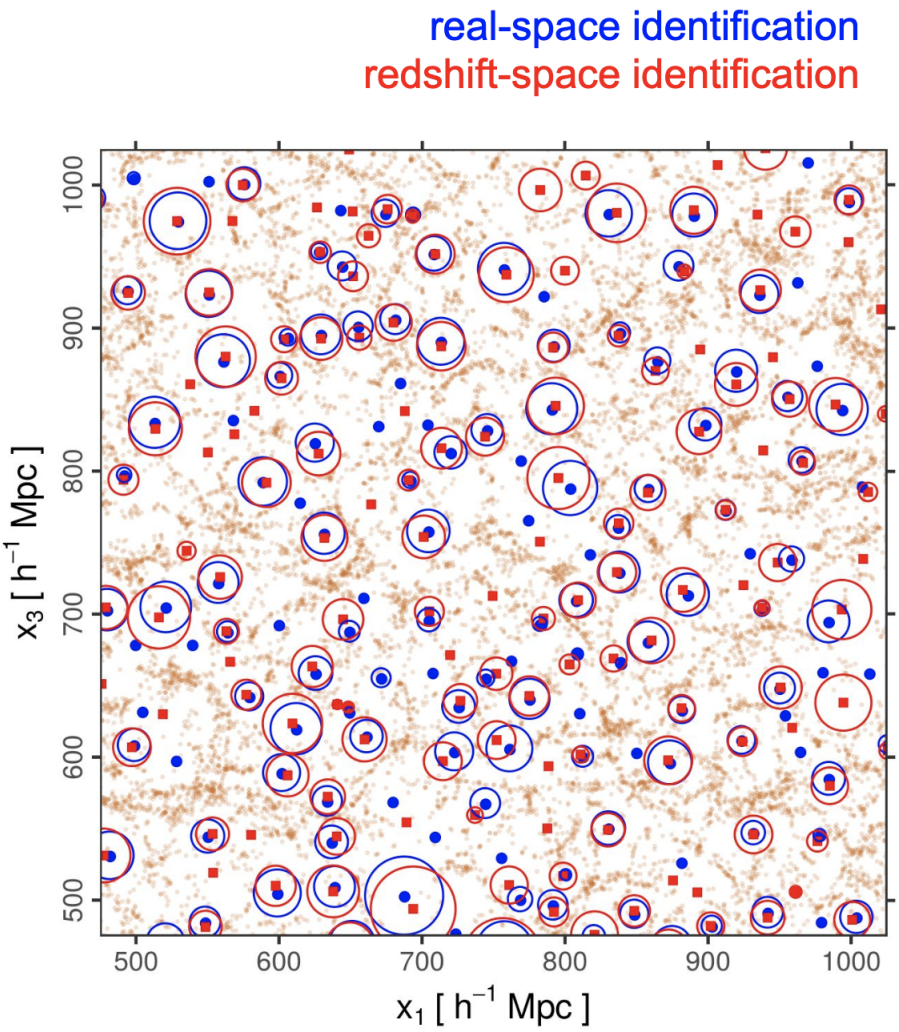

Figure 2.1: Figure courtesy of [58], wherein they obtained voids using a spherical void-finder in the Millenium XXL simulations. Here, they used `void-finder` to obtain voids in a simulation box, with a slice thickness of 10 h^{-1}Mpc, using the real-space coordinates (blue) and the redshift-space coordinates, as estimated using Equations 4.3,4.4. x_3 is chosen as the line-of-sight direction.

[49, 41, 162], for a discussion of the subtleties arising from varying void-centre definitions. We will focus our discussion on the subtleties between `zobov` and `voxel` watershed void-finders, and illustrate the differences in their application in the literature.

ZOBOV

ZOnes Bordering On Voidness (`zobov`) is an implementation of the watershed algorithm [159]. `zobov`'s algorithm is defined by trying to minimize the number of restrictions in finding voids from a given set of points.

The overdensity field is estimated by using a Vornoi Tessellation Field Estimator (VTFE). Voronoi tesselation is a process by which a distribution of points is divided into volumes, proportionate to the local distribution of neighbouring points. Each Voronoi cell is defined as the volume within the catalogue that is closest to a given particle as opposed to any other particle. The density is then estimated using these Voronoi volumes, $1/V(i)$. Thus, a catalogue of galaxies or dark matter particles can be converted into a catalogue of Voronoi cells, whose volume and density is defined by their relative proximity to each other.

These minima are used in the watershed stage as the starting points. Each minima's cell starts by adding the next deepest neighbouring cell, combining neighbouring cells sequentially. So long as the minimum density of this collection of cells $\rho_{\rm min}$ is less than the density of the next cell ρ_l, the watershed will continue to grow. The growth will stop if $\rho_{\rm min} > \rho_l$, i.e. the growth causes the watershed to spill into another region. Once this stopping criterion has been reached for all neighbouring cells in a given collection of cells, this region is marked as a void. Naturally, the deepest of all minima will keep growing to encompass the entire volume, this is the *raw* version of the `zobov` void-finding algorithm.

`zobov` voids are post-processed by using a trimming condition that is imposed based on the density contrast of a given void and neighbouring voids in a catalogue. First, the density contrast is computed by taking the minimum void-density of a given void, $\rho_{\rm min}$ and comparing it with its neighbouring void density, specifically, we look at the minimum particle on a ridge beyond which holds a deeper void. The deeper the neighbouring void, the higher this ratio between ρ_l and $\rho_{\rm min}$. The distribution of these neighbouring density contrasts is modelled as a Poisson-Voronoi distribution, $P(r)$. $P(r)$ is an estimate of the likelihood of a void with a given density contrast, $r = \frac{\rho_l}{\rho_{\rm min}}$ arising from Poisson noise. The deeper this contrast, the larger the spillover between watersheds. The subvoids found in the watershed stage, are therefore split by computing the density contrast for these regions and selecting volumes that have a sufficiently high contrast ratio, this is often set

to $r = 2 - 3$ [159], as higher contrasts imply that the void is significantly more likely to not be a product of Poisson noise, and therefore a true void. Lowering this requirement means that increasingly smaller voids, more likely to arise from Poisson noise are included in the final catalogue, while increasing it means more zones (with higher density contrast) are added to each void, with the limit being a single supervoid spanning the catalogue.

`zobov` is a computationally intensive process, producing voids that are uniquely non-spherical. This focus on minimizing the number of restrictions in the void-finder leads to it producing a catalogue of voids that are nonspherical, and is immensely sensitive to shot-noise. After selecting out low-significance voids, the void-finder has often found a similar number of voids as other void-finders. Its hierarchical classification of voids, akin to galaxy clusters, and the ability to distinguish voids by the void-finder's statistical significance (i.e. the probability a given void does not arise from Poisson noise) makes it a powerful tool. `zobov` has seen use in almost all corners of the literature - spanning void-galaxy clustering (e.g. [116, 158, 234]) to void-size measurements and even less common applications including weak lensing with `zobov` voids [41], modelling halo bias [229], and in studying galaxy evolution in a variety of environments, (e.g. [63]).

VOXEL

`voxel` is a refinement of the `zobov` void-finder focused on speeding up the void-finding stage. `voxel` uses a discrete set of points to estimate the density field, with the catalogue divided into a series of discrete volume pixels (voxels) in a grid to estimate the field. Voxel voids employ a version of the watershed algorithm, but the differentiating factor lies in how the resultant voids are identified and characterized. Similar to `zobov`, no explicit assumptions are made about the shape in the density field. The density field is estimated by particle-mesh interpolation. For periodic boxes, such as N-body simulations populated with galaxies, the density field is estimated by computing the average number of galaxies per voxel cell. For survey data and cutsky (mock catalogues trimmed to match the geometry of galaxy-redshift surveys) mocks, this is computed by estimating the number of randoms in each cell relative to the average number of randoms. The field is then smoothed with a Gaussian kernel, with a smoothing scale prescribed by the void-finder, R_s.

Once the field has been smoothed, the density of each voxel can be calculated. From there, the watershed process is executed in the same manner as `zobov`. The lowest density voxels in the field are each minima, which are used to seed the growth of the watershed. Individual neighboring voxels are added into each void, until a critical point is reached, causing the watershed process to stop. Once this process has been completed for all voids,

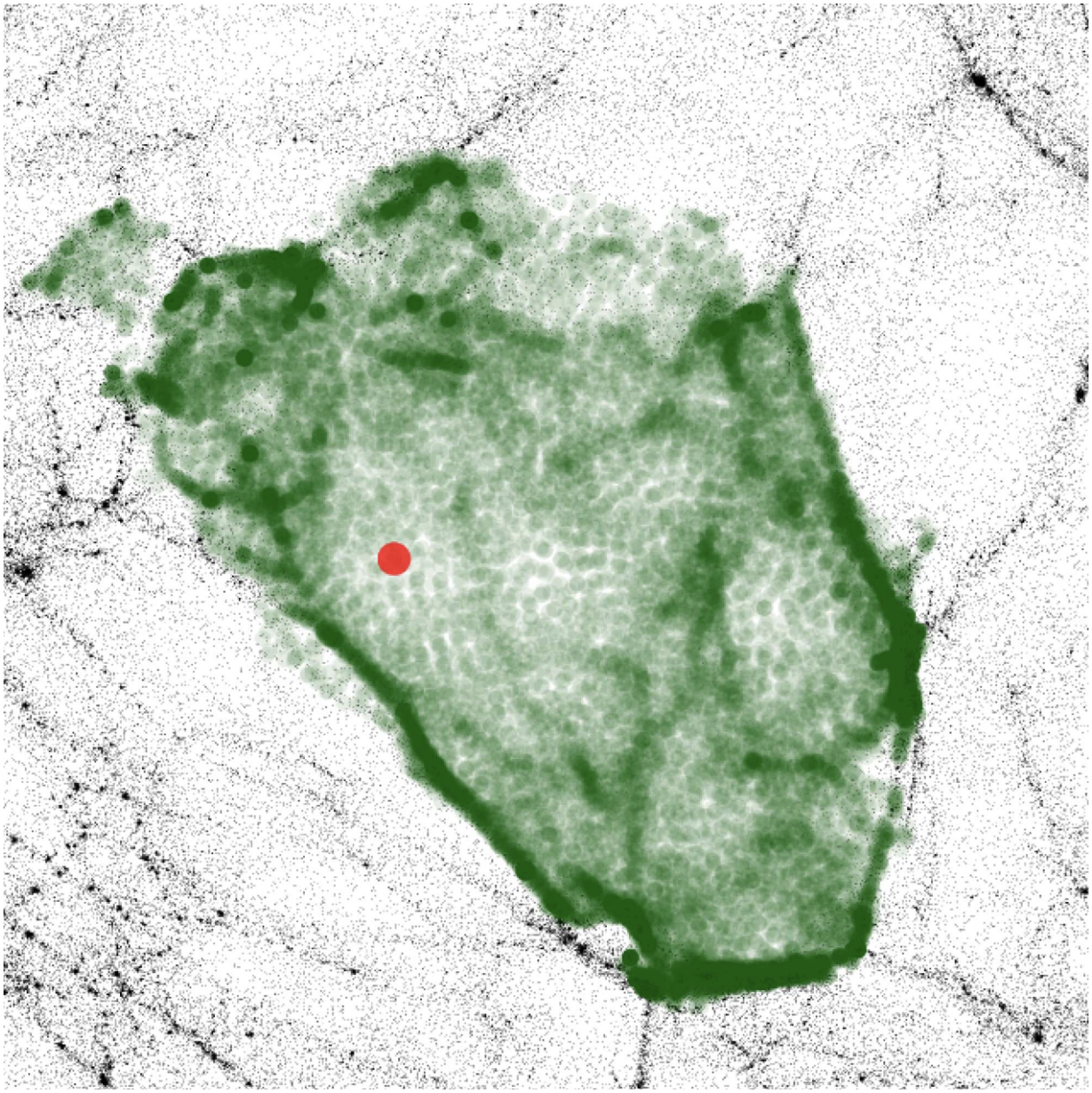

Figure 2.2: Figure courtesy of [159], illustrates the extent of a large void found in an N-body simulation using the `zobov` void-finder. The red dot is the core/minimum density particle, the green particles are constituent of the void, and the black particles are not.

the center of the lowest density voxel in each void is defined as the *void-centre*. These irregularly shaped voids are assigned a characteristic radius, that defines the irregular volume as a sphere of a given radius $R_V = (3N_{voxel}V_{voxel}/4\pi)^{1/3}$. There are no exclusions made based on the probability of voids being emergent from Poisson noise, i.e. spurious voids are included. Instead the choice of smoothing scale R_s and voxel-size s_{voxel} influences this. By shrinking the smoothing scale, the number of spurious voids in a given catalogue will increase, and by increasing it, the number of smaller (likely spurious) voids will decrease. The `voxel` algorithm is easily parallelizable, making it attractive for obtaining a high volume of watershed void catalogues at a relatively lower resource cost. This algorithm was originally programmed for cubic periodic boxes and galaxy-redshift surveys, but was updated in [86] to accommodate non-cubic box volumes.

The final catalogue consists of a list of voids with characteristic radii R_V and positions for the void-centres. A common practice in void literature is to remove smaller voids, and we conducted our analysis by adhering to this, where we removed all voids with radii below $R_V = 20h^{-1}$Mpc, to remove small, spurious voids that arose due to shot noise.

2.3 Measuring void-galaxy clustering

Void-finders allow us to define cosmic voids in terms of a *void-centre* and a void radius, R_V. This allows voids to be used as a tracer when computing cosmological summary statistics. As alluded to above, the choice of void-finder, and consequently, the choice of void-centre utilized does have downstream effects on the resultant summary statistic computed. For this thesis, this was the VG-CCF. Here, we will review the fundamentals of the estimators of the VG-CCF, their method of computation and their physical interpretation.

2.3.1 A review of estimators of the two-point correlation function

AP and RSD effects on voids can be measured through our primary statistic, the VG-CCF, which also describes the void-profile. The VG-CCF is an implementation of the two-point correlation function, which measures the clustering of different populations. The void-galaxy cross-correlation function tells us how galaxies cluster around voids. The two-point galaxy correlation function tells us how galaxies cluster with respect to each other, and as discussed in Section 1.4.4, this clustering signal depends on a number of physical processes in the Universe, including those that produce the BAO feature.

For this thesis, we will discuss correlation function $\xi(s)$.

Galaxy-redshift surveys contain complicated three dimensional information about the large-scale structure (sheets, filaments, clusters, voids). These surveys are also characterized by selection functions, which makes it difficult to quantify the exact region surveyed. To quantify the extent of the region that is observed, catalogues of Poisson sampled random points are generated, where these catalogues match the angular and radial distribution of the survey. In cubic periodic boxes, the expected density is known at every point, so these randoms are used exclusively for surveys and mimicking the survey geometry. These randoms catalogues are quite involved for survey geometry, as they need to account for missing redshifts, bright star exclusions, and extinction, fiber-assignment and object selection.

The two-point correlation function is estimated by computing the distribution of the pairwise distance between tracers of a density field within the field's region. For the galaxy two-point correlation function, this is the pairwise distance between galaxies, while for the VG-CCF, this is the pairwise distance between galaxies and void centres. These pair separations are binned by their distance (real-space r, redshift-space s) and μ, the cosine of the angle between the line-of-sight and the separation vector. This gives us the tracer paircounts

$$D_v D_g = \frac{2}{N_g N_v} \sum_i^{N_g} \sum_j^{N_v} w_i w_j n_{g_i} n_{v_j} \tag{2.1}$$

which we compare to the paircounts computed with the randoms catalogue.

$$R_g R_v = \frac{2}{N_{R_g} N_{R_v}} \sum_i^{N_{R_g}} \sum_j^{N_{R_v}} w_i w_j n_{r_{g_i}} n_{r_{v_j}} \tag{2.2}$$

To further debias this estimate of the two-point correlation function, several weighting schemes have been explored by the literature (e.g. Davis and Peebles $\xi_{DP}(s,\mu)$ [66], Peebles and Hauser, [117], Landy and Szalay $\xi_{LS}(s,\mu)$ [141]). Data-random pairs, DR are used to improve the bias and variance properties of the estimator (e.g Hamilton [117], Davis & Peebles [66]). This is crucial, as failure to do so properly will have a significant impact on the inferred cosmological parameter estimate, as discussed in [127, 225]. The data-random pairs we used are defined as:

$$D_v R_g = \frac{1}{N_v N_{R_g}} \sum_i^{N_v} \sum_j^{N_{R_g}} w_i w_j n_{v_i} n_{r_{g_j}} \tag{2.3}$$

D_gD_v, R_gR_v, D_vR_g paircounts can then be used to estimate the correlation function using any of the aforementioned estimators

$$1 + \xi(s,\mu)_{\rm N} = \frac{D_gD_v}{R_gR_v}\frac{\bar{n}_{r_g}\bar{n}_{r_v}}{\bar{n}_g\bar{n}_v} \tag{2.4}$$

$$1 + \xi(s,\mu)_{DP} = \frac{D_gD_v}{D_vR_g}\frac{\bar{n}_{r_g}}{\bar{n}_g} \tag{2.5}$$

$$1 + \xi(s,\mu)_{LS} = \frac{D_gD_v\left(\frac{\bar{n}_r}{\bar{n}_g}\right)^2 - D_gR_v\left(\frac{\bar{n}_{r_v}}{\bar{n}_g}\right) - D_vR_g\left(\frac{\bar{n}_{r_g}}{\bar{n}_v}\right) + R_gR_v}{R_gR_v} \tag{2.6}$$

where $n_{g,v}, n_{r_g,r_v}$ are the number densities of the data points and the randoms respectively.

Equation (2.6) is the Landy-Szalay estimator, and it is favoured in the literature [127], as it has minimal bias when correctly implemented (i.e. it will closely estimate the underlying two-point correlation function). However, for the VG-CCF, this means that there are terms in Equation (2.6) that involve a galaxy randoms catalogue, R_g and a voids randoms catalogue, R_v. The latter catalogue can be estimated, but void-finders are very sensitive to the boundaries of surveys. Consequently, this means that the R_v randoms catalogue is not only expensive to produce (multiple void-finding runs required for a single randoms catalogue) but also sensitive to survey geometry. In order to minimize the potential bias this might introduce in the VG-CCF, we used the Davis-Peebles estimator Equation (2.5) for this thesis. These two-dimensional estimates of the correlation functions can be decomposed using Legendre polynomials

$$\xi(s)_\ell = \int_0^1 \xi(s,\mu)\mathcal{L}_\ell(\mu)d\mu \tag{2.7}$$

We note that the limited volume of the survey increases the noise at larger scales, leading to the emergence of bias at very large scales, which motivates favouring the Landy-Szalay estimator over the Davis-Peebles estimator, but for the void-galaxy cross-correlation function, we were interested in clustering up to scales of $s < 150\ \ h^{-1}$Mpc. Consequently, for this thesis the effective difference between the Davis-Peebles estimator's clustering and the Landy-Szalay estimator is small, as is their bias.

For this thesis, and for the majority of the contemporary VG-CCF literature, the two-dimensional VG-CCF is decomposed into the first two non-zero multipole moments, ξ_0, the monopole, and ξ_2, the quadrupole. At various points, the monopole can be understood as the radial average of the VG-CCF, a radial profile of the stack of voids, with a pronounced

underdensity at smaller scales, increasing to an overdensity peak (where the boundary galaxies are found), and dropping close to zero at large scales, matching the background overdensity. The quadrupole has been used to model RSD, as its non-zero amplitude is in part dictated by the anisotropies that arise from RSD (more pronounced along the line-of-sight, as opposed to transverse of it). The shot noise from the catalogue dominates higher order multipole expansion terms, so our analysis was focused on the ξ_0 monopole and the ξ_2 quadrupole.

2.4 Cosmic voids in previous analyses using galaxy-redshift surveys

2.4.1 Linear RSD and the linear bias approximation

Performing the AP test with voids [144], required careful modelling of RSD. We will detail the broad chronology of the AP test with voids in Sections 2.4.2 and 2.4.3, but we will start with the first group of models, that tried to map the redshift-space VG-CCF back to the real-space VG-CCF. This took the linear model of RSD in [131] for the two-point correlation function and applied it to galaxies around voids. This linear RSD treatment allowed for the VG-CCF to be linked to the growth rate of large-scale structure, $f(z)$ in addition to using voids as an AP test.

In our overview of linear RSD applied to voids, we will adopt the notation of Nadathur & Percival [157]. For this treatment, and subsequent discussions of RSD in voids, we will adopt the approximation that the peculiar velocity is small, i.e. $|v/raH| \ll 1$, where the peculiar velocity v adjusted by the scale factor a, is dwarfed by the comoving distance of the observer r:

$$s = r + \frac{v}{aH} \tag{2.8}$$

In the literature, it is common to assume that the peculiar velocity is linearly coupled to the density field, meaning that the overdensity field in redshift-space $\delta(\boldsymbol{r})^s$ and real-space $\delta(\boldsymbol{r})$ can be related through the Jacobian of the coordinate transformation of r and s. In following linear perturbation theory [170], which was discussed in Section 1.4.1, the density field in redshift-space is related to the real-space density field and the gradient of the radial component velocity field around the void projected along the line-of-sight.

$$\delta^s(\boldsymbol{r}) = \delta(\boldsymbol{r}) - v_r'(r) \tag{2.9}$$

This projected gradient $v_r'(r)$ can be written as

$$v_r'(r) = (1-\mu^2)\frac{1}{aH}\frac{v(r)}{r} + \mu^2\frac{1}{aH}\frac{\partial v_r(r)}{\partial r} \tag{2.10}$$

imposing a requirement for volume conservation of the number of galaxies,

$$(1+\xi^s(\boldsymbol{s}))d^3s = (1+\xi^r(\boldsymbol{r}))d^3r \tag{2.11}$$

yields, alongside equations 2.9, 2.10 an expression relating the void-galaxy cross-correlation function in real space and redshift space:

$$(1+\xi^s(\boldsymbol{s})) = (1+\xi^r(\boldsymbol{r}))\left[1 + \frac{v_r}{raH} + \frac{(v_r' - v_r/r)}{aH}\mu^2\right]^{-1} \tag{2.12}$$

For linear perturbation theory [170], we can derive that $v_r(r)$ can be written in the form

$$v = -\frac{1}{3}afHr\Delta(r) \tag{2.13}$$

and define the the integrated density contrast as

$$\Delta(r) = \frac{3}{r^3}\int_0^r \delta(r')r'^2\mathrm{d}r' \tag{2.14}$$

Adopting these results from linear theory, Equation 2.12 can be rewritten as

$$1+\xi^s(\boldsymbol{s}) = (1+\xi^r(r))\left[1 - \frac{f}{3}\Delta(r) - f\mu^2\left(\delta(r) - \Delta(r)\right)\right]^{-1} \tag{2.15}$$

To apply this model to a survey, we would require some knowledge of how $\xi^r(r)$ is mapped onto the overdensity field $\delta(\boldsymbol{r})$. The issue is that $\delta(\boldsymbol{r})$ is not known a-priori unless cosmological simulations are employed. A popular assumption was to introduce a linear bias approximation. Similar to how the galaxy bias was obtained from large-scale clustering to map the distribution of dark matter halos to galaxies, it was assumed that $\xi^r(r) = b\delta(\boldsymbol{r})$, $\bar{\xi^r}(\boldsymbol{r}) = b\Delta(r)$, using the same linear galaxy bias.

In some parts of the literature, this approximation was used to factorize f from the multipoles of the two-dimensional VG-CCF. Cai et al. used this [38], and obtained the expression:

$$G(f) = \frac{2f}{3+f} = \frac{\xi_2^s(r)}{\xi_0^s(r) - \frac{3}{r^3}\int_0^r \xi_0^s(r')r'^2 \mathrm{d}r'} \tag{2.16}$$

As Nadathur & Percival [157] explained, the above equation arose from a partial expansion of Equation (2.15). If Equation (2.15) was accurately expanded while using the linear bias approximation, then the factorization of f from the VG-CCF was shown to be more complicated. It could still be used to constrain f, provided that the real-space VG-CCF was known. They modified the velocity field by adding a dispersion for galaxy velocities around the line-of-sight, given by

$$\boldsymbol{v} = v_r \hat{\boldsymbol{r}} + v_\parallel \hat{\boldsymbol{X}} \tag{2.17}$$

where $v_\parallel$ is given by a probability distribution, $P(v_\parallel)$ given by

$$P(v, r) = \frac{1}{\sqrt{2\pi}\sigma_v(r)} \exp\left(\frac{-v^2}{2\sigma_v(r)^2}\right) \tag{2.18}$$

This was then applied to the dispersion model by the equation

$$(1 + \xi(s, \mu)) = \int (1 + \xi(r)) P(v_\parallel, r) \mathrm{d}r \tag{2.19}$$

which links the velocities to the correlation functions in real- and redshift-space.

This bears a similarity to the radial velocity profile modelling [150] and quasi-linear treatments of RSD [171] and the Gaussian Streaming Model [192, 191, 231] that have been explored in the literature. The RSD model in Nadathur & Percival produces strong reductions in error for $f(z)$, as long as $\delta(r)$ and $\sigma_{v_\parallel}$ are known, see: Section 2.4.3.

Cai [38], Nadathur & Percival [157], Hamaus [116, 113, 112] encompass a broader swathe of strategies that have been adopted to model the VG-CCF in real-space $\xi^r(r)$, either through fitting the correlation function to a model, de-projecting the redshift space VG-CCF $\xi(s)$, or through iterative reconstruction methods that aim to map $\xi(s)$ back to $\xi(r)$. What has been clear is that appropriate modelling or removal of RSD from void-finding is a core part of any model of the VG-CCF.

2.4.2 Template modelling

Modelling the behaviour of the density and velocity fields around voids through N-body simulations gave rise to the template method. The template method focuses on using a series of mocks to estimate the density and velocity templates, for a stacked set of cosmic voids that can then be combined to measure the RSD and AP signals in galaxy-redshift surveys.

This model relied on computing $\Delta(r)^{fid}$ in a mock with a known RSD model (σ_v, which can be parameterized) and cosmology (which can be set by the mock catalogues). This allowed the amplitude of this template profile to scale with σ_8, as established by Nadathur et al. [155], in Equation 2.20. The outflow velocity variance $\sigma_{v_\parallel}$ was also modelled. The templates are constructed by iterating over different velocity parameterizations.

$$\Delta(r;z) = \frac{\sigma_8(z)}{\sigma_8^{\mathrm{BigMD}}(z)}\Delta^{fid}(r;z_{\mathrm{ref}}) \tag{2.20}$$

The best model is subsequently chosen by comparing these different parameterizations, and the resultant VG-CCF in real and redshift space, per Equation (2.19). The AP parameters presented in Equation (1.70) are used to rescale the multipoles:

$$\xi^s(s_\perp, s_\parallel) = \xi^{s,\mathrm{fid}}(\alpha_\perp s_\perp^{\mathrm{fid}}, \alpha_\parallel s_\parallel^{\mathrm{fid}}) \tag{2.21}$$

The fiducial cosmology that Δ^{fid} is computed in is one that is close to the best-fit from the data. More exotic cosmological models can be explored, but multiple iterations may be required. Once the above condition is satisfied, then an analytic model describing the RSD and AP scaling effects is incorporated. The resultant best-fitting template is used to fit the data to fit the amplitudes of the VG-CCF.

Putting this together, the template method can be executed over the following steps: First, measure the real-space and redshift-space multipoles in a series of mocks, proximated to a redshift bin that closely matches the data. These mocks allow σ_8, $v_\parallel$ to vary. Then with these templates, the relationship between $\xi(s)$ and $\xi(r)$ can be established using Equations (2.19, 2.18). Specifically, the $\xi(r)$ real-space VG-CCF, dispersion of the velocity field $\sigma_{v_\parallel}$ and the integrated density contrast $\Delta(r)$ can be iterated over until a resulting best-fit for the RSD in survey can be found. Using an MCMC sampler, the best fit for $f\sigma_8$ in the data that best matched the template mocks is found. Selecting the best fit then amounts to comparing template mocks, and obtaining a fit for the AP ratio $\alpha_\perp/\alpha_\parallel$ and $f\sigma_8$.

This method has been successfully applied to the BOSS and eBOSS data from SDSS-III and SDSS-IV, with the constraints in $\Omega_{\mathrm{m}}, f(z)\sigma_8(z)$ ([155, 234]) outlined there. These

are constraints that were frequently competitive with measurements obtained from galaxy-clustering, and presented an immense opportunity for a reduction in errors. The same can be said of the follow-up application that focused on constraining w_0, w_a and the curvature of the universe in addition to these parameters [235].

2.4.3 Cosmological constraints from the VG-CCF

The VG-CCF has seen extensive use in the literature as a summary statistic in large-scale structure cosmology. Directly modelling a void-finder and the population it identifies is difficult. The literature has focused on modelling the VG-CCF, and its sensitivity of the voids in these catalogues to the RSD and AP effects. Below, we will summarize the cosmological constraints obtained by key papers that utilized the VG-CCF to obtain constraints on the AP parameters and RSD. This is not meant to be an exhaustive list, but an illustration of the extent to which the modelling and constraining power have evolved since the inception of the VG-CCF as a summary statistic.

Using a stack of voids (i.e. the VG-CCF) was first proposed as a measurement by Lavaux & Wandelt [143, 144]. Using the isotropic properties of the real-space universe meant that a catalogue of voids, would, when stacked, produce a spherical object in real-space, an ideal candidate for AP measurements.

Sutter et al. [221] followed this programme, and measured voids in the SDSS DR7 luminous red galaxy (LRG) sample, with a median catalogue redshift of $z = 0.31$. They employed the `zobov` watershed void-finder, identifying voids up to a redshift of $z = 0.36$. They were able to identify the AP effect, but the sparsity of voids in the catalogue limited their ability to quantify it.

Cai et al. [38] examined the redshift-space VG-CCF. As discussed in Section 2.4.1, they adopted a linear RSD model to describe the VG-CCF, and also assumed a linear bias approximation. The latter assumption allowed them to link the VG-CCF multipoles ξ_0, ξ_2 to the growth rate, $f(z)$. They measured their voids using a modified version of the spherical void-finder presented in Padilla et al. [161]. They used 1 $(\mathrm{Gpch}^{-1})^3$ mocks, with a fiducial cosmology of $\Omega_\mathrm{m} = 0.24$, $\Omega_\Lambda = 0.76$, $h = 0.73$, and $n_s = 0.958$ and $\sigma_8 = 0.80$ to test their model. They obtained constraints to f/b with 5% precision in a $3(h^{-1}\mathrm{Gpc})^3$ simulation volume, which was competitive with existing analyses with galaxy catalogues. The multipole decomposition of the 2D VG-CCF into the monopole ξ_0 and the quadrupole ξ_2 moments, and their focus on the redshift space information, enabled comparisons to subsequent papers that measured ξ_0, ξ_2.

A milestone in VG-CCF modelling came with the joint analysis of the VG-CCF that was done by Hamaus et al. [113], where they applied the template modelling procedure identified above to constrain the linear growth rate $f(z)$ and the AP parameters $\alpha_{\parallel}, \alpha_{\perp}$. Hamaus et al. [113] selected galaxies from the Sloan Digitial Sky Survey (SDSS) DR11 release, with a median catalogue redshift of $z = 0.57$. This followed the template procedure laid out in [116, 115], wherein voids were found in the 2HOT Nbody simulation and through mocks that were crafted to match the HOD parameters of the survey. In their letter, they obtained the first joint fit of the AP and RSD parameters, which they used $\frac{(D_{\rm M}/D_{\rm H})_{true}}{(D_{\rm M}/D_{\rm H})_{fid}} = 1.003 \pm 0.0012$, $f(z)/b = 0.417 \pm 0.089$, in assuming a ΛCDM cosmology to obtain a constraint on $\Omega_{\rm m} = 0.281 \pm 0.031$. This work can be compared to Cai et al. [38], as both papers characterized the influence of RSD on a stack of voids. This fuelled subsequent papers, where RSD either was a feature to be removed through reconstruction algorithms to recover the real-space void positions, or to be modelled, and incorporated into the description of voids found in redshift space.

Hamaus et al. [111] and Pisani et al. [180] built on the modelling laid out in Hamaus et al. [113], where they measured the redshift-space voids in the SDSS DR12 sample, and used galaxies from the LOWZ and CMASS samples, a region of $0.15 < z < 0.70$, with well over a million galaxies in total. The authors used the watershed void-finder, VIDE, a modification to the `zobov` watershed void-finder. Similar to [38] and preceding analyses, the void catalogues were trimmed by eliminating voids below a specific radius, this removed spurious voids that would have been more sensitive to noise and would have polluted analyses that examine small-scale ($< 20h^{-1}$Mpc) effects like RSD. Armed with a catalogue of voids centered at redshifts $z = 0.32, 0.54$ from the LOWZ and CMASS samples, they used a linear RSD model to link the redshift-space and real-space multipoles through the RSD parameter $\beta = f/b$. With their model, they were able to obtain competitive constraints on β for each of their samples, $\beta(\bar{z} = 0.32) = 0.599^{+0.134}_{-0.124}$ and $\beta(\bar{z} = 0.54) = 0.457^{+0.056}_{-0.054}$.

Nadathur et al. 2019 [154] adopted an iterative reconstruction method to "remove" the RSD from galaxy catalogues. A velocity template was still adopted, but the corresponding template mocks were also subject to reconstruction. They noted that the improvement from incorporating the (reconstructed) VG-CCF into their pipeline, and adopting a ΛCDM cosmology did not significantly alter the mean value of $\Omega_{\rm m}$, and gave rise to a significant reduction in errors. They characterised this (inexpensive) reduction in errors as the avenue for using the VG-CCF in large-scale structure analysis. This reduced the errors in $\Omega_{\rm m}$, where their measurement went from: $\Omega_{\rm m} = 0.3105 \pm 0.0064$ to $\Omega_{\rm m} = 0.3093 \pm 0.0057$. This was equivalent to an 11% reduction in errors for $\Omega_{\rm m}$ and H_0.

A follow-up joint analysis came from Hamaus et al. 2020 [112] which analysed voids

in the BOSS DR12 release and examined the real-space void shape without RSD, derived from first principles. They obtained voids in a catalogue with a mean redshift of $z = 0.51$. They utilized a combined sample of the LOWZ and CMASS samples of the survey, and assumed a flat ΛCDM model, and they used the Planck 2013 cosmological parameters when they calibrated this model with `Patchy` mocks. They utilized the `VIDE` void-finder. In their model, they defined a pair of nuisance parameters that account for anisotropies arising from selecting voids in redshift space, and from inaccuracies in the deprojection of the voids, which are treated as amplitude modifications to the quadrupole and monopole of the VG-CCF. They assumed a flat ΛCDM cosmology and they translated their constraints in the AP and RSD parameters to obtain a measurement of $\Omega_{\rm m} = 0.312 \pm 0.020$. This analysis underscores that a stack of voids (VG-CCF) is a competitive AP test, while the adjacent RSD constraints are highly model dependent.

Woodfinden et al. 2022 [234] also adopted a template-based approach to their model of the VG-CCF. They utilized templates constructed using the Big MultiDark simulation boxes, containing 3840^3 dark matter particles, bounded in a box $2.5{\rm h}^{-1}$ Gpc across, with dark matter halos proximated at redshifts $z = 0.1, 0.32, 0.52$, to match them to the MGS sample, LOWZ and CMASS galaxy samples from BOSS respectively, and using the HOD parameterization to match the galaxy halo connection in each of these samples. They constructed their samples using the `Revolver` void-finder, using it to obtain `zobov` voids in the reconstructed galaxy catalogues. They utilized reconstruction to uphold the isotropy of the void distribution; a stack of voids in a redshift-space distorted catalogue will be asymmetric, and dependent on the line-of-sight orientation. In turn, these distortions will bias the voids selected by the void-finding algorithm, as underdensities aligned along the light-of-sight will appear more strongly stretched, and in turn appear to have a lower density, making it more likely that they are selected as the void-centre in a given catalogue. Reconstruction aims to remove this redshift-space distortion, through a series of iterative Fast Fourier Transforms (FFT), using the galaxy bias b and large-scale structure growth parameter f as inputs.

They found the voids in each of these template mocks, and in each of the SDSS catalogues from the BOSS and eBOSS samples, creating a `zobov` void catalogue for $0.07 < z < 1.0$. By splitting these void-catalogues across five redshift bins, and supplementing it with results from [158] they can obtain a best fit for the velocity dispersion parameter σ_v. They then utilized their templates to identify the best-fitting velocity profile, which they fit to the VG-CCF and measured $f(z)\sigma_8(z)$ and $\alpha_\perp/\alpha_\parallel$, which in a flat ΛCDM model correspond to a matter density parameter of $\Omega_{\rm m} = 0.337^{+0.026}_{-0.029}$, on par with those of contemporary works (e.g. [112]). In a similar fashion, these results were translated into cosmological parameter constraints by assuming a model – they assumed a flat ΛCDM model, and obtained a

constraint: $\Omega_{\rm m} = 0.337^{+0.026}_{-0.029}$.

An extension to this analysis was done in Woodfinden et al. 2023 [235]. Their aim was to obtain cosmological constraints on dark energy parameters and the energy-density of curvature Ω_k by using the redshift evolution of the VG-CCF. They incorporated the aforementioned void results from their previous paper, and added the BAO measurements from SDSS-II, SDSS-III and SDSS-IV. This grouping of measurements was used to construct a cross-covariance matrix, which contained both voids from the aforementioned 6 different redshift bins, and the BAO information across 4 different redshift bins, and they used this to quantify the complementary constraints obtained from these tracers.

They augmented the scope of their analysis by extending the parameter space of ΛCDM cosmologies to allow for non-flat cosmologies (e.g. oCDM) and for non-Λ dark energy models (e.g. wCDM, $w(z)$CDM). This information was combined with constraints from the SDSS DR16 BOSS BAO measurement, Planck 2018 CMB measurements, and Pantheon SN Ia sample. Voids were measured at multiple redshift bins, as the AP ratio D_M/D_H and $f(z)\sigma_8(z)$ contains information about the expansion history of the universe at late-times. They obtained templates from the BigMultiDark mocks with different cosmological parameterizations. They constructed these templates so their pipeline could constrain the cosmological parameters for $\Omega_{\rm m}$, Λ, the energy density of curvature Ω_k and $w, w(z)$ the dark energy equation of state parameters.

When they assumed a ΛCDM model, they found $\Omega_{\rm m} = 0.354 \pm 0.029$, supplementing this with constraints from galaxy clustering, they found this improves to $\Omega_{\rm m} = 0.356 \pm 0.024$. Combining this with CMB and SNIa constraints boosts it even further: $\Omega_{\rm m} = 0.3170 \pm 0.0069$.

When they extended their parameter space to allow for wCDM cosmologies, the redshift evolution of their void slices allowed for them to obtain $\Omega_{\rm m} = 0.330^{+0.11}_{-0.033}$ and $w = -1.17^{+0.64}_{-0.31}$. The addition of galaxy clustering further constrains these cosmological parameters to $\Omega_{\rm m} = 0.391^{+}_{-}$, $w = -1.50^{+0.43}_{-0.28}$.

Extending the parameter space for open curvatures owCDM gives rise to $\Omega_{\rm m} = 0.385^{+0.082}_{-0.13}$, $\Omega_k = -0.04^{+0.22}_{-0.15}$, $w = -1.29^{+0.58}_{-0.32}$

Extensions incorporating $w(z)$CDM, oCDM and $ow(z)$CDM were also explored. Overall, they found strong agreement with flat spatial curvatures, and strong support for a ΛCDM cosmological model. They find a 43% reduction in errors for $\Omega_{\rm m}$ compared to galaxy clustering in ΛCDM, while for wCDM it is a 44% and 45% reduction in errors for $\Omega_{\rm m}$ and w respectively. The reduction in errors for Ω_k and $w(z)$ dark energy parameters are also significant, pointing to enormous potential of using multiple redshift bins of the VG-CCF to extract such cosmological information. See Figure 2.3 for an illustration of

the complementarity of these void constraints compared to those obtained from galaxy clustering.

2.4.4 Biases in the void-finder

Void-finders are greatly influenced by a number of variables across galaxy catalogues; the density of tracers (galaxy number density); the geometry of the survey region; and the distortions arising from AP and RSD greatly affect which regions are classified as voids and selected by the void-finder.

Accurately modelling these effects will help limit the selection bias of voids in galaxy catalogues. One major bias that can be introduced into these void analyses comes from analyzing void catalogues in redshift-space as opposed to real-space. In the literature, this bias in void-selection has to be factored when creating real-space velocity and density templates. The options have been to either extend the modelling of void-finding and the resulting void-profile (VG-CCF) to redshift-space [111] or to (approximately) remove RSD through reconstruction of the real-space galaxy positions. The latter was advocated for in [154].

The goal of reconstruction is to recover the real-space positions of galaxies in a survey prior to void-finding, using information only available from a survey. Reconstruction uses the iterative Fast-Fourier Transform method (FFT) [33] to solve the Zel'Dovich equation in redshift-space [239], for the Lagrangian displacement field, Ψ in terms of the galaxy overdensity in redshift-space, the linear galaxy bias, and the growth rate, per [239]:

$$\nabla \cdot \vec{\Psi} + \frac{f}{b} \nabla \cdot (\vec{\Psi} \cdot \hat{r}) \hat{r} = -\frac{\delta_g}{b} \tag{2.22}$$

Solving this equation using FFT for the full displacement field, Ψ yields the RSD component $\vec{\Psi}_{\rm RSD} = f(\vec{\Psi} \cdot \hat{r})\hat{r}$, which we use to shift the galaxy positions by $-\Psi_{\rm RSD}$ to (approximately) remove the linear RSD for each galaxy. This procedure requires only f, b. The voids are found using the reconstructed positions of galaxies, but the VG-CCF is computed using the original redshift space positions (i.e. the velocities are only removed for void-finding, and reintroduced when computing the profile), as the aim is to remove the bias arising from void-finding in redshift space.

To further complicate this picture of void-finding, the effect of AP distortions on the void-finder needs to be carefully considered. If voids are found in a catalogue with their true cosmology, and then their *void-centres* are shifted using the AP parameters Equation (1.70)

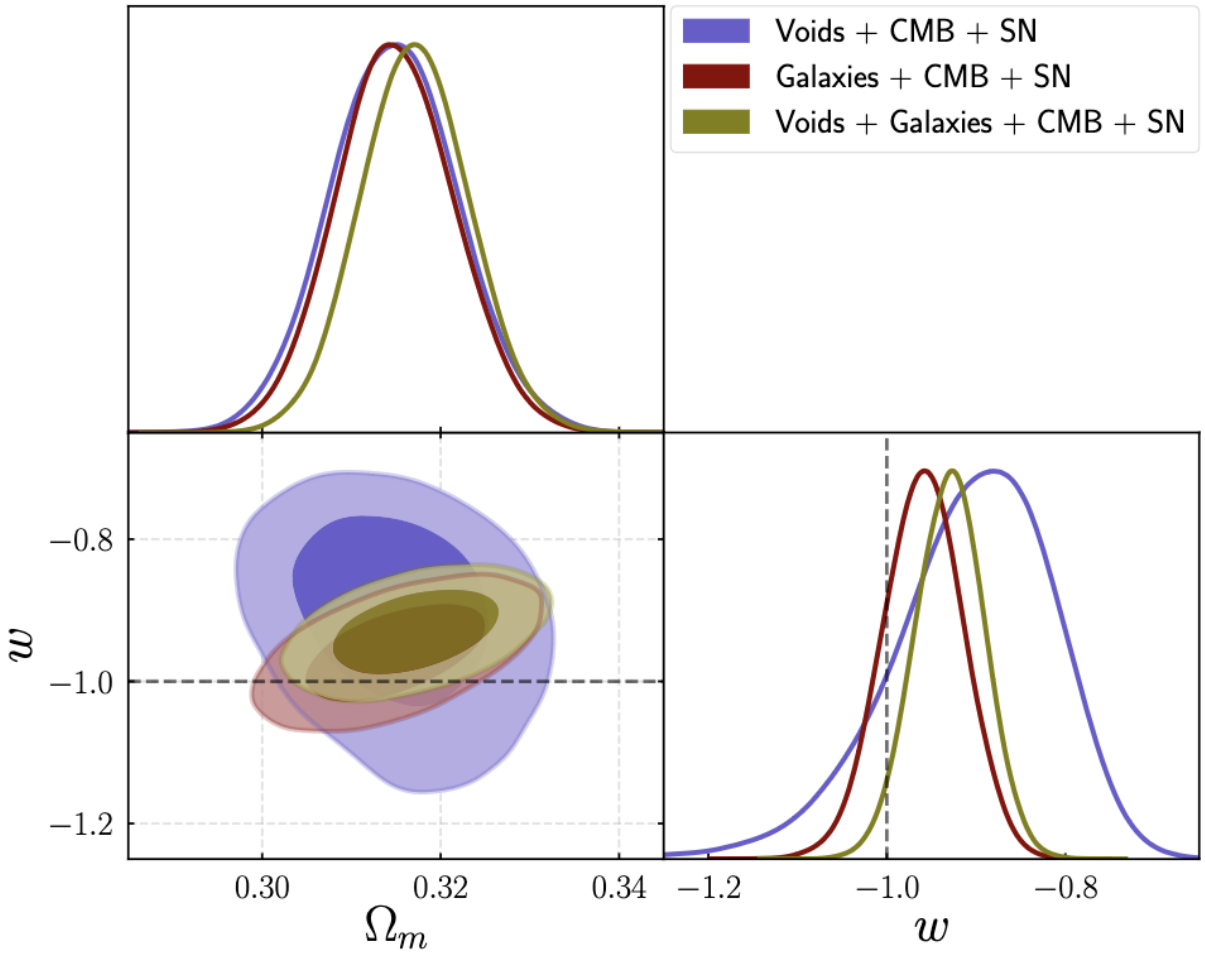

Figure 2.3: Image credit: [235]. Here, the template method was applied to measure voids in 5 different redshift bins from $0.07 < z < 1.0$ in BOSS and eBOSS. The resultant measurements in $D_{\rm M}/D_{\rm H}$ and $f\sigma_8$ were then used to constrain $\Omega_{\rm m}$ and w upon assuming a cosmological model (e.g ΛCDM, wCDM, oCDM). In combining this information from the VG-CCF with galaxy clustering (read: BAO) and CMB and SNIa information, they found that the voids constraints are orthogonal to galaxy-clustering constraints when combined with CMB and SNIa information. This highlights that the voids contain of wealth of information that should be exploited with DESI, Euclid and LSST data.

that are used to shift galaxy positions, the resulting VG-CCF is incongruent with one that would be obtained in a catalogue that was distorted with AP parameters prior to void-finding. This treatment of AP is critical as it means that the voids found in the true catalogue cannot be approximated as the ones found with even a mild AP distortion, as it ends up significantly biasing the cosmological constraints. This was the focus of [187] and a part of [86], as it was shown that applying the AP distortion parameters to the void-centres is not equivalent to void-finding in an AP distorted catalogue. Figure 2.4, taken from [187] illustrates the extent to which it biases the VG-CCF. We also discuss this in Chapter 4, and illustrate that this bias is not unique to template models of the VG-CCF, and showed it had a significant effect on the inferred AP and RSD parameter, even though a completely different model was used.

2.4.5 Template-free modelling of the void-galaxy cross-correlation function

One branch we explored in the development of a model of the VG-CCF was through applying a physically motivated assumption to the VG-CCF to obtain AP and RSD information $f\sigma_8(z), D_{\rm M}/D_{\rm H}(z)$. This model focused on the quadrupole moment of the VG-CCF, which changes drastically between voids found in a real-space catalogue and voids found in a redshift-space catalogue. The quadrupole of the VG-CCF experiences a significant change in its amplitude at scales $10 < s < 80\ h^{-1}$Mpc, reflecting the anisotropy in the VG-CCF due to RSD. Conversely, in real-space, the VG-CCF's quadrupole has a much smaller amplitude, with it being centered around zero.

The reconstruction algorithm can successfully remove most of the RSD in a field of galaxies, enabling us to find voids in a catalogue that is much more similar to the real space catalogue. Therefore, if we could correctly infer f by finding the value of f that produces a VG-CCF quadrupole closest to zero, we could have used this model without having to construct an array of velocity templates. Instead of templates, we would have simply iterated over $f(z)$ until we obtained a quadrupole that was closest to zero. We would have then adopted that as our estimate of f, and subsequently estimated the RSD parameter $f\sigma_8(z)$ and consequently have estimated the AP parameters. This is an idea that has been explored in detail in [68].

The first set of tests we ran was to obtain the fiducial value of f from a series of mocks with a known fiducial cosmology. We ran reconstruction on multiple realizations of these catalogues for different values of $f(z = 0.5)$ chosen. We averaged each of these realizations, to obtain a mean ξ_2 quadrupole for each of the values of $f(z)$ chosen for reconstruction. We

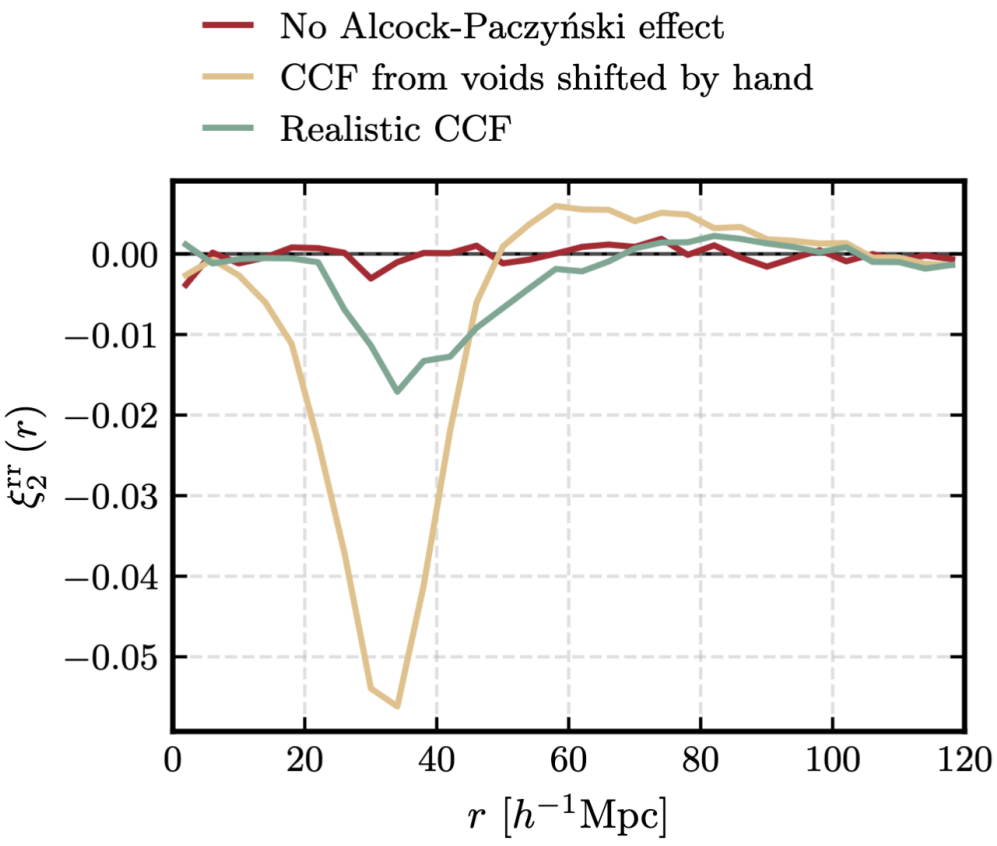

Figure 2.4: Figure courtesy of [187], where the quadrupole ξ_2 of the real-space VG-CCF was computed under three scenarios: **i) - (red)** from the VG-CCF computed in the true cosmology, where the void-centres were found using the real-space galaxy positions, i.e. $\alpha_\perp, \alpha_\parallel = 1$, **ii) - (beige)** from the approximation of the VG-CCF, where the voids were found in a fiducial cosmology such that $\alpha_\perp, \alpha_\parallel \neq 1$, but the galaxy positions were not distorted by the AP parameters for void-finding, and instead the void-centres and galaxy positions were distorted by the AP parameters during the computation of the VG-CCF (by hand), **iii) - (green)** from a realistic VG-CCF, which describes the VG-CCF obtained from finding the voids in a fiducial cosmology where $\alpha_\perp, \alpha_\parallel \neq 1$, the galaxy positions have already been adjusted by the AP parameters, and the voids were found in this distorted catalogue, with the CCF being computed subsequently. The amplitude and features of the quadrupole are shifted, which greatly affects their measurements. As co-authors of [187], we do a similar comparison in [86], where we prepared mocks for cases **ii)** and **iii)** and compared the model predictions, and uncovered a similar bias.

also computed a covariance matrix using the mocks to estimate the errors in Figure 2.5. We found that for a value of $f(z = 0.5) = 0.784$ as the fiducial linear growth rate, this template-free method consistently underestimated $f(z)$. In a case where the AP effect was not present, we obtained a best fit of $f(z) = 0.744$. Moreover, the mean was not only underestimated, but poorly constrained, as this method yielded a 27% constraint on $f(z)$. This result also came from incorporating all voids in the VG-CCF, a practice that is typically avoided as smaller spurious voids have been known to affect the VG-CCF. Imposing a size cut worsened the constraining power further. We attempted to extend the modelling of Seo et al. 2015 [209], where they modelled the reconstructed BAO in Fourier space, to the case of the VG-CCF. Such modifications did not produce a satisfactory model of the VG-CCF post-reconstruction.

These tests were extended to include the AP parameter ratio, $D_{\rm M}/D_{\rm H}$ in tandem with $f(z)$, but the errors were wider and the constraining power more limited. We found that while the original concept of potentially inferring $f(z)$ from reconstructing a catalogue and evaluating the ξ_2 in physically motivated manner to be exciting, the execution with a reconstructed galaxy field was quite difficult. We kept searching for an alterative to templates, but we note that template-free approach is an idea that has been successfully pursued by other groups [68]. This illustrates that modelling the VG-CCF is costly and difficult. We will now outline our rationale for continuing our search for a model of the VG-CCF that does not use the previously discussed velocity templates.

2.4.6 Limitations of previous modelling

The template model has been successful in producing cosmological constraints from the VG-CCF that are complementary and competitive with those from galaxy clustering, illustrating the depth of information obtainable from void catalogues. This model is built on the assumption that the voids in these catalogues, and subsequently the VG-CCF are unbiased. To ensure this, careful modelling of the RSD's effect on the VG-CCF (e.g. [111]) or the removal of it (e.g. [154]) was required, as RSD would introduce a bias into the stack of voids obtained from the catalogue. Appropriate modelling of AP distortions in the catalogue are also necessary to uphold this, see: [187].

Future galaxy-redshift surveys will utilize larger volumes, and aim to explore wider cosmological parameter spaces, in light of growing tensions in the dark energy equation of state. The templates utilized for VG-CCF modelling are centered on a fiducial cosmology, which may or may not be close to the true cosmology of the Universe. The requirements of higher precision, larger parameter spaces being explored and the need for the void-catalogues to be unbiased for template modelling will present challenges that were not

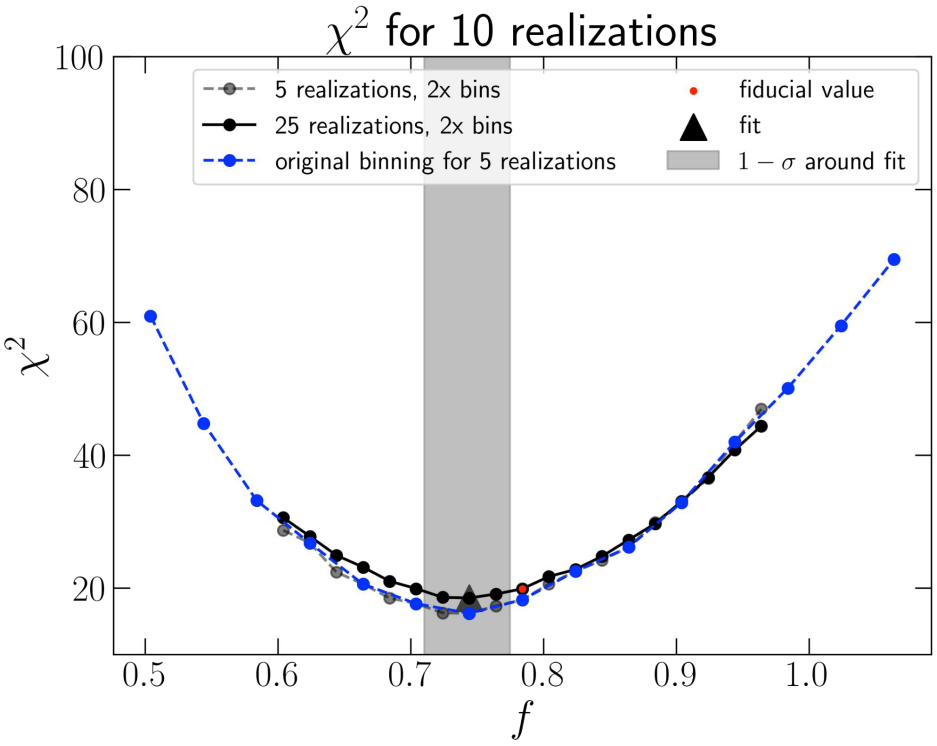

Figure 2.5: An implementation of the template-free method. We tested our model using 5 realizations (grey dashed, blue dashed) and 25 realizations (black solid) of the `AbacusSummit` mocks populated using different values of f, with varying bin width (grey vs blue). Even in a case where there is no AP distortion, even with denser binning for f, the model struggled to detect the fiducial value for f.

broached in SDSS, as this was not of concern during that era (e.g. [234, 235]). This might be something that would bear consideration with ongoing and future surveys (e.g. Euclid [110, 53, 186]), but this has not been explored.

In particular, while the effects of RSD on void-finding can either be modelled or mitigated through reconstruction, they each have challenges. While reconstruction is a powerful method, incomplete reconstruction, where RSD still leaves small amplitude artefacts in the VG-CCF's monopole and quadrupole moments could present an enormous challenge with an already sensitive template model. Given that the effects of RSD also modulate the AP distortions, this inhibits the constraining power of templates further.

Consequently, we needed a model for the VG-CCF that models the effects of RSD on the amplitude and shape of the VG-CCF's monopole and quadrupole $\xi_0(s), \xi_2(s)$, i.e. successful modelling of the consequence of void-finding in a redshift-space catalogue. We have also shown that great care needs to be taken when modelling void-finding in AP-distorted catalogues, lest we significantly bias the model. Accurately modelling these effects for a watershed void-finder is very difficult. We opted to use simulations to model these effects and quantify the biases arising from this. This is a fact we will discuss in the next chapter.

We also needed this model to allow us to explore the cosmological sensitivity of the void-finder and the VG-CCF to various cosmological parameters, as being able to explore a large and varied parameter space would be the keystone to making this new model competitive with existing models of the VG-CCF.

With this simulation-based approach to our model, we implemented a neural network emulator that made use of the simulations we planned to use to model the biases in void-finding arising from RSD and AP, and we used this to explore a large cosmological parameter space while modelling the full shape of the VG-CCF monopole and quadrupole. In the following chapter, the ingredients for the model we have developed that overcomes these challenges will be discussed. In Chapter 4, we will outline how this model can produce constraints that are competitive with the template models of the VG-CCF.

Chapter 3

Emulators

3.1 Emulators

To tackle the challenges arising from modelling the VG-CCF, we used an emulator to model the effects of RSD and AP on the void-finder and the subsequent VG-CCF. Since prior modelling of the VG-CCF had already made use of cosmological simulations to develop a model, we used cosmological simulations to construct our emulator. This emulator allowed us to explore a large cosmological and galaxy-halo connection parameter space. Emulators are constructed either through training a neural network, or through training a Gaussian Process model. We will outline differences between the two, their application to large-scale structure cosmology, and then outline the requirements for our VG-CCF model.

3.1.1 Emulation of large-scale structure summary statistics

Large-scale structure cosmology, as discussed in Chapter 1 is an immensely valuable model that becomes even more informative when it is extended to describe higher-order clustering (e.g.[40, 102, 25]). These models offer a theoretical framework for two-point summary statistics and extensions to higher orders (e.g. galaxy bispectrum modelling: [206, 207, 205]; with its applications to survey data: [9, 101, 178, 179]). However, exploring these extensions through cosmological simulations is prohibitively expensive, especially for extending summary statistics to incorporate different tracers of structure into domains that are not described by perturbations. The cost of running more N-body simulations to develop these models across a vast parameter space is resource intensive. One way to

overcome this resource cost is to interpolate between the measurements obtained for different simulations for different cosmological or galaxy-halo parameters. Interpolation in the form of Gaussian Processes (GP) and neural networks have been increasingly used - ranging from the galaxy-halo connection [139, 236, 78] to the nonlinear power spectrum [200, 232, 44]. This class of interpolation methods are referred to as **emulators**.

Gaussian Process Emulators

Gaussian Processes (GP) are a class of stochastic models that describe a probability distribution of the infinite set of possible functions that fit a set of data points, given some kernel. The mean of this probability distribution corresponds to the most probable characterization of the mapping between the input data X and the function $f(X)$. The probabilistic nature of this model, and the completeness of Gaussian statistics to second-order means that in addition to the model's mean (prediction), uncertainty can be assigned to these predictions, allowing us to estimate the uncertainties of the model and its contributions from the individual functions. This latter feature makes it an attractive tool for the task of emulation. Typically, this class of models starts by imposing a mean $\mu = 0$, centering the data. The kernel, or covariance matrix $\boldsymbol{K}$ is then chosen, its form is dependent on the regression or modelling task at hand. Finally, a set of priors for our model is chosen. Adding training data reshapes the joint distribution $P(X, f(X))$, which, by using the properties of multivariate Gaussian distributions, can be used to obtain the conditional $P(X|f(X))$. As a regression model, they are limited by the prior range, but they can provide a robust estimate of the mean and uncertainty of the noisy mapping of data X to observations $f(X)$. Optimizations can be made to the GP, to simplify the matrix multiplications used while emulating the function and inferring the parameter constraints. Much like neural networks, the choices in the hyper-parameters and the kernels utilized are model dependent, and identifying the correct kernel for modelling is one of the major limitations in using GPs. Furthermore, the computational intensity, even when using sparse matrices or decomposed data in training GPs presents its own set of challenges, especially in the face of leveraging large datasets for emulation. We suggest for interested readers to review the works of [188, 23, 148, 97] for an overview of the methodology and use of Gaussian Processes.

Gaussian Processes have been extensively used in the literature. They were first applied to extract cosmological parameters, with robust predictions obtained in [105] and [204]. They extracted robust predictions for cosmological parameters based on a relatively small number of accurate simulations. This was done by constructing a sample of simulated power-spectra, exploring 128 input parameter configurations over their five input

parameters by using the Latin Hyper Cube (LHC) sampling algorithm [222] to construct the training sample. These power-spectra were further compressed using a Singular Value Decomposition (SVD), decomposing the measurement from each simulation into a series of principal components. The resulting emulator was a probability model that described the simulation output of this compressed power-spectrum at an untested parameter combination. In effect, the emulator is an interpolation. They then tested this emulator by comparing the model's best fit for untested parameter combinations, and sequestered 3/128 mocks as part of a hold-out test. The latter was done by selecting three of their mocks, excluding it from the modelling stage, and interpolated at their points in the parameter space, which tested the emulator for all 128 simulations. They also compared the shifts in accuracy when using 64 and 128 simulations, and showed drastic improvement in constraining power (down to a few percent in accuracy). They noted that their emulator could be easily extended to very large parameter spaces to incorporate CMB and LSS measurements, which, in addition to the inexpensive means of obtaining these constraints, and relatively conservative estimate of uncertainty, has spurred a fruitful development of these techniques. They described a framework that is foundational to practically every emulator that has followed in the literature: they used a small, tractable number of expensive (N-body) simulations to create an accurate emulator, replacing a full suite of simulations; they used an optimal sampling strategy of simulation runs to obtain their best possible performance; they integrated the various uncertainties from emulator predictions into the parameter constraints; they constructed this model without imposing a heavy cost on computational resources. [119] offered extensions of this model, and focused on using simulations to estimate the cosmological parameters, and ran additional tests of the GP model calibration, and in estimating noisy-covariance matrices, the latter has become a popular application of emulators (e.g. [73, 152]).

3.1.2 Neural network emulators

Neural networks have been increasingly used to emulate large-scale structure summary statistics. They are primarily oriented around solving the inverse parameter problem, where for a given observable $\boldsymbol{y}$, the model is trained to predict $\boldsymbol{\theta}$, the set of cosmological or galaxy-halo parameters that give rise to it. Neural network emulators are favoured when using larger datasets, as the trained model is less memory intensive, and are more flexible from a modelling perspective than Gaussian Processes. The difficulty lies in generating a training set that is sufficiently dense in parameter space, to match the precision of Gaussian Processes. Over time, neural network emulators have become more popular than their GP counterpart.

Neural network emulators were first applied by [10] to emulate the relationship between the non-linear matter power spectrum and cosmological parameters. This method required a high density of sample points in the parameter space, using a series of simulations with parameterizations chosen by Latin Hypercube Sampling (LHC) [222]. A significant development in neural network emulators came in the form of applying them to conduct cosmological parameter inference, and to solve the inverse problem.

The literature highlights cases where neural networks, including neural network emulators have produced biased estimations - where the mean prediction was precise, but the uncertainty estimation was biased. [62] outlined the framework used in [163, 164, 86] and this thesis for error estimation. The emulator's bias was treated as an additive term to the model's covariance matrix. This model followed the prescriptions outlined by Cranmer et al. [60], which outlined the use of emulators for simulation-based inference, wherein neural network emulators interpolated the relationship between θ, the input parameters, which are translated into a series of latent states z_i, which in turn produced a vector $x \sim p(x|\boldsymbol{\theta}, z)$. In the context of cosmological neural networks, $\boldsymbol{\theta}$ describes the initial cosmological or physical parameters of a simulation, and x describes the observations - ranging from a handful of points to a higher dimension representation (i.e. an array of tracers in a volume).

Often, a neural network is a tool that is reserved for applications where an analytical model is difficult to obtain, and emphasis is placed to appropriately quantify the prediction. Neural networks are capable of rendering a precise mean prediction, but a significant area of the literature has been focused on their interpretability and uncertainty estimation. Developments in these uncertainty estimations have led to the growing application of neural networks in large-scale structure cosmology. The requirement for precise measurements of cosmological parameters from a suite of observables means that a premium is placed on obtaining more information from existing catalogues, especially if these model predictions can be adequately assessed.

Prior to neural network emulators, this relationship between $\boldsymbol{\theta}$ and x was, in the absence of a coherent likelihood, inferred by utilizing Kernel Density Estimates of the likelihood surface or even Approximate Bayesian methods to obtain this, e.g. [13, 128, 109, 160]. The challenge is posed by $\boldsymbol{\theta}$'s dimensionality, as a dense sampling of the parameter space is required for these methods to properly characterize the likelihoods without adversely affecting the inference of x. Additionally, the introduction of new data or simulations requires the entire analysis to be run again, putting even more strain on CPU and memory. Developments in neural networks paved the way to incorporating the derivatives calculated by these networks into the likelihood estimation through interpolating them. This represents a step forward from prior neural network emulators, as the incorporation of this information into the inference of parameters, particularly in solving the inverse problem

(i.e. obtaining $\boldsymbol{\theta}$ given some summary statistic x). This lays the groundwork for more interpretability in neural network emulator models, and for their derivatives to be used in estimating the parameters in cosmological simulations from a given observable. Our neural network drew from these developments as the basis for the model we trained. We will outline the fundamentals that go into building a neural network, applying it as an emulator, and then modelling specifics we were interested in.

Fundamentals of Neural Networks

Neural networks are a family of algorithms that can iteratively evolve to model the non-linear relationship between the inputs $\boldsymbol{x}$ and outputs $\boldsymbol{y}$ through the function $\boldsymbol{y} = \phi(\boldsymbol{x})$. This encompasses a wide gamut of models: from Multilayer Perceptrons, to Convolution Neural Networks, to Transformer Networks. All of these have seen significant use in cosmology, as well-defined, closed-form, relationships between cosmological parameters and observable quantities are rare. A significant challenge in employing neural networks in cosmology lies in their iterative and stochastic qualities, turning them into a black-box, bereft of any explicit causal relationship being defined between the variables. Great care must be taken to adequately model the error and biases of such models, as well as the causal relationship between the parameters being modelled and the prediction.

One common model, which we used in our thesis is a Fully-Connected Neural (FCN) network (see: [64, 96, 240] for a detailed review of machine-learning concepts). FCNs aim to approximate the function $\boldsymbol{y} = \phi(\boldsymbol{x})$ as $\hat{\boldsymbol{y}} = F(\boldsymbol{x})$, where the input $\mathbf{x}$ is mapped to a prediction $\hat{\boldsymbol{y}}$ through Equation (3.1), and each layer of the network is represented by a function f_ℓ for L layers. These functions f_ℓ are called activation functions of individual *neurons* in the network, defined as $n_{\ell,m}$, where ℓ is the corresponding layer, m is the number of neurons in the given layer. Activation functions stem from the biological motivation surrounding artificial neural networks, and describe a given neuron's response to input data, in order to map the input vector $\boldsymbol{x}$ into an output $\hat{\boldsymbol{y}}$. Activation functions are often non-linear, but restricted to a range of $y \, \epsilon \, (-1, 1)$.

The network is constructed by applying a sequence of non-linear activations f_ℓ to each input vector $\boldsymbol{x}$, transforming it with a series of trainable weights $\boldsymbol{W}$ and biases $\boldsymbol{b}$, where the bias vector for each layer ℓ is an $(1, h^\ell)$ tensor, where h^ℓ is the number of hidden units or neurons for a given layer ℓ, and $\boldsymbol{W}$ is an $(h^\ell, h^{\ell+1})$ tensor, applying linear combinations of trainable weights to the previous layer's output $h^{\ell-1}$, constructing the vector h^ℓ, the current layer's output. This is the standard structure of a fully-connected or artificial neural network. Figure 3.1 is an illustrative diagram of a network with two hidden layers,

and a scalar output. A mathematical formulation of this model is given by the following equations:

$$\hat{\boldsymbol{y}} = F(\boldsymbol{x}) = f_L \circledast f_{L-1} \circledast ... \circledast f_1(x) \text{ for } \ell = 1, ..., \text{L} \tag{3.1}$$

$$\boldsymbol{z}^\ell = \boldsymbol{W}^{T,\ell}\boldsymbol{h}^\ell + \boldsymbol{b}^\ell \tag{3.2}$$

$$h^\ell(\boldsymbol{z}^\ell) = f_\ell(\boldsymbol{z}^\ell) \tag{3.3}$$

$$h^L = \hat{\boldsymbol{y}} \tag{3.4}$$

$$h^0 = \boldsymbol{x} \tag{3.5}$$

$$\nabla_{\boldsymbol{x}}\boldsymbol{z} = \left(\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}\right)^T \nabla_{\boldsymbol{y}}\boldsymbol{z} \tag{3.6}$$

The operations described in Equations 3.2-3.5 will allow us to produce an initial prediction of the data for a given sample, $\hat{\boldsymbol{y}}$. To evaluate the quality of this guess across the training data, we evaluate the partial derivatives of the $\mathcal{L}(\hat{\boldsymbol{y}}, \boldsymbol{y})$ training loss with respect to the input $\boldsymbol{x}$. This is done using the numerical derivatives of the weight and bias tensors, through an operation called backpropagation, summarized in Equation (3.6).

Careful choice of the activation function, f_ℓ is necessary, to ensure the computation of stable numerical derivatives. For this thesis, we explored a variety of activation functions (see Equations 3.10-3.14 later in this Chapter), but other varieties have been used in the literature with success. A detailed overview of the best practices for choosing activation functions and designing neural networks can be found in [77] and [138].

With the elements of a simple FCN outlined, we will briefly enumerate the steps involved in training the neural network. This begins with taking the inputs x, and computing the result of Equations 3.4-3.5, to calculate the initial prediction, $\hat{\boldsymbol{y}}$. The weights are initialized as $W = 1, b = 0$ This is the *feed-forward* step. With the initial prediction, we can evaluate the quality of the prediction $\hat{\boldsymbol{y}}$ compared to the ground truth $\boldsymbol{y}$ in terms of the loss function, $\mathcal{L}$. *Backpropagation* is the second stage of training a neural network, and it allows us to use the loss function we chose to evaluate the quality of the neural network's modelling of $\phi(\boldsymbol{x})$ with respect to its prediction, $\hat{\boldsymbol{y}}$. This works by computing the partial derivatives of the final layer, h^L, and iteratively computing the gradients of the previous layer, h^{L-1}, computing the gradients of the loss $\mathcal{L}$ in terms of all the weights and biases for all layers. These gradients are then saved for the following step of *Stochastic Gradient Descent*. *Stochastic Gradient Descent* is where these gradients are applied to update all of the weights and biases in the model, which can be likened to "learning". There exists

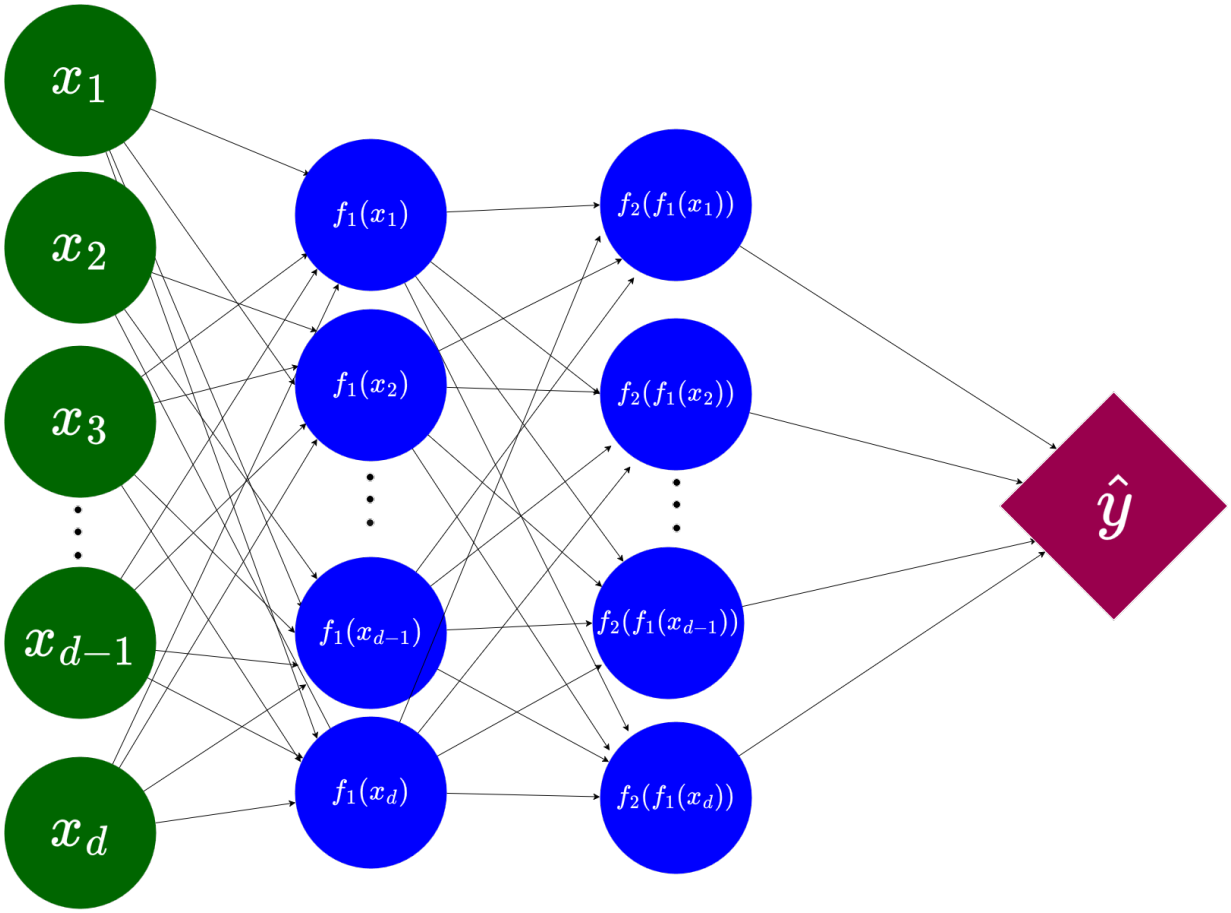

Figure 3.1: A fully connected neural network with two hidden layers, with an input dimension of d, producing a single dimensional output $\hat{\boldsymbol{y}}$. Each layer, ℓ takes the previous layer's activation function of $f_{\ell-1}(x)$ as its input, for each neuron. For a detailed example, see Equation 3.1.

a whole family of gradient descent algorithms, they all focus on calculating the updated parameters, $\Gamma = \boldsymbol{W}, \boldsymbol{b}$ in the interest of minimizing the $\mathcal{L}$ loss. The most common in use is stochastic gradient descent, which updates the parameters of the model for each training sample. It is parameterized as follows:

$$\Gamma \leftarrow \Gamma - \boldsymbol{\eta} \cdot \nabla_{\boldsymbol{\Gamma}} \mathcal{L}(\Gamma, x^i, y^i) \tag{3.7}$$

where the parameters $(\boldsymbol{W}, \boldsymbol{b}) = \Gamma$, such that the parameters are updated by computing the gradient with respect to Γ of the loss $\mathcal{L}$ given the previous set of network parameters, Γ, some input x^i from the training sample, and some target y^i value from the training set, where x^i, y^i are individual entries from the training set or *batches*, small subsamples of the training data. The parameters Γ update with respect to the gradients modified by the learning rate, η. The neural network's loss function $\mathcal{L}$ is again a design choice that can be made when building neural networks.

The change to the weights is governed by our choice in learning rate, η, a hyperparameter that describes the network. The weights and biases are updated using Equations 3.2-3.6 with Section 3.1.2. When the weights are updated by this learning rate, this completes one step in the training of the network, called an epoch (consisting of the aforementioned stages: *feed-forward, backpropagation, gradient descent*). We iterate over multiple epochs until the loss computed across the validation set is minimized and converges.

In developing a neural network, there are various decisions outlined above that can be re-evaluated, and the network will be retrained until the mean prediction is considered satisfactory. These are parameters than can be altered across training runs, such as: the learning rate; the activation function; the loss function; the number of neurons per layer (architecture); the probability a given neuron is omitted from calculation during a training step (dropout); whether optimzation algorithms are used to update the weights, biases and learning rate; whether the data is normalized or pre-processed prior to training. All of these are hyperparameters. To expedite the selection of these hyperparameters, we utilize the [1]`optuna` package to iterate over different hyperparameter configurations in a single epoch, choosing a configuration with the lowest validation loss.

For an emulator, the model we trained learned the mapping of the input cosmological parameters $\boldsymbol{\theta}$, to the summary statistic $\boldsymbol{y}$, obtained from a cosmological N-body simulation. The trained model compressed this mapping into a series of weights and biases, that can be applied to interpolate and predict $\boldsymbol{y}$ as $\hat{\boldsymbol{y}}$ given a set of cosmological parameters $\boldsymbol{\theta}$, that lies within the parameter ranges defined by the set of simulations used in training the model. This model was used to also solve the inverse problem, obtaining $\boldsymbol{\theta}$ with $\boldsymbol{y}$, the observable, obtained from a survey. In the next subsection, we will specify the caveats that are specific to the VG-CCF, that must be considered in preparing the training data for the model.

[1] https://github.com/optuna/optuna

3.1.3 Modelling objectives in emulating the VG-CCF

In prior treatments of the VG-CCF, an unbiased sample of voids was necessary. For the template method, if the VG-CCF was modelled post-reconstruction (removing most of the RSD from the galaxy positions), it required rescaling of the VG-CCF by the assumed RSD model in the form of $\beta = f/b$ and AP distortion parameters $\alpha_{\parallel}$, $\alpha_{\perp}$. If the VG-CCF was modelled in redshift space, then extensions to the VG-CCF model were required. In either case, proper modelling of the AP distortions and RSD model were required for an unbiased treatment. See [187] and Figure 2.4 for a discussion of this.

We wanted an emulator that models the influence of RSD and AP distortions on the void-finding process, and the subsequent VG-CCF. We also wanted our model to be capable of exploring a broad region of cosmological parameter space, modelling the cosmological dependence of the VG-CCF on $\Omega_{\rm m}$, σ_8. An emulator can neatly incorporate this information, and the selection of RSD and AP as required. Specifically, we wanted a model that solved the inverse problem, which mapped the VG-CCF (ξ_0, ξ_2) obtained from voxel voids found in a galaxy-redshift survey back to a set of cosmological parameters $\boldsymbol{\theta}$. Special emphasis was placed on quantifying the uncertainties in our model, and on the cosmological constraints we obtained with it.

In the next section, we will outline the ingredients in our neural network emulator model. We will focus on the cosmological parameters that vary in the `AbacusSummit` mock catalogues, our training data. These mocks were processed to generate the training data using both cubic periodic boxes (Chapter 4) and cutsky mocks that matched the survey volume (Chapter 5). In line with previous models of the VG-CCF, we excluded voxel voids with void-radii smaller than $R_V = 20\ h^{-1}$Mpc, as they were more likely to be spurious.

3.2 Ingredients for a neural network emulator of the VG-CCF

Having outlined how a neural network can be trained to emulate summary statistics, demonstrating the general fundamentals of neural networks, their applications in cosmology and in measuring and emulating the large-scale structure, and having also articulated the requirements of our specific model, we will now outline the ingredients that go into training this model. Every neural network emulator needs a clearly defined training, testing and validation sample, to train the network, evaluate the quality of its mean predictions and in our case, estimate the model uncertainty. We extended this treatment of the uncertainties

by describing which mocks we used in estimating the uncertainties of our model, how the covariance matrix was constructed, and how we estimated the covariance arising from the model. Following this, we then trained our neural network. The final set of decisions were centered around the training and evaluation of different iterations of neural network emulator models. These models can be differentiated by their **hyperparameters**, some of which can be quantified (e.g. learning rate, depth, number of neurons) and others which are centered around the activation functions used in the model and the loss function used when evaluating it. Once we have appropriately shown how these are selected, we can then discuss the implementations in BOSS and DESI.

3.2.1 Preparation of training data

To meet our requirements of modelling the VG-CCF across a broad swathe of parameter space, we selected the `AbacusSummit` mocks. The `AbacusSummit` simulations [149] consist of a suite of N-body cosmological simulations run with the `Abacus` code [92], designed to meet the simulation requirements of the Dark Energy Spectroscopic Instrument [69, 70]. The suite is comprised of simulations of different sizes and resolutions, as well as multiple cosmological models. The *base* simulations follow the evolution of 6912^3 dark matter particles in periodic boxes of $2\,h^{-1}$Gpc on a side, with a mass resolution of $2 \times 10^9\,h^{-1}\mathrm{M}_\odot$. Dark matter halos are identified using `CompaSO` algorithm [106], which is a hybrid Friends-of-Friends and spherical overdensity algorithm.

In this thesis, we used 85 out of the 97 available cosmologies, which covered an extended w_0w_aCDM parameter space, characterized by the physical baryon and cold dark matter densities (ω_b and ω_cdm, respectively), the amplitude of matter fluctuations in spheres of $8h^{-1}$ Mpc (σ_8), the spectral index and its running (n_s and α_s, respectively), the effective number of massless relics (N_eff), and the dark energy equation of state parameters w_0 and w_a, which govern the parameterization $w(a) = w_0 + w_\mathrm{a}(1-a)$, with a being the scale factor. In this model, ΛCDM is recovered when $w_0 = -1$ and $w_\mathrm{a} = 0$. The different simulations can be summarized as follows.

- `c000`: Baseline ΛCDM cosmology, with parameters that match the Planck 2018 best-fit model assuming base-ΛCDM [11].
- `c001`–`c004`: Secondary cosmologies, including variations with a low ω_cdm, a thawing w_0w_a model, a high N_eff, and a low σ_8.
- `c013`: A base-ΛCDM cosmology matching the Euclid Flagship 2 simulation [20].

- `c100`–`c126`: Linear derivative grid with single-parameter variations around the baseline.

- `c130`–`c181`: Emulator grid with a wider coverage of the parameter space.

All cosmologies assume a flat geometry, and the value of the dimensionless Hubble parameter h is a derived parameter, calibrated to match the Cosmic Microwave Background (CMB) angular acoustic scale, θ_*, measured by Planck 2018 [11]. These dark matter simulations were populated using a 5-parameter halo occupation distribution (HOD) model articulated in Chapter 4, the baseline HOD model defined by [242], with a 7-parameter extension explored in Chapter 5. We will discuss how the dark matter halo `AbacusSummit` mocks were populated with galaxies in those chapters. These `AbacusSummit` mocks are split into two classes: *base* and *small*. The training data was generated from *base* boxes, which are $2\,h^{-1}$Gpc on a side.

The *small* simulations, which were used for the estimation of the covariance matrix, are run with the same resolution as the *base* simulations, within periodic boxes that are 500 h^{-1}Mpc a side. These 1883 smaller boxes are only available for the baseline cosmology, `c000`, with the same resolution as the *base* catalogues, consisting of 1728^3 particles. These mocks were also run for different realizations, i.e. different random seeds, to estimate the effect of cosmic variance.

We began by splitting the populated HOD catalogues into a training, validation, and test set. We reserved the `c000-c004` and `c013` cosmologies for the test set, as they corresponded to locations in the parameter space where we expected the emulator to fail most strongly. The test sample did not need to be representative, and this should therefore be considered conservative.

As part of the preparation of these `AbacusSummit` *base* mocks, in addition to selecting HOD parameters to create catalogues with similar number densities and clustering to the survey data, we also impose selection criteria on the voids obtained from these catalogues.

Namely, for all of our catalogues, we selected voxel voids with a void radius of $R_V \geq 20h^{-1}$ Mpc. We chose this cut to remove small, spurious voids that arose due to shot noise [159, 59], and to faciliate comparisons with previous studies in the literature [234, 235]. Another consideration was the influence of the galaxy number density on the resultant VG-CCF in the catalogue. In this thesis we explored strategies to keep this fixed, to ensure that our emulator models learned the cosmological parameter relationships in the VG-CCF, without the number density of catalogues influencing it. We will discuss the efficacy of this in Chapters 4 and 5.

Parameter	Range	Interpretation
$\omega_{\rm cdm}h^2$	$(0.1032, 0.140)$	Physical cold dark matter density
$\omega_{\rm b}h^2$	$(0.0207, 0.0243)$	Physical baryon density
$n_{\rm s}$	$(0.9012, 1.025)$	Tilt of the primordial power spectrum
N_{ur}	$(1.188, 2.889)$	Number of ultra relativistic (massless neutrino) species
α_s	$(-0.038, 0.038)$	Running scalar spectral index
σ_8	$(0.678, 0.938)$	r.m.s of matter fluctuations in spheres of 8 h^{-1}Mpc
w_0	$(-1.22, -0.726)$	Present-day dark energy equation of state parameter
$w_{\rm a}$	$(-0.628, 0.621)$	Redshift evolution of the dark energy equation of state
$\log M_{\rm cut}$	$(12.19, 13.39)$	Minimum halo mass to host a central galaxy
$\log M_1$	$(13.44, 14.32)$	Typical halo mass to host one satellite galaxy
α	$(0.43, 1.71)$	Power-law index of the expended number of satellites
κ	$(-1.0, 3.8)$	Regulation of the minimum halo mass to host a satellite
$\log \sigma$	$(-0.23, 0.65)$	Slope of the transition from zero to one central galaxy
α_c^*	$(0.09, 0.57)$	Central velocity bias
α_s^*	$(0.52, 1.08)$	Satellite velocity bias

Table 3.1: Parameter space spanned by the `AbacusSummit` cosmologies used in our training data. Starred parameters are not considered when developing the training set for the BOSS CMASS emulator, and are instead set to their default values ($\alpha_c = 1, \alpha_s = 0$). For all of our parameter inference, our references to α_s will purely focus on the HOD parameter, not the cosmological parameter with the same notation, which is fixed to the fiducial value of $\alpha_s = 0$.

3.2.2 Covariance matrices

One of the challenges inherent with a neural network emulator is in thoroughly accounting for the model-driven uncertainty. Previous work with neural network emulators, particularly those that solved the inverse problem of emulating the relationship between a summary statistic $\boldsymbol{y}$ and cosmological parameters $\boldsymbol{\theta}$ found that the uncertainties estimated by the neural network were under-estimated [62, 164]. This motivated the development of a modification to the covariance matrix. This approach aims to be more conservative in estimating the amplitude of uncertainties in the predicted VG-CCF by the model.

The covariance of the VG-CCF, or any two-point summary statistic can be estimated through calculating the statistic in a series of mocks with the same cosmology and galaxy-halo connection parameters, across different realizations (i.e. different random seeds). The resultant covariance term can be estimated as

$$\mathsf{C}^{\rm sim} = \frac{1}{(N_{\rm s}-1)} \sum_{\rm i=1}^{N_{\rm s}} (\bar{\xi} - \xi_i)(\bar{\xi} - \xi_i)^T \tag{3.8}$$

In our model, we followed [62], where they calculated an additional term in their covariance, defined by Equation (3.9). The covariance in our model was computed as the sum of simulation covariance C^{sim} from a collection of mock catalogues (`AbacusSummit` *small* for BOSS, `EZmocks` for DESI), in addition to this term.

Since our emulator was trained from `AbacusSummit` *base* boxes with a fixed phase realization, its predictions did not factor in the contributions from cosmic variance due to the finite simulation volume. The simulation covariance, $\mathsf{C}^{\rm sim}$ was computed to account for this, and used several simulations with the same cosmological model and fixed HOD model, to quantify the uncertainty that arose from cosmic variance. The simulation covariance $\mathsf{C}^{\rm sim}$ was typically defined in terms of the mean of the data $\bar{\xi}$ across all samples, N_s, with respect to an individual datum ξ_i, per Equation 3.8.

To study the systematic error in our model due to an imperfect emulation, we examined six test cosmologies with 100 HODs each, a sample of 600 mocks in total. The test cosmologies included `c000`, a baseline ΛCDM model using the Planck 2018 parameters; `c001`, a ΛCDM model using the WMAP9+ACT+SPT parameters from [39]; `c002`, a wCDM model with thawing parameters; `c003` a ΛCDM model with a higher $N_{\rm eff}$; `c004` a baseline ΛCDM model with low σ_8; and `c013`, a simulation run with the Euclid Flagship 2 ΛCDM cosmology. For each of these mocks, we made model predictions using the true parameters of the simulations, and compared them against the measured multipoles from the simulations.

The emulator covariance is a term that arose from the observation in [62] that uncertainty in the predictions were consistently underestimated by the emulator. To compensate for this underestimation, a more conservative definition for emulator uncertainty was defined. To quantify the emulator's contribution to the total covariance, we computed the covariance matrix associated with the difference between the model predictions and the test measurements, $\Delta\xi = \xi^{\rm pred} - \xi^{\rm data}$, and $\xi = (\xi_0^{vg}, \xi_2^{vg})$, and concatenated the monopole and quadrupole into a single data vector:

$$\mathsf{C}^{\rm emu} = \frac{1}{n_{\rm test}} \sum_{i=1}^{n_{\rm test}} \left(\Delta\xi - \overline{\Delta\xi}\right) \left(\Delta\xi - \overline{\Delta\xi}\right)^T \tag{3.9}$$

Here, an overline denotes the average data vector across all 600 mocks, following [62]'s conservative definition. The variance of these residuals calculated across the full set of test simulations was sensitive to outliers in the distribution, which could have appeared due to the emulator being less accurate at extreme regions of the parameter space that were not densely covered by the training data. This potentially overestimated the true emulator uncertainty in regions of the parameter space that were well sampled by our simulations.

3.2.3 Choice of model architecture

Once the training data and covariance mocks had been selected, the neural network model could be trained. What remained was the selection of several parameters that underpinned the model. The pre-processing and formatting of this data (which parts of the VG-CCF were sampled, what cosmological parameter space was considered) and the neural network's architecture were both crucial design decisions. Extensive work was required to converge on an architecture that minimized the loss function, and also passed the tests we devised for this emulator, which were defined case-by-case for BOSS in Chapter 4 and DESI DR1 in Chapter 5. We will summarize the broad steps of this method below.

To begin, we first selected an activation function for our neural network. Recall from our overview in the previous section, that neural networks can employ a wide zoology of activation functions, as these are a broad class of differentiable nonlinear functions, often spanning $y \, \epsilon \, (-1, 1)$. Here, we will enumerate the various activation functions we considered in the models for this thesis and used in the development of these models: Sigmoids,

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{3.10}$$

Learned Sigmoids,

$$\sigma_L(x) = \frac{\beta(1-\beta)x}{1+e^{-\alpha x}} \qquad \alpha, \beta \; \epsilon \mathcal{R} \tag{3.11}$$

rectified linear units (`ReLUs`) and sigmoid linear units (`SiLUs`),

$$ReLU(x) = \begin{cases} x & , \text{if } x \geq 0 \\ 0 & , \text{if } x < 0 \end{cases} \tag{3.12}$$

$$SiLU(x) = x\sigma(x) \tag{3.13}$$

and `tanh`.

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{3.14}$$

This is not an exhaustive list, merely an overview of the common forms of activation functions. Activation functions often have the property of having well-defined derivatives $\forall\ x$, and $y\ \epsilon\ \Re, \forall\ x$.

The appropriate choice of activation function varies across implementations of neural networks, hence the focus on iterating across different models until we obtained one that, in tandem with the other modelling decisions outlined below, satisfied our requirements for accurate modelling of the VG-CCF.

With the training data and activation function chosen, we had enough information to complete the *feed-forward* step, but to evaluate the quality of our predictions, we needed a loss function, $\mathcal{L}$, to evaluate the proximity of the prediction $\hat{\boldsymbol{y}}$ to the truth $\boldsymbol{y}$. The loss function is the last major decision that can be made in the construction of a neural network. Much like with activation functions, the optimal choice of a loss function is entirely dependent on the training data, and how it should be scored. We explored a handful of loss functions for the work in both Chapter 4 and Chapter 5. These included the L1 loss,

$$\mathcal{L} = \sum_{i=1}^{N} |y - \hat{y}| \qquad \text{L1 loss} \tag{3.15}$$

the L2 or Euclidean loss,

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} (y - \hat{y})^2 \qquad \text{L2 loss} \tag{3.16}$$

and the Gaussian loss function, where var and ϵ are prescribed as hyperparameters.

$$\mathcal{L} = \frac{1}{2}\left(\log(\max(var, \epsilon) + \frac{y - \hat{y}}{\max(var, \epsilon)}\right) \text{ Gaussian loss} \tag{3.17}$$

Once a loss function was chosen, we executed the *backpropagation* step, where the partial derivatives are computed. These gradients were then applied during *stochastic gradient descent*, alongside the learning rate, η, a hyperparameter. This sequence of *feed-forward, backpropagation, gradient descent* encompasses a single iteration of the training loop. The training loop is iterated until the loss function computed with the validation sample was sufficiently minimized. Our neural network used a few more hyperparameters, that each affected different parts of the training loop.

There are several more quantifiable hyperparameters explored which included the dropout rate (probability that a given a neuron in the network is excluded from a training-epoch); the number of neurons per layer; and the whether the number of neurons per layer stayed fixed. All of these altered the architecture of the network, and were optimized by using the `optuna` [14] package. In our model, each of these hyperparameter combinations was randomized, the loss was calculated, and the architecture with the lowest overall validation loss in each epoch was chosen as the model. We iterated over hundreds of epochs until we converged on an architecture that was most likely to produce an accurate prediction of the mean, informed by minimization of the loss.

We will then specify in each of the following chapters the best fitting models; their activation functions, their loss, and describe the specific differences in how different mocks were used in the training, testing and evaluation of the model for both the BOSS and DESI data.

Chapter 4

Emulation of the VG-CCF in BOSS

In this chapter and the subsequent chapter, we will outline the methodology and the results of implementing a neural network emulator of the VG-CCF, and the considerations that arose from utilizing the BOSS CMASS sample (this chapter) and DESI Y1 (next chapter).

4.0.1 SDSS-III: BOSS CMASS

This analysis used a subsample of galaxies obtained from the Sloan Digital Sky Survey (SDSS) [81]. SDSS-III was the third iteration of a series of spectroscopic surveys undertaken using the Sloan Foundation's 2.5m telescope at Apache Point Observatory in New Mexico, USA. The telescope utilized five optical bands *ugriz* across a 7 deg^2 field-of-view. SDSS-III utilized two newly upgraded spectograph units, where each of them were fed with 500 fibres to measure the spectra of 1000 objects at once. The positioning of the fibres was pre-determined using the tiling process described in [28], where individual targets were assigned to a specific fibre on a specific *plate*, with a series of aluminum *plates* manufactured to allow for the redshifts of these objects to be measured (see Figure 4.1 as an example). The threading of these fibres, and the subsequent treatment of the fibre positioners in DESI will have an effect on the systematics of this survey (fibre-collision).

One of the primary science objectives of SDSS-III was to obtain a precise measurement of the BAO feature [81]. The Baryon Oscillation Spectroscopic Survey (BOSS) program was an observing program undertaken over the course of six years, from Summer 2008 to Spring 2014, to obtain almost 1.5 million galaxy redshifts across nearly $10,000$ deg^2 of the sky, split across two hemispheres: the North Galactic Cap (NGC) and the South Galactic Cap (SGC). These galaxies were split across two samples based on their target selection

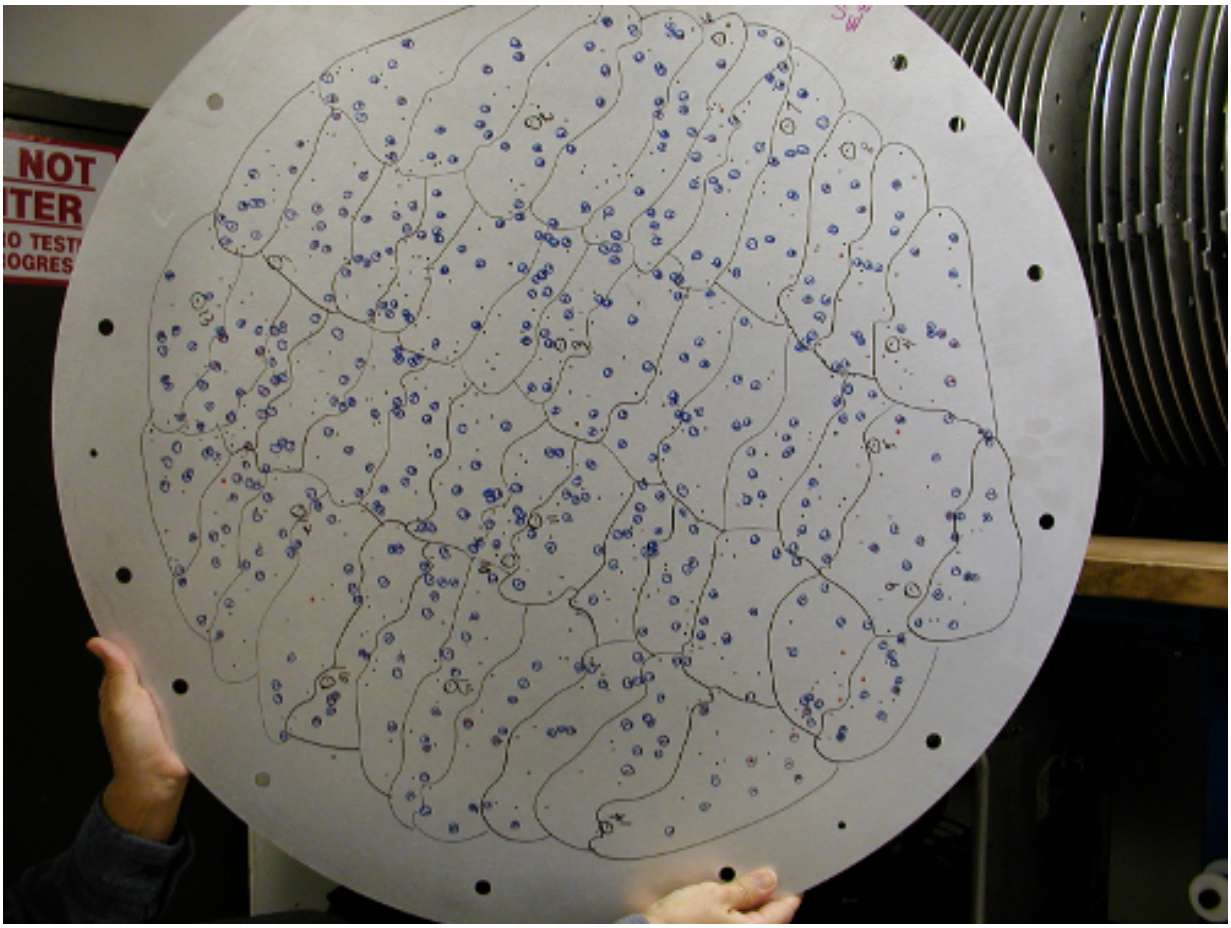

Figure 4.1: An aluminum fibre plug plate from SDSS-III. Each plate was manufactured with a unique configuration of holes that would contain fibres to observe several hundred objects simultaneously. Image credit: [147]

algorithm: LOWZ ($0.2 \leq z \leq 0.4$) and CMASS ($0.4 \leq z \leq 0.6$) [80, 189]. The CMASS sample was constructed to select a population of luminous red galaxies (LRGs) that had a consistent mass limit of $M > 10^{10} M_{\odot}$. This sample was constructed with a high and consistent completeness [189]. The definitions for *completeness* vary across surveys, but they each describe the number of objects observed in a given angular region of the survey compared to the number of objects that were expected to be observed. The CMASS sample had a mean completeness of 98.8% (as defined by [67]), with little angular variation, as Figure 4.2 shows.

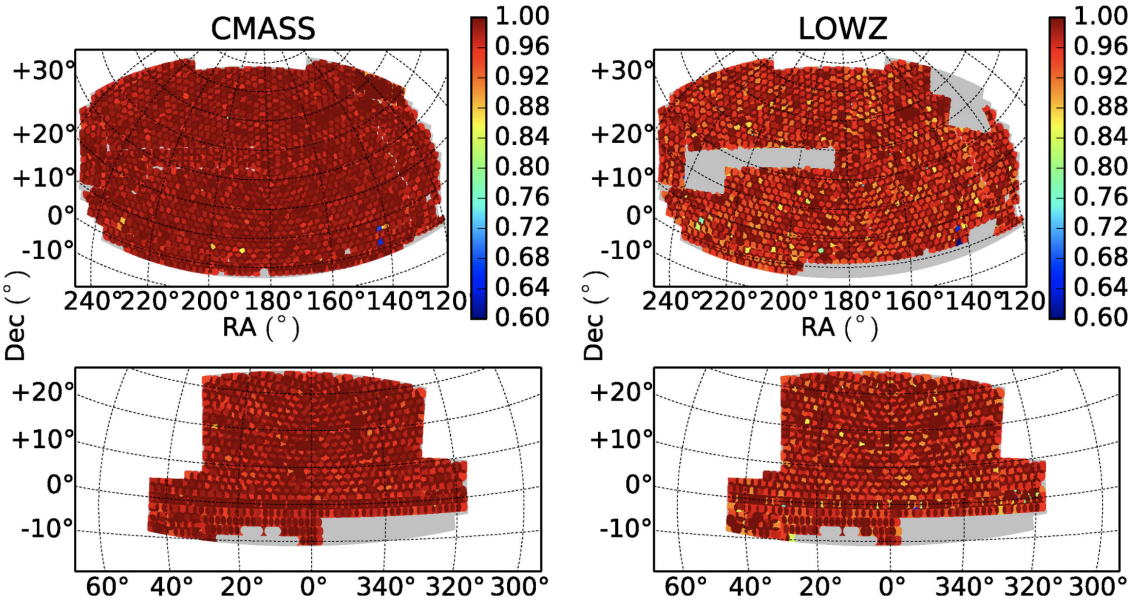

Figure 4.2: Completeness maps of the CMASS and LOWZ samples from BOSS program. Each patch of colour corresponds to a *plate* from the survey. BOSS utilized well over 2000 unique plates [67].

The BOSS CMASS sample, in this thesis, refers to the population of LRGs obtained from the CMASS sample within the BOSS program. This sample includes both the North and South galactic caps, and spans a redshift of $0.4 < z < 0.6$, a subset with a relatively flat number density. The full BOSS CMASS catalogue was described in DR12 of the SDSS [67, 189]. The relatively uniform number density, and the high completeness were significant factors in constructing our training sample.

4.0.2 Preparation of BOSS CMASS training data

Our emulator training data used the `AbacusSummit` mocks, which are a suite of N-body simulations containing dark-matter halos. We utilized 85 different cosmologies, and populated them with a halo occupation distribution (HOD) model. Populating these `AbacusSummit` mocks with an HOD model creates a catalogue that contains the comoving positions of galaxies (x_{real},y_{real},z_{real}), given in h^{-1}Mpc, and the peculiar velocity components, v_x,v_y,v_z. This information will be used in Equation (4.4), Equation (4.3) to construct a redshift-space catalogue by selecting a line-of-sight direction.

We also used the `Nseries` mocks to test the robustness of our model's predictions prior to analyzing the BOSS data (as they were constructed to match the BOSS footprint). The data covariance $\mathsf{C}^{\rm data}$ was estimated using `Patchy` mocks.

`Nseries` mocks

The `Nseries` suite of simulations [95] consist of a series of seven independent cubic periodic box mocks and 84 pseudo-independent cutsky mock catalogues designed to match the footprint, clustering and radial selection of the BOSS DR12 CMASS galaxy sample. These cubic box mocks were run in a periodic box 2.6 h^{-1}Gpc a side with a mass resolution of 1.5×10^{11} $M_\odot$. The halo catalogues are populated with an HOD model, with parameters chosen to best resemble the BOSS DR12 CMASS sample. The boxes are trimmed and rotated to produce 12 cutsky mocks that match the geometry of the CMASS sample, yielding 84 catalogues, and are run through the fiber assignment code used for BOSS to incorporate the effects of fiber collision [108]. The `Nseries` mocks adopt a flat ΛCDM cosmology with $\Omega_{\rm m} = 0.286$, $\Omega_{\rm b} = 0.0470$, $h = 0.70$, $\sigma_8 = 0.82$, and $n_{\rm s} = 0.96$.

`Patchy` mocks

We used the BIG MULTIDARK-PATCHY suite of mock catalogues, which we refer to as `Patchy` mocks [133, 134, 198] to construct the data covariance. The approximated lightcone mocks are constructed through using Augmented Lagrangian Perturbation Theory (ALPT), which combines Lagrangian perturbation theory for larger scales and a spherical collapse halo model at smaller scales to obtain a halo catalogue. These halos are then populated using a subhalo abundance matching (SHAM) code `Patchy` which is calibrated to match the 1,2, and 3-point clustering and bias of the BOSS LRG survey and the BIG MULTIDARK simulations. The mocks adopt the flat ΛCDM Planck2013 fiducial cosmology, given as:

$\Omega_m = 0.307115$, $\Omega_b = 0.048$, $\sigma_8 = 0.8288$, $n_s = 0.9611$, $h = 0.6777$. These periodic boxes are then transformed into lightcone mocks as described by [134], and have been configured to match the angular footprint, halo occupation and redshift distribution of the BOSS DR12 survey. The lightcone mocks are then assembled by using the process described in [198], wherein each of the redshift snapshots transformed to match the angular footprint and redshift distribution of BOSS DR12, with the same completeness and fiber collision weight codes applied. This was done to create mock lightcones of the LOWZ sample, the LOWZ+CMASS sample, and the CMASS sample, for both galactic caps (NGC, SGC). In this work, we used the CMASS sample of mocks. These 2048 mocks include both stellar mass completeness weights and fiber collision weights. These mocks were used to compute our data covariance, for both the `Nseries` cutsky void-galaxy cross-correlation function (VG-CCF), and for the CMASS void-galaxy cross-correlation function, across the redshift window of $0.45 < z < 0.60$.

Galaxy-halo connection

We linked between dark matter haloes and galaxies using a halo occupation distribution (HOD) prescription. The HOD is a statistical model that assigns galaxies to dark matter halos based on properties of the host halos, most importantly the halo mass. Dependencies on other secondary halo properties can be included, such as concentration and local environment.

In this work, we adopted the base halo model from [242], in which the expected number of central galaxies on halos of mass M can be written as

$$\langle N_c \rangle(M) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{\log M - \log M_{\mathrm{cut}}}{\sqrt{2}\sigma}\right)\right), \tag{4.1}$$

where $\mathrm{erf}(x)$ is the error function, M_{cut} is the minimum mass required to host a central galaxy, and σ is the slope of the transition from having zero to one central galaxy. The expected number of satellite galaxies is given by

$$\langle N_s \rangle(M) = \langle N_c \rangle(M)\left(\frac{M - \kappa M_{\mathrm{cut}}}{M_1}\right)^{\alpha}, \tag{4.2}$$

where κM_{cut} corresponds to the minimum mass required to host a satellite galaxy, M_1 is the typical halo mass to host a single satellite, and α is the power law index for the number of satellites.

We generated 8,500 HOD parameter combinations in a Latin hypercube, and we assigned 100 parameters to each of the `AbacusSummit` cosmologies. We then populated the

`AbacusSummit` halos with galaxies using `AbacusHOD` [236], which is a highly efficient and modular `python` code that implements multiple HOD prescriptions.

Whenever possible, our HOD catalogues were constructed to match the observed number density of galaxies from the CMASS sample at $0.45 < z < 0.6$, which is close to $3 \times 10^{-4}\ h^3\mathrm{Mpc}^{-3}$. When a certain combination of HOD parameters gives a number density of galaxies that is above our target, we randomly subsampled the catalogue to match the target, invoking an incompleteness fraction f_{ic}. HODs with number densities below the target were kept as is, but they constitute a small fraction of the total number of HODs in our training sample.

Populating the mocks

Since our goal was to train an emulator from simulations that reproduced the clustering patterns of the real data, we incorporated RSD and the AP effects into the catalogues prior to void-finding.

For RSD, since we are using cubic boxes, we created three different RSD catalogues for each mock. We calculated the coordinates with RSD by using Equation (4.3) and Equation (4.4) for all three light-of-sight (LOS) directions $(\hat{x}, \hat{y}, \hat{z})$.

$$x_{\mathrm{rsd}} = x_{\mathrm{real}} + \frac{v_{x,\mathrm{pec}}}{a(z)H(z)} \quad LOS = \hat{x} \tag{4.3}$$

$$y_{\mathrm{rsd}} = y_{\mathrm{real}} \tag{4.4}$$

$$z_{\mathrm{rsd}} = z_{\mathrm{real}} \tag{4.5}$$

For AP, we manually distorted the box coordinates using the factors from Equation (1.70), before void-finding. More specifically, we rescaled the box dimensions and galaxy coordinates across and along the light-of-sight by a factor $q_\perp$ and $q_\parallel$, respectively. These factors are determined for each mock and depend on the mock true cosmology and redshift, and our choice of fiducial cosmology, which is taken to be the `c000 AbacusSummit` cosmology. A natural consequence of this stretching was that the simulation boxes were not necessarily cubic anymore, although they maintained periodicity across each face, with a period that was now given by the box length across that dimension times the corresponding AP stretching factor. Once the boxes were distorted, we only ever used the fiducial cosmology to perform subsequent actions, mimicking the process applied to actual data, where only a single fiducial cosmology is used.

We performed void-finding using the `voxel` algorithm implemented in the [1]`Revolver` code. `voxel` identified underdense regions in simulations or real data without making assumptions about the shapes of voids, while still maintaining computational efficiency. The algorithm started by painting galaxy positions to a rectangular grid[2] in comoving Cartesian coordinates, with a cell resolution $R_{\rm cell}$, which we set to $5h^{-1}$Mpc. This was converted to an overdensity grid, either by dividing by the average number of galaxies per cell in the periodic-box case, or by a grid that was painted with a random catalogue that follows the survey geometry when working with observational data and the `Nseries` mocks. In the latter case, we masked those cells for which the number of random particles is below 75% of the average number of randoms per cell. The overdensity field was then smoothed with a Gaussian filter of a characteristic length R_s, which we set to 10 h^{-1}Mpc in this work. A watershed algorithm was then applied to classify the cells into different watershed regions. At this stage, each of those watershed regions was considered an individual void.

The center of a void was defined as the center of a voxel that has the lowest density among all voxels comprising the void. We also defined an effective radius for each void, which is the radius of a sphere that has the same volume as the sum of the volumes or all voxels or cells that composed the void. The final void catalogue was then constructed by removing those voids with radii below 20 h^{-1}Mpc. We had chosen this cut to remove small, spurious voids that arise due to shot noise [159, 59], and to ease comparison with previous studies in the literature [234, 235].

4.0.3 Void-galaxy cross-correlation

Our main observable was the cross-correlation function between the void centers and the redshift-space galaxy positions, which we call the void-galaxy cross-correlation function (VG-CCF). This was calculated using [3]`pycorr`, which is a wrapper around a modified version of the `CorrFunc` pair-counting code [213]. The VG-CCF was estimated using the natural estimator ξ_N for the periodic simulation boxes where we can analytically calculate the expected density, and the Davis-Peebles estimator ξ_{DP} [66] for survey-like cutsky mocks and the CMASS data, Equation (2.5). We used the Davis-Peebles estimator rather than the Landy-Szalay estimator to avoid having to create an additional random catalogue with a distribution matching that of the voids. We binned the VG-CCF in s and μ, where

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

[2] When applying the void-finder to the CMASS data or the HOD mocks that have been stretched by the AP effect, this grid was generally non-cubic. We have modified the void-finder to handle non-cubic grid configurations, maintaining the periodic boundary conditions across each axis in the simulations case.

[3] https://github.com/cosmodesi/pycorr

s is the scalar void-galaxy pair separation, and μ is the cosine of the angle between the light-of-sight and the vector that connects the pair. We used 30 linearly-spaced bins in s from 0.7 h^{-1}Mpc to 120 h^{-1}Mpc, and 240 linearly-spaced bins in μ from -1 to 1. We then decomposed the correlation function into multipoles, using Equation (1.60).

We calculated the VG-CCF for each of the `AbacusSummit` catalogues described in the previous section. When doing so, we applied the RSD along the three simulation axes and measured the VG-CCF for each of them separately, and averaged the clustering measurements to increase the signal-to-noise ratio, as in [215]. We additionally measured it for the `Nseries` and `Patchy` mocks, as well as the CMASS data.

4.0.4 Neural network model design

We built an emulator of the VG-CCF using the HOD catalogues described in Section 3.2.1. We trained a fully-connected neural network to learn the dependence of the VG-CCF on the cosmological and HOD parameters using the `sunbird`[4] package [62]. We began by splitting the HOD catalogues into a training, validation, and test set. We reserved the `c000-c004` and `c013` cosmologies described in Section 3.2.1 for the test set, as they correspond to locations in the parameter space where we expected the emulator to fail most strongly. The test sample did not need to be representative, and this should therefore be considered conservative. From the remaining cosmologies, we randomly split the simulations, leaving 74 of them for training, and 5 for validation. The inputs of the neural network are the cosmology and HOD model parameters, normalized to be within 0 and 1, and the output are the concatenated monopole and quadrupole moments of the VG-CCF, also normalized. We used Sigmoid Linear Units (SiLU) as the activation function and adopted a mean squared error as the loss function (Equation (3.16), where N is the batch size, which we set to 256). We optimised the neural network hyperparameters using `optuna`[5], varying the learning rate, number of layers, number of hidden units per layer, dropout rate, and the weight decay, and selected the model that minimizes the validation loss. These hyperparameters were used in regularization routines that help improve a neural network's generalization ability by reducing overfitting, and used in this implementation and in Chapter 5.

[4] https://github.com/florpi/sunbird
[5] https://github.com/optuna/optuna

Parameter	Prior	Interpretation
$\omega_{\rm cdm}$	$\mathcal{U}(0.1032, 0.140)$	Physical cold dark matter density
$\omega_{\rm b}$	$\mathcal{U}(0.0207, 0.0243)$	Physical baryon density
$n_{\rm s}$	$\mathcal{U}(0.9012, 1.025)$	Tilt of the primordial power spectrum
σ_8	$\mathcal{U}(0.678, 0.938)$	r.m.s of matter fluctuations in spheres of 8 h^{-1}Mpc
w_0^*	$\mathcal{U}(-1.22, -0.726)$	Present-day dark energy equation of state parameter
$w_{\rm a}^*$	$\mathcal{U}(-0.628, 0.621)$	Redshift evolution of the dark energy equation of state
$\log M_{\rm cut}$	$\mathcal{U}(12.4, 13.3)$	Minimum halo mass to host a central galaxy
$\log M_1$	$\mathcal{U}(13.2, 14.4)$	Typical halo mass to host one satellite galaxy
α	$\mathcal{U}(0.7, 1.5)$	Power-law index of the expended number of satellites
κ	$\mathcal{U}(0.0, 1.5)$	Regulation of the minimum halo mass to host a satellite
$\log \sigma$	$\mathcal{U}(-3.0, 0.0)$	Slope of the transition from zero to one central galaxy

Table 4.1: Prior distributions used during the parameter inference stage. Parameters marked with a star are fixed to their default base-ΛCDM values when not considering these model extensions.

4.0.5 Covariance estimation

In order to estimate the covariance between the data and the emulator-based model, we followed [62] and split the covariance into three components that arose from: the ability of the emulator to reproduce the same simulations of realization used for training $\mathsf{C}^{\rm emu}$, the difference between the simulations of single-realization and the true model $\mathsf{C}^{\rm sim}$, and the effects of cosmic variance and shot noise in the data $\mathsf{C}^{\rm data}$. Our covariance was calculated as the sum of these three terms, per Equation (4.6). We will elaborate on the definitions presented in Section 3.2.2, as each iteration of the emulator had its nuances.

$$\mathsf{C} = \mathsf{C}^{\rm emu} + \mathsf{C}^{\rm sim} + \mathsf{C}^{\rm data} \tag{4.6}$$

Data covariance

To construct the data covariance for the BOSS emulator, we used the BIG MULTIDARK-PATCHY suite of mock catalogues, which were referred to as `Patchy` mocks [133, 134, 198].

We computed the variance associated with the data vector comprising sample variance and shot noise by using the 2048 realizations of the BIG MULTIDARK-PATCHY lightcone mocks, and computed the monopole and quadrupole of each realization, across a redshift window of $0.45 < z < 0.60$. These mocks are made to match the clustering, angular footprint and number density of the BOSS DR12 samples, and we used the CMASS NGC

sample. We computed the redshift-space VG-CCF with Davis-Peebles estimator and used their random catalogue with 50 times the galaxy number density. We followed the binning of void-galaxy separation of [234]: 30 bins for $0.7 \leq s \leq 120$ h^{-1}Mpc. The resulting covariance $\mathsf{C}^{\rm data}$ is displayed in the right panel of Figure 4.3.

Simulation covariance

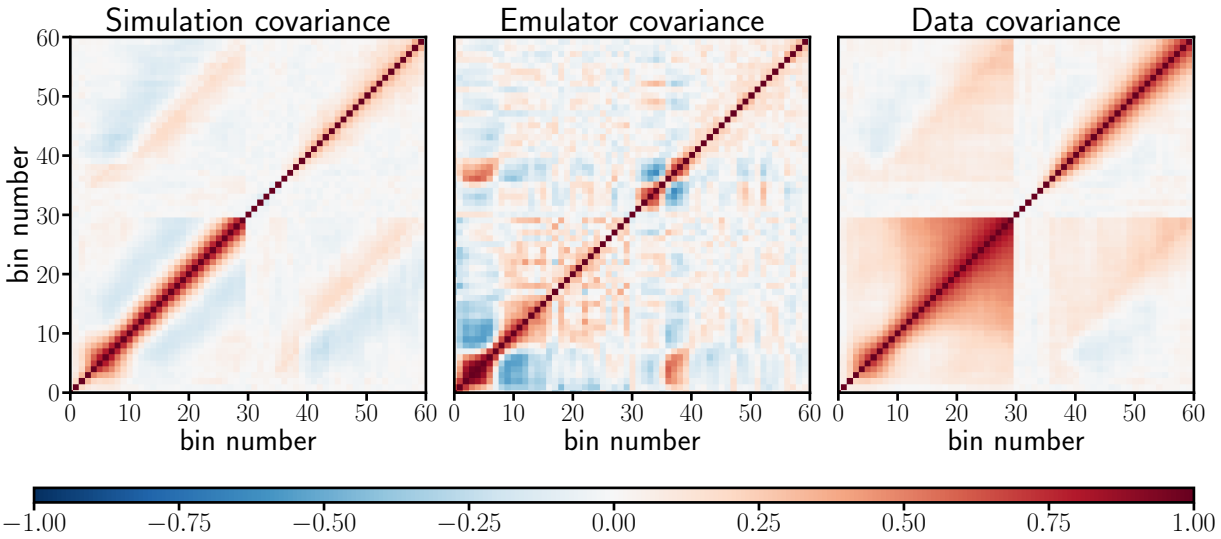

Figure 4.3: Correlation matrices associated to the different components that contribute to the total error budget in our likelihood analysis. The simulation covariance (left) accounts for cosmic variance due to the finite volume of the `AbacusSummit` simulations used to train our model. The emulator covariance (center) quantifies the systematic error in our model due to an imperfect emulation of the multipoles of the void-galaxy cross-correlation function. The data covariance (right) was computed from the monopoles and quadrupoles of the void-galaxy cross-correlation function, obtained from 2048 `Patchy` mocks $0.45 < z < 0.60$.

We used the `AbacusSummit` *small* boxes, which were run with the baseline `AbacusSummit` cosmology and included 1883 different phase realizations. We calculated the VG-CCF multipoles from each realization using a fixed HOD model, with parameters calibrated to minimize the χ^2 between the predicted multipoles and those measured from the BOSS CMASS sample. The procedure for measuring the CCFs followed the description from Section 4.0.3, averaging along each of the three lines of sight for each box. The covariance was then calculated using the standard method Equation (3.8), shown in the left panel of Figure 4.3.

The covariance was computed using the concatenated monopole and quadrupole of the VG-CCF, hence the divisions in the structure at bin 30. Before appending this component into the overall covariance matrix, we rescaled the covariance to match the volume of an `AbacusSummit` *base* box, dividing the term by a factor of 64. While such a volume scaling is approximately correct, it neglects some nuances such as the super-sample covariance, because both smaller boxes and larger boxes have mean overdensity set to exactly the expected background level, rather than allowing for fluctuations in this [123]. At the level of our analyses, we do not expect these issues to be important.

Emulator covariance

To study the systematic error in our model due to an imperfect emulation, we examined six test cosmologies with 100 HODs each, giving us a sample of 600 mocks in total. These test cosmologies were outlined in Section 4.0.5. For each of these mocks, we made model predictions using the true parameters of the simulations, and compared them against the measured multipoles from the simulations. Figure 4.4 shows how well the model reproduced the data, where we have averaged the measurements for each cosmology across all 100 HOD models. We noted the presence of small scale differences between the data and the model exceeding 3σ for a few bins, particularly in `c002, c003, c013`. These variations will be accounted for in the emulator covariance, which we discuss below.

In Figure 4.5, we show the 68 and 95 percentiles for the difference between the model prediction and the test data across all mocks. To assess how this systematic model uncertainty compared to the errors from the real observations, the vertical axis is shown in units of the standard deviation of the CMASS multipoles, as measured from the `Patchy` covariance mocks. Our emulator achieved high accuracy at most scales, with the emulation error being smaller than the error from the data across a wide scale range. The error budget was dominated by the systematic emulator uncertainty at small scales ($s \leq 20\ h^{-1}$Mpc) and limiting the improvement from the increased data precision in this regime. The fractional accuracy of the emulator predictions remained similar across the full scale range, and therefore, it becomes dominant with respect to the data error as we approach smaller separations. Future work might be able to improve on this aspect by applying non-linear transformations to the input training data or weight scales differently during the training process.

To quantify the emulator's contribution to the total covariance, we computed the covariance matrix associated with the difference between the model predictions and the test measurements, given by Equation (3.9), concatenating the monopole and quadrupole into a single data vector.

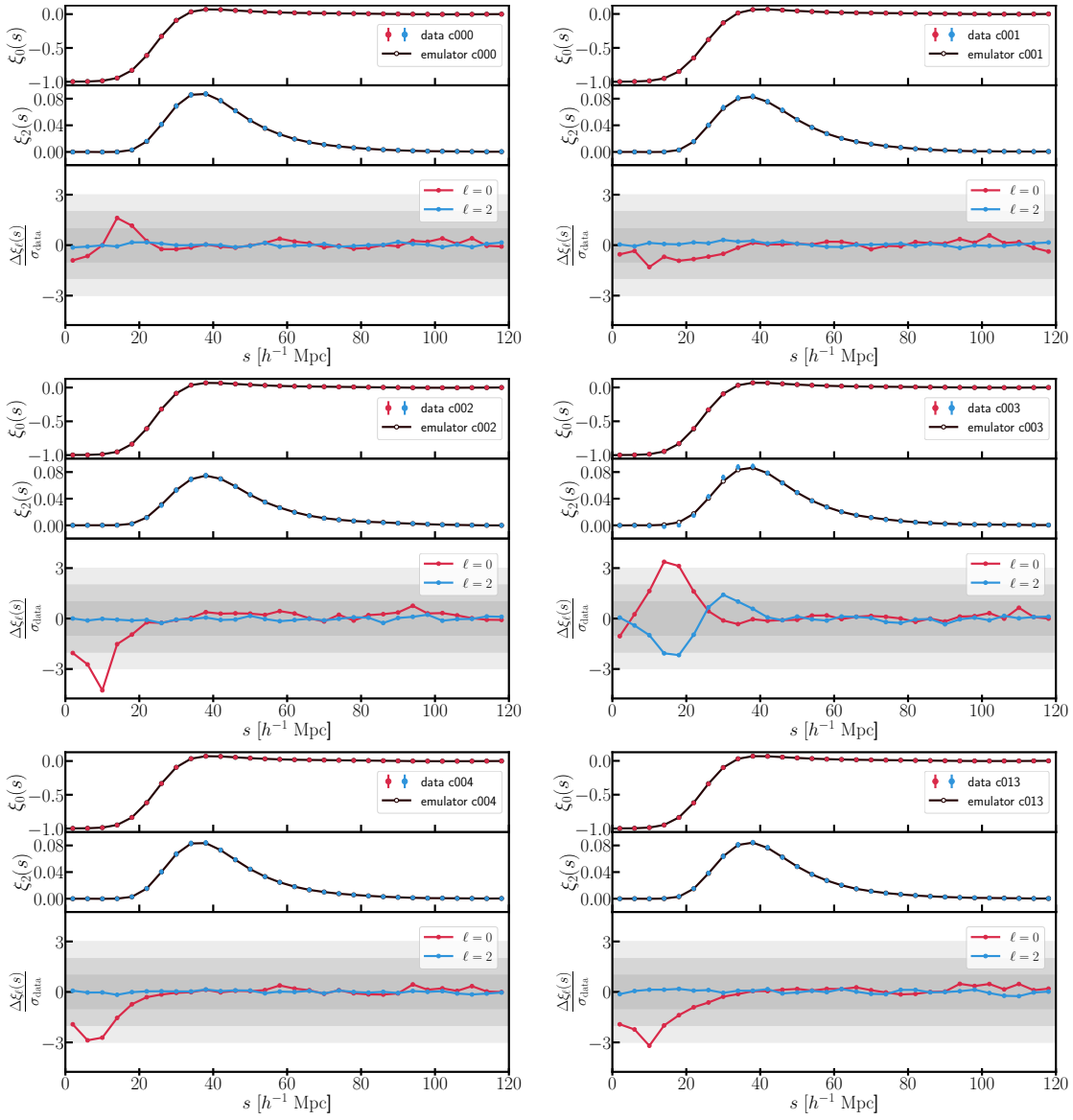

Figure 4.4: Emulator predictions of the monopole and quadrupole, for each of our six test cosmologies and averaged across 100 HODs, with their residuals in the panel below. We compute the covariance from the volume-adjusted `AbacusSummit` 500 h^{-1}Mpc covariance box, scaled to match a single measurement of the monopole and quadrupole from a 2 h^{-1}Gpc box. Because we stack voids this conflated the interior of larger voids with the edges of smaller voids, and which obscured the behaviour of individual voids of different sizes from this plot.

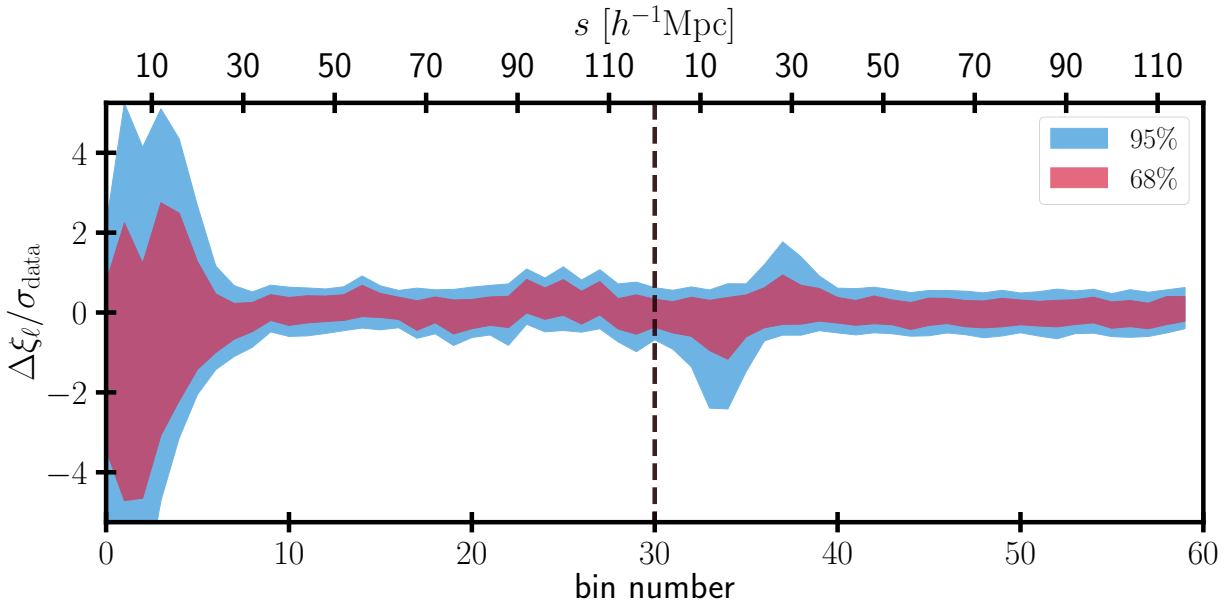

Figure 4.5: 68 and 95 percentiles (red and blue, respectively) associated to the difference between the emulator predictions and the test data for the VG-CCF monopole (bin numbers from 0 to 30) and quadrupole (30 to 60), measured across 600 test simulations. This is shown in units of the error bars from the CMASS data, calculated from the `Patchy` covariance mocks. The bin numbers associated to each multipole cover a scale range $0.7 \leq s \leq 120$ h^{-1} Mpc, shown on the upper x-axis. We note that the performance of the emulator is worse for bins below 25 h^{-1}Mpc for the monopole and 30 h^{-1}Mpc for the quadrupole. We will discuss the emulator error in subsequent sections and its importance to future emulator work.

The resulting covariance matrix is shown in the middle panel in Fig. 4.3. While this matched the procedure followed in [62], this estimation of the emulator error can be deemed conservative. The variance of these residuals calculated across the full set of test simulations is sensitive to outliers in the distribution, which appeared to be due to the emulator being less accurate at extreme regions of the parameter space that were not densely covered by the training data. This potentially overestimated the true emulator uncertainty in regions of the parameter space that were well sampled by our simulations.

4.0.6 Inference with the emulator and MCMC

We used the Markov Chain Monte-Carlo (MCMC) technique to explore the posterior surface for the parameters of interest. We used the [6]`numpyro` implementation of the adaptive Hamiltonian Monte Carlo (HMC) sampler. We cross-referenced the posteriors obtained with HMC with other samplers, and find strong agreement.

For the cosmology and HOD parameters, we kept our priors uniform, as specified by Table 4.1. For the cosmological parameters, these match the bounds defined by `AbacusSummit` for their simulations. We evaluated the likelihood

$$\log \mathcal{L} = \left(\mathrm{X}^{\mathrm{obs}} - \mathrm{X}^{\mathrm{theo}}(\mathcal{C}, \mathcal{G})\right) \mathsf{C}^{-1} \left(\mathrm{X}^{\mathrm{obs}} - \mathrm{X}^{\mathrm{theo}}(\mathcal{C}, \mathcal{G})\right)^{\mathrm{T}} , \tag{4.7}$$

where $\mathrm{X}^{\mathrm{obs}}$ is the measured data vector. $\mathrm{X}^{\mathrm{theo}}(\mathcal{C}, \mathcal{G})$ is the model predicted by the emulator for cosmological parameters $\mathcal{C}$ and HOD parameters $\mathcal{G}$. The covariance C was adjusted by applying the correction proposed by [173] to each component of the covariance according to how it was calculated.

4.0.7 Angular CMB scale prior

Each `AbacusSummit` simulation used to train our emulator adopted a value of dimensionless Hubble parameter h that gave the same angular CMB peak position θ_*, matching the best-fit value ($100\theta_* = 1.0435133$) obtained by Planck 2018 [11]. We imposed a prior on θ_* to ensure that our emulator is used within the same parameter space as it was trained. Importantly, using `AbacusSummit`, we could only emulate cosmologies that matched this value of θ_*. Consequently, we did not have h as a free parameter in our fits, but determined it at each chain point from the other parameters and the fixed value of θ_*.

[6] https://github.com/pyro-ppl/numpyro

sim	ω_b	ω_{cdm}	N_{ur}	n_s	α_s	σ_8	w_0	w_a
c000	0.02237	0.1200	2.0328	0.9649	0.0	0.807952	-1.0	0.0
c001	**0.02242**	**0.1134**	2.0328	**0.9638**	0.0	**0.776779**	-1.0	0.0
c002	0.02237	0.1200	2.0328	0.9649	0.0	**0.808189**	**-0.7**	**-0.5**
c003	**0.02260**	**0.1291**	**2.6868**	**0.9876**	0.0	**0.855190**	-1.0	0.0
c004	0.02237	0.1200	2.0328	0.9649	0.0	**0.749999**	-1.0	0.0
c013	**0.02200**	**0.1206**	2.0328	**0.9600**	0.0	**0.813715**	-1.0	0.0

Table 4.2: Table of cosmologies used in the test set. Bold values highlight changes from the fiducial `c000` case. The full description of these simulations, the cosmologies used and the way these simulations are obtained can be found in [92].

When we compared the emulator's posteriors with those obtained from the template method of [234, 235], we had to apply the same prior on θ_* to the template's MCMC chains. In both the case of the emulator and the template method, the inclusion of CMB information through θ_* is a significant source of constraining power. This meant that our constraints are not exclusively sourced from late-time void-galaxy clustering, but also from the CMB. No CMB information other than the constraint on θ_* was included in this analysis, and we note that the prior on θ_* is well-motivated, as it is well constrained by Planck to within 0.03% precision, and is robust under different cosmological models [11].

4.1 Testing the pipeline using mock catalogues

We trained the emulator as specified in Section 3.2, and quantified the emulator's ability to match the concatenated monopole and quadrupole for the `AbacusSummit` test cosmologies through $\mathsf{C}^{\mathrm{emu}}$, as described in Section 4.0.5.

We considered the ability of the pipeline to recover the cosmological parameters of interest. We wanted to examine to what extent a successful emulation of the monopole and quadrupole of the VG-CCF would shape the inferred cosmological parameters. We started this by considering the `AbacusSummit` simulations used to test the emulator in Section 4.0.5, and then extended this with fits to the `Nseries` cutsky mocks, which included BOSS fiber-collision effects and geometry — properties which were not part of our training set. Finally, once we had shown that the cosmological recovery stages of the pipeline were robust, we applied the same pipeline to the BOSS CMASS DR12 NGC sample (see Section 4.2).

4.1.1 AbacusSummit based tests

For each AbacusSummit mock considered, we inferred the posterior of the cosmological and HOD parameters; where the mock data vector is compared to those generated by the emulator. We considered the same six test cosmologies c000, c001, c002, c003, c004, c013 used to construct $\mathsf{C}^{\rm emu}$ in Section 4.0.5. For each, we selected HOD parameters that best fit the BOSS CMASS sample. We utilized the same binning as the CMASS sample utilized in [234], with 30 separation bins for the monopole and quadrupole, for 0.7 h^{-1}Mpc $\leq$ s $\leq$ 120 h^{-1}Mpc. We ran our inference with the priors prescribed by Table 4.1.

We first fitted c000, the baseline cosmology of AbacusSummit. The full posterior distribution is shown in Figure 4.6. We found that the sensitivity of the emulator to HOD parameters was limited—all five parameters hit the limits of our priors. We concluded that the VG-CCF is uninformative in terms of constraining the galaxy-halo connection, and we marginalized over these parameters for the remainder of this analysis. Our model did obtain constraints for n_s, but since this quantity is highly constrained by the CMB and we do not expect much sensitivity from the VG-CCF to it, most of the constraints we presented throughout the rest of the thesis are marginalized over it. Our analysis focused on obtaining cosmological parameters and derived quantities that pertain to the large scale structure evolution at late times. We found that for all six of our test cosmologies unbiased constraints could be obtained for $\omega_{\rm cdm}$ and σ_8. This was underscored by Figure 4.7, where marginalized constraints for $\omega_{\rm cdm}$ and σ_8 are shown for four test cosmologies that vary these parameters, with the marginalized posteriors fell within 2σ of the true parameter values. This figure explicitly showed that despite adopting a fiducial cosmology that does not resemble the one used in the box (e.g. c002) we successfully recovered the correct cosmological parameter values for $\omega_{\rm cdm}$, σ_8.

4.1.2 Nseries based tests

The recovery tests we ran using the Nseries mocks were similar to those described in the previous section, with a few key differences: i) there is a single Nseries test cosmology and HOD model we inferred; ii) the Nseries mocks were run with a different N-body code and galaxy-halo connection model than one used in our training data; iii) the Nseries mocks are cutsky mocks, not cubic boxes; iv) the cutsky mocks include approximate fiber-collision and completeness weights [190]; v) the radial number density is not uniform. Points iii)-v) are particularly salient, as they helped us test the robustness of the pipeline with data that

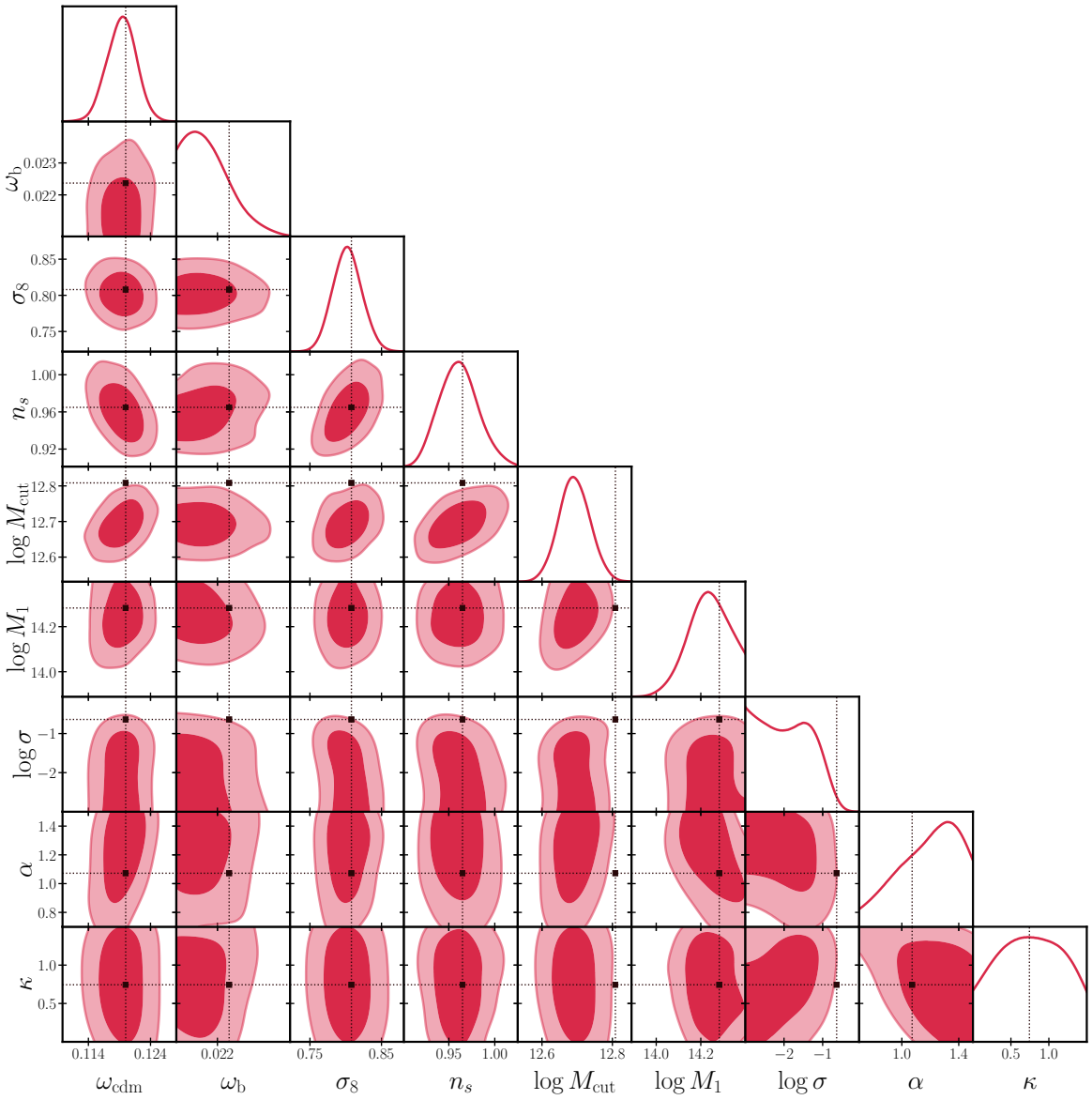

Figure 4.6: Full posterior of the `c000` test cosmology. True values for the parameters are shown by the solid black lines. Overall, we found that the posterior range for the HOD parameters $\log M_{\rm cut}$, $\log M_1$, $\log \sigma$, α, κ consistently hit the limits of our priors and were uncorrelated with our cosmological parameters. The full posteriors for the other test cosmologies can be found in the appendix of [86]. We then marginalized over these HOD parameters for the remainder of our analysis.

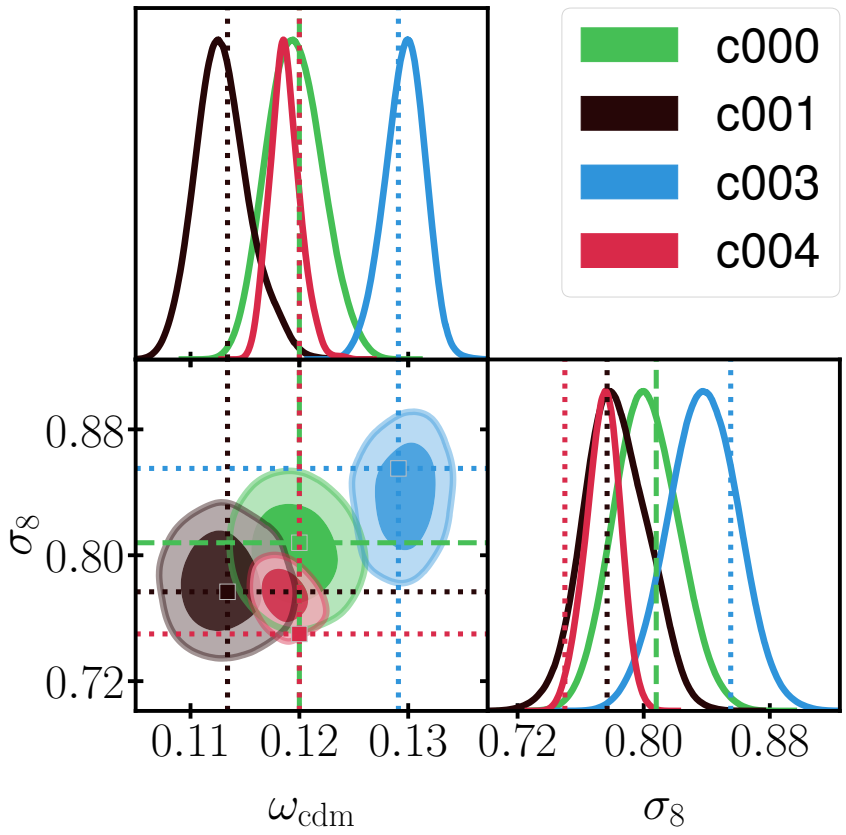

Figure 4.7: Posteriors obtained from the emulator, marginalized over $\omega_{\rm b}$, $n_{\rm s}$ and the HOD parameters, using the priors described in Table 4.1, with the test cosmologies and their true values for $\omega_{\rm cdm}$, σ_8 from Table 4.2. This figure confirmed that our emulator could successfully recover the true cosmological parameters after emulating the VG-CCF of our test set. Notably, these test cosmologies were chosen locations in the parameter space where we expected the emulator to fail most strongly.

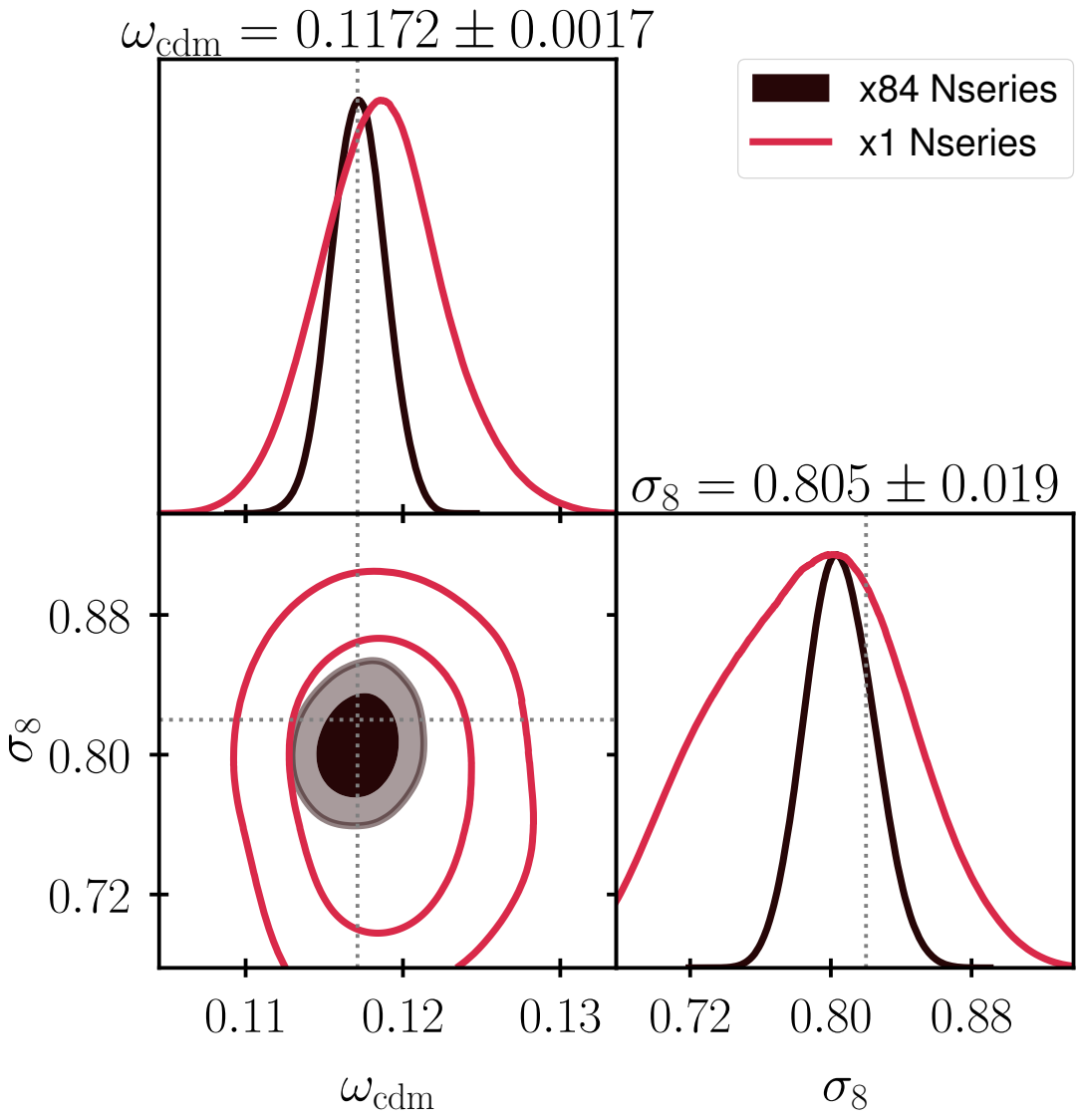

Figure 4.8: Marginalized posteriors of the emulator's predictions of the `Nseries` cosmology, with the errors scaled to the BOSS volume (black) and 84 times the BOSS volume (red), with the best fit for the x84 chain annotated above. We found that in both cases the emulator successfully predicted $\omega_{\rm cdm}$ and σ_8 to within 1 σ. This confirmed that the emulator's predictions were robust—and it successfully recovered cosmological parameters even when survey holes, geometry and fiber collision weights were not included in the training data.

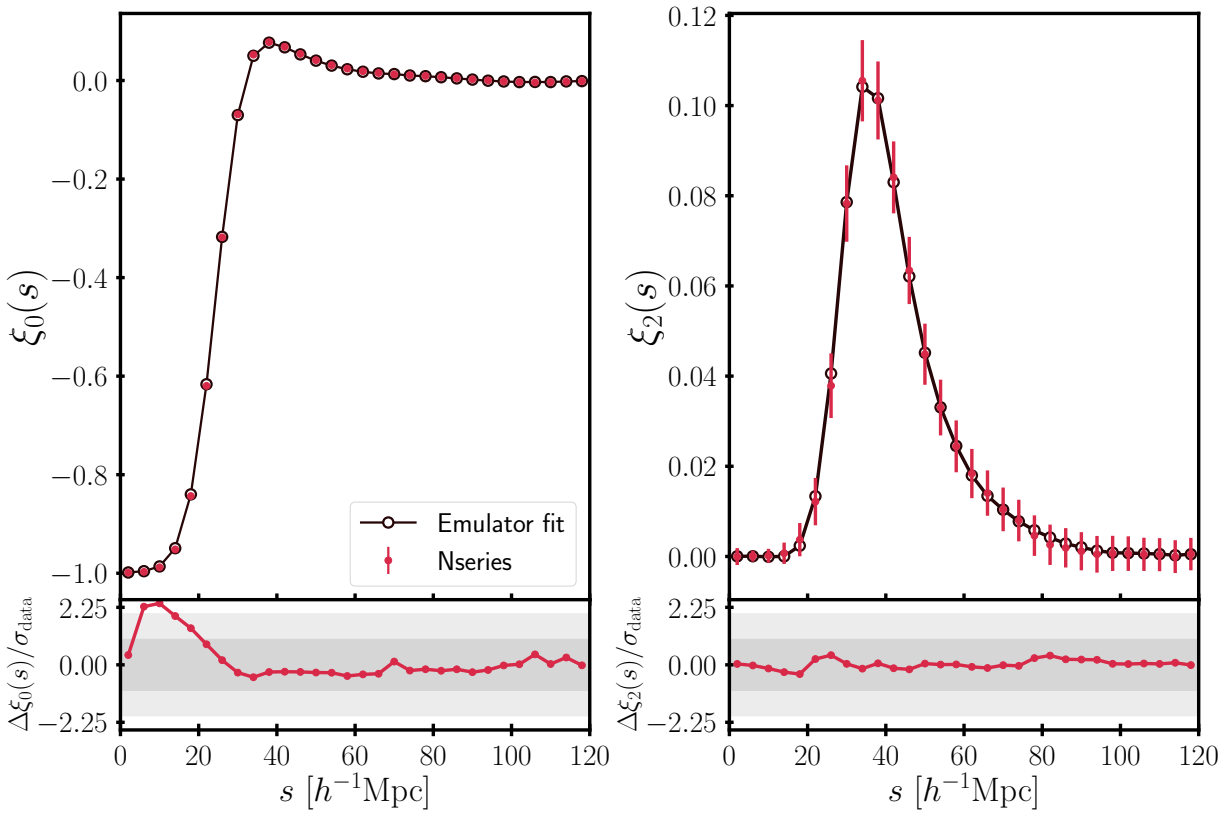

Figure 4.9: Predicted monopole and quadrupole from the `Nseries` cutsky mocks, with an x84 volume scaling applied to limit cosmic variance. Overall, we found that even with some variation in the recovered cosmological parameters, where we obtained $\Omega_{\rm m}$ to 1.3σ of the `Nseries` value and σ_8 to within 0.8σ of the `Nseries` value, the monopole and quadrupole from the emulator's prediction closely matched the monopole and quadrupole obtained from running the void-finder on the 84 `Nseries` cutsky mocks.

more closely resembles spectroscopic survey data. The `Nseries` mocks are not full light-cones, and so the growth is stepwise corresponding to the boxes from which the mocks were created. The geometry does match that of the BOSS sample, and the AP effect smoothly varies across the mocks with redshift, rather than being fixed at an effective value. Thus we consider that fitting to these offered a rigorous test of the method.

We fitted the mean of the `Nseries` data vector, and considered errors scaled for either the total volume (labelled as x84 in Figure 4.8) or the volume of a single mock (labeled as x1 in Figure 4.8). The same priors were used as in the `AbacusSummit` inference, and we obtained the 2D posterior distribution, marginalizing over all parameters except $\omega_{\rm cdm}$ and σ_8. We then obtained h, $\Omega_{\rm m}$ and computed $D_{\rm M}/D_{\rm H}$ and $f\sigma_8$, along the chain, and compared these to the values obtained from the true `Nseries` cosmology. The 2D posterior is shown in Figure 4.8. We found that even when rescaling by the total volume (84 times the `Nseries` mock volume), the true values were within the 2σ marginalised credible intervals. For the cosmological parameters, we obtained $\Omega_{\rm m}$ to within 1.3σ of the `Nseries` value and σ_8 to within 0.8σ of the `Nseries` value.

The residuals between the emulator's maximum likelihood solution and the mock multipoles are shown in Figure 4.9. The residuals are shown in units of the standard deviation of the data vector, derived from the `Patchy` mocks. We found the residuals for the monopole and quadrupole lie within 2σ, with the exception of two points, both in the monopole at scales below 20 h^{-1}Mpc. The bias seen at smaller scales was similar to the one seen in the tests with the `AbacusSummit` data. Taken together with the results of the cosmological recovery, this indicated that our emulator was robust to changes in survey geometry, weights and N-body codes, with the caveat that the model might have had a mismatch the clustering amplitude at very small scales.

4.2 CMASS data fits

Having tested the emulator on two sets of mocks, and shown that our pipeline is robust and accurate, we then fitted our model to the VG-CCF obtained from the BOSS CMASS DR12 galaxy catalogue.

For template based models, fits are typically made to the multipoles using a compressed set of parameters quantifying the AP and RSD effects [234]. For our emulator-based approach, we instead had to work from the cosmological parameters that define the simulations used by the emulator. The comparison between approaches therefore had to be undertaken by comparing posterior surfaces for the cosmological parameters, having matched priors, including the θ_* prior discussed in Section 4.0.7.

4.2.1 Baseline ΛCDM

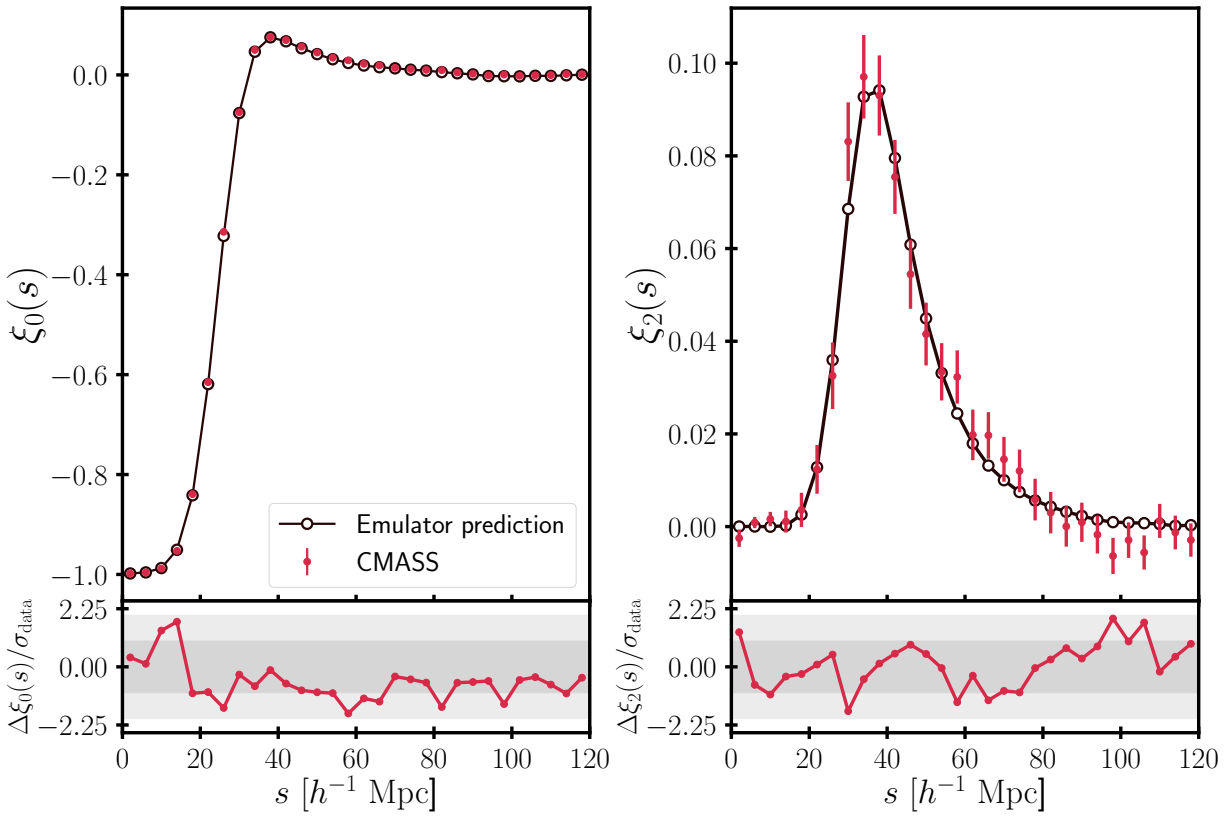

Figure 4.10: Emulator's fit of the monopole and quadrupole of the void-galaxy cross-correlation function from the CMASS sample (top panels), with the residuals between our fit and the measured monopole and quadrupole (bottom panels). With 1 and 2 σ regions shaded in.

We first considered a flat ΛCDM model, and fit to the data using the methodology outlined in Section 4.0.6. The multipole moments of the maximum likelihood solution are compared to the data in Figure 4.10, with the residuals shown in the bottom panel. We found that residuals were not significantly offset at small scales compared to larger scales, in contrast to the result seen in the `AbacusSummit` and `Nseries` tests. We found the greatest deviation in the residuals in the monopole at scales below 20 h^{-1}Mpc. While for the quadrupole, we found some deviation at larger scales, but within 2 σ. The bias was added in quadrature to the covariance to account for it. The projected posterior with

cosmological constraints on ω_c, ω_b, σ_8, n_s (cosmological free parameters) and Ω_m, H_0 is shown in Figure 4.11.

Figure 4.12 shows the 2D posterior for D_M/D_H and $f\sigma_8$ derived from, and marginalised over, the ΛCDM parameters. This was compared to the results reported by [234] and [155] using the template method. We compared against the measurements of D_M/D_H and $f\sigma_8$ obtained from [234] at the $0.4 < z < 0.5$ and $0.5 < z < 0.6$ redshift bins that encompass the CMASS data. We called this sample "CMASS template – 2 bins". Using [155], we also compared against previous measurements made assuming a single redshift bin from [155], which covers a wider redshift range: $0.4 < z < 0.73$. We called this sample "CMASS template – 1 bin". We repeated this procedure including measurements from all five redshift bins of [234], called "SDSS template". Although the bins used in both [234] and [155] cover a slightly larger redshift range than used in our analysis, this did not affect our conclusions significantly (see Section 4.3). Since our emulator was constructed using a single time slice, we did not further adjust the redshift range of the data considered in this work, in order to ensure that the emulator predictions remained valid throughout the sample. To match the analysis using the template-based results to those from the emulator, we used `cobaya` to find parameter constraints for adopting either a ΛCDM or wCDM model given the set of $f\sigma_8$ and D_M/D_H measurements and the same priors used in the emulator-based approach, including the θ_* prior. Narrowing the prior range to match the emulator's affected the constraints obtained from these template models. We then recomputed derived parameters D_M/D_H and $f\sigma_8$ at the effective redshift $z_{\rm eff} = 0.52$, within the context of these cosmological models. Note that the derived parameters, even from the template-based approach, will be dependent on the cosmology adopted and, in general, will be more constrained than the direct measurements. Adopting this procedure allowed for a fair comparison between template and emulator based approaches, and allowed us to compare derived parameter combinations from both methods with similar assumptions. [112] had previously considered the choice of fiducial Ω_m as a systematic. They found similar constraints suggesting this was either not important, or its effect could be absorbed by their nuisance parameters. They found constraints of $\Omega_m = 0.312 \pm 0.012$, and $D_M/D_H = 0.588 \pm 0.004$, within 1.5σ and 3.1σ of ours, respectively.

We reported the results from the template method in comparison to our emulator in Tables 4.3 and 4.4. The cosmological parameters are given in Table 4.3, while the derived parameters are presented in Table 4.4. We saw a reduction in errors with the emulator method compared to the template-based approach for both the cosmological parameters, Ω_m, σ_8 and the derived AP and RSD parameters D_M/D_H and $f\sigma_8$. Specifically, we saw at least a 25% reduction in errors for the AP parameter, and a 8.4% reduction in errors for the RSD parameter depending on the CMASS sample used. These arose from at least a

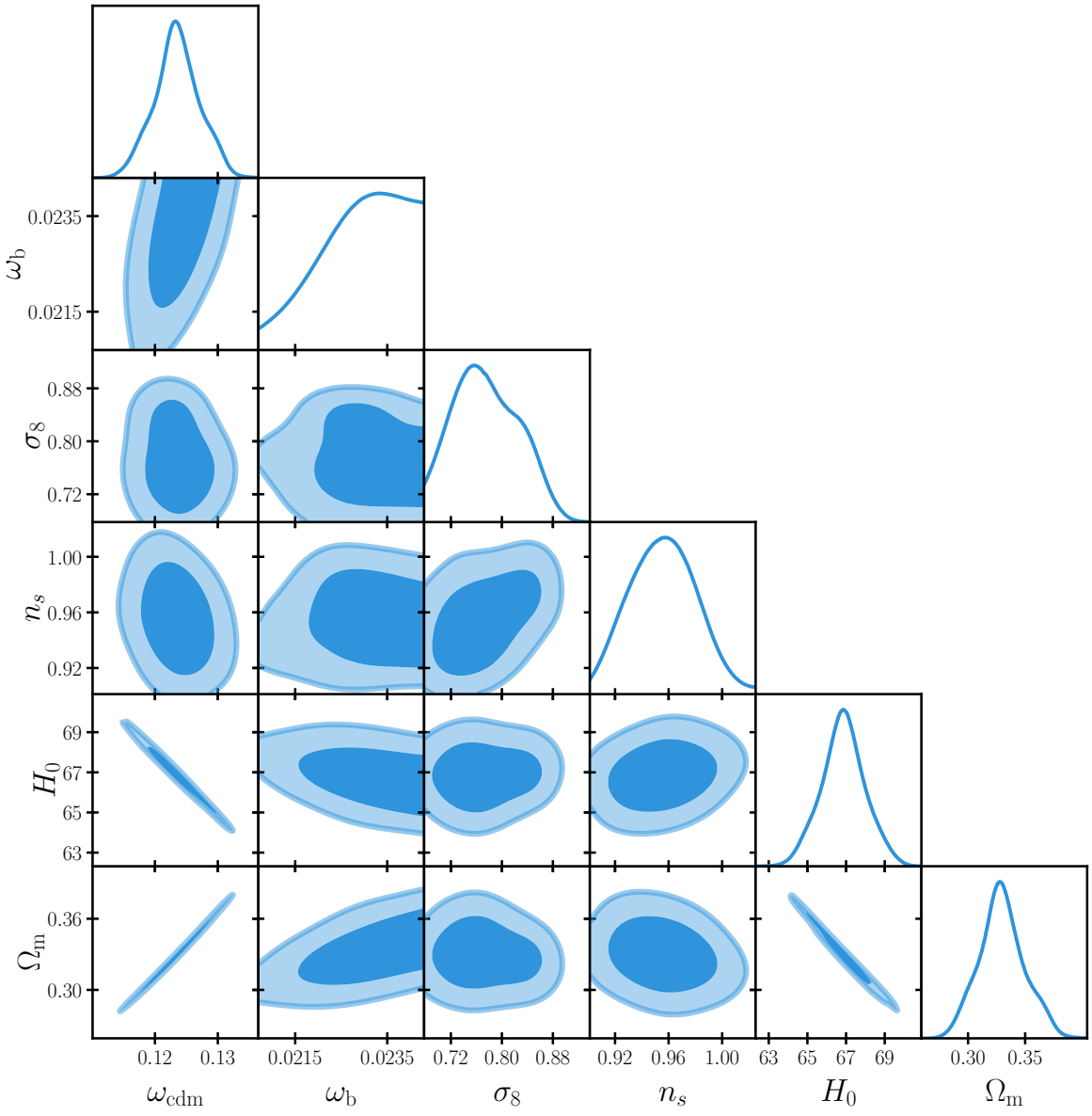

Figure 4.11: Posterior from fitting CMASS data using the emulator for a flat ΛCDM model, where $\{\omega_{\rm b}, \omega_{\rm cdm}, \sigma_8, n_{\rm s}\}$ were our cosmological free parameters and where we had marginalized over the HOD parameters. H_0 was derived from fixing θ_*. Consequently, $\Omega_{\rm m}$ can be computed from our free parameters $\omega_{\rm cdm}$ and $\omega_{\rm b}$. Fixing θ_* will also influence the baryon fraction, and subsequently $\omega_{\rm b}$.

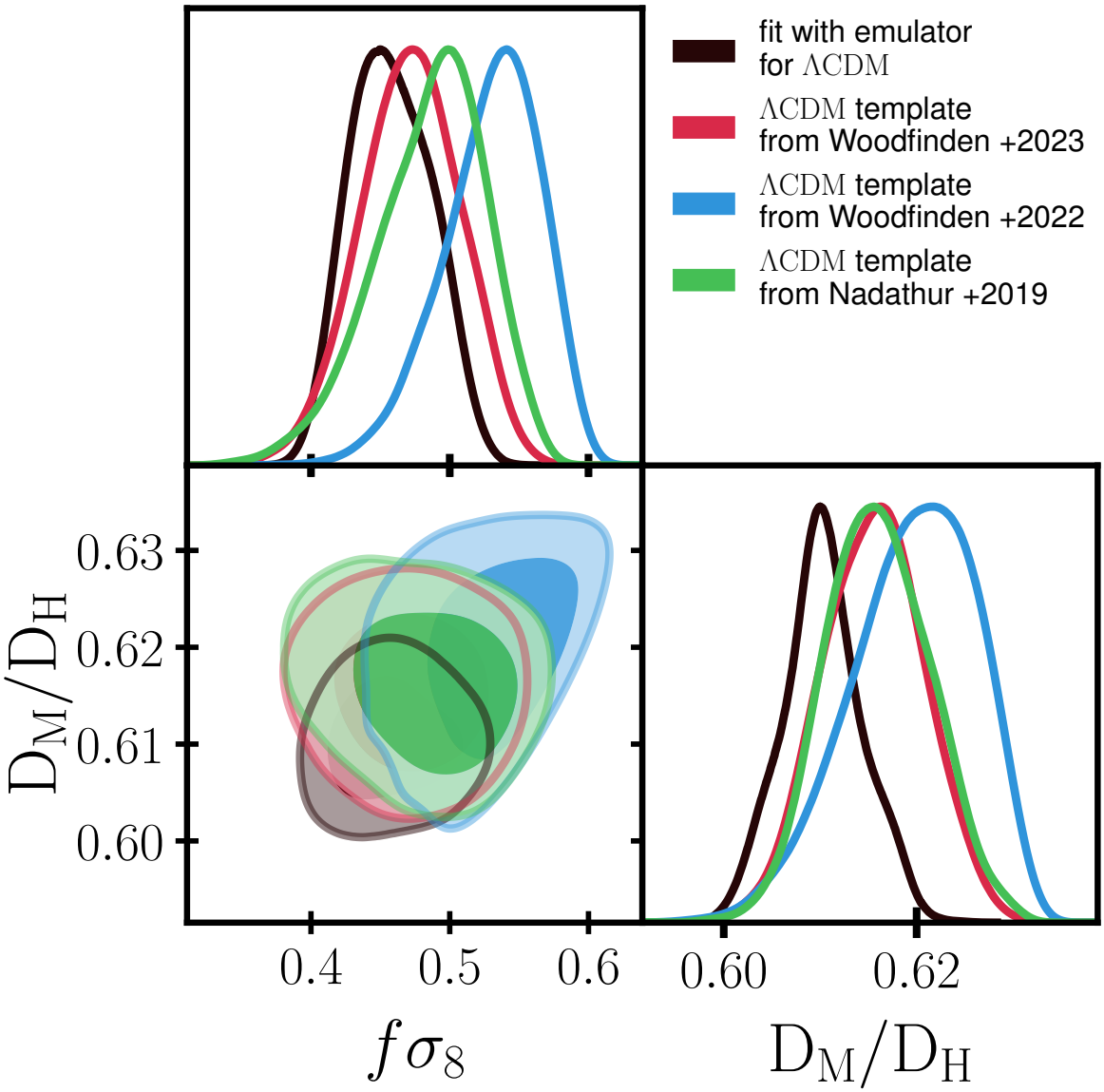

Figure 4.12: Since we were able to obtain Ω_m due to fixing θ_* we then derived the Alcock–Paczyński parameter D_M/D_H and the structure growth rate $f\sigma_8$ at the effective redshift $z_{eff} = 0.52$. These results (black) were then compared to the values computed from the template method, using the SDSS sample (red) [235], the CMASS galaxy sample (blue) [234] and from an earlier implementation of the template method with CMASS galaxies [155] (green), where the same constraint on θ_* has been included. The templates were fit using the same (narrower) priors as the emulator, which has a minor effect on the constraints produced compared to what is quoted from the respective papers.

Result	$\omega_{\rm cdm}$	σ_8	w_0
ΛCDM (emulator)	0.1235 ± 0.0037	$0.777^{+0.047}_{-0.062}$	—
wCDM (emulator)	$0.1236^{+0.0062}_{-0.0051}$	$0.770^{+0.039}_{-0.056}$	$-1.008^{+0.079}_{-0.10}$
ΛCDM (SDSS template [234])	0.1271 ± 0.0048	$0.792^{+0.071}_{-0.058}$	—
wCDM (SDSS template [234])	$0.1255^{+0.0089}_{-0.0053}$	0.784 ± 0.069	-0.99 ± 0.15
ΛCDM (CMASS template – 2 bins [234])	$0.1305^{+0.0069}_{-0.0039}$	$0.877^{+0.066}_{-0.040}$	—
wCDM (CMASS template – 2 bins [234])	$0.1279^{+0.0095}_{-0.0049}$	$0.874^{+0.075}_{-0.043}$	$-0.935^{+0.18}_{-0.094}$
ΛCDM (CMASS template – 1 bin [155])	0.1273 ± 0.0047	$0.814^{+0.082}_{-0.055}$	—
wCDM (CMASS template – 1 bin [155])	$0.1243^{+0.0087}_{-0.0060}$	$0.802^{+0.089}_{-0.064}$	$-0.96^{+0.19}_{-0.11}$

Table 4.3: Table of credible intervals for cosmological parameters from various fits. "CMASS template – 2 bins" describes the constraints obtained with a template using the two redshift bins of [234] covering most of the CMASS galaxies: $0.4 < z < 0.5$ and $0.5 < z < 0.6$. "CMASS template – 1 bin" indicates the results from applying the template method using the CMASS sample to a single redshift bin $0.4 < z < 0.73$ from [158], while "SDSS template" results include all 5 redshift bins. Note that 2D posteriors for both fits hit the prior walls, defined by the emulator, for ω_b and σ_8.

28% reduction in errors for $\Omega_{\rm m}$ and at least a 1.7% reduction in errors for σ_8. Because we added the θ_* prior to the template results, this was a clear indication that the improvement in constraining power did not only come from this prior, but also from the full-shape of the VG-CCF, alongside a proper treatment of the Alcock–Paczyński effect in terms of modelling the void-galaxy cross-correlation function, and the adoption of an emulator model over a simulation-based template model for velocity fields around voids. This methodology produced 1D marginalized constraints on $\Omega_{\rm m} = 0.330 \pm 0.020$ and $\sigma_8 = 0.777^{+0.047}_{-0.062}$, which was the main result of this work. Additionally, the emulator produced a constraint on $n_{\rm s} = 0.956 \pm 0.024$.

We also compared the credible intervals obtained from the posteriors to those derived using the template method. A similar improvement in constraining power was seen for both the AP and RSD parameters, and for $\Omega_{\rm m}$ and σ_8. This further illustrated that a full-shape analysis of the VG-CCF can produce unbiased and competitive constraints on $\Omega_{\rm m}$ and σ_8.

Result	D_M/D_H	$f\sigma_8$	Ω_m	σ_8
ΛCDM (emulator)	0.6103 ± 0.0042	$0.458^{+0.029}_{-0.033}$	0.330 ± 0.020	$0.777^{+0.047}_{-0.062}$
wCDM (emulator)	$0.6089^{+0.0061}_{-0.0073}$	0.463 ± 0.033	$0.329^{+0.019}_{-0.024}$	$0.778^{+0.048}_{-0.059}$
ΛCDM (SDSS template [234])	0.6157 ± 0.0055	$0.472^{+0.041}_{-0.034}$	$0.357^{+0.026}_{-0.029}$	$0.792^{+0.071}_{-0.058}$
wCDM (SDSS template [234])	0.6159 ± 0.0056	0.473 ± 0.037	$0.348^{+0.049}_{-0.034}$	0.784 ± 0.069
ΛCDM (CMASS template – 2 bins [234])	$0.6199^{+0.0080}_{-0.0051}$	$0.529^{+0.039}_{-0.029}$	$0.378^{+0.040}_{-0.028}$	$0.877^{+0.066}_{-0.040}$
wCDM (CMASS template – 2 bins [234])	0.6210 ± 0.0069	$0.523^{+0.049}_{-0.033}$	$0.363^{+0.055}_{-0.033}$	$0.874^{+0.075}_{-0.043}$
ΛCDM (CMASS template – 1 bin [155])	0.6160 ± 0.0055	$0.486^{+0.046}_{-0.033}$	$0.358^{+0.025}_{-0.030}$	$0.814^{+0.082}_{-0.055}$
wCDM (CMASS template – 1 bin [155])	0.6159 ± 0.0055	$0.487^{+0.050}_{-0.034}$	$0.341^{+0.047}_{-0.038}$	$0.802^{+0.089}_{-0.064}$
Density Split (emulator [163])	—	0.462 ± 0.020	0.306 ± 0.012	0.792 ± 0.034
Galaxy 2PCF (emulator [236])	—	0.444 ± 0.016	$0.309^{+0.018}_{-0.017}$	0.762 ± 0.024
WST BOSS (emulator [228])	—	0.469 ± 0.012	0.322 ± 0.007	$0.834^{+0.058}_{-0.039}$

Table 4.4: Table of derived parameters and σ_8, with their means and standard deviations shown. The density-split clustering and the galaxy two-point correlation function (2PCF) emulator were fitted to the same data, the BOSS CMASS NGC catalogue with similar redshift binning: $0.45 \leq z \leq 0.60$ for Density Split and $0.46 < z < 0.60$ for the 2PCF emulator. The wavelet scattering transform (WST) analysis was run across a redshift range of $0.46 < z < 0.57$, with $z_{\rm eff} = 0.515$ used for the $f(z)\sigma_8(z)$ calculation.

4.2.2 Extending to wCDM cosmologies

We extended our analysis by considering a wCDM cosmology where the dark energy equation of state w is constant in time, but allowed to have $w \neq 1$. The 2D posterior of the wCDM parameters can be found in Figure 4.13, where we marginalized over the HOD parameters. We found significant improvement over the template method in the constraining power of Ω_m, σ_8 and w with the emulator model, as seen in Figure 4.14. Our emulator model favoured lower values of σ_8 and Ω_m when compared to the template model. The results are presented in Table 4.3 & Table 4.4. For wCDM models, we ended up with a significantly wider posterior for D_M/D_H than for ΛCDM models, with the upper 1σ bound enlarged by a factor of 1.5, and the lower 1σ bound enlarged by a factor of 1.7. We also found a 0.3σ shift in the mean of the posterior towards a smaller D_M/D_H and a 0.2σ increase in the mean $f\sigma_8$, with the posterior widths remaining similar. The means and standard deviations in Ω_m and σ_8 were largely unchanged.

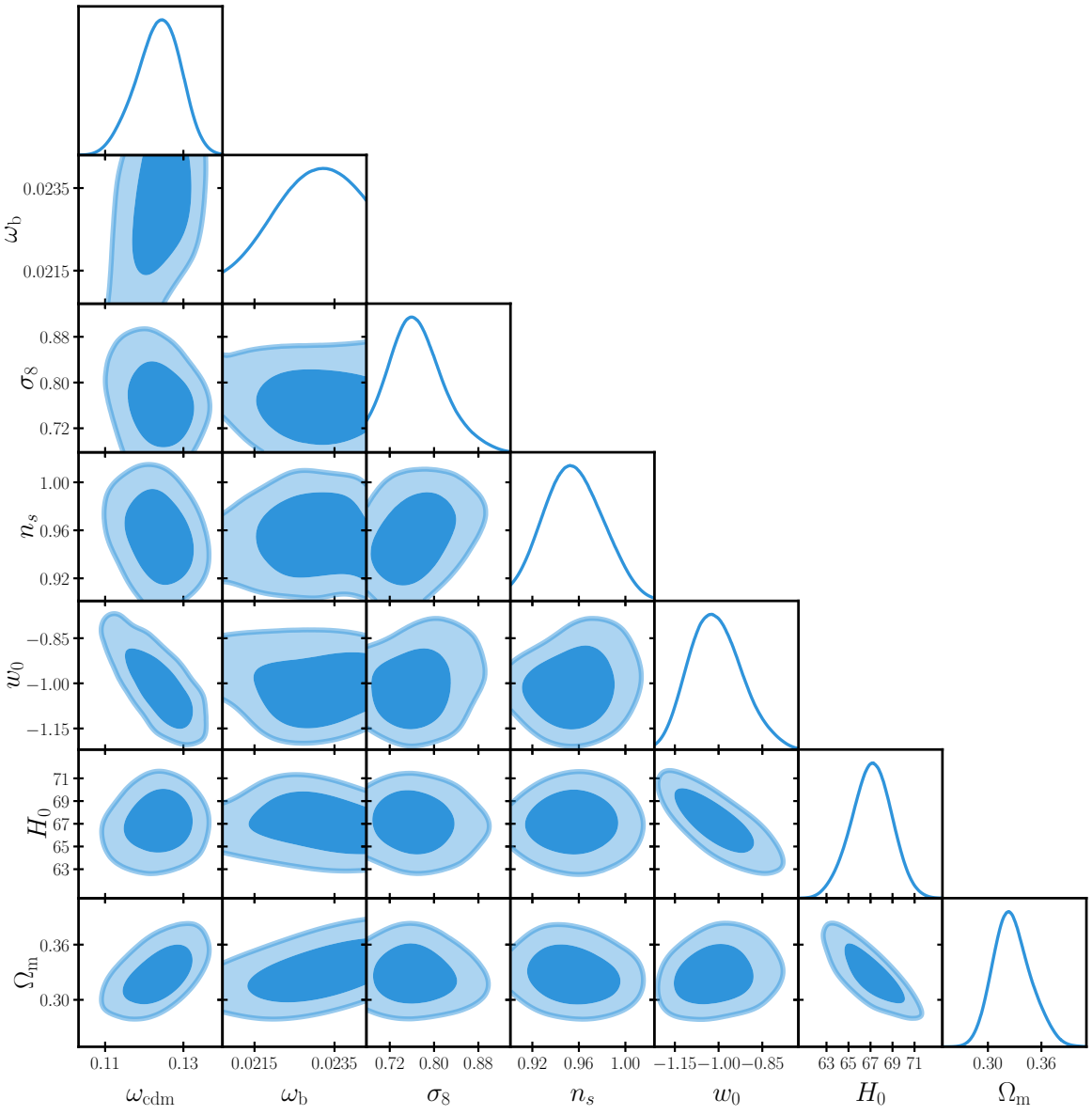

Figure 4.13: Posterior from fitting CMASS data using the emulator for a flat wCDM model, where $\{\omega_\mathrm{b}, \omega_\mathrm{cdm}, \sigma_8, n_\mathrm{s}, w_0\}$ were our cosmological free parameters and where we have marginalized over the HOD parameters. H_0 is derived from fixing θ_*. Consequently, Ω_m can computed from our free parameters ω_cdm and ω_b. Fixing θ_* will also influence the baryon fraction, and subsequently ω_b.

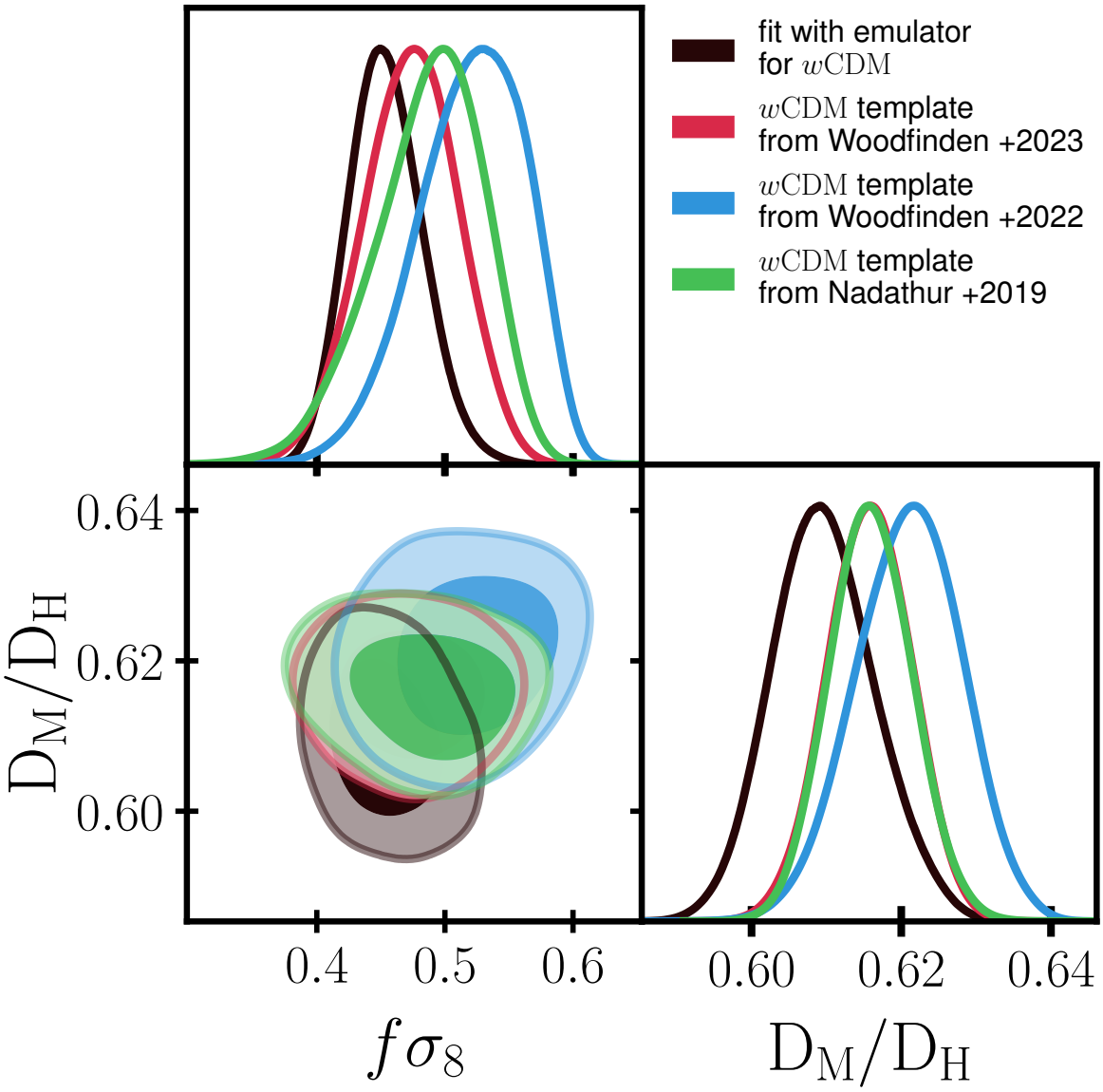

Figure 4.14: We again derived the Alcock–Paczyński parameter $D_{\rm M}/D_{\rm H}$ and the structure growth rate $f\sigma_8$ from our emulator's free parameters and $\Omega_{\rm m}$, this time for wCDM cosmologies. These values (black) were compared to the results of using the template method, as applied to the SDSS sample (red) by [235], using a subset of the CMASS galaxy sample (blue) by [234] and an earlier analysis of CMASS galaxies (green) by [155] where in all of these templates the same constraint on θ_* has been included.

4.2.3 Testing the parameters of the fits

Figure 4.15 shows the change in constraints on the primary cosmological parameters, $\Omega_{\rm m}$ and σ_8, constrained by the voids, after we made a number of changes to the method. After providing the baseline ΛCDM and wCDM results, the rows show the results of using a simplified model of the AP effect in the void-galaxy cross-correlation function: void centres were assumed to be shifted by the AP parameters, equivalent to the ones used with galaxies. Other fits of the model were explored, which included fitting to the monopole or quadrupole only, and reducing the covariance by removing components as described in Section 4.0.5. We also considered reducing the scales fitted, fixing the HOD parameters, fitting reduced sets of cosmological models, and fixing either ω_b or n_s to their Planck 2018 values [11]. We found our results were remarkably robust to all of the changes considered, and could outline the driving factors in our emulator's biases.

The monopole provides more information than the quadrupole for $\Omega_{\rm m}$, indicating that the anisotropic stretching of the cross-correlations may not be as important as the spherically averaged shape for this parameter. The result was more balanced for σ_8. Our interpretation of this was that the σ_8 constraint was limited by the RSD measurement, while the dependence of $\Omega_{\rm m}$ on the AP effect was more nuanced. This matches the observation that, for $\Omega_{\rm m}$ unlike σ_8, there remains significant information on large scales, suggesting information is being extracted from the cosmological large-scale clustering. Also, we would expect that fixing the HOD or alternative cosmological parameters increases the amount of signal from the shape of clustering, and this also improves the $\Omega_{\rm m}$ error, but has a weaker effect on σ_8.

The sub-covariances discussed in Section 4.0.5 arising from the data and emulator add roughly equally to the combined covariance, which suggested a match between the current size of the `AbacusSummit` simulation suite and the data. For future surveys where the data covariance is improved, emulators with a denser sampling of parameters would be required to reduce the errors associated with the emulators, such that they are sub-dominant.

4.3 Discussion

The results presented in Section 4.2 relied on several underlying decisions for our methodology, particularly the choice of voids to include, priors, and scales fitted, which all have a downstream effect on the constraining power arising from the monopole and quadrupole. To compare with the work of [234], we used a similar treatment for our void catalogues, and stacked all voids with radius $> 20\ h^{-1}$ Mpc for subsequent analysis. Note that the

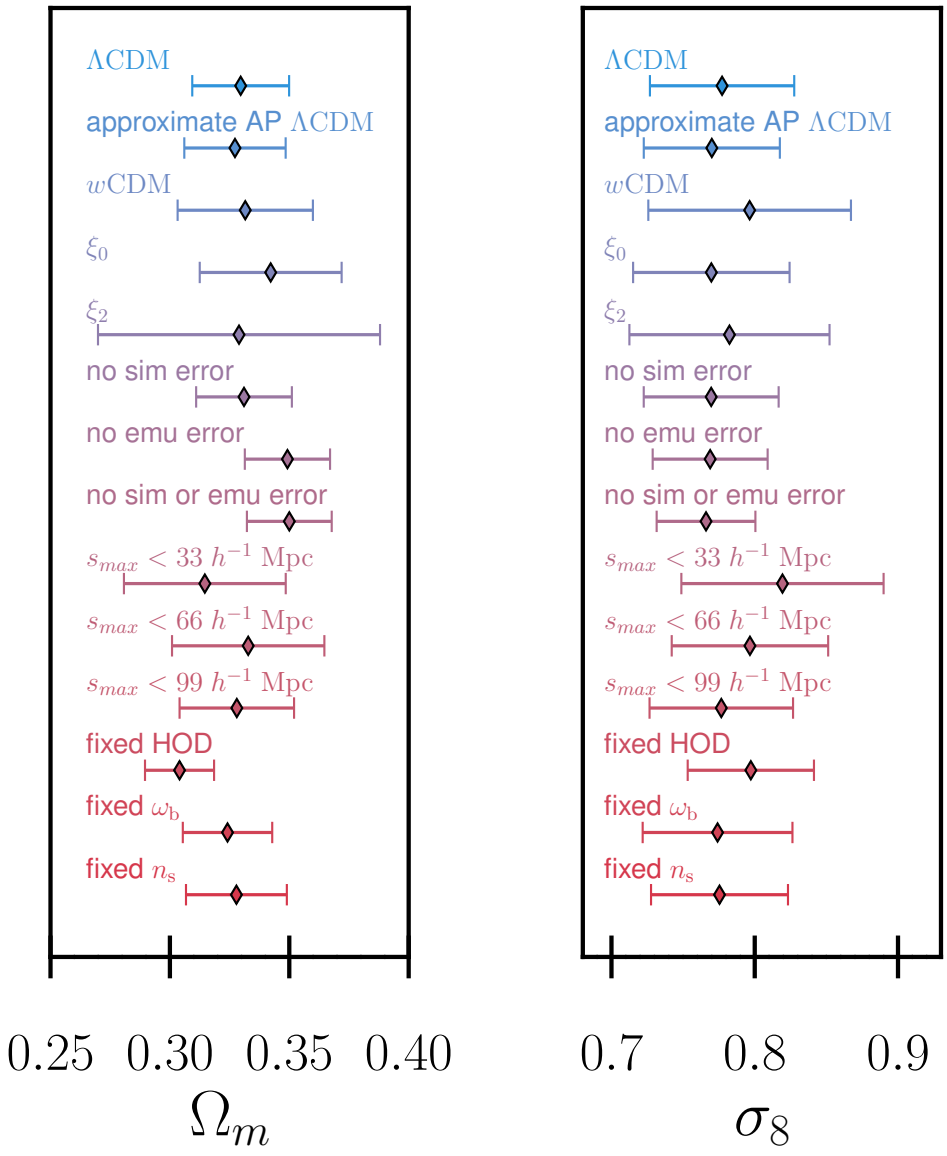

Figure 4.15: Marginalized constraints on $\Omega_{\rm m}$ (left) and σ_8 (right) from the BOSS CMASS data, under different settings for the emulator. We explored extensions to flat ΛCDM, different implementations of the Alcock–Paczyński effect to the void-finder, fitting with either the monopole or quadrupole only, as well as the effects of removing different parts of our covariance or limiting the scale of emulation $s_{\rm max}$, as well as fixing $\omega_{\rm b}, n_{\rm s}$ and the HOD parameters.

voids were not the same, as the emulator method used voids found in redshift-space, while [234] used voids found post-reconstruction (as described in Chapter 2). We also used the `voxel` void-finder, while [234] used `ZOBOV` [159], although figure 11 of [187] showed that the voids found were similar for both approaches. An important caveat was that the emulator's parameter space is fundamentally limited by the mocks used in training. Predictions made by the emulator outside of the range of the `AbacusSummit` mocks cannot be used in our model. This informed our choice of priors when we compared with the template model, where we had to narrow our prior ranges used in `cobaya` to match the `AbacusSummit` prior range.

In addition, our training data should mimic all aspects of the observations. In the analysis presented, we instead used simplified mocks, testing the robustness to observational effects. We believe that the tests presented justified these simplifications. However, further testing and adopting training mocks using full lightcones with the survey geometry would make the method more robust for upcoming analyses. In that sense, our analysis was the first step along this path.

One of the effects that would be captured in lightcones and our cutsky Nseries tests is the redshift evolution of the AP effect. Our training data had a single AP stretch parameter, computed for the simulation's snapshot redshift at $z = 0.5$, whereas the AP shifts varies with redshift. In the case of a serious mismatch between the fiducial and the true (box) cosmology, this could cause a mismatch in the average behaviour of void-galaxy pairs. To test for this, we averaged the AP shifts for a random set of void-galaxy pairs from $0.45 < z < 0.60$, and compared this with the expected shift at $z = 0.5$, corresponding to the simulation's snapshot redshift for ten different cosmologies chosen within the 2σ of our $\omega_{\rm cdm}$ and σ_8 measurements. We found that the average vs. expected AP shifts differed by 0.15% in $q_\parallel$ and 0.11% in $q_\perp$ at most. These were significantly below our errors—our measurement of the AP factor has an error of 0.7% for ΛCDM models. Larger survey volumes and training sets will necessitate using lightcone mocks as the errors in $D_{\rm M}/D_{\rm H}$ will shrink.

4.3.1 Impact of the θ_* prior

As discussed, the `AbacusSummit` simulations were used to train our emulator all have the same θ_*. This limited our analysis to only consider models with the same projected CMB peak position. In order to be self consistent, and not use the emulator to predict models with very different θ_* we had to either only consider constraints including additional CMB data, or include a prior as we have done here. We chose the latter route in order to limit

the information to voids as much as possible. While theoretically justified as the CMB observations are very strong [11], it would have been simpler in terms of interpretation were this constraint not present in the training sample of mock catalogues.

For a ΛCDM model at fixed physical baryon density the θ_* prior is equivalent to a prior on $\Omega_{\rm m}h^{3.4}$ [174]. The dependence on the baryon fraction is weak, such that even broad bounds on $\omega_{\rm b}$, as given in Table 4.1, led to only a small weakening of this degeneracy. While the primary information coming from the voids remains the AP and RSD parameters, the emulator also recovered information from the shape of the clustering signal. This information, together with the θ_* prior explained why our ΛCDM fit was able to constrain H_0 and σ_8 (or $\log A$), as seen in Figure 4.12, for example, breaking a number of degeneracies present in previous void-only analyses. This is underscored by the marginalized constraints shown in Figure 4.15, where the monopole and quadrupole both contribute to constraining $\Omega_{\rm m}$ and σ_8.

To compare to the results from [234, 235], we had to add the same θ_* prior to their results, and only compared within the set of cosmological parameters for a given model, rather than the measurements directly, as described in Section 4.2.1. The interpretation of the cosmological constraints was complicated by having to disentangle the measurements from this prior.

4.3.2 The Alcock–Paczyński effect

The emulator method easily allowed the AP effect on void-finding and on the void-galaxy cross-correlations to be fully included by shifting the galaxies before void-finding, as would be the case for an analysis of data. To test this, we emulated the monopole and quadrupole moments of the redshift-space void-galaxy cross-correlation, with the void-finder obtaining void-centres under the following two scenarios: i) Voids were found in the true cosmology of the training data and the void centres and galaxies are shifted by the AP stretching. ii) The AP shift was applied to galaxies before the void centres were found, a scenario that matched the process applied to data. (i) was an assumption commonly used in the literature (e.g. [234]) as it simplified the analysis.

If the fiducial cosmology exactly matched the truth, we would not expect a systematic offset in the best-fit position for (i) versus (ii), but would still see a change in the derived credible intervals. Figure 4.16 shows how our results actually change between these two scenarios. The fiducial cosmology we adopted lies near the centre of the credible intervals derived from our fits, as shown in Figure 4.16. Under both scenarios, the best-fit for $f\sigma_8$ and $D_{\rm M}/D_{\rm H}$ were similar. However, the constraining power from the two different training

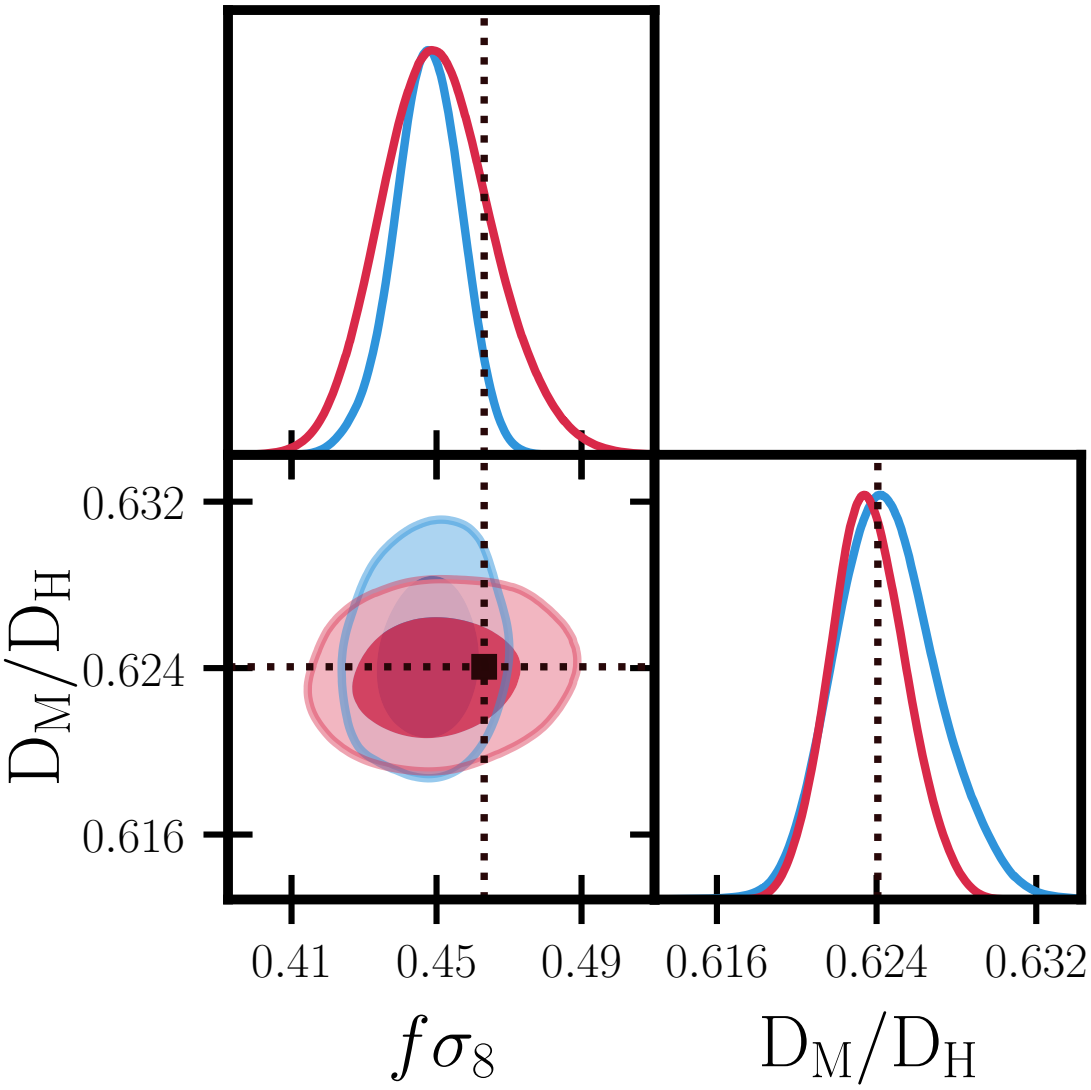

Figure 4.16: A comparison of the constraints obtained on $f\sigma_8$ and D_M/D_H through two sets of training data generated with two different assumptions for the Alcock–Paczyński (AP) effect applied to voids, one where the AP effect was applied after void-finding (red), and the better match to the process used for analysing real data, where the AP effect was applied before void-finding (blue). This figure was generated by fitting a mock with the same HOD within the `c002 AbacusSummit` non-fiducial cosmology by emulators trained with the different methods.

samples (i) and (ii) differed. For (i), we saw an improvement in the constraining power of $D_{\rm M}/D_{\rm H}$, while $f\sigma_8$ is less well constrained. This suggested that the simplified approach previously used had underestimated the errors in $D_{\rm M}/D_{\rm H}$.

The change in the sizes of the credible intervals was significant, and was discussed further in [187]. When we modelled the AP effect in the void-galaxy cross-correlation function from the same simplistic base assumptions of previous studies it did change the emulator-based models as the AP deviation became stronger, reducing the overall impact of the AP effect and resulting in wider credible intervals. [187] offered a thorough exploration of the assumption and how its effect translates into differences in the void-galaxy cross-correlation function. They showed that this was a fundamental consequence of the way that the AP effect affects void-finding, and not a consequence of the emulator-based approach. Our analysis agreed with [187]'s conclusion that void-centres' response to the AP distortions differs significantly from that of galaxies. While our results agreed with the explanation provided by [187], they did not by themselves prove this assertion. The emulator method did improve a common issue in void analyses: the cosmology being tested will change the topology of the density field and thus the distribution of spurious voids. As voids are found in these shifted catalogues in the emulator method, exactly as they would be in the data as if the cosmology was correct, then the number of spurious voids expected in the data matches that in the model. The only caveat to this was that our emulator is built on simulation cubes and not cutsky mocks, but we have tested the impact of this.

4.3.3 Comparison with other simulation-based clustering analyses

Previous works have also used mock galaxy catalogues to train simulation-based models of various clustering statistics. In particular, among those works who have used `AbacusSummit` to calibrate their models and applied to BOSS data, [236] emulated the small-scale galaxy two-point correlation function to obtain constraints on the halo occupation and ΛCDM parameters. [163] emulated density-split clustering statistics to obtain constraints on ΛCDM and single-parameter extensions, including $N_{\rm eff}$ and w_0. These constraints are summarized in Table 4.4. Additionally, beyond two-point clustering with voids has been used to enhance cosmological constraints. Notably [224] used the void-size function, the void galaxy cross power spectrum, and the galaxy auto power spectrum to constrain neutrinos masses in BOSS. Our work and [163, 1, 224] show that going beyond the two-point clustering can lead to significant improvements in constraining power. An important fraction of this constraining power is found at smaller scales, $s < 20\ h^{-1}$Mpc.

The error of our emulator was more dominant at small scales, where this signal is coming from, and highlighted the importance of further improvements in emulation techniques and sets of training data.

Characterizing the galaxy-halo connection in greater detail would also be another crucial ingredient in this methodology — we used a simple HOD model, without characterizing assembly bias and marginalizing over baryonic effects and velocity bias. The treatment of these effects might prove important in future work; the work of [217] with the marked power spectrum and probability underdensity function found that degeneracies between halo bias parameters and cosmological parameters can be broken by alternative clustering statistics. This potential is noted by [163] with a more detailed HOD model using the CMASS data. This motivated exploring if the shape of the VG-CCF can produce similar degeneracy breaking, either under a more detailed HOD model, or with different galaxy-halo connection models incorporated into the training (e.g. SHAM mocks). Another question raised is how the shot noise from our voxel voids might limit these constraints, in comparison to other low-density summary statistics.

There is significant information to be gained beyond 2-point statistics, in the form of the initial information suppressed by non-linear evolution, and from projection effects [202]. Each method discussed above can be considered a compression of this information, and will be sensitive to different cosmological processes, and hence parameter combinations [1]. Using voids, we limited the range of densities considered and focused on AP and RSD information. While we therefore provided less information overall, using only voids has the benefit of limiting the reliance on correctly modelling the full non-linear behaviour of the galaxy field at high overdensities.

Chapter 5

Void Galaxy Cross-Correlation Function in DESI DR1

The Dark Energy Spectroscopic Instrument (DESI) collaboration is an ongoing project primarily funded by the United States Department of Energy (DoE) to create a spectroscopic map of the universe $0 < z < 4$, and in doing so, measure the growth rate of large-scale structure, the equation of state of dark energy through the BAO peak, and measure the neutrino masses [69, 70]. The DESI instrument is a fibre spectrograph mounted onto the 4.2m Mayall telescope at Kitt Peak National Observatory in Arizona, USA [52]. The DESI instrument uses 5000 robotic fibres, split across 10 wedge-shaped clusters (*petals*, see Figure 5.1) to obtain 5000 simultaneous spectroscopic measurements. These fibres feed their spectra into one of the ten spectrographs, in turn splitting their output across 3 different cameras/colour bands (b, r, z). This field-of-view for the instrument covers a solid angle of 8 deg^2. This array of fibres allows for tens of thousands of spectra and their redshifts to be measured in single night, across a massive portion (14,000 square degrees) of the sky [212]. Each fibre is assigned a celestial target for observation, and a single configuration of all 5000 fibres is called a *tile* [153]. The same *tile* can be exposed multiple times, as part of different programs, which will be discussed below.

The DESI instrument has measured 5 years of spectroscopic data since March 2020, across a handful of survey campaigns. Prior to the collection of science data, there were three successive Science Verification survey campaigns (SV1-3). We utilized measurements from SV3, also known as the OnePercent survey (as it covered approximately one percent of the 14,000 deg^2 footprint) later in this chapter. The OnePercent survey was utilized to construct the priors for our HOD model. The main sample we utilized for this thesis comes from the Year One (Y1) survey, which comprises of redshifts measured across 13

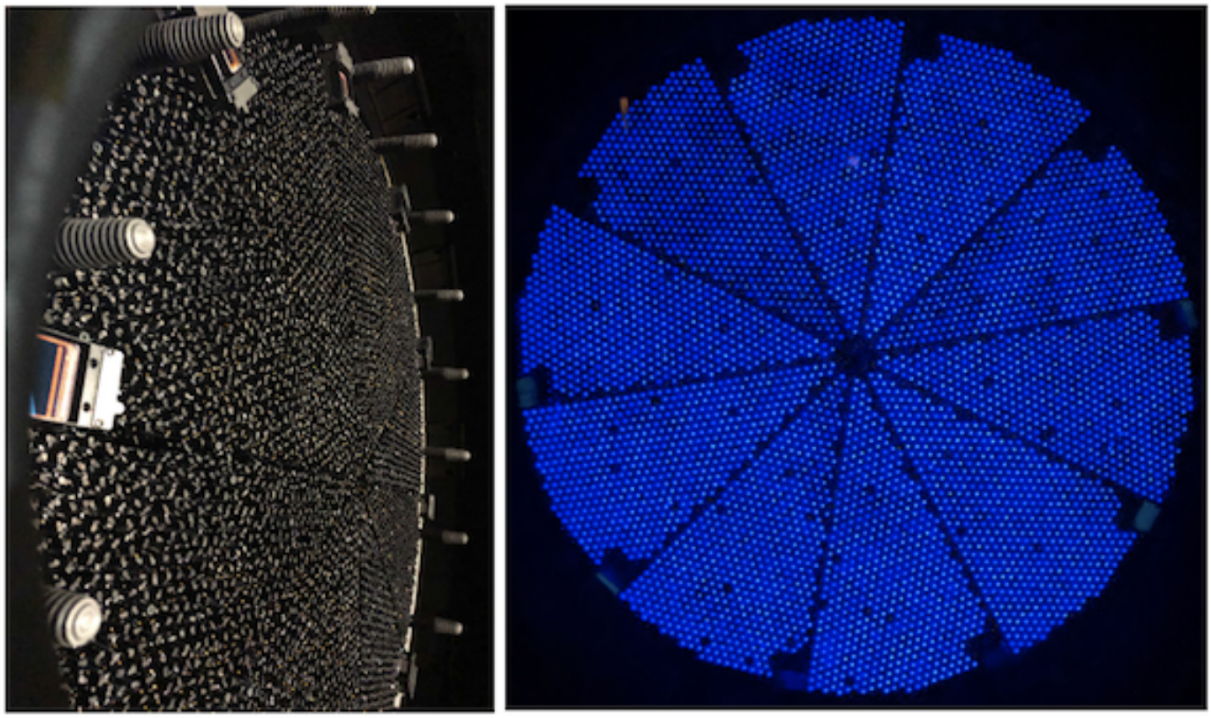

Figure 5.1: (Figure credit: [52]) A view of the 5000 fibres mounted onto the Mayall 4.2m telescope (Left). Backlit image of the fibres mounted onto the instrument, grouped into *petals* (Right).

months of operation which are now publicly available as DR1 [6]. The measurements are classed under three programs: `BRIGHT`, `DARK`, and `BACKUP`. For each of these programmes, *targets* are chosen using photometry from the DESI Legacy Surveys and Gaia [52, 71, 153]. The resultant list of *targets* is added to the Merged Target Ledger (MTL). Within each of these programs, targets are classified by their spectroscopic properties - Milky Way Survey (MWS), Bright Galaxy Survey (BGS), Luminous Red Galaxy (LRG), Emission Line Galaxy (ELG) and Quasars (QSO) [153]. The MWS and BGS targets are part of the `BRIGHT` program, while LRG, ELG and QSOs are part of the `DARK` program. The targets are assigned to the aforementioned tiles by using the MTL, and selecting the highest priority target within a fibre's patrol radius. `BRIGHT` tiles are exposed for an effective exposure time of $180s$, while `DARK` tiles are exposed for an effective exposure time of $1000s$. The allocation of these `DARK` and `BRIGHT` targets is done dynamically, see: [52]. These

programs are maintained for science verification (SV) and the main survey, which is split into successive data releases: DR1 (Y1 or year one), DR2 (Y3 or years 1-3), and DR3 (Y5 or years 1-5).

For the DR1 main survey, there were 2275 `BRIGHT` tiles, and 2744 `DARK` tiles observed in this period, which amounted to a coverage of 9739 deg^2 and 9528 deg^2 respectively. The survey area is split into galactic hemispheres, featuring the North Galactic Cap (NGC) and South Galactic Cap (SGC). The coverage of these tiles and the *completeness*, which measures the fraction of objects observed in a given tile is shown in Figure 5.2. The definition for completeness (or assignment completeness) in the DESI survey is outlined in [199] and summarized in [7], but can be broadly understood as the ratio between the number of objects observed in a *tile* and the number of objects in that *tile* that were chosen as targets. Successive passes (up to 7 per *tile* in DR1) increase the completeness.

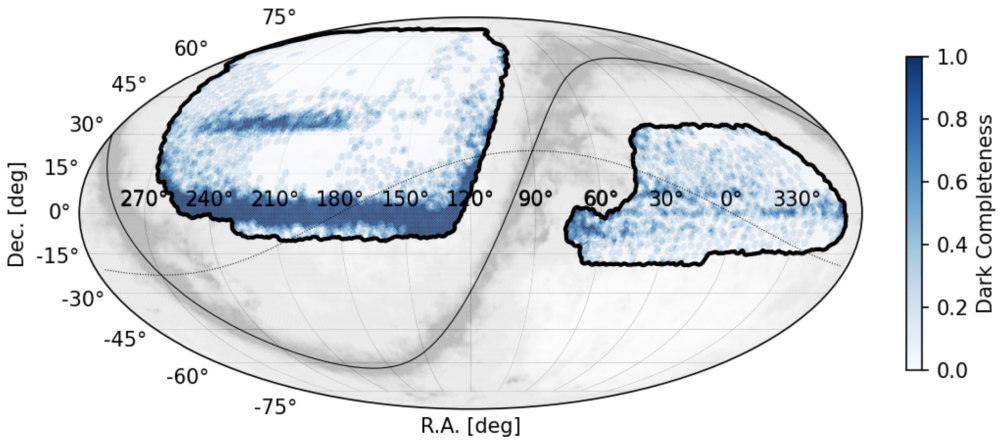

Figure 5.2: The DESI DR1 footprint, split across the north and south galactic caps (left, right), with each blue dot representing a `DARK` tile, the higher the opacity, the higher the completeness of a given tile.

The redshift measurements from a given set of *tiles* are then processed into large-scale structure catalogues, which are used to measure the BAO and in other LSS analyses, including this one. The processing of these catalogues entails selecting galaxies and computing the survey weights that arose from observational effects (e.g. extinction, seeing, dust) and from hardware specific systematics. Regarding the latter, the fibre positioning robots pose a problem for small-scale clustering measurements, as pairs of objects can easily have small angular separations that cause neighbouring fibres to collide with each other, limiting the measurement [199]. This is *fibre-collision*, and the effect it has on void-galaxy clustering is unknown but worth consideration, given its influence on galaxy-galaxy clustering.

Once the weights for these catalogues were computed, they were compared with catalogues that were created to model these effects. These catalogues consist of several cubic and lightcone mocks that mimic the survey's footprint and selection function in the fiducial cosmology, `c000`. These entail `AbacusSummit` cubic and lightcone catalogues that mimic the fibre assignment and fibre-collision, which maintain a high resolution for small-scale clustering, and catalogues which are used in computing the covariance. These catalogues are discussed in greater detail in the following subsections.

For this thesis, we selected galaxies from the LRG catalogue [245, 244], as part of the `DARK` observing program, for the DR1 (Y1) sample. We selected galaxies in the North Galactic Cap for `LRG-1` $0.4 < z < 0.6$. The methodology outlined below can be applied to the `LRG-2` $0.6 < z < 0.8$ and `LRG-3` $0.8 < z < 1.1$ slices of the `LRG` catalogue to analyze the VG-CCF in those redshift windows, provided a sufficient amount of computing resources and time, as will be discussed in the next subsection.

5.1 Boundary effects in DESI footprint on void-finder

As a consequence of the relatively uniform and high completeness of the BOSS CMASS sample, the angular variation of the survey's completeness meant that the VG-CCF obtained from the `AbacusSummit` mocks was reasonably close to the VG-CCF from CMASS. This feature does not extend to the DESI DR1 sample. Void-finding in cubic box mocks and the same mock converted into a geometric cutsky mock that broadly mimics the survey geometry (see: Section 5.2.2) yielded significantly different void catalogues, with notable differences in their void-size function (VSF). These differences translated into differences in the VG-CCF, meaning that a VG-CCF measured in a cubic box, even with a matching HOD and cosmology as its geometric cutsky counterpart, was not similar. These differences made it harder to construct a training sample for the DESI DR1 data. As Figure 5.3 shows, a larger fraction of voids in the catalogue obtained from cubic mocks are at larger radii than those from the corresponding geometric cutsky mock.

Training a neural network emulator for the VG-CCF assuming one distribution in the void population, and applying it to a sample with a significantly different void population would have meant that the emulator's predictions would have been significantly biased. In order to rectify this, we tried to reweight the void catalogue based on the relative differences in the VSF. This was done by computing the VSF for a target geometric cutsky catalogue and a series of cubic mocks with the same cosmological and HOD parameters.

We computed the VSF by binning the voids by their size $20 \leq R_V \leq 60h^{-1}$Mpc across

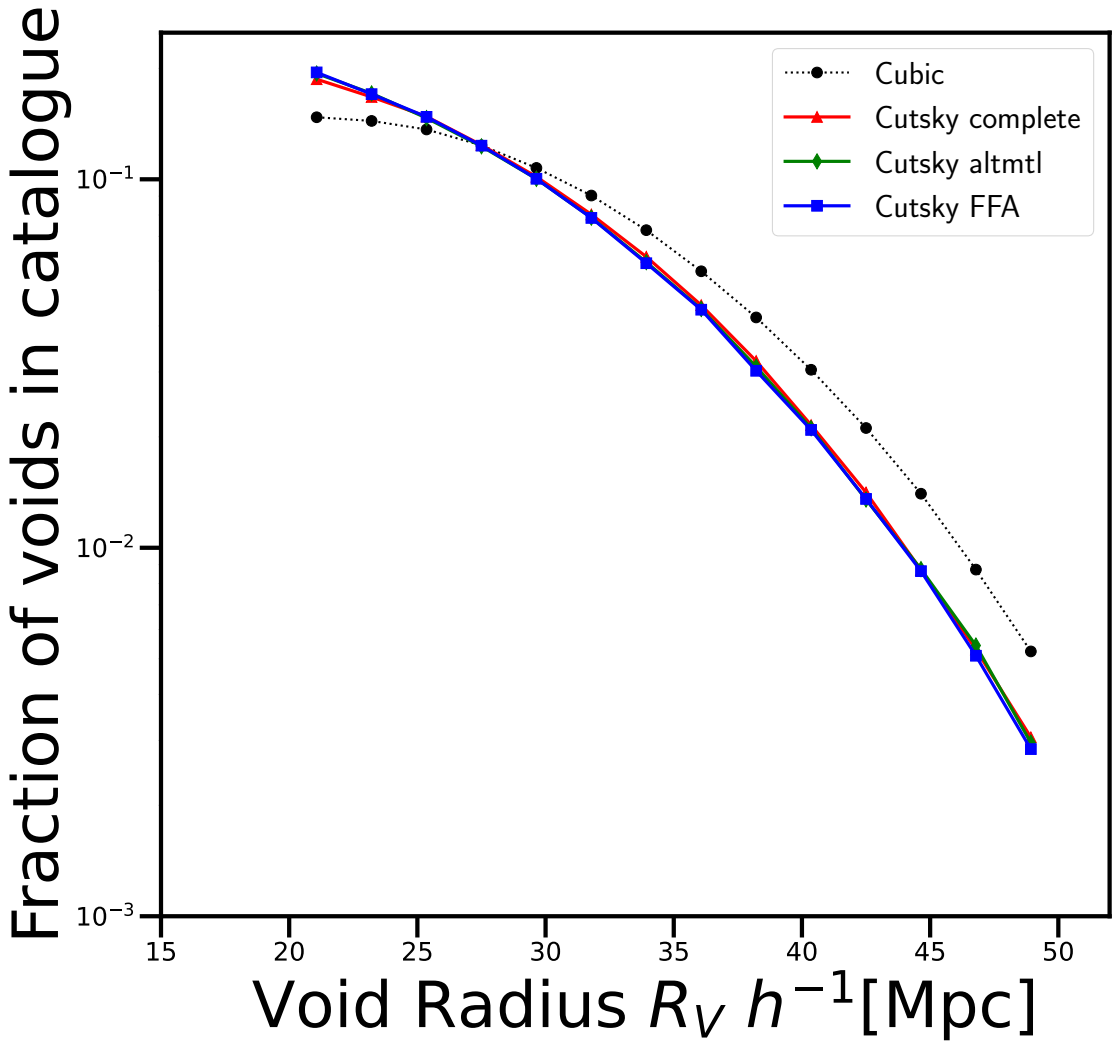

Figure 5.3: Void size function, scaled by the fractional number of voids per size bin, for `AbacusSummit` cubic mock catalogues and `AbacusSummit` fibre assigned (`altMTL`, `FFA`) cutsky catalogues, and `AbacusSummit` cutsky complete catalogues. Each of the VSF computed here were averaged across the 25 realizations available.

15 bins. The VSF obtained from the voids in the geometric cutsky and cubic mocks were then used to compute weights as:

$$w_{VSF}(R_V) = \frac{N_V(R_V)_{cutsky}}{N_V(R_V)_{cubic}} \tag{5.1}$$

Consequently, each void in the catalogue was assigned a weight based on the size-bin it falls into for the VSF. This method culminated in computing the ξ_{DP} VG-CCF using these VSF weights. This was tested with the 25 realizations of `AbacusSummit` fibre assigned cutsky `altMTL` and cubic mocks, as well as our own `AbacusSummit` DR1 mock and cubic mocks. The key difference in the latter set of mocks compared to the former is that our training data would only have one realization per cosmology, and that these tests were only done with the default cosmology `c000`, and at a fixed HOD. The model was tested using only the survey cosmology. Consequently AP distortions, which had been shown to alter the resulting void catalogue were not explored alongside this effect. This result should be treated as preliminary.

As Figure 5.4 shows, the introduction of void-size weights into the VG-CCF does not reconcile the differences at small scales, particularly for the monopole. The effect of implementing these weights was a marginal shift at smaller scales in the VG-CCF computed from cubic mocks compared to the DR1 mock VG-CCF. At larger scales, the two VG-CCF are nearly identical. The lower panels of Figure 5.4 illustrate this as well: where the residuals are largest at small scales and are still significant even after these weights were implemented. This was compounded by the small scale bias identified in the emulator, both in the one trained with cubic boxes, and later DR1 mocks, which primarily affected the monopole. This meant that implementing this reweighting scheme would only add to the deviation of the emulator's prediction from the data, and frustrate our analysis even further. Instead of adopting this reweighting scheme, we tackled this problem by constructing our own DR1 mocks. We note that this method of reweighting could yield a similar result in the future.

One test to explore in future work is through reweighting not only the void catalogue by the void size function obtained from DR1 mocks, but to explore how the D_vD_g paircounts might change between a cubic mock and a DR1 mock. This could be done by computing a set of weights derived from DR1 mocks and their precursor cubic mocks. We propose the following form for such weights:

$$w_{VG} = \frac{(D_vD_g/\sum D_vD_g)_{DR1}}{(D_vD_g/\sum D_vD_g)_{cubic}} \tag{5.2}$$

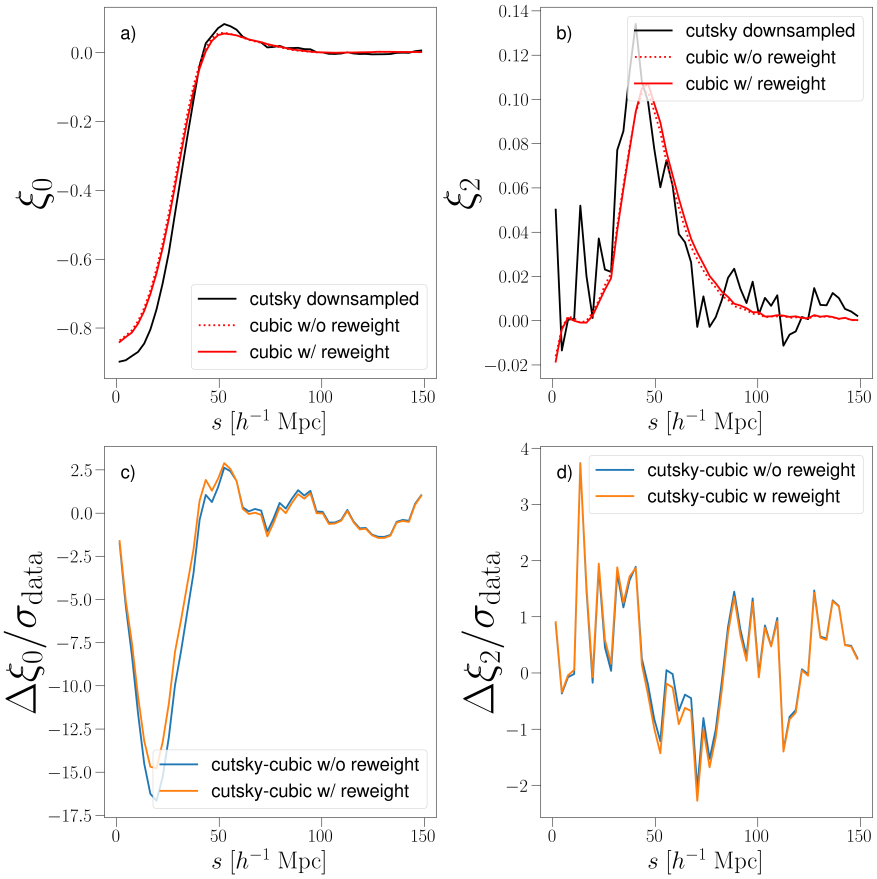

Figure 5.4: a) The VG-CCF monopole computed from a DR1 (cutsky) mock (black) that encompasses the DESI DR1 LRG $0.4 < z < 0.6$ region, and the VG-CCF computed from the cubic box used to construct it (red). The VG-CCF is reweighted by binning voids in the VSF, and assigning a weight to each void in the catalogue per Section 5.1. b) The VG-CCF quadrupole, computed for both the DR1 and cubic box mocks. The solid red line is the VG-CCF after the VSF weights were applied to it, and the dotted line is the original VG-CCF before the weights were computed. Subfigures c) and d) illustrate the difference between the monopole and quadrupoles of the DR1 and cubic mocks, scaled by the covariance from the `EZmock` catalogue. In subfigure c), it can be seen that the VSF reweighting does reduce the mismatch between the DR1 and cubic monopole at small scales, but only to the magnitude of a few σ, and in some bins, there is still a difference of $> 5\sigma$, meaning that constructing an emulator with mocks prepared with this method would only further exacerbate the existing small-scale bias. In subfigure d) the reweighting has a demonstrably smaller effect, but the difference in the DR1 mock and cubic box mock VG-CCF quadrupoles is far less pronounced.

These weights could be used to re-compute the VG-CCF of a cubic mock, in an effort to include the effect of survey boundaries on the number and distribution of voids (as was done with the VSF weights) but to also account for the resultant computation of the VG-CCF, which will have a tangible effect on void-galaxy pairs found at the edge of a survey.

5.2 Preparation of training data

The general steps of creating and training a neural network emulator were largely similar to the BOSS-CMASS version, albeit with several key adjustments made in the preparation of the galaxy catalogues. These adjustments were made in consideration of the limitations encountered from creating void-catalogues from cubic box mocks that had included AP and RSD. As the previous section has shown, the differences between cubic box mocks and DR1 mocks are not trivial when it comes to the VG-CCF, namely in the number density and boundary effects, both of which have an impact on the void-finding and the subsequent VG-CCF computation. Efforts had been made to bridge this gap through reweighting the VG-CCF, but had not sufficiently matured at the time of writing. It was therefore crucial to model how the void-finder responds to changes in the survey window. Failure to do so would have significantly biased our results, primarily through the training data being incongruent with the DESI sample.

We reconciled this difference between the cubic and survey VG-CCF by creating a mock pipeline to make DR1 mock catalogues that mimicked the angular and radial footprint of DR1 of the DESI survey LRG galaxy sample, in the $0.4 < z < 0.6$ bin, using the North Galactic Cap (NGC) subsample. We will describe the different stages in creating these mocks over the next few sections.

5.2.1 Creating cubic box mocks for DESI DR1 LRGs

For each cosmology, we first selected a range of HOD parameterizations that fall within the parameter space given by Table 5.1, which are the HOD parameter constraints identified by the DESI OnePercent survey (aka SV3). The mocks were created by generating a series of cubic box mocks using the `AbacusHOD` code from [1]`abacusutils`, which works by using the `CompaSO` halo finder to group simulation particles dark-matter halos, and thus populate the halos following an HOD model. The simulation boxes are 2 h^{-1}Gpc across,

[1] https://github.com/abacusutils

Parameter	Range	Interpretation
$\log M_{\rm cut}$	$(12.19, 13.39)$	Minimum halo mass to host a central galaxy
$\log M_1$	$(13.44, 14.32)$	Typical halo mass to host one satellite galaxy
α	$(0.43, 1.71)$	Power-law index of the expended number of satellites
κ	$(-1.0, 3.8)$	Regulation of the minimum halo mass to host a satellite
$\log \sigma$	$(-0.23, 0.65)$	Slope of the transition from zero to one central galaxy
α_c	$(0.09, 0.57)$	Central velocity bias
α_s	$(0.52, 1.08)$	Satellite velocity bias

Table 5.1: HOD Parameter space spanned by the DESI OnePercent survey. For all of our parameter inference, our references to α_s will purely focus on the HOD parameter, not the cosmological parameter with the same notation, which is fixed to the fiducial value of $\alpha_c = 0, \alpha_s = 1$. This model was introduced by [237], where α_c, α_s modulate the peculiar velocities of central and satellite galaxies by a Gaussian-distributed line-of-sight component.

and with a target number density scaled to match the LRG target number density for the `AbacusSummit` SecondGen mocks, with a number density of $\bar{n} = 7.36 \times 10^{-4}\ h^3\text{Mpc}^{-3}$.

Out of these parameterizations, the HOD parameter combinations that were kept were ones that resulted in the cubic box having a fiducial number density of $\bar{n_g} = 7.36 \times 10^{-4}\ h^3\text{Mpc}^{-3}$, matching the number density of the official DESI cutsky galaxy catalogues created for the DR1 analysis. The target number density was adjusted by the AP parameters, accounting for the distortion of the cubic box when going from a true cosmology to the fiducial cosmology `c000`, given in the Equation (5.3) below:

$$\bar{n}_g^{\rm fid} = \bar{n}_g^{\rm true} / \alpha_\parallel^2 \alpha_\perp \tag{5.3}$$

Failure to do so would have resulted in training data whose number densities were diluted/concentrated by a factor of $\alpha_\perp \alpha_\parallel^2$, which would have biased the emulator.

100 HOD parameterizations that matched these requirements for number density were selected, for each of the 85 cosmological models. We produced a total of 8500 mocks whose fiducial number density consistently matches the DESI DR1 number density target.

5.2.2 Creating DR1 DESI mocks

The conversion of our cubic mock catalogue creation pipeline to one that creates DESI DR1 mocks required a consistent treatment of the number density as computed in the

fiducial cosmology, the survey boundaries, and the resultant fibre assignment. To obtain the mocks, two libaries built by members of the DESI collaboration were used. They first converted the mock catalogues by replicating them and tiling them across the angular and radial coordinates of the DR1 survey region, and then masked them to match the survey characteristics of the LRG sample. Following this initial tiling, the resultant mock catalogue can be referred to as a *cutsky* mock catalogue, but with the various stages in this pipeline, we delineate their meanings below.

The first of these modules used in creating these mocks is [2]`desi-cutsky-mock` which takes the cubic box mock and decomposes the target wedge into a series of redshift slices, $10h^{-1}$Mpc thick. The box coordinates were then reshaped and assigned velocities based on the redshifts of each of these slices. Stacking them together, and excluding specific regions in RA and Dec to mimic the approximate DR1 footprint would then result in the final *geometric* cutsky mock being produced. This *geometric* cutsky mock matches the footprint, but does not have any of the survey weights, fibre assignment information or photometric or veto masks that define it. The mock produced in this step has an angular distribution that resembles the top left panel of Figure 5.5, i.e. broadly mimicking the survey footprint. Throughout the rest of this chapter we will refer to the following flavours of *cutsky* mocks:

- *geometric* cutsky - a mock catalogue that has been converted from a cubic box, and sliced into a series of wedges using the `desi-cutsky-mock` pipeline to mimic the footprint of DESI DR1 as discussed above.
- *forFA* cutsky - a mock catalogue that has reformatted the geometric cutsky mock by filtering out galaxies that are not in the bitwise and photometric mask (top right of Figure 5.5).
- `complete` cutsky - a mock catalogue that has had a potential fibre assignment file created, and selected galaxies from the *forFA* cutsky mock that could be assigned to a DESI fibre: this is equivalent to setting the survey completeness to $w_{comp} = 1$, and contains no fibre assignment information (bottom left of Figure 5.5).
- DR1 mock/downsampled mock - a mock catalogue that has been run through the entire pipeline, i.e. the bottom right panel in Figure 5.5. The mock catalogue matches the geometry of the DESI DR1 `LRG-1` NGC sample and also mimics the effects of fibre assignment. This is identical to the downsampled mocks that are discussed in the next section

[2] https://github.com/cosmodesi/desi-cutsky-mock

5.2.3 Adding fibre assignment information to cutsky mocks

Having populated cubic period mock, we then trimmed it to match the survey geometry (geometric cutsky), we then faced the task of adding the required fibre assignment information by utilizing the [3]`LSS` pipeline that was developed for the DESI collaboration. The incorporation of fibre assignment information was the final step to creating these DR1 mocks.

Independent of our pipeline were the `AbacusSummit` fibre assigned mocks, which had been previously constructed with the survey's fiducial cosmology (`c000`), and had been run through the `desi-cutsky-mock` and `LSS` pipelines to construct three variants of fibre assigned mocks:

- `complete` - catalogues that contain all targets in the DESI target list that remain after masking out bright objects, defective fibres, and vetoed objects. Each of these targets is assumed to be "observed", and hence have a completeness of $w_c = 1$. This is not realistic, as it does not account for the fibre-collisions, which are modelled in the remaining flavours:
- `altMTL` - mock catalogues that used the alternative Merged Target Ledger (`altMTL`) to assign fibres to each of the potential targets in the survey.
- `FFA` - fast fibre assigned mocks, which utilize a shallow learning model to emulate the fibre assignment algorithm used in the survey (`altMTL`) by using the local angular density and the friends-of-friends linking length between nearby pairs of targets.

The pipeline we constructed, using `LSS`, works by taking an existing geometric cutsky catalogue, and works to add in the veto, photometric and bitwise (bitwise masks select targets based on the type of target, differentiating between LRG, QSO, BGS, etc.) masks. We used [4]`brickmask` to rapidly assign these masks, and `LSS` for pre-processing the fibre assignment information. Once this mask was constructed, the `fibreassign` algorithm was applied to select galaxies that would be physically detectable by the survey. This potential assignment mask was combined with the photometric, bitwise and veto cuts to create a complete cutsky mock that contains galaxies that *can* be observed by the survey, meaning a fibre can be assigned to the galaxy and a spectrum can be measured, irrespective of fibre collision. The resultant catalogue is what we will refer to as a `complete` catalogue, i.e. a

[3] https://github.com/desihub/LSS/tree/main
[4] https://github.com/cheng-zhao/brickmask/tree/desi

catalogue where the completeness of all targets is one. The weights were also calculated by the `LSS` module in this step.

However, while the `complete` cutsky catalogue contained the galaxies and randoms needed to find voids, it had a significantly higher number density than the survey, and crucially, there was no fibre assignment done to the galaxies. This means that the fibre-collision that occurs when observing targets with small-scale angular separation was not captured by the `complete` cutsky mock, which necessitated the use of `altMTL` and `FFA` mock catalogues to model this for the galaxy two-point correlation function. The extent of fibre-collision on the VG-CCF cannot be analytically modelled, and this is another effect that had to be modelled in our emulator. Implementing either the `altMTL` or FFA models for fibre-collision would have been costly. The former taking days to produce 25 $2\,h^{-1}$Gpc mocks, the latter taking several hours. This compounds onto the existing computational cost of creating thousands of DR1 cutsky mocks, which made either of these existing flavours untenable for this thesis. To expedite the production of this training data, the `complete` cutsky mocks were downsampled to match existing `altMTL` mocks.

5.2.4 Downsampling

We needed to create mocks that match the angular and radial footprint of the fibre assigned catalogues of the DESI survey and the corresponding `AbacusSummit` mocks that have been used to test the pipeline. However, utilizing either the survey's fibre assignment algorithm (`altMTL`) or a neural network emulation of it (`FFA`) was prohibitively expensive given the volume of catalogues (8500) required. We instead opted to downsample the catalogues, by first pixelating the $0.4 < z < 0.6$ `complete` galaxy sample produced by our pipeline, and downsampled it to match a pixelation of the `altMTL` catalogue. We converted the angular positions of the galaxies and randoms into angular *healpixels* using `healpy`. We set the resolution such that an individual healpixel encompasses 3.4354 minutes of arc, equivalent to $nside = 1024$. Each of these galaxies and randoms were binned by healpixel, and were used to compute the downsampling fraction.

We first calculated the ratio between the number of galaxies in a given healpixel in the complete catalogue, and the corresponding number of galaxies (or randoms) in the `altMTL` catalogue. Once this ratio was properly computed, we then iterated over each healpixel in the array, and used the downsampling ratio to compute the probabilities of keeping a given galaxy in its pixel.

The computation of this downsampling ratio was complicated by the fact that the `complete` and `altMTL` catalogues have differences in the angular and radial distributions for

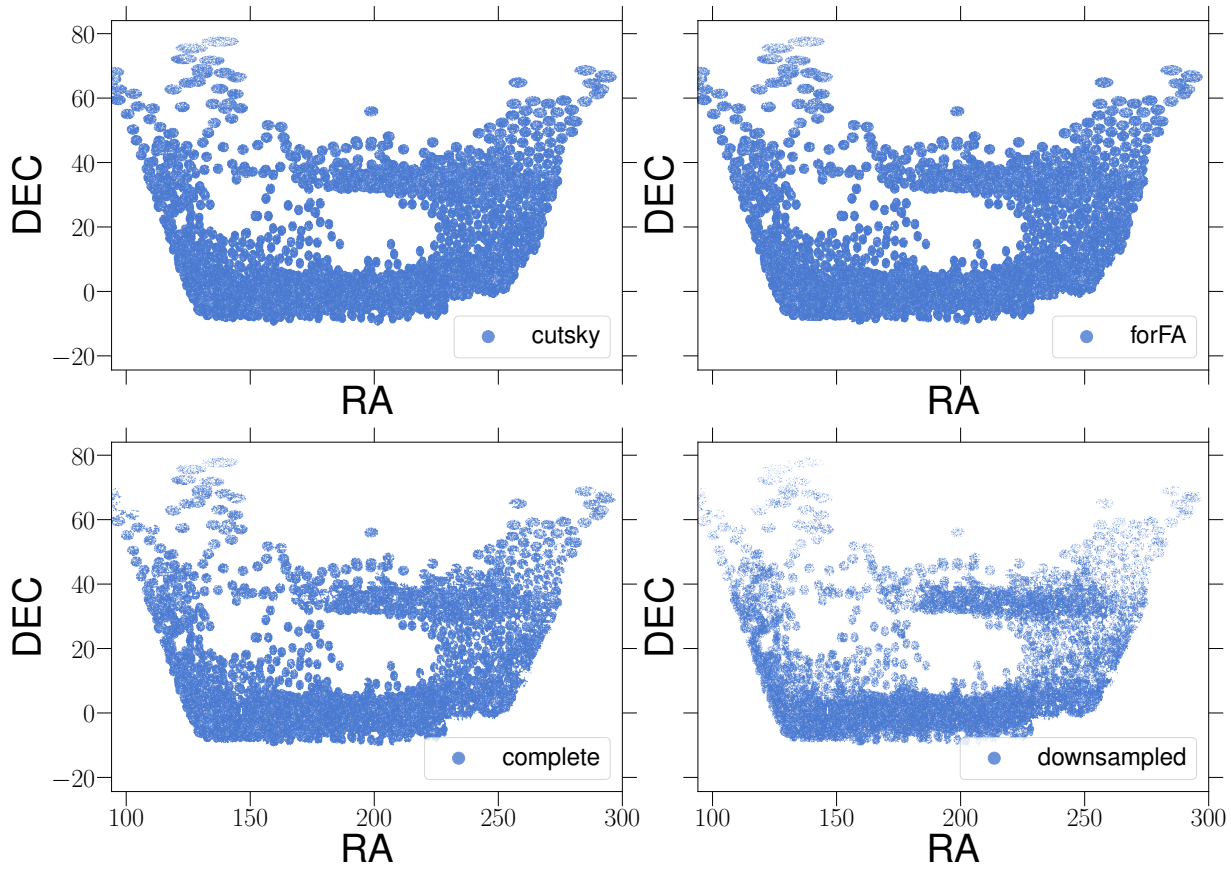

Figure 5.5: Each panel illustrates the distribution and number of galaxies at different stages in our pipeline that converts cubic boxes into geometric cutsky mocks (top left), the resultant distribution of galaxies after applying the imaging mask of the survey (top right), the subsequent catalogue once all targets that can be assigned a fibre are chosen (`complete` cutsky, bottom left), and the final catalogue obtained once we downsample the galaxies using healpixel angular bins to match the `altMTL` catalogue (bottom right) to obtain a DR1 mock.

their galaxies and their randoms. We normalized the distribution of galaxies in the `altMTL` and complete catalogues before computing the downsampling ratio, using the following steps.

$$F_{\rm norm} = \frac{\sum_{px} N_{r,complete,w}(px)}{\sum_{px} N_{r,altMTL,w}(px)} \times \sum_{px} N_{g,altMTL,w}(px) \tag{5.4}$$

The normalization factor can be calculated using the target distribution and the chosen random catalogue. We binned the galaxies and randoms by their pixel, and computed the number of galaxies or randoms in each pixel, weighting this total by the `WEIGHT` column, calculated during the previous step (i.e. when the `complete` mocks were created). Equation (5.4) consists of the ratio in the (weighted) sum of the number of `altMTL` galaxies divided by the (weighted) sum of the number of `altMTL` randoms, multiplied by the (weighted) sum of the number of randoms in the complete catalogue.

$N_{r,complete,w}(px)$ = the (weighted) total number of randoms in a given healpixel from the complete catalogue

$N_{g,complete,w}(px)$ = the (weighted) total number of galaxies in a given healpixel from the complete catalogue

This normalization factor, $F_{\rm norm}$ would be divided by $N_{g_{w,complete}}$; we isolated this term because it changes with the mocks utilized, i.e. the ratio is altered by the AP factor introduced earlier in the pipeline. Our goal is to ensure the number of galaxies in the final catalogue are similar in total, angular and radial distribution for all cosmologies, meaning that the $N_{g_{w,complete}}$ term had to be updated for each cosmology, and consequently F_{norm} had to be appropriately rescaled. This would otherwise distort the normalization factor by $\alpha_{\parallel}\alpha_{\perp}^2$.

In practice, we can use this normalization factor to compute the pixel ratio PR(px) by splitting it between a normalization factor, determined by the distribution of randoms between the precursor `complete` and target sample `altMTL` and the distribution of galaxies in both of these mocks. This normalization, $F_{\rm norm}$ was then multiplied by the pixel ratio, $PR(px)$, which was used to compute the full array of downsampling probabilities.

$$PR(px) = \frac{F_{\rm norm}}{\sum_{px} N_{g,complete,w}(px)} \times \frac{N_{r,altMTL,w}(px)}{N_{r,complete,w}(px)} \tag{5.5}$$

This pixel ratio was used to downsample the complete randoms catalogue and the complete galaxy catalogue produced by the previous steps of the pipeline (Section 5.2.3). Once these downsampled randoms and galaxy catalogues were produced, we also re-computed the FKP weights, which will be used in the calculation of the VG-CCF. This resulted in the production of galaxy catalogues that closely matched the `altMTL` galaxy catalogues with only a fraction of the computational cost (See Figure 5.6).

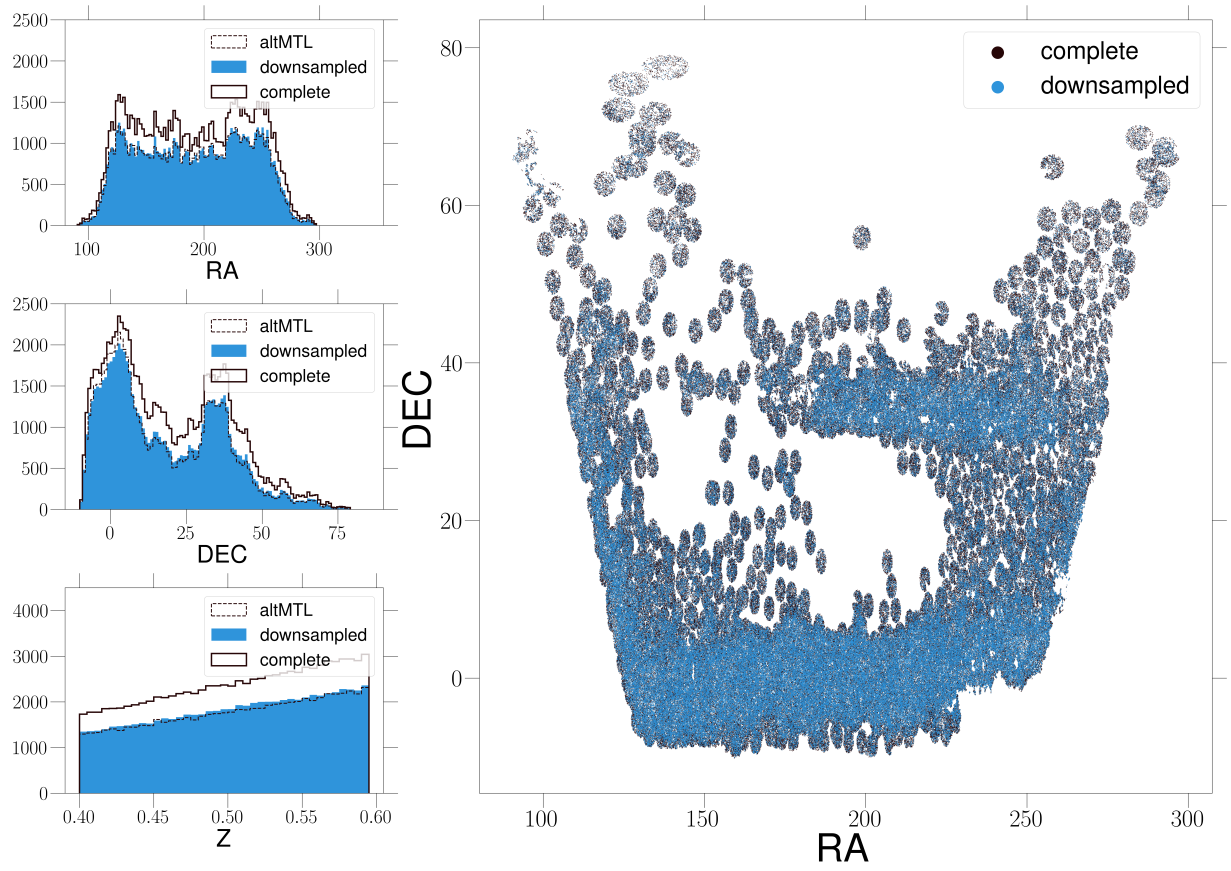

Figure 5.6: (Left - top to bottom) A series of histograms showing the distribution of galaxies in an `altMTL`, complete and downsampled catalogue, following through the downsampling stage of the pipeline. (Right) a scatter plot showing the distribution of galaxies in the `complete` (black) and downsampled (blue) mocks in our cutsky pipeline, for the DR1 NGC, $0.4 < z < 0.6$

5.2.5 Void-finding

In this chapter, we again utilized the `voxel` void-finding algorithm implemented in `Revolver`, with the overall implementation broadly matching our procedure in Chapter 4. `voxel` voids were found in the cutsky mocks produced by this pipeline, producing 8500 void catalogues that spanned a wide range of cosmological and HOD parameters (see Table 4.1). The voxel size was set to $s_{\rm voxel} = 5.0\ h^{-1}$Mpc, and we set the Gaussian smoothing scale of the density field to be $R_s = 17.31 h^{-1}$Mpc. We again implemented a void-size cut, and kept voids that were $R_v \geq 20.0\ h^{-1}$Mpc, to minimize the influence of spurious voids and ease comparison with previous work.

Just as we had in Chapter 4, the VG-CCF was calculated using `pycorr`, a wrapper for the `CorrFunc` pair-counting code. In all cases, the VG-CCF was estimated using the Davis-Peebles estimator, ξ_{DP}. For the DESI DR1 emulator, we binned the VG-CCF in s and μ, to slightly larger scales, using 30 linearly-spaced bins from $0.7h^{-1}$Mpc to $150h^{-1}$Mpc, and 240 linearly-spaced bins for μ, from -1 to 1, followed by decomposing the VG-CCF into the monopole and quadrupole moments using Equation (1.60).

We also measured the VG-CCF in the `EZmock` covariance mocks, `AbacusSummit FFA` and `altMTL` mocks, and the DESI DR1 data.

5.3 Data covariance

The BAO analysis for DR1 of the DESI survey estimated the covariance by using a combination of semi-analytic [177] covariance and simulation-based covariance mocks [6, 7]. They utilized the `EZmock` method to generate 1000 independent covariance mocks for each redshift snapshot of the simulation and for each tracer, split across both galactic caps (NGC and SGC). The `EZmock` catalogues are not as accurate as the `AbacusSummit` fibre assigned cutsky mocks: they trade off small-scale clustering accuracy to produce nearly two orders of magnitude more realizations (1000 vs 25), while minimizing the influence of cosmic variance. This method has been successfully used in estimating the covariance, notably in the eBOSS collaboration [241].

Each of the `EZmock` catalogues for each galactic cap (NGC, SGC or NGC+SGC) for each snapshot and tracer combination started with cubic mocks that were constructed using 2000 $(6h^{-1}\text{Gpc})^3$ boxes with 1728^3 dark matter particles, each of these mocks had independent Gaussian initial conditions, and used the c000 cosmological model as their fiducial cosmology. The mocks were populated using a combination of the Zel'Dovich

approximation applied to the evolution of the dark matter field and a flexible bias model [46, 241]. Adding galaxies into the dark matter mocks entailed using the galaxy-halo parameters that were obtained from calibrating using the two-point and three-point galaxy correlation functions from the `AbacusSummit` mocks at each of the snapshots, which in turn utilized the HOD parameters obtained from the DESI OnePercent survey [238].

For each of these $(6h^{-1}\text{Gpc})^3$ cubic boxes, the positions of the galaxies were mapped to sky-coordinates, constructing a wedge centered around the snapshot redshift. The volume of these cubic boxes is significantly larger than the `AbacusSummit` counterparts, meaning that the process of mapping the Cartesian coordinates to sky coordinates did not require any replication of galaxies from the box to fill the volume. This is beneficial, as it minimized cosmic variance that may arise from incorrect replication. Each of these mocks were concatenated together, creating a pseudo-lightcone that spans the depth of the survey. Finally, each of these pseudo-lightcones were run through the FFA pipeline to account for fibre-collision and fibre assignment effects. This produced a library of 1000 mock catalogues for a specified combination of galactic caps (NGC, SGC, NGC+SGC) and tracers (LRG, ELG, QSO, BGS) that can be used to estimate the covariance.

There is one caveat, as when these mocks were being utilized in the DR1 analysis, they found that the `EZmock` covariance matrices systematically underestimated the variance, while the semi-analytical model had a higher variance estimate. This is relevant to our analysis, as it required a rescaling of the `EZmock` covariance for any covariance estimate going forward. The rescaling factor is calculated for different tracer bins, and we found that for our corresponding tracer, `LRG-1`: $0.4 < z < 0.6$, required a rescaling of the covariance by a factor of 1.15.

5.4 Model

We trained our neural network to learn the dependence of the VG-CCF on cosmological and HOD parameters using the `sunbird` and [5]`acm` packages [99]. We utilized the `AbacusSummit` DR1 mocks discussed in Section 5.2 as our training data and the `EZmock` covariance mocks to estimate our covariance matrix Section 5.3. We explored a variety of model architectures, loss functions and activation functions. To evaluate the quality of our model's predictions, we examined the best-fitting multipoles of the VG-CCF ξ_0, ξ_2, the covariance matrix, and the model's ability to solve the inverse problem: its estimation of cosmological parameters from the best fitting VG-CCF.

[5] https://github.com/epaillas/acm

We trained the neural network across 600 epochs, testing a variety of loss functions (Gaussian, L1 and L2 loss) and activation functions (SiLU, PReLU, Learned Sigmoid). We determined the best fit through its performance on the tests discussed in the next subsection. The best-performing model used a Learned Sigmoid activation function, with L1 (mean absolute error) loss. The trained network was a FCN with two hidden layers (with 75 neurons in the first layer and 448 neurons in the second), which made it a relatively shallow neural network, and regularly obtained a smaller loss than networks that had deeper architectures.

5.4.1 Covariances

We trained our emulator as specified in Section 3.2.1, to match the concatenated monopole and quadrupole moments of the VG-CCF $\xi_{vg} = (\xi_0, \xi_2)$ in our training set. We utilized the `sunbird` package to train the neural network and utilized the `acm` package to infer the cosmological parameters and test the pipeline. The uncertainties in our model were estimated by using the following equation for our covariance:

$$\mathsf{C} = \mathsf{C}^{\mathrm{emu}} + \mathsf{C}^{\mathrm{data}} \tag{5.6}$$

Unlike in the BOSS chapter, adding a $\mathsf{C}^{\mathrm{sim}}$ term would double count the covariance contribution from the `EZmock` catalogues. This parameterization of the covariance utilized the `EZmock` catalogues to estimate the covariance from the clustering, $\mathsf{C}^{\mathrm{data}}$ while the test cosmologies, `c000`, `c001`, `c002`, `c003`, `c004` and `c013` were utilized to construct the aforementioned $\mathsf{C}^{\mathrm{emu}}$. We constructed $\mathsf{C}^{\mathrm{emu}}$ following Equation (3.9). The resultant terms in the covariance can be found in Figure 5.7.

We then extended these tests to incorporate fibre assigned mocks that were constructed to test the DESI BAO pipeline before we fitted our DESI voids to the model. These fibre assigned mocks are the `altMTL` and `FFA` fibre assigned mocks, which each consist of 25 realizations of a mock with the `AbacusSummit` `c000` cosmology, and were constructed with either the `altMTL` or `FFA` fibre assignment regime. Once we showed this pipeline could potentially recover the cosmological parameters of not only the `AbacusSummit` DR1 mocks that we had prepared, but also the externally sourced `altMTL` and `FFA` mocks, we then applied the same pipeline to the DESI DR1 `LRG-1` NGC sample.

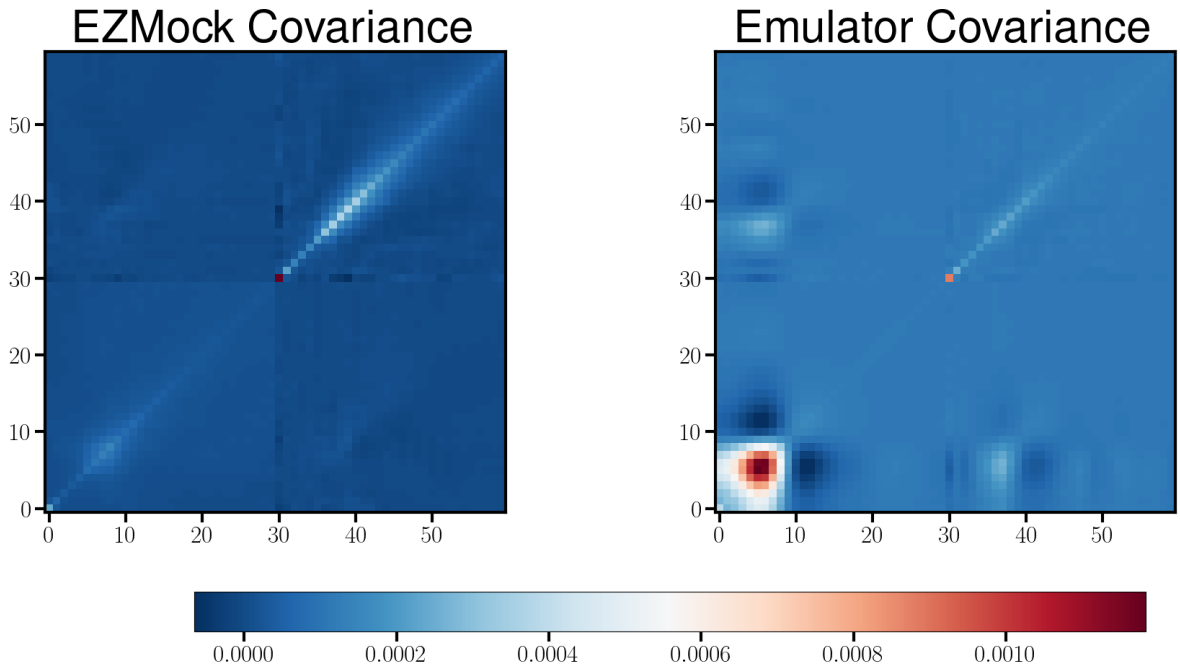

Figure 5.7: The components of our covariance matrix, as described by Equation (5.6) and Section 4.0.5. In this version of our model, we compute the covariance matrix from the `EZmock` mocks and multiply it by the aforementioned volume-scaling factor (left). We also compute the emulator covariance by using the emulator to predict the multipoles (ξ_0, ξ_2) for six different cosmological models, `c000` – `c004`, `c013`, which consist of 600 mocks as part of our test set. We note the presence of a bias in the covariance at small scales (right).

5.5 Testing the pipeline with `AbacusSummit` DR1 mocks

For each `AbacusSummit` DR1 mock considered, much as we did in Section 4.1.1, we inferred the posterior of the cosmological and HOD parameters: comparing the mock data vector ξ_0, ξ_2 to those generated by the emulator. As always, we considered the same six test cosmologies as we had utilized in the BOSS CMASS emulator, albeit with the changes described earlier in this chapter. We took the ξ_{vg} measured for the following test cosmologies `c000`, `c001`, `c002`, `c003`, `c004`, `c013` in redshift-space, and used them to construct C^{emu}, the emulator covariance. We utilized a consistent binning for the `AbacusSummit` and DESI data, where we calculated the multipole moments ξ_0, ξ_2, and split them into 30 separa-

tion bins each, for $0.7h^{-1}\text{Mpc} \leq \text{s} \leq 150h^{-1}\text{Mpc}$. We ran our inference using the priors described in Table 3.1.

We first fit `c000`, the baseline cosmology of `AbacusSummit`: The full posterior distribution is shown in Figure 5.8.

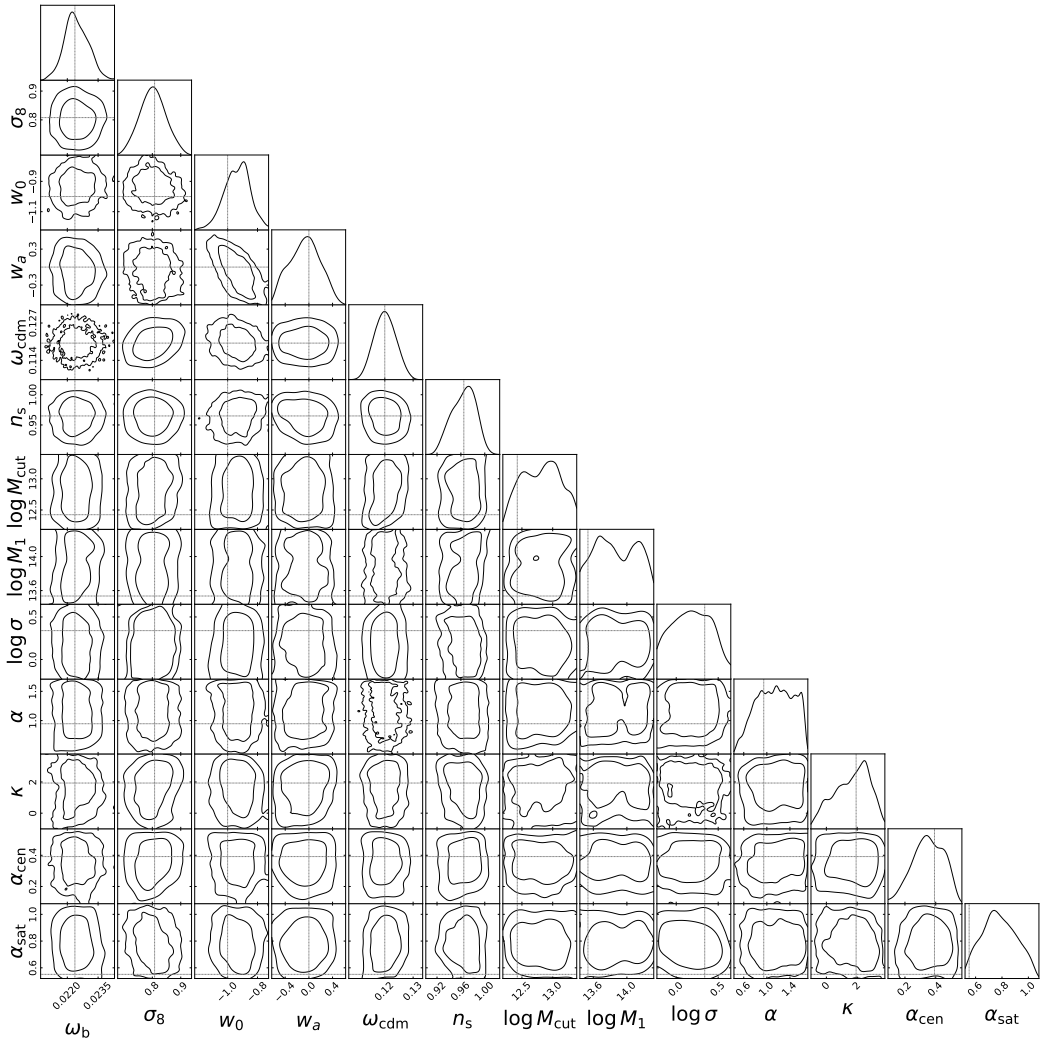

Figure 5.8: Full posterior obtained from the emulator for the `c000` cosmology and HOD parameters. While the cosmological parameters such as σ_8 and ω_{cdm} are well constrained, the same does not apply for the HOD parameters, continuing a trend seen in Chapter 4, where the VG-CCF is largely agnostic to galaxy-halo connection models.

Much like the CMASS emulator, we found that the full posterior shows limited to non-existent constraining power for all seven HOD parameters, as all seven parameters hit the limits of the priors.

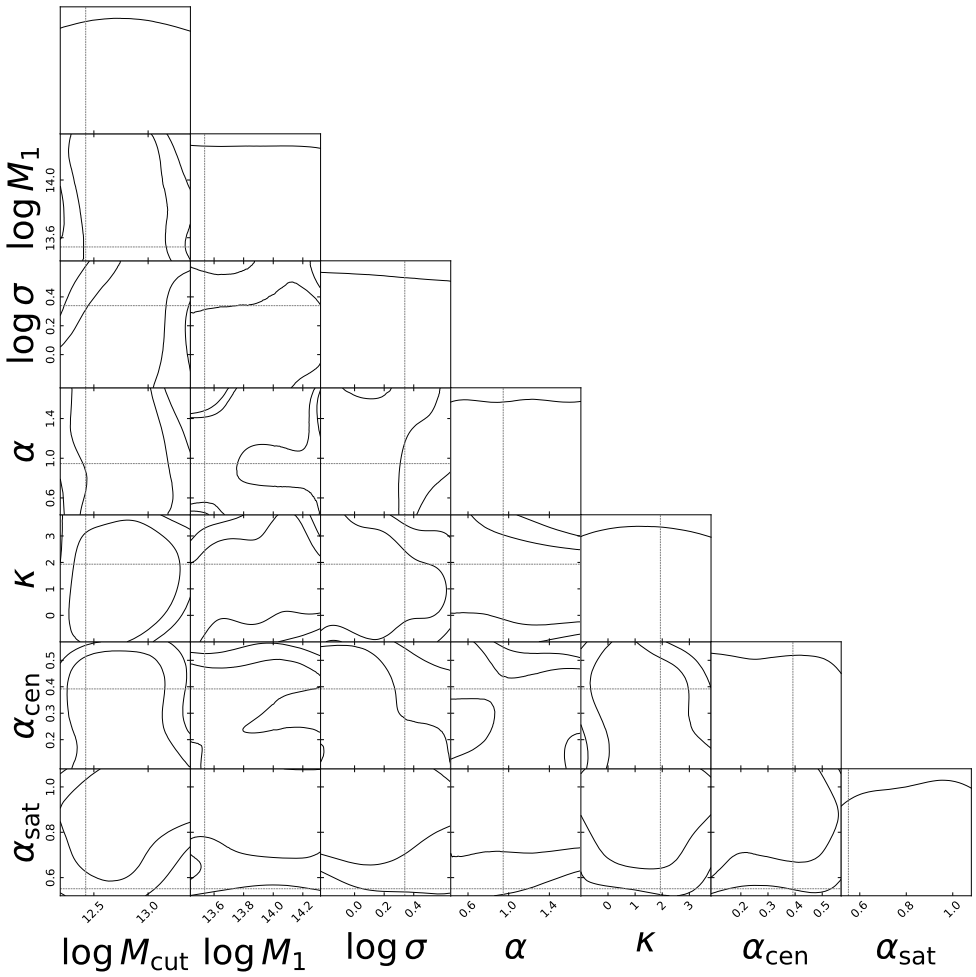

Figure 5.9: Building off of the results of Figure 5.8, we explore the constraining power of the VG-CCF exclusively for the HOD parameters in our model. We find the VG-CCF is agnostic to galaxy-halo connection models.

Figure 5.9 further exemplified this, as even limiting our parameter space to the seven parameter HOD model further underscored how the VG-CCF was agnostic for galaxy-halo connection parameters. As we did in 4.1.1, we marginalized over the HOD parameters

for the remainder of this work. Our model obtained constraints for various cosmological parameters, but we note that n_s, ω_b are tightly constrained by the CMB. Consequently, we placed emphasis on the constraints on $\omega_{\rm cdm}$, w_0, w_a, σ_8 and their derived quantities.

We first tested the emulator's recovery of the VG-CCF for one of test cosmologies, `c000`, the fiducial cosmology of the DESI survey. We found the emulator's prediction of VG-CCF was unbiased and accurate, with no part of our emulator's fit to `c000` being more than 2σ away from the measured VG-CCF, as Figure 5.10 shows.

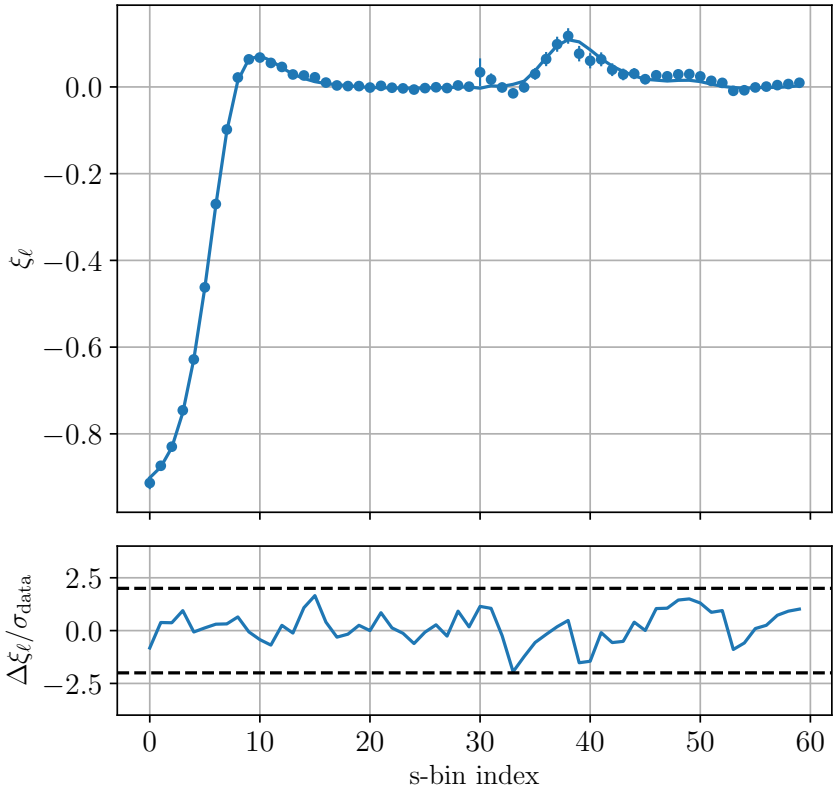

Figure 5.10: The emulator's predictions of the concatenated monopole and quadrupole of the void-galaxy cross-correlation function from the `AbacusSummit` test cosmology `c000`, averaged across 100 HODs, with their residuals in the panel below.

We evaluated our model's robustness by obtaining the best fits for the multipoles of the VG-CCF for `c000`, `c001`, `c002`, `c003`, `c004`, `c013`, the `AbacusSummit` test cosmologies, and then focused on our model's ability to obtain cosmological parameters that pertain to the

large-scale structure evolution at late times.

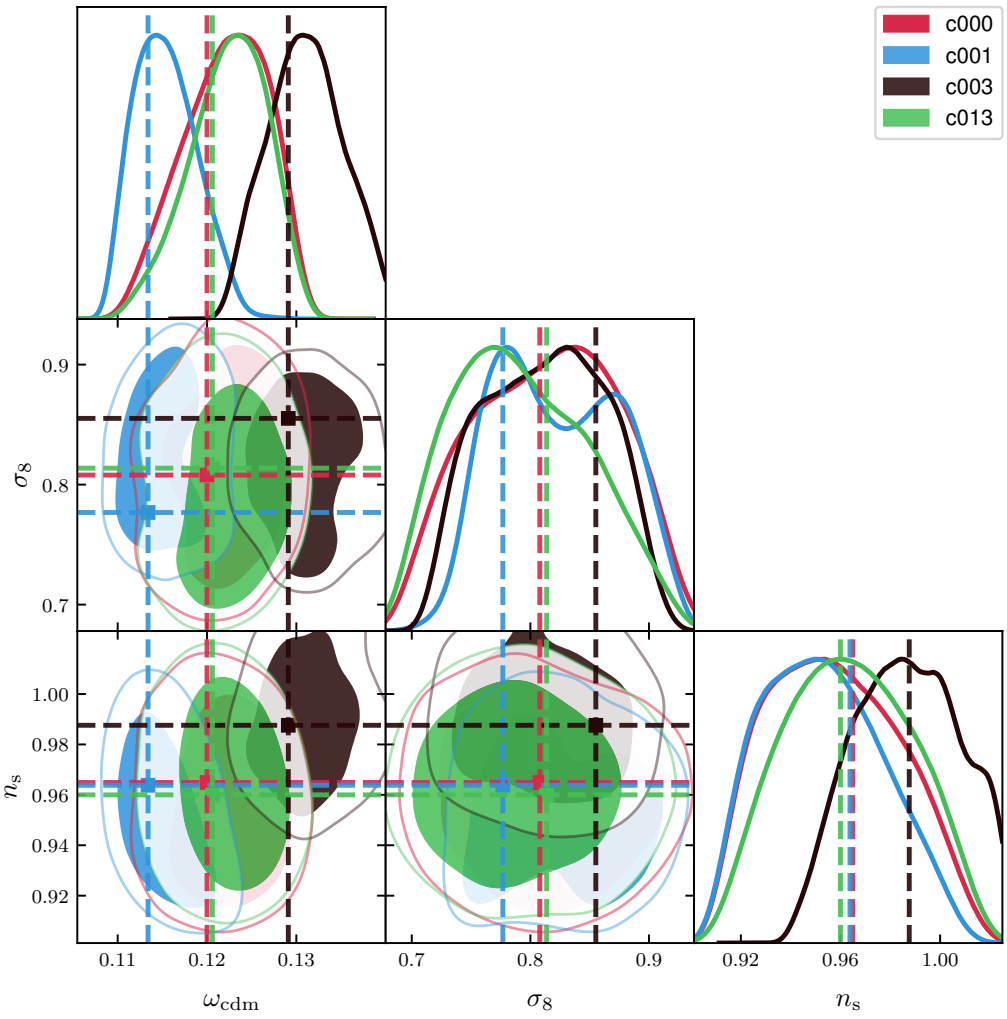

Figure 5.11: Posteriors obtained from the emulator, marginalized over ω_b and the HOD parameters, using the priors described in Table 4.1, with the test cosmologies and their true values for σ_8, n_s and $\omega_{\rm cdm}$. As discussed in the previous chapter, θ_* is fixed in the `AbacusSummit` simulations, allowing for $\Omega_{\rm m}$ to be obtained from $\omega_{\rm cdm} = (\Omega_c h^2), \omega_{\rm b} = (\Omega_b h^2)$.

As seen in Figure 5.11, we obtained marginalized constraints for $\omega_{\rm cdm}$ for most test cosmologies, while for σ_8 and n_s our constraints were limited by their prior walls. Compared

to our BOSS emulator, we note that the errors are notably larger, likely due to the DR1 footprint (varying completeness), the uniform binning in the VG-CCF, and the expanded HOD parameter space, which made it ill-equipped to model the small-scale features. Figure 5.11 shows the marginalized constraints for $\omega_{\rm cdm}$ and σ_8 and n_s fell within 2σ of the true parameter values but were limited by prior walls. Moreover, this constraining power in $\omega_{\rm cdm}$ was shown to apply to derived quantities, particularly $\Omega_{\rm m} = (\omega_{\rm cdm} + \omega_b)/h^2$. This was a crucial first step in the development of this DESI emulator, as even with significant changes to our pipeline, parameter space and model, we still demonstrated the possibility of recovering cosmological parameters.

5.6 Fibre assigned cutsky mock tests

The parameter recovery tests we performed using the `AbacusSummit` DR1 mocks with `altMTL` and `FFA` are quite similar to those we performed Section 5.5, but with a few key differences that motivate them. For one, even though the HOD parameter space subtended by our training sample is derived from the DESI One-Percent survey, none of our mocks used the same HOD parameters as the `altMTL` and `FFA` mocks, which used the best-fitting HOD parameters. Moreover, while the pipelines used to create these mocks and the ones discussed in Section 5.2 are the same, the purpose of these `altMTL` and `FFA` mocks is to test the DESI BAO pipeline, and consequently they each contain 25 realizations, compared to the mocks Section 5.2 that were created with a single realization. This means the `altMTL` and `FFA` multipoles were averaged by the number of realizations (25), while the `AbacusSummit` realizations were noisier. Finally, these DR1 mocks were downsampled to mimic fibre assignment, while the fibre assigned mocks utilized either the same fibre assignment regime as DESI (`altMTL`) or `FFA`. This latter point made for a crucial test of the pipeline - to what extent did fibre assignment affect the monopole and quadrupole of the VG-CCF, and its constraining power, relative to a downsampled mock?

We tested this fitting the mean of the `altMTL` and `FFA` mocks, with the errors scaled for a single volume (equivalent to the DESI DR1 NGC volume), and calculated the errors using the `EZmock` catalogues. The same priors were used as in the `AbacusSummit` section. We obtained the 2D posterior distribution, where we marginalized over all parameters except for $\Omega_{\rm m}$ and σ_8, seen in Figure 5.12. We calculated $\Omega_{\rm m}$ by obtaining h, and computing it along the chain, using a fixed value of $\omega_{\rm ncdm}$ and the varying ω_b, $\omega_{\rm cdm}$ values, marginalizing over them. We compared these to the survey cosmology, `c000`.

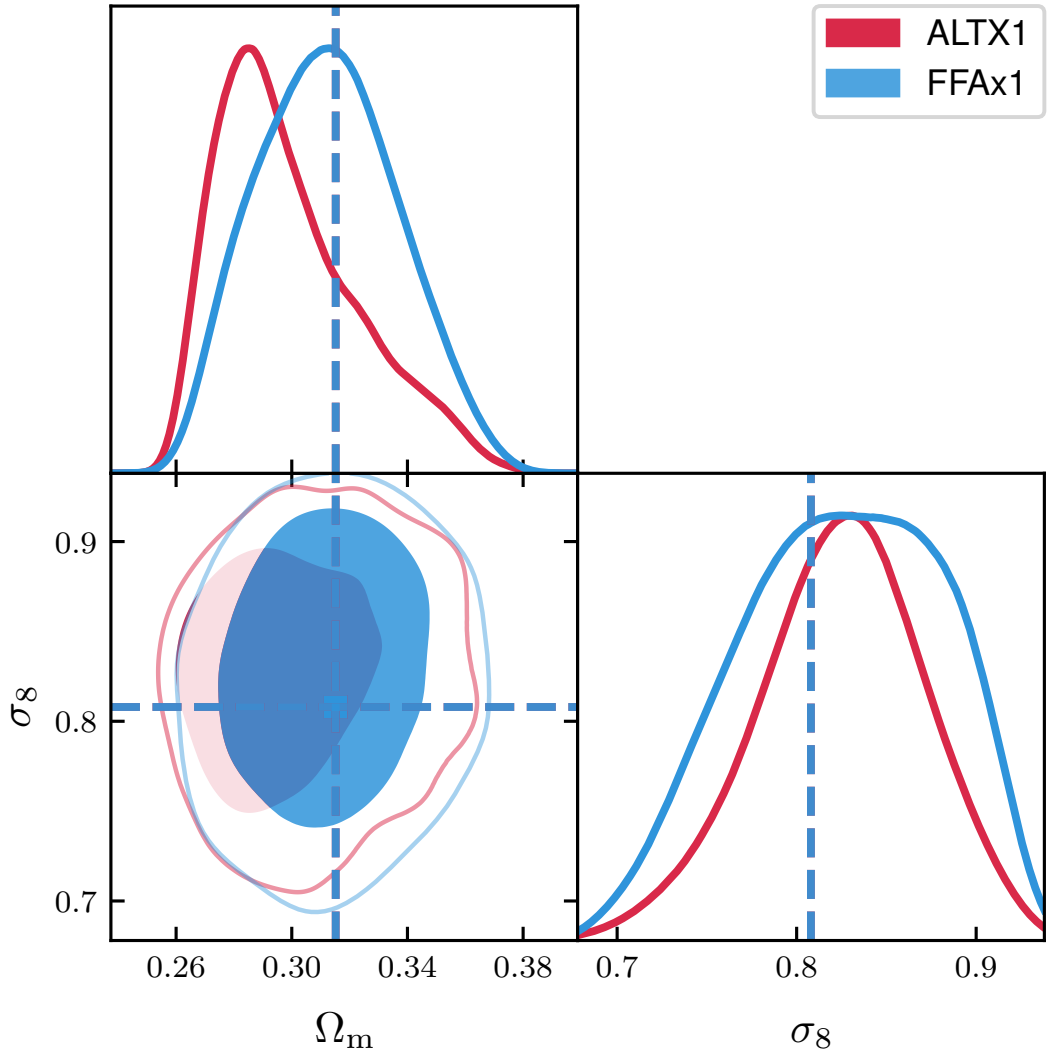

Figure 5.12: The marginalized posteriors of the emulator's predictions for the `altMTL` and `FFA` cutsky mocks prepared from the `AbacusSummit` cubic boxes using the `lss` pipeline. We averaged the 25 realizations and scaled the errors to the DESI NGC DR1 volume. We find that in both cases, the emulator can successfully predict Ω_m, σ_8 to within $1-\sigma$, across two sets of mocks that had two different fibre assignment treatments. The fiducial cosmology used to prepare these mocks is overplotted as dashed lines in this corner plot.

In the `FFA` mock, we found that the residuals between the emulator's maximum likeli-

hood solution and the mock's multipoles, shown in Figure 5.13, with the residuals in the lower panel scaled by the errors, σ_{data}, deviated from each other at very small scales, for the first 8 bins in the monopole, nearly 3σ apart.

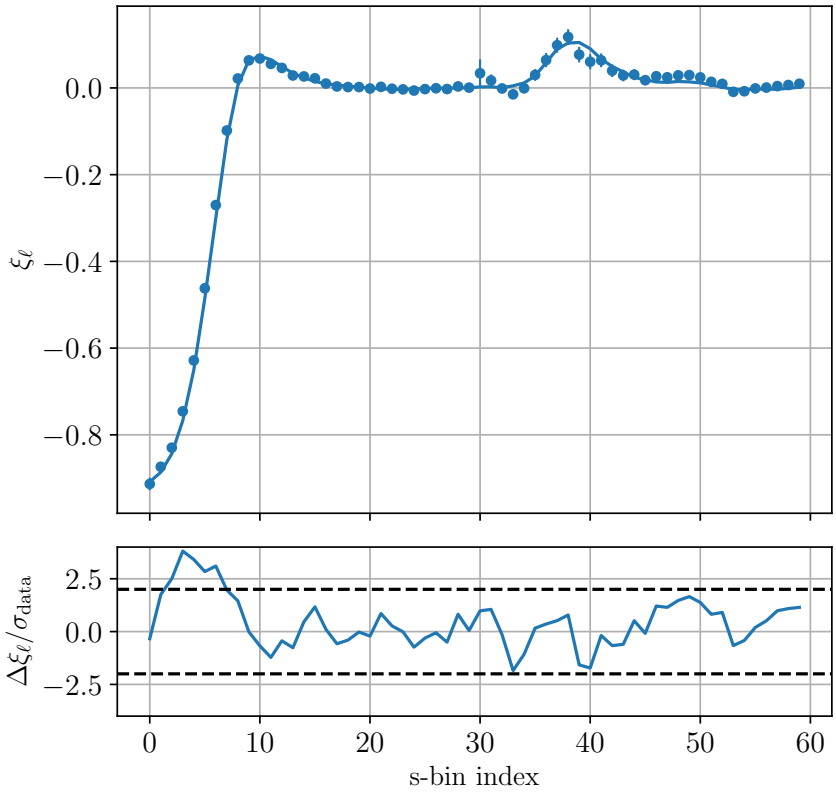

Figure 5.13: The emulator's predictions of the monopole and quadrupole of the VG-CCF from the mean of the VG-CCF for 25 `AbacusSummit` FFA mocks, with their errors scaled to match a single DESI LRG-1 volume. The residuals are shown in the lower panel.

At larger scales for the monopole, and for all scales for the quadrupole, this deviation was within 2.5σ, with the peak in the quadrupole's amplitude being offset from the maximum likelihood solution by a few bins. This is similar for the residuals of the multipoles in `altMTL`, seen in Figure 5.14, but here the small-scale bias is more pronounced, exceeding 3σ for 2 bins, with a similar feature in the quadrupole.

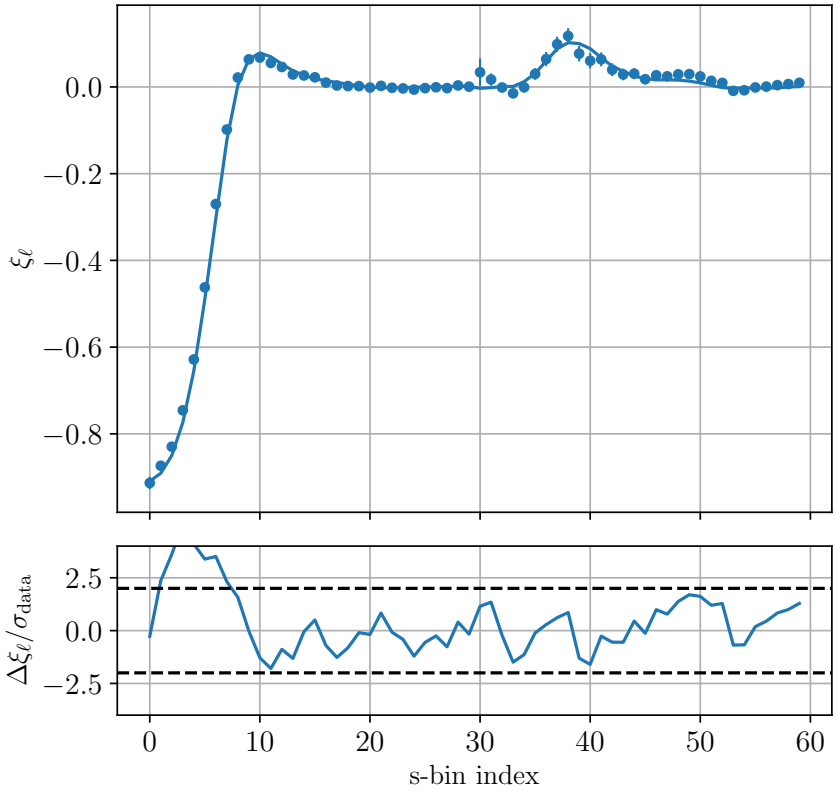

Figure 5.14: The emulator's predictions of the monopole and quadrupole of the VG-CCF from the mean of the VG-CCF for 25 `AbacusSummit` `altMTL` mocks, with their errors scaled to match a single DESI LRG-1 volume. The residuals are shown in the lower panel.

Figure 5.12 shows that the marginalized constraints for $\Omega_{\rm m}$ and σ_8 fell within 1σ of the true cosmological values used to construct these mocks, but the σ_8 constraints were prior-limited. We obtained $\Omega_{\rm m} = 0.312^{+0.023}_{-0.028}$ and $\sigma_8 = 0.825^{+0.068}_{-0.050}$ when using the VG-CCF from the `FFA` mocks, compared to $\Omega_{\rm m} = 0.299^{+0.013}_{-0.032}$ and $\sigma_8 = 0.824^{+0.050}_{-0.043}$ using the VG-CCF from the `altMTL` mocks. The marginalized 1D constraints for these parameters found Ω_m within 1.3σ of the true value for `altMTL`, and 0.170σ for `FFA`, and found σ_8 within 0.32σ of the true value for `altMTL`, and 0.236σ of the true value for `FFA`. Our best fit was closer to the true cosmology using the `FFA` mocks than it was with `altMTL`, but in both cases, the errors were similar: for $\Omega_{\rm m}$, the `FFA` constraint was 13% wider than `altMTL`'s constraint, while for σ_8 the `FFA` constraint was nominally 16% wider. Notably, for the `altMTL` mocks, which were used to construct the DR1 mocks through downsampling, we saw a more pronounced

small-scale bias relative to the FFA mocks, and also noticed the altMTL fits resulted in a less accurate recovery of Ω_m and σ_8, however, the recovered parameters were within 1σ of the true cosmology used to construct these mocks.

Broadly speaking, with this recovery of the cosmological parameters from voids in the altMTL and FFA mocks, we concluded that our emulator could feasibly recover the cosmological parameters, but we note the presence of a small-scale bias and of mild differences in the fibre assignment regimes. With these caveats in mind, we applied the model to the DESI DR1 LRG-1 sample, which we did in the following sections.

5.7 DESI data fits

Having tested our emulator model on DR1 mock catalogues generated on a wide cosmological parameter space to evaluate its ability to solve the inverse problem, and then tested them on multiple realizations of DESI fibre assigned mocks, we then wanted to evaluate the model with the VG-CCF obtained from the survey data. We tested this model using the VG-CCF obtained from DR1 of the DESI survey and selected the NGC portion of the sample from the LRG-1 bin $0.4 < z < 0.6$. These fits served as a demonstration of the feasibility of this technique, with the fibre-collision and selection effects of the DESI DR1 survey window incorporated. We first examined the cosmological parameter recovery for a ΛCDM cosmology, and explored the full posterior. We then demonstrated how a future wCDM and w_0w_aCDM analysis could be conducted, once multiple redshift windows are added to the training set, which would allow the redshift evolution of the VG-CCF to be trained into the emulator.

The comparison of these results to the DESI BAO cosmological parameter constraints and other void analyses come with the caveat that the AbacusSummit mocks are constructed with a fixed θ_*. The implications have been previously discussed in Chapter 4.

5.7.1 Baseline ΛCDM

We first considered a flat ΛCDM model, and fit to the data using the same methodology discussed above. The VG-CCF is obtained from DR1, and our emulator model obtained the best fitting VG-CCF monopole and quadrupole, which we then used to obtain cosmological parameter constraints. The multipole moments of this maximum likelihood solution are compared to the data in Figure 5.15, with their residuals shown in the lower panel.

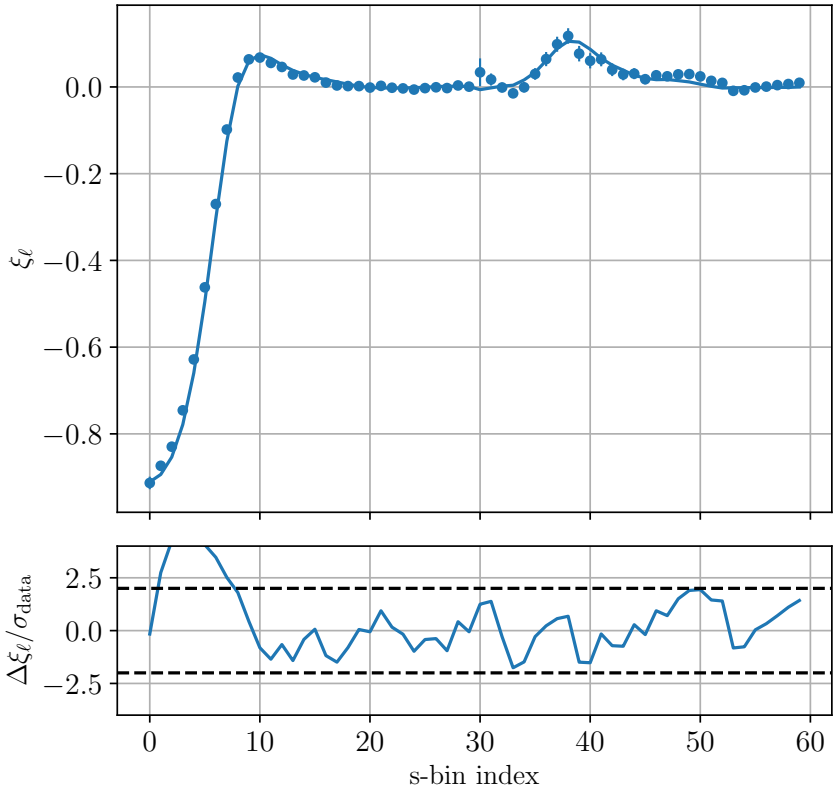

Figure 5.15: ΛCDM best-fit for the VG-CCF obtained from the emulator when the DESI DR1 sample is analyzed with it. The monopole and quadrupole predictions lie within 2.5σ of each other, save for the first 6 bins of the monopole, where the small-scale bias of the emulator is apparent.

The most crucial difference between this fit and the CMASS BOSS fit for the data is that the small-scale bias is still readily apparent when we fit the model using the survey data.

We found that the residuals are significantly offset at small scales compared to larger scales, similar to what we had seen in tests in the previous sections. The greatest deviation occurs at scales below $20h^{-1}$Mpc for the monopole, while for the quadrupole, there was a small deviation (2.5σ) in the bins around the scale of $100h^{-1}$Mpc. The bias was added in quadrature to the covariance to account for this. This bias was detected in the BOSS CMASS emulator, but with a smaller amplitude. For the DESI DR1 emulator, there could be a few factors influencing the amplitude. First, the DESI DR1 emulator was

trained with mocks that have varying angular completeness, while the BOSS emulator never incorporated this information in its training data, as its angular completeness was far more uniform. Second, these DR1 mocks also modelled the selection of voids from the survey, including fibre-collision, a small scale effect, while the BOSS mocks did not. Third, the binning of the VG-CCF was uniform, meaning there would only be a handful of small-scale bins to capture this information. Finally, emulator covariance estimation in both was an aggressively conservative estimate, so it is possible that this bias is partially driven by the emulator covariance.

The cosmological parameter constraints are wider than the BOSS emulator, likely driven by these differences in the model. This can be inferred from the projected posterior with cosmological constraints on σ_8, and Ω_m shown in Figure 5.16. We note that the σ_8 marginalized constraint is partially limited by the prior wall for $\sigma_8 > 0.938$. With our emulator, we obtained a ΛCDM marginalized constraint of $\Omega_\mathrm{m} = 0.291^{+0.011}_{-0.027}$ and $\sigma_8 = 0.842^{+0.046}_{-0.036}$ (prior limited).

We compared this Ω_m constraint to the BAO analysis of DR1 from DESI, and (notionally with σ_8) to the full-shape analysis of DESI DR1. This set of comparisons, shown in the first rows of Table 5.3, illustrated the constraining power obtained from this emulator with respect to the BAO measurement. The DESI DR1 BAO analysis obtained $\Omega_\mathrm{m} = 0.293 \pm 0.015$. When we compared this to our CMASS emulator, which utilized cubic boxes as its training data, that pipeline yielded a 5% reduction in Ω_m error, and a 32% reduction in σ_8 error.

While the DR1 mock pipeline incorporated fibre assignment and DR1 selection effects into the emulator training sample in order to improve the fidelity of the training sample with respect to the VG-CCF from DESI DR1, the constraining power was limited compared to the emulator model discussed in Chapter 4. When we compared the DESI void emulator's Ω_m and σ_8 parameter constraints to best fit of the BAO analysis of DESI DR1 using redshift bins, our errors in Ω_m were 23% larger than the constraint from the BAO analysis of DESI DR1.

We extended this comparison of our emulator to the DR1 full-shape analysis, limited to the same redshift bin $\texttt{LRG} - \texttt{1}$. The full-shape analysis of DESI DR1 using the data from $\texttt{LRG} - \texttt{1}$ led to the following constraints of: $\Omega_\mathrm{m} = 0.297 \pm 0.201$ and $\sigma_8 = 0.844^{+0.083}_{-0.095}$. Our emulator's constraints, in comparison, led to a 10% reduction in errors in Ω_m and halved the errors in σ_8. The mean constraint for Ω_m was within 0.5σ of the full-shape `LRG-1` bin result and within 0.05σ of the full-shape `LRG-1` bin result for σ_8. The results were recorded in Table 5.2.

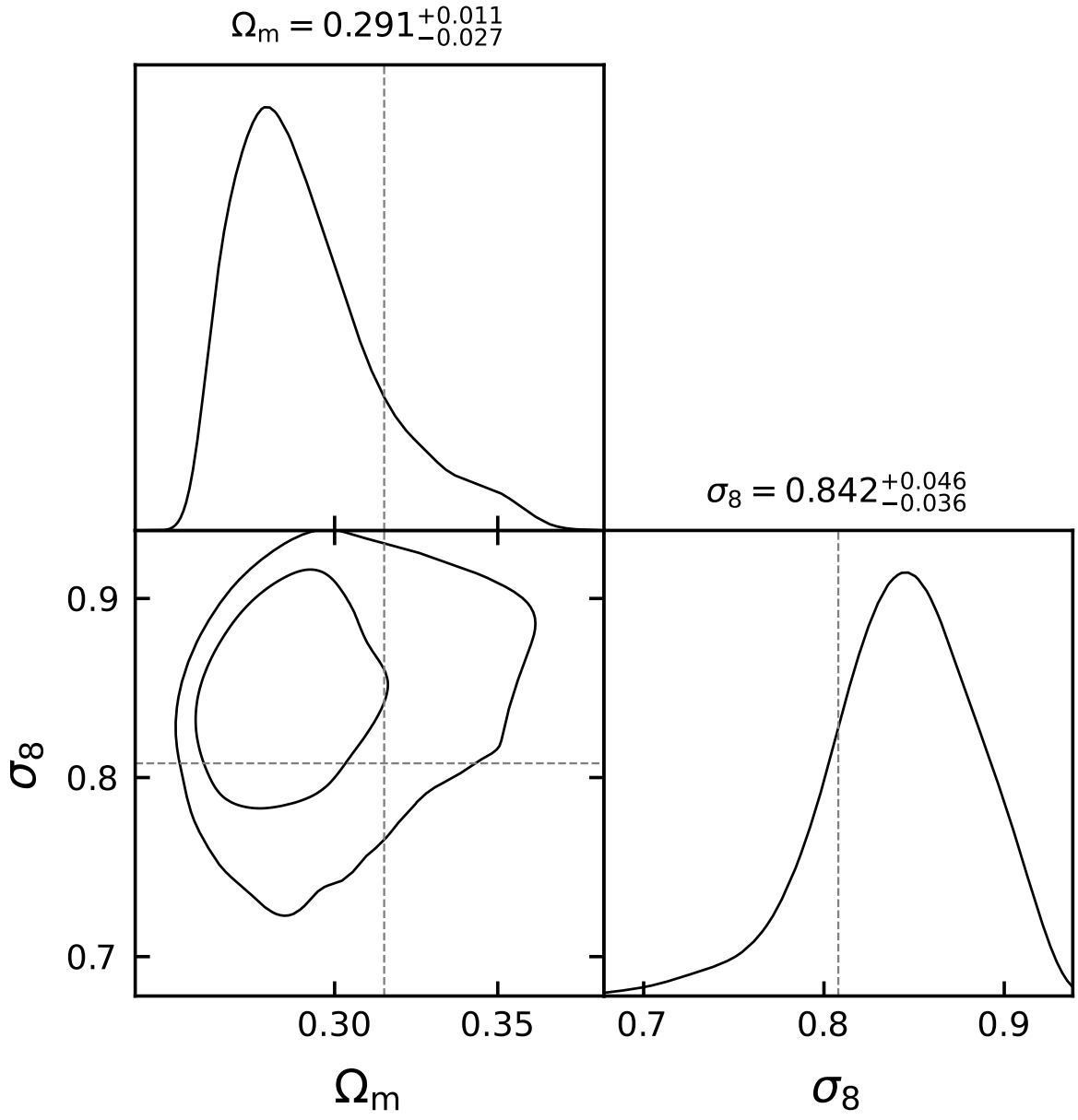

Figure 5.16: The marginalized posterior distribution of ΛCDM cosmological parameters inferred from fitting the emulator to the DESI DR1 data. We left ω_{cdm}, ω_b, n_s,σ_8 as free parameters and kept all other cosmological and HOD parameters fixed before marginalizing to obtain σ_8 and $\Omega_{\rm m}$ parameter constraints. We overplot the survey's fiducial cosmological parameter values, `c000`.

This demonstrated that the emulator obtained constraints on $\Omega_{\rm m}$ and σ_8 that could potentially be comparable if these limitations in the model's sensitivity to σ_8 are addressed. These constraints illustrated that the VG-CCF emulator can produce cosmological constraints that are complementary to the information obtained from a BAO analysis. It should be noted that these results used only a single redshift bin, and would not contain the redshift evolution information that the BAO and Full-Shape analyses of DESI DR1 contained. This lack of tomographic information made obtaining constraints on the dark energy equation of state parameters for wCDM and w_0w_aCDM cosmologies infeasible. The void emulator obtained ΛCDM cosmological constraints, primarily for $\Omega_{\rm m}$, using a single redshift bin, $0.4 < z < 0.6$ — a subsample of the DESI DR1 catalogue. While we compared our constraints in that section to the cosmological constraints obtained from the BAO analysis of DESI DR1, this came with the caveat that the BAO contains tomographic information; the feature was analysed at multiple redshift bins.

5.7.2 wCDM results

Our analysis in this section and the subsequent one was a proof-of-concept of joint constraints that could be obtained using the information from an emulator of the VG-CCF combined with the BAO and BAO+SNIa datasets from DESI DR1. Extensions to multiple redshift bins should make a thorough execution of this concept feasible and interesting. The figures and constraints obtained from the joint likelihood analysis discussed below are merely a proof-of-concept of what constraints might be obtained if a future VG-CCF emulator modelled multiple redshift bins.

Obtaining standalone cosmological parameter constraints with the VG-CCF emulator strained the model, as the parameter space used in the ΛCDM analysis, $(\Omega_{\rm m}, \sigma_8, w_0)$, had encountered difficulties when sampling σ_8 and w_0, as the posterior distribution was prior-limited, as seen in Figure 5.17. For w_0, this was expected, as fitting the DESI DR1 VG-CCF from a single redshift bin with the emulator would not contain extensive redshift evolution which would have been needed to constrain these quantities. As such, obtaining constraints on wCDM and w_0w_aCDM cosmologies was done by utilizing our void emulator as a complementary constraint to existing datasets.

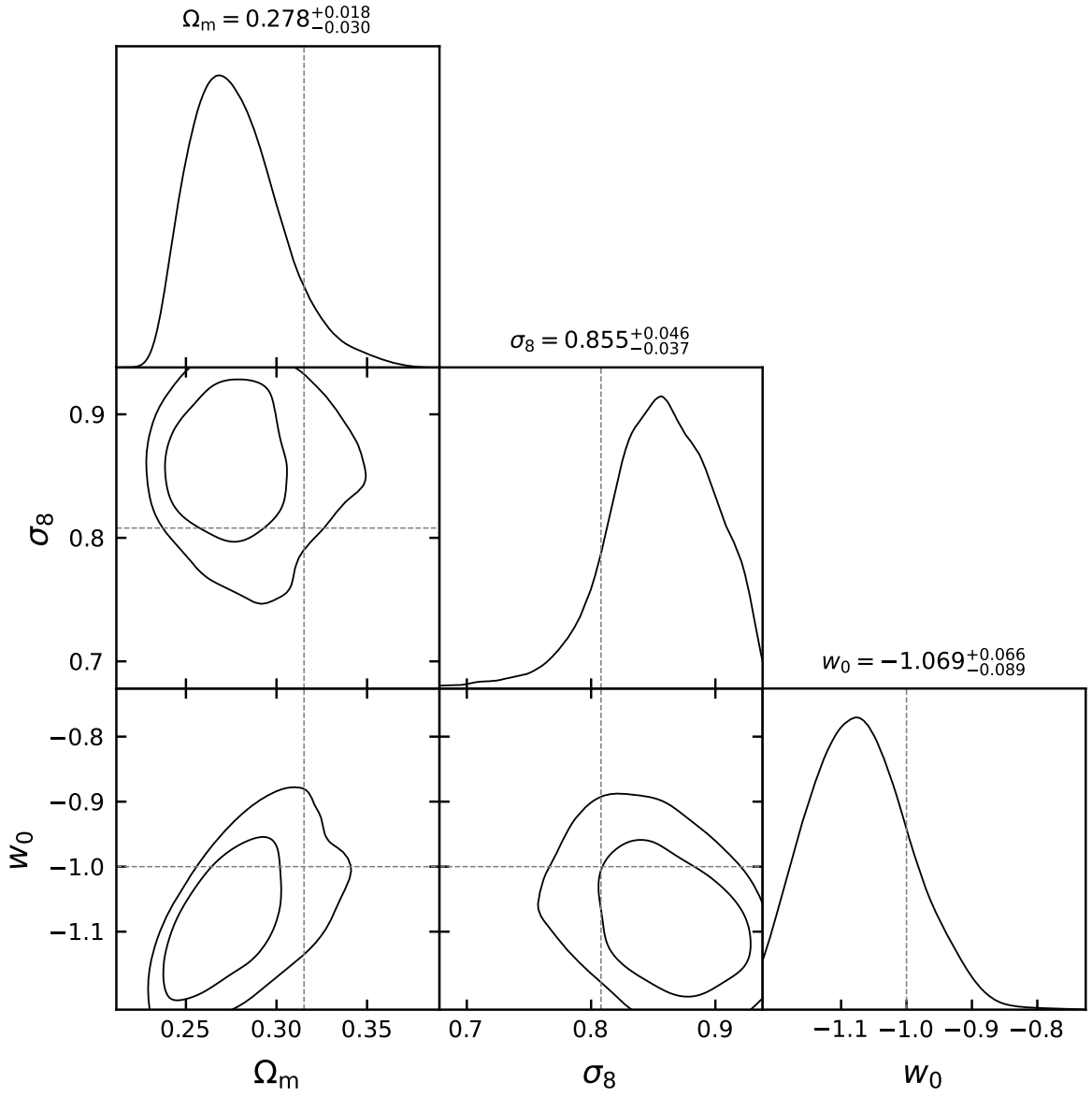

Figure 5.17: The marginalized posterior distribution of wCDM cosmological parameters inferred from fitting the emulator to the DESI DR1 data. Leaving ω_{cdm}, ω_b, n_s, σ_8 and w_0 as free parameters and kept all other cosmological and HOD parameters fixed. The survey's fiducial cosmological parameter values, `c000` were overplotted. Unlike the CMASS emulator, the DESI VG-CCF emulator is challenged by extending the parameter space to include w_0, limiting its ability to obtain any information about dark energy.

We did this by employing a joint likelihood analysis. We took advantage of the independence of the VG-CCF with respect to the BAO feature, which has been demonstrated in past literature, notably [235]. This independence of measurements allows for DESI DR1 BAO chains to be importance sampled, and the joint likelihood of combining the BAO and VG-CCF information to be easily extracted. We utilize the fact that the chains share a set of common cosmological parameters, θ, and used a Gaussian kernel density estimate (KDE) to interpolate the likelihood surface using the weights and parameters of the BAO chain. We then inferred the new joint likelihood by using the sample points of the VG-CCF chain. We utilised this analysis in order to obtain cosmological parameter constraints for wCDM and w_0w_aCDM models. These results came with a few additional caveats. For one, the BAO chain utilizes information from multiple redshift bins, while our void emulator was trained using the voids found in a single redshift bin (`LRG-1`, $0.4 < z < 0.6$). Additionally, while this assumption has only explored for the VG-CCF and BAO measurements, but not for a full-shape analysis. Furthermore, the priors of the BAO chains were significantly larger than the emulator's prior walls.

We employed this resampling to obtain constraints on $\omega_{\rm cdm}$ and w_0 in Figure 5.18. We obtained $\Omega_{\rm m} = 0.290^{+0.012}_{-0.014}$ and $w_0 = -1.033 \pm 0.062$ as our marginalized constraints. Our joint likelihood analysis favoured lower values of w_0 and lower values of $\Omega_{\rm m}$ than the DESI BAO. For $\omega_{\rm cdm}$, we obtained $\omega_{\rm cdm} = 0.1151^{+0.0023}_{-0.0035}$ and $w_0 = -1.029 \pm 0.065$. Compared to the standalone constraints on wCDM obtained using the CMASS emulator, we found that our means were similar ($-1.008^{+0.079}_{-0.10}$ vs -1.033 ± 0.062), and the errors with this joint analysis were about 66% the size of the BOSS emulator's errors. This comparison illustrated that the emulator used in Chapter 4 is a demonstration of the methodology, but does not include the modelling of fibre-collision and the survey's variation in completness that were added in Chapter 5, which likely offer a more realistic picture of the VG-CCF constraining power.

As a proof-of-concept, we also explored the cumulative effect of combining the BAO and SNIa datasets (Pantheon Plus, Union3 and DESY5) with the VG-CCF constraints from this emulator. We did this using importance sampling of the MCMC chains obtained from sampling the void-emulator's model for w_0CDM and w_0w_aCDM cosmologies. For the former, this analysis is still inhibited by prior walls even with the joint constraint. It should be emphasized that in this proof-of-concept, only some information can be obtained from emulation of the VG-CCF, since most of the constraining power in w_0w_aCDM cosmologies using large-scale structure observables comes from their redshift evolution. These SNIa+BAO constraints for w_0w_aCDM cosmologies are explored and discussed in the next section.

When we explored this proof-of-concept, we combined the BAO information with the

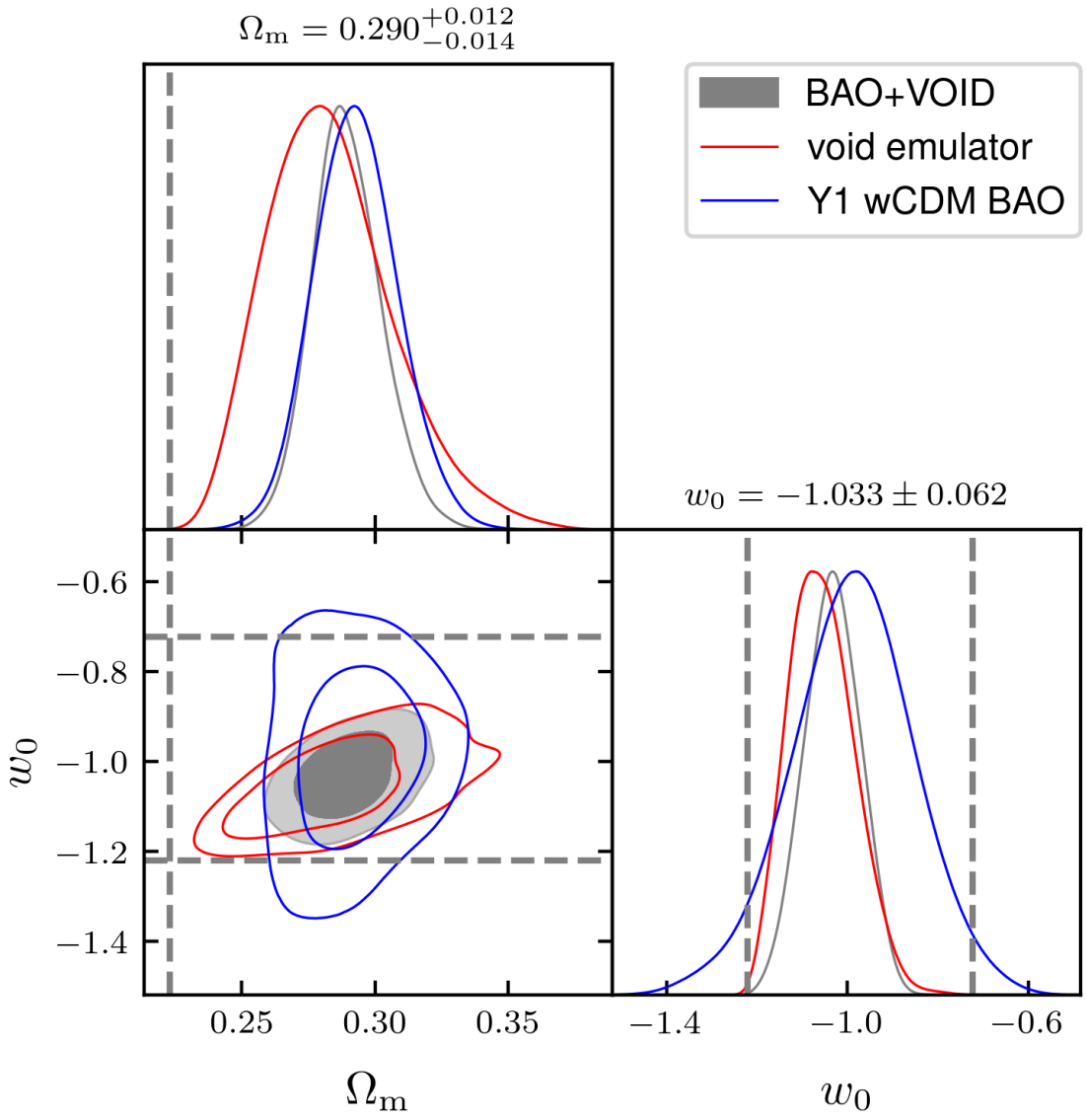

Figure 5.18: Joint likelihood analysis utilizing the DESI DR1 BAO dataset (blue) and the void emulator (red) to obtain joint constraints on $\Omega_{\rm m}$ and w_0 (grey).

constraints from our void emulator that led to a drastic reduction in errors for $\Omega_{\rm m}$ and w_0, and it was further bolstered by incorporating SNIa cosmological parameter constraints, leading to a reduction in errors for w_0, where the DESI DR1 BAO wCDM result for $w_0 = 0.99^{+0.15}_{0.13}$ shifts to $w_0 = -1.029 \pm 0.065$ when combining the void and BAO information resulting in a 46% reduction in errors. Combining the SNIa+BAO datasets with the void emulator's predictions also had a pronounced impact on the inferred dark energy equation of state parameter, and it is here where we can see the void emulator is more effective in constraining $\Omega_{\rm m}$ (or $\omega_{\rm cdm}$) than w_0.

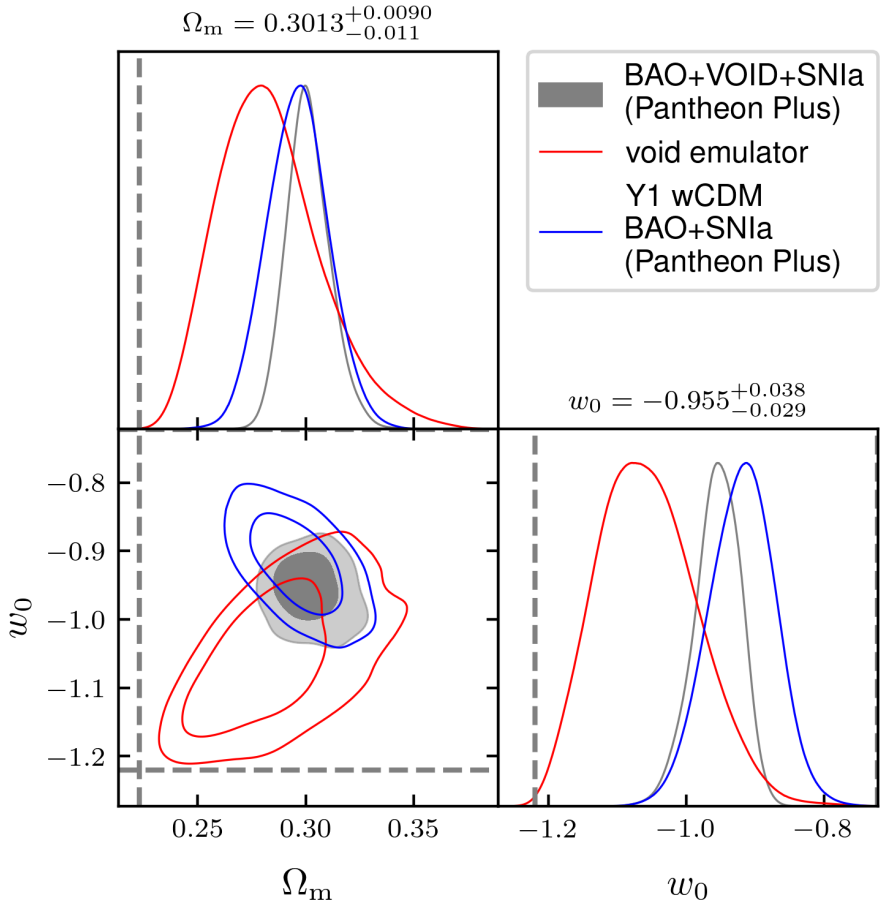

Figure 5.19: Joint likelihood analysis utilizing the DESI DR1 BAO+SNIa (Pantheon Plus) dataset (blue) and the void emulator (red) to obtain joint constraints on $\Omega_{\rm m}$ and w_0 (grey).

A full comparison to the DESI DR1 BAO and DESI DR1 full-shape result is included in Table 5.2. Additional constraints for DESI BAO+SNIa analyzed with the void emulator are enclosed in Appendix A, which also included the results of our joint analysis using a parameter space of ω_{cdm}, w_0, w_a instead of Ω_m, w_0, w_a, as we wanted to explore how much these constraints would change as we swapped between the two parameterizations. For reference, our emulator chains allowed ω_b and ω_{cdm} to vary.

Measurement	$\Omega_{\rm m}$	σ_8	w_0	w_a
flat ΛCDM				
VG-CCF LRG-1 NGC	$0.291^{+0.011}_{-0.027}$	$0.842^{+0.046*}_{-0.036}$		
DESI Full-Model LRG-1 [8]	0.297 ± 0.021	$0.844^{+0.083}_{-0.095}$	–	–
DESI Full-Model all bins [8]	$0.284^{+0.010}_{-0.011}$	0.839 ± 0.034	–	–
DESI FS+BAO+BBN [5]	0.2962 ± 0.0095	0.842 ± 0.034	–	–
DESI DR1 BAO [6]	0.293 ± 0.015	–	–	–
flat wCDM				
DESI DR1 BAO [6]	0.293 ± 0.015	–	$-0.99^{+0.15}_{-0.13}$	–
VG-CCF+BAO	$0.290^{+0.012}_{-0.014}$	–	-1.033 ± 0.062	–
VG-CCF+BAO+Pantheon	$0.3013^{+0.0090}_{-0.011}$	–	$-0.955^{+0.038}_{-0.029}$	–
flat w_0w_aCDM				
DESI DR1 BAO [6]	–	–	$-0.55^{+0.39}_{-0.21}$	<-1.32
DESI FS+BAO+BBN+Pantheon [5]	0.3084 ± 0.0089	0.820 ± 0.035	-0.875 ± 0.072	$-0.61^{+0.42}_{-0.36}$
VG-CCF+BAO	$0.292^{+0.013}_{-0.016}$	–	-1.001 ± 0.090	$-0.13^{+0.21}_{-0.28}$
VG-CCF+BAO+Pantheon	$0.3048^{+0.0093}_{-0.011}$	–	-0.912 ± 0.048	$-0.24^{+0.16}_{-0.23}$

Table 5.2: Table of marginalized constraints of $\Omega_{\rm m}$, σ_8,w_0 and w_a obtained from emulation of the VG-CCF using DESI DR1 LRG-1 voids, and from the DESI BAO and Full-Shape analysis of the DR1 sample. Where possible, we select results from the LRG-1 redshift bin, which corresponds to the redshift bin used in our analysis ($0.4 < z < 0.6$). The dashed rows contain the results of the joint analysis with DESI DR1 BAO and BAO+SNIa. Starred quantities denote prior-limited constraints.

5.7.3 w_0w_aCDM results

As above, the void emulator with a single redshift bin of the VG-CCF from DESI DR1 was not be expected to produce reasonable w_0 and w_a constraints. We again utilised the joint-likelihood analysis discussed previously to obtain joint constraints on cosmological parameters. This emulator could be feasibly extended to include redshift evolution by including more than one redshift bin, a fact we will address in the discussion.

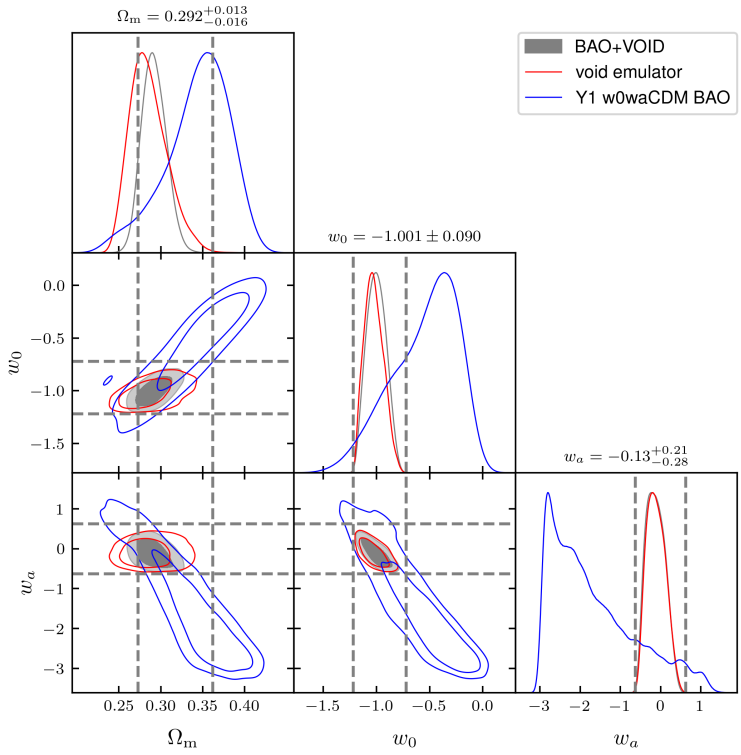

Figure 5.20: Cosmological parameter constraints from our joint likelihood analysis. Importance sampling was used to construct a joint likelihood for the BAO and void emulator chains. The combined information from the BAO (blue) and void emulator (red) does significantly reduce the marginalize constraints (grey).

This analysis was conducted with a w_0w_aCDM cosmological model, exploring the combined constraining power of our void emulator in tandem with the BAO and BAO+SNIa

datasets used in the DESI DR1 analysis. The full-shape chains were excluded from this analysis, as the assumption of independence has only been tested with voids and BAO, but not with full shape modelling.

Combining the BAO w_0w_aCDM likelihood with the void emulator chain yields the constraints in Figure 5.20. We obtained $\omega_{\rm cdm} = 0.1161^{+0.0023}_{-0.0033}$, $w_0 = -0.981 \pm 0.080$,$w_a = -0.21^{+0.15}_{-0.25}$. We also obtained $\Omega_{\rm m} = 0.3048^{+0.0093}_{-0.011}$, $w_0 = -0.912 \pm 0.048$, $w_a = -0.24^{+0.16}_{-0.23}$. Crucially, the constraints we obtained for w_0 and w_a were enclosed within the prior walls of our emulator pipeline. Our errors for w_0 were 3.75 times smaller than the DESI DR1 BAO.

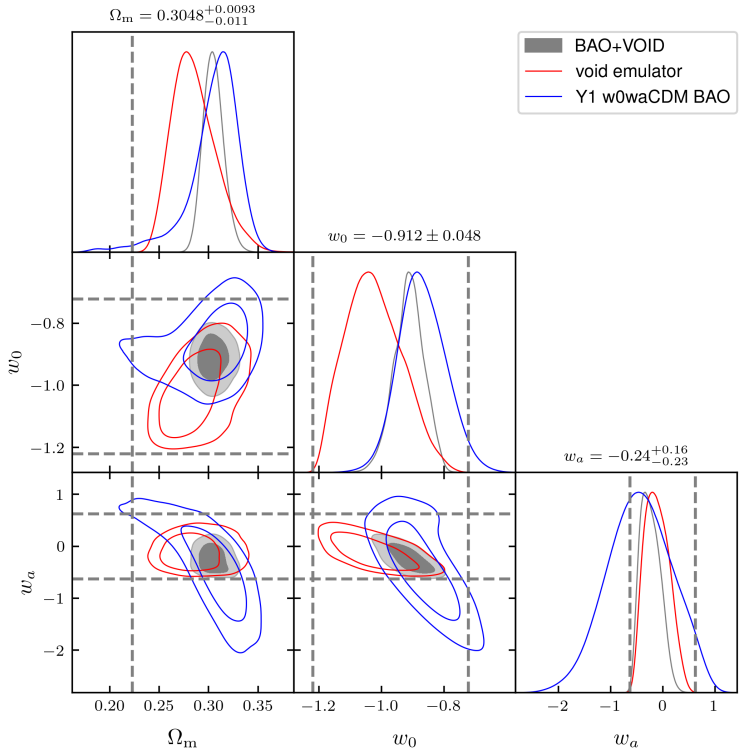

Figure 5.21: Joint likelihood analysis combining the BAO+Pantheon result (blue) with the void emulator (red). The grey dashes demarcate the prior walls for the void emulator, and the grey contour denotes the joint likelihood.

Extending our analysis to include SNIa datasets (Union3, Pantheon Plus and DESY5) can be seen in Figures 5.21, A.7 and A.8. Combining the BAO with SNIa datasets to obtain a joint void-BAO+SNIa likelihood did seem to reinforce the existing support for a w_0w_aCDM cosmological model, shifting the mean of w_0 further away from its fiducial value of -1, and also shifting the mean of w_a further away from its fiducial value of 0.0 as well. Most of the improvement in constraining power came from the void emulator being particularly sensitive to $\omega_{\rm cdm}$, and the narrower parameter range, which might have narrowed the existing constraints. Future work would be required to explore the constraining power of a void emulator as a way to boost existing constraints in w_0 and w_a as obtained from BAO and SNIa measurements. The cosmological parameter constraints from this section and the previous are summarized in Table 5.3.

Measurement	$\omega_{\rm cdm}$	w_0	w_a
flat wCDM			
VG-CCF+BAO	$0.1151^{+0.0023}_{-0.0035}$	-1.029 ± 0.065	–
VG-CCF+BAO+DESY5	$0.1158^{+0.0021}_{-0.0035}$	$-0.922^{+0.040}_{-0.035}$	–
flat $\mathbf{w_0w_a}$CDM			
VG-CCF+BAO	$0.1161^{+0.0023}_{-0.0033}$	-0.981 ± 0.080	$-0.21^{+0.15}_{-0.25}$
VG-CCF+BAO+DESY5	$0.1173^{+0.0022}_{-0.0035}$	-0.871 ± 0.045	$-0.32^{+0.11}_{-0.21}$
VG-CCF+BAO+Pantheon	$0.1168^{+0.0023}_{-0.0036}$	$-0.921^{+0.052}_{-0.044}$	$-0.26^{+0.13}_{-0.23}$
VG-CCF+BAO+Union3	$0.1169^{+0.0022}_{-0.0039}$	$-0.888^{+0.068}_{-0.046}$	$-0.30^{+0.12}_{-0.22}$

Table 5.3: Table of marginalized constraints for $\omega_{\rm cdm}$, w_0 and w_a utilizing a joint-likelihood analysis of the void emulator with BAO and SNIa, exploring wCDM and w_0w_aCDM cosmological models.

5.8 Discussion

This implementation of the VG-CCF emulator as applied to the DESI DR1 NGC sample demonstrates the feasibility of this technique for obtaining cosmological constraints, but this is dependent on the implementation of this model. In our implementation, the size of the parameter space explored, and the density of our training data likely influenced the emulator's ability to recover constraints on $\Omega_{\rm m}$ and σ_8. Furthermore, the demonstrated necessity to incorporate the effects of fibre assignment, fibre-collision and reshaping the training data to match the window of DR1 could have introduced challenges for this model. The transition in the preparation of training data from Chapter 4 to Chapter 5 illustrated that there were many subtleties to consider when designing these models.

This technique could also be applied to cosmological models beyond the standard ΛCDM model, either through combining the complementary constraining power of the void emulator with existing SNIa and BAO results, or through expanding the training sample to include the redshift evolution of the VG-CCF, facilitating a standalone constraint on w_0 and w_a in future work.

5.8.1 Effects of fibre-collision and DR1 mock window on emulator

In Chapter 4, we had emphasized that the training data for emulation of the VG-CCF should mimic all aspects of the observations. Between the BOSS emulator discussed in Chapter 4 and the DESI emulator discussed in Chapter 5, the training sample had to be significantly altered. This meant building a library of galaxy catalogues that matched the number density, survey footprint, and to a lesser extent, the completeness and fibre assignment that was congruent with the DESI DR1 sample. While the DESI DR1 footprint has a radically different completeness compared to the more mature CMASS sample, the need for a robust pipeline that would incorporate the pecularities of this survey was warranted. This came with a significantly higher computing cost - populating the `AbacusSummit` cubic box mocks at a snapshot redshift was only the first step in a much longer process, requiring hundreds of node hours of computing time on NERSC to produce the full set of 8500 mocks.

These changes in the pipeline came with a price in our constraining power. This did not lead to a dramatic reduction in errors as seen in Chapter 4. The small-scale bias in our model persisted and directly influenced the constraints in $\Omega_{\rm m}$, and might have even dulled the VG-CCF's sensitivity to σ_8, in considering the discussion in Chapter 4 around the constraining power of the monopole and quadrupole. While our mean estimate for $\Omega_{\rm m}$ remained consistently within 1σ compared to the BAO analysis of DESI DR1 and full-shape analysis of the DESI DR1 results, this does highlight a pressing objective for future work that utilizes emulators to extract more information from the VG-CCF and other summary statistics: through masking out smaller scales (i.e. training above a certain separation scale, s), where the bias estimate seems to be overly conservative, modifying the binning of the summary statistic to train the model on smaller-scale variations, or to modify the $\mathsf{C}^{\rm emu}$ term such that it is less conservative in its error estimation. Adding more voids, or even more redshift bins to this VG-CCF emulator could drive down errors in σ_8 and w_0, but as Figure 4.15 in Chapter 4 showed, $\Omega_{\rm m}$ largely influences small-scales in the VG-CCF monopole. Furthermore, to obtain constraints that are comparable to those from

Chapter 4, a larger sample of mock catalogues would need to be constructed, to ensure our emulator at minimum could consistently extract $\omega_{\rm cdm}$ and σ_8 from its test cosmologies while handling the selection effects discussed in this chapter.

The tests that were run in Section 5.6 of this chapter illustrate that the healpix down-sampling algorithm did mimic fibre assignment for void-finding and the VG-CCF, and an emulator trained with this sample was able to infer cosmological constraints when fitting to `altMTL` and `FFA` mocks, the latter with more accuracy and slightly less precision. This further underscores the need to appropriately model the emulator's small-scale bias, and that amplitude of this bias has a direct impact on our cosmological parameter inference.

5.8.2 Emulator bias estimation

The road forward for improving this class of neural network emulators lies in tackling the problem of emulator small scale bias. This is not a problem unique to this implementation, as earlier results from Chapter 4, and work external to this thesis (e.g [62, 164]) both pointed to the small-scale bias, and the estimation of the emulator error as confounding factors that cannot be suppressed through using more training data or changes in neural network architecture (c.f. Fig. 4.15). While significant information can be extracted from the VG-CCF at large scales, as discussed in Chapter 4, more can be gained if this bias is appropriately estimated. As mentioned above, this overly conservative treatment of the emulator bias has been shown to affect our emulation of the VG-CCF, and presented challenges in model selection and analyzing the quality of our model's fit. It is expected for this to continue being a limitation in neural network emulators, unless a more thorough understanding of the emulator bias can be gained. We plan to explore the ubiquity of this bias across different neural network architectures and models in a future work.

5.8.3 w_0w_a constraining power

As noted before, the constraining power of this model is wholly dictated by the training sample of the emulator. The `AbacusSummit` simulations encompass a large cosmological parameter space, but the parameter range of w_0 and w_a was limited compared to the DESI BAO parameter space explored. Moreover, the emulator in its current state would not hold enough redshift evolution information to meaningfully constrain w_0 and w_a. We stress that the joint likelihood analysis illustrated two crucial points: it was a proof-of-concept of the constraining power of cosmic voids as a low-cost, high value-added piece of cosmological information that is complementary to existing constraints from large-scale structure (e.g.

BAO) and that there remains a wealth of information yet to be fully utilized with these emulator models beyond this simple proof-of-concept.

This limitation in dark energy constraining power arises from a few different factors. For one, the void-size function is typically more sensitive to changes in w_0 and w_a, and is favoured over the VG-CCF. These constraints in w_0 and w_a are also stymied by the fact that our observable, the VG-CCF was measured across a single redshift bin, $0.4 < z < 0.6$, while the information it was combined with, the BAO and SNIa datasets, span multiple redshift bins. Nonetheless, the additional constraining power obtained from the void emulator's inference of $\omega_{\rm cdm}$ does help narrow the parameter space significantly, and offers a glimpse of how far this analysis could go if additional redshift bins are added to the emulator in the future. These joint constraints could go even further, if the VG-CCF could be added to the existing BAO or full-shape likelihood utilized in the DESI DR1 analysis. An additional caveat comes in the form of the parameter space of the `AbacusSummit` simulations, as training over a larger parameter space could help rectify the model's struggles with w_0 and w_a cosmological parameter recovery. Doing so would allow the void emulator to potentially obtain standalone constraints for the dark energy equation of state parameters, enabling an analysis pipeline similar to the DESI BAO. This in turn could also elucidate how the VG-CCF could be sensitive to changes in w_0 and w_a and possibly even allow us to differentiate between different void sizes or the contribution of the ξ_0 monopole and ξ_2 quadrupole contributions to our models' ability to constrain w_0 and w_a, allowing for greater interpretability of the neural network emulator.

In the void emulator pipeline's current form, this would necessitate producing thousands more mock catalogues, but the improvements in constraining power would promise to be considerable, and could influence the emerging debate around w_0w_aCDM cosmologies, with the VG-CCF emerging as yet another low-cost observable (much like the VSF) to tackle this problem.

5.8.4 Roadmap for future VG-CCF analyses

The sample utilized in this chapter could be a starting point for a far more ambitious emulator in future work. Most of the information that could be obtained about w_0 and w_a comes from the redshift evolution of voids, and thus the VG-CCF at multiple redshift bins would likely contain a wealth of information that we did not explore in this thesis. One question to explore would be to ascertain how the joint likelihood constraints improve once multiple redshift bins are employed in our void analysis, making the void model more analogous to the BAO and SNIa datasets. Expanding this analysis to include the results

of the full-shape analysis would also be ideal, as it could augment the existing constraints. Adding a VG-CCF emulator likelihood, or a likelihood term utilizing the VG-CCF could also be fruitful, as it would allow for an even more thorough analysis of the VG-CCF result than had been previously explored, and allow for a greater test for voids as a complementary constraint than what has been currently seen.

Significant improvements can also be made in the form of the training sample, as the existence of N-body simulations that span a wider parameter space would help explore a wider range of cosmological models, allowing for us to probe the constraining power of the VG-CCF for spatial curvature, or combine it with existing analyses that modelled the masses of neutrinos using the VG-CCF and VSF. Training a new emulator that made use of this information might not necessarily produce the tightest constraints on their own, but we again can turn towards a joint likelihood analysis to augment the existing measurements from DESI and beyond. Additionally, more realizations per cosmology would allow us to drive down cosmic variance, potentially making it easier to improve cosmological constraints in the existing parameter space. The development of N-body mocks that had the limitations of emulators like this one in mind (i.e. trading off parameter space for number of realizations), and the streamlining of pipelines to make the creation of cutsky mocks less expensive could allow emulators to be pushed even further. We are not alone in prescribing a larger space for w_0 and w_a, as more emulators are being trained with constraining these parameters in mind (e.g. [45]).

Another avenue to explore would be neural network architectures. This thesis primarily employed FCNs, a family of neural networks that has been extensively utilized in prior emulator literature, and has been an instrumental component in their adoption. However, if one is to emulate observables, whose bins are highly correlated with each other, it might make more sense to explore neural networks that preserve that information and capture that contextual information more effectively. One obvious candidate would be a convolutional neural network (CNN), which would convolve neighbouring bins in the VG-CCF with each other, allowing the emulator to learn the cosmological parameter sensitivity of the VG-CCF more effectively. These convolutional layers group together local information into successively larger and larger features, and have been successful in relational tasks such as image recognition. It would be relatively simple to re-work the existing FCN used in this emulator as a CNN, and explore it akin to an image recognition problem: given some VG-CCF, can a trained CNN interpolate the VG-CCF to match some region of the cosmological parameter space, placing extra weight on small scale features in the monopole of the VG-CCF, and larger scale features in the monopole and quadrupole. Another potential neural network architecture future work could target would be under the family of Graph Neural Networks, which specialize in retaining the relational information of data

set. These models have been successful in physics, and could be more competitive than the model we used here (e.g. [243, 211, 230]). This would require a more expansive change to the underlying `sunbird` and `acm` libraries, but it could be a valuable test. Training a graph neural network, convolutional neural network and fully-connected neural network using the same training set (e.g. `AbacusSummit` cubic box or cutsky mocks) could also help probe the nature of the small-scale bias in the emulator covariance, or at least shed light into the sensitivity of the VG-CCF to $\Omega_{\rm m}$ and σ_8 and help ground the results found here.

In the final chapter, we will outline how this pipeline can be leveraged to the DESI DR2 sample, and what might be expected from this future analysis, and then summarize the arguments presented in this thesis.

Chapter 6

Conclusions and future work

6.0.1 Summary

Cosmology has been undergoing an exciting period, where the standard model that has shaped it for the past half-century, ΛCDM, is now facing a growing set of inconsistent observations. These inconsistencies arise in the form of statistical tensions emerging between early-time and late-time measurements of H_0 and S_8. Successive generations of CMB and SNIa measurements have drastically shrunk the statistical uncertainties in our measurements of the components of the universe, down to the level of sub-percent precision by the end of this decade. This is a remarkable achievement, and this evolutionary picture of the universe, centered on the evolving energy density of matter, radiation and dark energy has allowed for a detailed history of the universe to be constructed from a handful of measurements at early- and late-times. The emergence of large-scale structure cosmology as another frontier to fill in the details of this history has also been remarkable, with the BAO now established as a powerful way to measure the dark energy equation of state. Following this success of the BAO, first with SDSS III, SDSS IV and now with DESI and Euclid, a new tension has emerged in the dark energy equation of state, further stressing the linchpin of the ΛCDM model.

Large-scale structure cosmology has been a frontier where successive generations of galaxy-redshift surveys have seen improvements in their constraining power, in particular through summary statistics that extract information from galaxy clustering. Compressing the wealth of information found in the late-time universe, the cosmic web, into an observable like the galaxy two-point correlation function means that there will be a loss of information. The galaxy two-point correlation function contains a wealth of information

about the universe's matter density and dark energy through the BAO feature. Significant work has been done over the last three decades to explore the information that can be extracted from the cosmic web by targeting summary statistics such as the galaxy bispectrum, weak-lensing and the galaxy four-point correlation function to paint a more comprehensive picture of the universe at very early times.

The cosmic web does not solely consist of galaxies, however, as it is spanned by vast empty regions between galaxies, cosmic voids, and substructures of gas and dust between clusters, filaments. Using the formalism of the two-point correlation function to explore the clustering of voids, galaxies, filaments, i.e. different density regions in the cosmic web, is not a novel technique. It is one that has gained increasing purchase as the tensions in ΛCDM have grown, either to bridge the gap between late-time and early-time measurements, or to further reinforce it. The growing volume of future galaxy-redshift surveys, and this broad swathe of summary statistics means that there are a wide variety of options in terms of extracting more information from a given galaxy-redshift survey (e.g. [178, 163, 217, 145, 205, 214, 233, 1]).

Cosmic voids have been particularly exciting in this context, as they are relatively inexpensive to obtain samples of, and the information gained from void-galaxy summary statistics is both nontrivial and complementary to the existing galaxy two-point correlation function [155]. As a prerequisite to our work, this thesis discussed the emergence of the void-galaxy cross-correlation function (VG-CCF) as a complementary statistic to the BOSS and eBOSS BAO measurements, and how appropriately modeling this void-galaxy cross-correlation function could impart significant improvements in the constraints of $\Omega_{\rm m}$. In this thesis we described the use of voids to solve these aforementioned problems and, in particular, applying them to measure cosmological parameters. We focused on the void-galaxy cross-correlation function as our summary statistic for this thesis, utilizing voids obtained using the `voxel` void-finder, but note that other observables (e.g. [21, 145, 42, 36, 29, 136]), namely the void-size function have also been successful in obtaining cosmological constraints [27, 57, 47, 55, 176, 114, 37, 76].

In this thesis, the nuances of void-finding have been explored, as cosmic voids can be found in a variety of ways, and each of those definitions imposes their own set of downstream effects on the resultant population. The `voxel` void-finder was applied and the void-galaxy cross-correlation function (VG-CCF) was selected as our summary statistic. The focus was then placed on extracting cosmological information through the AP effect [15], and its liability to becoming biased from a variety of selection and observational effects. In particular, the development of the VG-CCF was traced as a proposed means of measuring the Alcock-Paczynski (AP) effect, to its first implementation [143, 144], and then its joint usage with a redshift-space distortion (RSD) model [113, 111, 180]. These

two effects overlap as the AP distortions arise from differences in distances due to the choice of the cosmological model, modifying the distance to redshift relation. Choosing an inappropriate fiducial cosmology distorts the stack of voids obtained from the VG-CCF, elongating or contracting the standard shape [187]. RSD can also modulate this signal, hence the emphasis on either modelling it or removing it through reconstruction (e.g. [154, 113, 38]). The need for an unbiased sample of voids led to a type of modelling that utilized cosmological N-body simulations to construct *templates* of the VG-CCF under different parameterizations of the RSD model, which were then used to fit the data (e.g. [111, 234, 158]). Analytically modelling the VG-CCF in redshift-space is challenging, and often groups have resorted to using N-body simulations to construct model templates. We argued that this simulation-based modelling was an opportunity to implement a simulation-based emulator, utilizing the N-body simulations to interpolate across different cosmologies.

These modelling requirements and the N-body simulations used for RSD VG-CCF templates meant that a neural network emulator could be trained. Emulators are models that take an existing set of N-body simulations, and interpolate across the parameter space defined by these simulations, learning the dependence of a particular summary statistic to changes within this parameter space. Emulators are often implemented either using a Gaussian Process [188, 23, 148], or neural networks [164, 62, 10, 60]. In this thesis, the motivation for emulators in cosmology was reviewed, as they have been applied to numerous parts of the field (e.g.[119, 10, 200, 217, 236, 139]). In particular, we discussed their application to solving the *inverse problem* [60], wherein an emulator is created to model the parameter dependence of a particular cosmological summary statistic, i.e. learning the dependence of ξ_{vg} on parameters θ [62, 164, 86]. For our model, we used a neural network emulator, as it is less resource intensive than a Gaussian Process and had seen significant improvements in its ability to solve the inverse problem. We discussed the latest developments in neural networks that have led to the adoption of this form of emulator, and discussed their suitability in solving problems laid out by summary statistics in cosmology. We described the necessary components for constructing a neural network emulator of the VG-CCF.

After laying out the the requirements of such a model, a neural network emulator of the VG-CCF was implemented and applied to voxel voids found in the BOSS CMASS sample ($0.45 < z < 0.6$), allowing for the first ever emulation of the full-shape of the void-galaxy cross-correlation function [86]. These findings were discussed in a published paper, which demonstrated the potential of this technique to extract competitive cosmological constraints on $\Omega_{\rm m}$ and σ_8. What made this result so exciting is that prior work with the VG-CCF used RSD templates to then extract the AP parameters, which in turn could then be used to obtain a measurement for $\Omega_{\rm m}$, while our model allowed for the inference

of $\Omega_{\rm m}$ and σ_8 without such steps, directly inferring it from the VG-CCF monopole and quadrupole. This allowed us to explore the parameter dependence of different parts of the VG-CCF. This emulator was constructed with a simplified training sample of cubic galaxy mock catalogues with the [92, 149] cosmologies, with AP and RSD added to the mock catalogues so that our emulator could learn the VG-CCF's response to it. While this was a simplified training set, we had shown that this was adequate for the BOSS sample [86].

For the DESI survey, such a simplification proved insufficient. Differences in the survey window, number density, completeness, and consequently the void selection effects meant that the DR1 VG-CCF was in tension with a cubic box counterpart with a matching number density $\bar{n}_g$. These tensions were at smaller scales, and the emulator had historically struggled to model the VG-CCF at these scales. Consequently, any attempt to model the VG-CCF with this training data for DESI DR1 would have likely been unsuccessful. Training a DESI emulator with the same training sample as BOSS CMASS, matching the number density to $\bar{n}_g$ of the DR1 sample, would have been grossly insufficient. We tried to modify the void catalogue to reflect the differences in the selection effects in a cubic box compared to a DR1 sample, but none of these proved to be fully applicable to the training sample.

This meant that in order to train the DESI DR1 emulator, a new set of training data was required to construct a set of DR1 mocks. As described in Chapter 4 and Chapter 5 of this thesis, in order to construct these DR1 mocks, a pipeline needed to be assembled that would convert the cubic mocks into a DR1 mock, to incorporate all of the selection effects and systematics that would come with the DESI DR1 survey footprint and fibre assignment. This required trimming the cubic boxes into wedges that match the survey geometry. Subsequently, this geometric *cutsky* mock then required the pixel and bitwise masks to be assigned, filtering out galaxies such that they correspond to the LRG tracers. Finally, these galaxies were selected by choosing those that could possibly have a fibre assigned to them, rendering it a *complete* cutsky mock catalogue. The final step in this pipeline entailed the fibre assignment process used in the DESI survey. Existing implementations: `altMTL` and Fast-Fiber Assign (`FFA`) are both too costly to implement for a large volume of mock catalogues to train an emulator, as even the `FFA` algorithm takes several hours to execute on a single mock catalogue [199]. However, the execution time was significantly reduced by using a downsampling algorithm, which binned the galaxies into healpixels and then downsampled them to match an existing set of `altMTL` catalogues. The FKP weights were then computed, and this new catalogue was used as a starting point for our void-finding. This lead to the new pipeline discussed in Chapter 5, which was implemented across a single redshift window, `LRG-1`, $0.4 < z < 0.6$, for the NGC portion of the DR1 sample. In implementing the pipeline described in Chapter 5 to this sample, it was demonstrated a

neural network emulator of the VG-CCF can be feasibly constructed to obtain cosmological measurements from a reduced DESI DR1 sample.

This emulator has demonstrated the feasibility of emulating the VG-CCF with the DESI DR1 galaxy sample, and with a limited subset of it, with some limited constraints being obtained from using only the VG-CCF emulator, and as a proof-of-concept, stronger constraints from a joint analysis utilizing the BAO information in tandem with the VG-CCF emulator. This pipeline has demonstrated that extending this emulator to use the VG-CCF from the full DR1 sample can yield competitive constraints on cosmological parameters, and if appropriately extended to include multiple redshift windows (i.e. `LRG-2` and `LRG-3+ELG-1`) this information could translate into dark energy constraints. This thesis has demonstrated that significant work has to be done in terms of constructing an emulator training sample that incorporates as much of the survey's selection effects and window function as possible (e.g. bitmasks, photometric masks, fibre assignment), and that while these emulators can feasibly produce standalone cosmological constraints, there still remain questions regarding the implementations of the neural network. In particular, the matter of the small-scale bias in these summary statistic emulators needs to be addressed. Future work could explore these questions in far greater detail.

All of this is now possible with the pipeline we have worked to create. Voids, and alternative tracers of the large-scale structure in general are rapidly being adopted as a means of extracting more information from the DESI survey, and this pipeline is not the only one in development. The demand for increased training volumes has led to the commissioning of more mocks with the aim of covering a larger w_0, w_a parameter space, the repurposing of the `LSS` pipeline towards one that is suited for thousands of resource-intensive runs, and an accelerated process in creating survey-approximate galaxy mock catalogues across a wide range of cosmologies. This will make the exploitation of this VG-CCF emulator pipeline increasingly rapid, and allow it to be leveraged across a wide array of redshift slices and survey volumes in the near future, as the most crucial components have been implemented. This technique could serve as an avenue for enhancing constraints on key cosmological parameters obtained from the large-scale structure, and thereby help resolve the statistical tensions that have arisen in our measurement of these parameters.

6.0.2 Roadmap for DESI DR2

Chapter 5 made the case for extending this emulator to larger samples in DR1 and for extending it to DR2. These extensions will ultimately allow us to take the existing joint results, and enhance them further, and even make standalone cosmological constraints on

σ_8, w_0, w_a feasible. As discussed in the last chapter, this pipeline was built through combining a handful of packages (`abacusHOD` [92], `desi-cutsky-mock`, `lss` [199]) to populate a large number of cosmological simulations, starting from 85 dark-matter halo simulations, which were then populated using HODs matching our specifications, with the resultant mock catalogues trimmed into DR1 mocks, matching the survey footprint and later imaging masks. These modules were not designed to be run thousands of times to produce 8500 mock catalogues, and the process of optimizing the steps of this pipeline (Figure 6.1) is ongoing. We estimated that converting a cubic mock to a DR1 mock that has potential fibre assignment information and FKP weights would take approximately 12 hours. The largest bottlenecks in the current pipeline lie in the `desi-cutsky-mock` scripts, which slice the cubic box into a cutsky mock broadly matching the target survey footprint, and the `pota2clus` stage of the `LSS` pipeline, which converts the `forFA` cutsky mocks into `complete` cutsky mocks. Specifically, it involves combing through the potential fibre assignments and trimming the mock to include galaxies that could only be observed with those fibres. Addressing such bottlenecks would drastically affect the estimated CPU node hour requirements (8500 mock catalogues currently require approximately 400 node hours of computing time), enabling substantial increases in the training sample over a shorter period of time. There are ongoing efforts to create a unified alternative clustering statistics pipeline to reach this goal, but it seemed useful to identify it explicitly as a technical bottleneck to larger training datasets.

While this emulator of the DESI DR1 VG-CCF did not encompass as expansive an analysis as the DESI DR1 BAO, it is worth articulating how the aforementioned pipeline can be leveraged to produce an increasingly robust model of the VG-CCF, and how these constraints are poised to augment the existing DESI DR1 and DR2 BAO constraints.

This could begin with the least resource intensive option - filling out the remainder of `LRG-1`'s volume by utilizing the SGC sample of galaxies in addition to the NGC sample that was used in the training and testing of the emulator in Chapter 5. This move only requires altering the downsampling script and downsampling both the SGC and NGC catalogues. While it may be expensive in terms of storing mocks prior to being processed into downsampled DR1 mocks, it is achievable, as it only requires running one of the least expensive stages of the cutsky mock pipeline. Doing so would increase the number of galaxies/tracers from $380,569$ to $508,976$, a 33% increase in tracers and a 32% increase in the volume, bringing the total volume encompassed by the sample up to 2.6Gpc^3. If the sample variance decreases as $\sqrt{N}$, then we can expect the errors to decrease by a factor of 1.2.

Another, marginally more expensive option would involve increasing the number of mocks in the training sample - the existing pipeline used 100 HOD mocks per cosmology

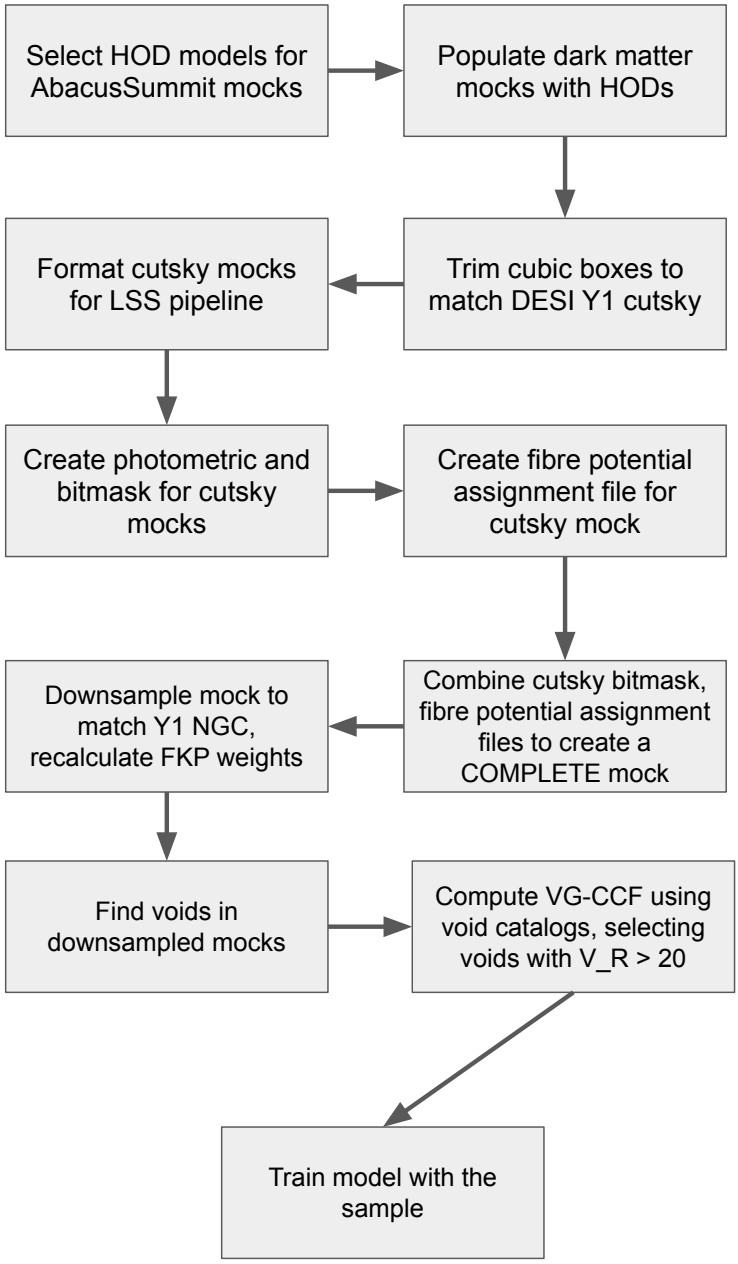

Figure 6.1: A diagram of the void catalogue pipeline, starting from dark matter simulations and culminating in voxel void catalogues.

to encompass a training sample of 8500 mocks. In the first few stages of the pipeline, where cubic mocks are created by selecting HOD parameterizations that fall within a target number density, we recorded 1000 HOD parameterizations per cosmology prior to filtering, and averaged over 200 HOD parameterizations per cosmology that were populated as cubic boxes for the first stages of the pipeline. Only the first 100 parameterizations were utilised for each cosmology. This means that we would only need to re-run the DR1 mock stage of the pipeline to double our training sample. DR1 mock creation is still costly, but the pipeline has been developed to facilitate its creation. Over the course of a few months, the training sample could be doubled, and encompass a 29% reduction in standard errors. The benefits would be that the new model would more densely sample this enlarged HOD parameter space, hopefully allowing the model to obtain tigther cosmological constraints, which could help improve the emulator's precision in cosmological parameter estimation. Moreover, this would also enable a denser sampling of the 6 test cosmologies used by the emulator to construct $\mathsf{C}^{\rm emu}$, allowing for a more thorough diagnosis of its reported conservatism.

One possibility to consider is the estimated improvement in constraining power that would come from scaling up this analysis. This can be done in two different ways: one is to incorporate more than a single redshift bin (i.e. including `LRG-2` and `LRG-3+ELG-1` in addition to `LRG-1`) and the other is in applying this analysis to DESI DR2, exploiting the higher completeness and larger galaxy sample in each of these redshift bins. For the first option, let us consider how much the BAO constraint on $\Omega_{\rm m}$ shrinks as we go from a single redshift bin `LRG-1`, $(0.4 < z < 0.6)$ to including the full redshift range. We found that $\sigma_{\Omega_{\rm m}}$ goes from $\sigma_{\Omega_{\rm m}} = 0.16$ to $\sigma_{\Omega_{\rm m}} = 0.015$ respectively, accounting for a factor of 10 narrowing in the constraint [6]. Furthermore, in DESI DR2, this latter constraint narrowed by another factor of 2, resulting in $\sigma_{\Omega_{\rm m}} = 0.0089$ [3]. Provided that VG-CCF constraints scale in the same way the BAO did as more redshift information is added through multiple bins and the effective volume increases, we estimate a factor of 10 improvement in constraining power if all redshift bins are included, and an additional factor of 2 improvement in constraining power as when we moved from DR1 to DR2. This is a baseline estimate in the potential improvement, and does not account for the additional information that would be gained from the redshift evolution of the VG-CCF being incorporated into the emulator. We anticipate that this method will allow voids to augment the significant cosmological constraints obtained from the BAO measurement in DESI.

6.0.3 Future work with VG-CCF emulators

Given the forthcoming improvements in available spectroscopic galaxy samples from DESI [69, 70] and Euclid [142] the work presented here will be of increasing importance in the future. In particular, the emulator-based approach allows us to use self-consistent profiles for the voids we are fitting, and allows a full inclusion of the AP effect. We have shown that both only have a small effect on the data analysed here, as demonstrated by the match between our results and those of [234].

We have also shown that these emulators are particularly sensitive to void-profiles, as the amplitude of the small scale bias discussed in Chapter 4 had a stronger effect on our model in Chapter 5. This is likely owing to the confluence of the error estimates in the DESI DR1 VG-CCF emulator becoming more conservative, and the change in the survey (fibre assignment, high variation in completeness) influencing the emulator's ability to recover the VG-CCF. This thesis has demonstrated that this technique is feasible, but as Chapter 5 showed, there are many subtle factors that need to be considered in the implementation of this model.

The independence of the VG-CCF with respect to the BAO, shown in [235] that allowed for our joint-likelihood analysis with the DESI BAO merits further exploration beyond its proof-of-concept, as even when voids may not produce standalone cosmological constraints, their joint constraining power with other probes of large-scale structure should be explored. The potential of a full-shape model of the VG-CCF (enabled by this emulator) used in concert with the BAO full-shape model merits exploration; in particular to explore how sensitive the two models may be to different sets of cosmological parameters.

Optimizations in the void-finder are poised to have significant downstream effects on improving the accuracy of the emulator, as the largest bottleneck in simulation-based emulator models stems from the preparation of galaxy catalogues used as the training data. These improvements could allow for a denser sampling of the cosmological and HOD parameter space, allowing us to explore some of the questions raised in Chapter 4. In particular, these questions concerning the sensitivity of the VG-CCF to $\Omega_{\rm m}$ and σ_8. A more thorough exploration of more complex galaxy-halo connection models and the potential constraining power offered by summary statistics exploring low density regions — as identified in [217, 163] is necessary.

Such optimizations might also be of assistance in quantifying the nature of the emulator small-scale bias detected at scales below $20h^{-1}$ Mpc in Chapter 4 and Chapter 5. The ubiquity of this feature should also be explored, as we have shown that it is factor that can readily affect the constraining power of the model. Whether this is a feature of models

constructed with dramatic differences in their architecture, as discussed in Chapter 5 or inherent to the definition of $\mathsf{C}^{\mathrm{emu}}$ is worth exploring.

The VG-CCF is determined by stacking voids of different sizes. Consequently, the differences between voids of different sizes would be smoothed out, potentially obscuring the information obtained from the void-galaxy cross-correlation function. Exploring this in tandem with the void-galaxy paircount reweighting raised in Chapter 5 might prove illuminating, allowing for a detailed comparison in the selection of void populations in terms of the galaxy catalogue utilised.

Another limitation of our current emulator is that it was trained on simulations at a single snapshot. This restricts our analysis to a relatively narrow redshift range, where the evolution of the void clustering statistics can be approximated as a constant. Future work should incorporate redshift evolution into this emulator model using the `AbacusSummit` lightcone simulations [107], as it will enable w_0 and w_a constraints to be potentially obtained, as discussed above.

Taking a broader view, this thesis demonstrated that cosmological constraints could be extracted from the VG-CCF using a neural network emulator. The neural network employed a relatively straightforward architecture. Future work should explore a myriad of neural network configurations, notably convolutional neural networks and graph neural networks, as they could allow even stronger constraints to be obtained if the problem is appropriately posed [211, 211, 230]. This thesis illustrated that the frontier of neural networks in large-scale structure cosmology is a rapidly evolving one, with countless possibilities to explore along this axis.

Void-galaxy clustering measured through the VG-CCF is a tool that allows for cosmological information to be leveraged from existing galaxy samples. Modelling the VG-CCF with a neural network emulator is a technique that allowed us to overcome the problems of statistical precision that have challenged other models of the VG-CCF. This thesis showed that such a technique can obtain precise cosmological parameter constraints, and the matters discussed in it lay out a blueprint for executing this analysis with precision. Implementing this with a wide range of galaxy samples obtained from DESI will serve as an important test in the model's ability to keep pace with the stringent requirements for precision in modern galaxy redshift surveys. The cosmological information obtained from these samples will be valuable as they can augment existing cosmological constraints even further. Extending this model will potentially allow for it to contribute to resolving the tensions emerging in S_8 and the dark energy equation of state.

Cosmic voids have been an invaluable tool, and in large-scale structure cosmology they have been invaluable in squeezing out more information out of existing large-scale struc-

ture datasets. Neural network emulators open up a new frontier to extend our modelling of summary statistics such as the VG-CCF, allowing for us to interpolate over a wide parameter space at a fraction of the cost of running the corresponding N-body simulations, to build models from a sparse dataset. While this pipeline and technique are in need of additional refinement, we have shown that the VG-CCF and cosmic voids still hold a wealth of information: an undiscovered country.

References

[1] Beyond 2pt Collaboration, Elisabeth Krause, Yosuke Kobayashi, Andrés N. Salcedo, Mikhail M. Ivanov, Tom Abel, Kazuyuki Akitsu, Raul E. Angulo, Giovanni Cabass, Sofia Contarini, Carolina Cuesta-Lazaro, ChangHoon Hahn, Nico Hamaus, Donghui Jeong, Chirag Modi, Nhat-Minh Nguyen, Takahiro Nishimichi, Enrique Paillas, Marcos Pellejero Ibañez, Oliver H. E. Philcox, Alice Pisani, Fabian Schmidt, Satoshi Tanaka, Giovanni Verza, Sihan Yuan, and Matteo Zennaro. A parameter-masked mock data challenge for beyond-two-point galaxy clustering statistics. 2024.

[2] T. M. C. Abbott et al. Dark Energy Survey: implications for cosmological expansion models from the final DES Baryon Acoustic Oscillation and Supernova data. 3 2025.

[3] M. Abdul Karim, J. Aguilar, S. Ahlen, S. Alam, L. Allen, C. Allende Prieto, O. Alves, A. Anand, U. Andrade, E. Armengaud, A. Aviles, S. Bailey, C. Baltay, P. Bansal, A. Bault, J. Behera, S. BenZvi, D. Bianchi, C. Blake, S. Brieden, A. Brodzeller, D. Brooks, E. Buckley-Geer, E. Burtin, R. Calderon, R. Canning, A. Carnero Rosell, P. Carrilho, L. Casas, F.J. Castander, M. Charles, E. Chaussidon, J. Chaves-Montero, D. Chebat, X. Chen, T. Claybaugh, S. Cole, A.P. Cooper, A. Cuceu, K.S. Dawson, A. de la Macorra, A. de Mattia, N. Deiosso, J. Della Costa, R. Demina, A. Dey, B. Dey, Z. Ding, P. Doel, J. Edelstein, D.J. Eisenstein, W. Elbers, P. Fagrelius, K. Fanning, E. Fernández-García, S. Ferraro, A. Font-Ribera, J.E. Forero-Romero, C.S. Frenk, C. Garcia-Quintero, L.H. Garrison, E. Gaztañaga, H. Gil-Marín, S. Gontcho A. Gontcho, D. Gonzalez, A.X. Gonzalez-Morales, C. Gordon, D. Green, G. Gutierrez, J. Guy, B. Hadzhiyska, C. Hahn, S. He, M. Herbold, H.K. Herrera-Alcantar, M.-F. Ho, K. Honscheid, C. Howlett, D. Huterer, M. Ishak, S. Juneau, N.V. Kamble, N.G. Karaçayl, R. Kehoe, S. Kent, A.G. Kim, D. Kirkby, T. Kisner, S.E. Koposov, A. Kremin, A. Krolewski, O. Lahav, C. Lamman, M. Landriau, D. Lang, J. Lasker, J.M. Le Goff, L. Le Guillou, A. Leauthaud, M.E. Levi, Q. Li, T.S. Li, K. Lodha, M. Lokken, F. Lozano-Rodríguez, C. Magneville, M. Manera, P. Martini, W.L. Matthewson, A. Meisner, J. Mena-Fernández, A. Menegas,

T. Mergulhão, R. Miquel, J. Moustakas, A. Muñoz-Gutiérrez, D. Muñoz-Santos, A.D. Myers, S. Nadathur, K. Naidoo, L. Napolitano, J.A. Newman, G. Niz, H.E. Noriega, E. Paillas, N. Palanque-Delabrouille, J. Pan, J.A. Peacock, M.P. Ibanez, W.J. Percival, A. Pérez-Fernández, I. Pérez-Ràfols, M.M. Pieri, C. Poppett, F. Prada, D. Rabinowitz, A. Raichoor, C. Ramírez-Pérez, M. Rashkovetskyi, C. Ravoux, J. Rich, A. Rocher, C. Rockosi, J. Rohlf, J.O. Román-Herrera, A.J. Ross, G. Rossi, R. Ruggeri, V. Ruhlmann-Kleider, L. Samushia, E. Sanchez, N. Sanders, D. Schlegel, M. Schubnell, H. Seo, A. Shafieloo, R. Sharples, J. Silber, F. Sinigaglia, D. Sprayberry, T. Tan, G. Tarlé, P. Taylor, W. Turner, L.A. Ureña-López, R. Vaisakh, F. Valdes, G. Valogiannis, M. Vargas-Magaña, L. Verde, M. Walther, B.A. Weaver, D.H. Weinberg, M. White, M. Wolfson, C. Yèche, J. Yu, E.A. Zaborowski, P. Zarrouk, Z. Zhai, H. Zhang, C. Zhao, G.B. Zhao, R. Zhou, and H. and Zou. Desi dr2 results. ii. measurements of baryon acoustic oscillations and cosmological constraints. *Physical Review D*, 112(8), October 2025.

[4] Bela Abolfathi, D. S. Aguado, Gabriela Aguilar, Carlos Allende Prieto, Andres Almeida, Tonima Tasnim Ananna, Friedrich Anders, Scott F. Anderson, Brett H. Andrews, Borja Anguiano, Alfonso Aragón-Salamanca, Maria Argudo-Fernández, Eric Armengaud, Metin Ata, Eric Aubourg, Vladimir Avila-Reese, Carles Badenes, Stephen Bailey, Christophe Balland, Kathleen A. Barger, Jorge Barrera-Ballesteros, Curtis Bartosz, Fabienne Bastien, Dominic Bates, Falk Baumgarten, Julian Bautista, Rachael Beaton, Timothy C. Beers, Francesco Belfiore, Chad F. Bender, Mariangela Bernardi, Matthew A. Bershady, Florian Beutler, Jonathan C. Bird, Dmitry Bizyaev, Guillermo A. Blanc, Michael R. Blanton, Michael Blomqvist, Adam S. Bolton, Médéric Boquien, Jura Borissova, Jo Bovy, Christian Andres Bradna Diaz, William Nielsen Brandt, Jonathan Brinkmann, Joel R. Brownstein, Kevin Bundy, Adam J. Burgasser, Etienne Burtin, Nicolás G. Busca, Caleb I. Cañas, Mariana Cano-Díaz, Michele Cappellari, Ricardo Carrera, Andrew R. Casey, Bernardo Cervantes Sodi, Yanping Chen, Brian Cherinka, Cristina Chiappini, Peter Doohyun Choi, Drew Chojnowski, Chia-Hsun Chuang, Haeun Chung, Nicolas Clerc, Roger E. Cohen, Julia M. Comerford, Johan Comparat, Janaina Correa do Nascimento, Luiz da Costa, Marie-Claude Cousinou, Kevin Covey, Jeffrey D. Crane, Irene Cruz-Gonzalez, Katia Cunha, Gabriele da Silva Ilha, Guillermo J. Damke, Jeremy Darling, James W. Davidson, Kyle Dawson, Miguel Angel C. de Icaza Lizaola, Axel de la Macorra, Sylvain de la Torre, Nathan De Lee, Victoria de Sainte Agathe, Alice Deconto Machado, Flavia Dell'Agli, Timothée Delubac, Aleksandar M. Diamond-Stanic, John Donor, Juan José Downes, Niv Drory, Hélion du Mas des Bourboux, Christopher J. Duckworth, Tom Dwelly, Jamie Dyer, Garrett Ebelke, Arthur Davis

Eigenbrot, Daniel J. Eisenstein, Yvonne P. Elsworth, Eric Emsellem, Michael Eracleous, Ghazaleh Erfanianfar, Stephanie Escoffier, Xiaohui Fan, Emma Fernández Alvar, J. G. Fernandez-Trincado, Rafael Fernando Cirolini, Diane Feuillet, Alexis Finoguenov, Scott W. Fleming, Andreu Font-Ribera, Gordon Freischlad, Peter Frinchaboy, Hai Fu, Yilen Gómez Maqueo Chew, Lluís Galbany, Ana E. García Pérez, R. Garcia-Dias, D. A. García-Hernández, Luis Alberto Garma Oehmichen, Patrick Gaulme, Joseph Gelfand, Héctor Gil-Marín, Bruce A. Gillespie, Daniel Goddard, Jonay I. González Hernández, Violeta Gonzalez-Perez, Kathleen Grabowski, Paul J. Green, Catherine J. Grier, Alain Gueguen, Hong Guo, Julien Guy, Alex Hagen, Patrick Hall, Paul Harding, Sten Hasselquist, Suzanne Hawley, Christian R. Hayes, Fred Hearty, Saskia Hekker, Jesus Hernandez, Hector Hernandez Toledo, David W. Hogg, Kelly Holley-Bockelmann, Jon A. Holtzman, Jiamin Hou, Bau-Ching Hsieh, Jason A. S. Hunt, Timothy A. Hutchinson, Ho Seong Hwang, Camilo Eduardo Jimenez Angel, Jennifer A. Johnson, Amy Jones, Henrik Jönsson, Eric Jullo, Fahim Sakil Khan, Karen Kinemuchi, David Kirkby, Charles C. Kirkpatrick IV, Francisco-Shu Kitaura, Gillian R. Knapp, Jean-Paul Kneib, Juna A. Kollmeier, Ivan Lacerna, Richard R. Lane, Dustin Lang, David R. Law, Jean-Marc Le Goff, Young-Bae Lee, Hongyu Li, Cheng Li, Jianhui Lian, Yu Liang, Marcos Lima, Lihwai Lin , Dan Long, Sara Lucatello, Britt Lundgren, J. Ted Mackereth, Chelsea L. MacLeod, Suvrath Mahadevan, Marcio Antonio Geimba Maia, Steven Majewski, Arturo Manchado, Claudia Maraston, Vivek Mariappan, Rui Marques-Chaves, Thomas Masseron, Karen L. Masters , Richard M. McDermid, Ian D. McGreer, Matthew Melendez, Sofia Meneses-Goytia, Andrea Merloni, Michael R. Merrifield, Szabolcs Meszaros, Andres Meza, Ivan Minchev, Dante Minniti, Eva-Maria Mueller, Francisco Muller-Sanchez, Demitri Muna, Ricardo R. Muñoz, Adam D. Myers, Preethi Nair, Kirpal Nandra, Melissa Ness, Jeffrey A. Newman, Robert C. Nichol, David L. Nidever, Christian Nitschelm, Pasquier Noterdaeme, Julia O'Connell, Ryan James Oelkers, Audrey Oravetz, Daniel Oravetz, Erik Aquino Ortíz, Yeisson Osorio, Zach Pace, Nelson Padilla, Nathalie Palanque-Delabrouille, Pedro Alonso Palicio, Hsi-An Pan, Kaike Pan, Taniya Parikh, Isabelle Pâris, Changbom Park, Sebastien Peirani, Marcos Pellejero-Ibanez, Samantha Penny, Will J. Percival, Ismael Perez-Fournon, Patrick Petitjean, Matthew M. Pieri, Marc Pinsonneault, Alice Pisani, Francisco Prada, Abhishek Prakash, Anna Bárbara de Andrade Queiroz, M. Jordan Raddick, Anand Raichoor, Sandro Barboza Rembold, Hannah Richstein, Rogemar A. Riffel, Rogério Riffel, Hans-Walter Rix, Annie C. Robin, Sergio Rodríguez Torres, Carlos Román-Zúñiga, Ashley J. Ross, Graziano Rossi, John Ruan, Rossana Ruggeri, Jose Ruiz, Mara Salvato, Ariel G. Sánchez, Sebastián F. Sánchez, Jorge Sanchez Almeida, José R. Sánchez-Gallego, Felipe Antonio Santana Rojas, Basílio Xavier Santiago,

Ricardo P. Schiavon, Jaderson S. Schimoia, Edward Schlafly, David Schlegel, Donald P. Schneider, William J. Schuster, Axel Schwope, Hee-Jong Seo, Aldo Serenelli, Shiyin Shen, Yue Shen, Matthew Shetrone, Michael Shull, Víctor Silva Aguirre, Joshua D. Simon, Mike Skrutskie, Anže Slosar, Rebecca Smethurst, Verne Smith, Jennifer Sobeck, Garrett Somers, Barbara J. Souter, Diogo Souto, Ashley Spindler, David V. Stark, Keivan Stassun, Matthias Steinmetz, Dennis Stello, Thaisa Storchi-Bergmann, Alina Streblyanska, Guy S. Stringfellow, Genaro Suárez, Jing Sun, Laszlo Szigeti, Manuchehr Taghizadeh-Popp, Michael S. Talbot, Baitian Tang, Charling Tao, Jamie Tayar, Mita Tembe, Johanna Teske, Aniruddha R. Thakar, Daniel Thomas, Patricia Tissera, Rita Tojeiro, Christy Tremonti, Nicholas W. Troup, Meg Urry, O. Valenzuela, Remco van den Bosch, Jaime Vargas-González, Mariana Vargas-Magaña, Jose Alberto Vazquez, Sandro Villanova, Nicole Vogt, David Wake, Yuting Wang, Benjamin Alan Weaver, Anne-Marie Weijmans, David H. Weinberg, Kyle B. Westfall, David G. Whelan, Eric Wilcots, Vivienne Wild, Rob A. Williams, John Wilson, W. M. Wood-Vasey, Dominika Wylezalek, Ting Xiao , Renbin Yan, Meng Yang, Jason E. Ybarra, Christophe Yèche, Nadia Zakamska, Olga Zamora, Pauline Zarrouk, Gail Zasowski, Kai Zhang, Cheng Zhao, Gong-Bo Zhao, Zheng Zheng, Zheng Zheng, Zhi-Min Zhou, Guangtun Zhu, Joel C. Zinn, and Hu Zou. The fourteenth data release of the sloan digital sky survey: First spectroscopic data from the extended baryon oscillation spectroscopic survey and from the second phase of the apache point observatory galactic evolution experiment. *The Astrophysical Journal Supplement Series*, 235(2):42, April 2018.

[5] A.G. Adame, J. Aguilar, S. Ahlen, S. Alam, D.M. Alexander, C. Allende Prieto, M. Alvarez, O. Alves, A. Anand, U. Andrade, E. Armengaud, S. Avila, A. Aviles, H. Awan, B. Bahr-Kalus, S. Bailey, C. Baltay, A. Bault, J. Behera, S. BenZvi, F. Beutler, D. Bianchi, C. Blake, R. Blum, M. Bonici, S. Brieden, A. Brodzeller, D. Brooks, E. Buckley-Geer, E. Burtin, R. Calderon, R. Canning, A. Carnero Rosell, R. Cereskaite, J.L. Cervantes-Cota, S. Chabanier, E. Chaussidon, J. Chaves-Montero, D. Chebat, S. Chen, X. Chen, T. Claybaugh, S. Cole, A. Cuceu, T.M. Davis, K. Dawson, A. de la Macorra, A. de Mattia, N. Deiosso, A. Dey, B. Dey, Z. Ding, P. Doel, J. Edelstein, S. Eftekharzadeh, D.J. Eisenstein, W. Elbers, A. Elliott, P. Fagrelius, K. Fanning, S. Ferraro, J. Ereza, N. Findlay, B. Flaugher, A. Font-Ribera, D. Forero-Sánchez, J.E. Forero-Romero, C.S. Frenk, C. Garcia-Quintero, L.H. Garrison, E. Gaztañaga, H. Gil-Marín, S.Gontcho A. Gontcho, A.X. Gonzalez-Morales, V. Gonzalez-Perez, C. Gordon, D. Green, D. Gruen, R. Gsponer, G. Gutierrez, J. Guy, B. Hadzhiyska, C. Hahn, M.M.S. Hanif, H.K. Herrera-Alcantar, K. Honscheid, C. Howlett, D. Huterer, V. Iršič, M. Ishak, R. Joyce, S. Juneau, N.G.

Karaçaylı, R. Kehoe, S. Kent, D. Kirkby, H. Kong, S.E. Koposov, A. Kremin, A. Krolewski, O. Lahav, Y. Lai, T.-W. Lan, M. Landriau, D. Lang, J. Lasker, J.M. Le Goff, L. Le Guillou, A. Leauthaud, M.E. Levi, T.S. Li, K. Lodha, C. Magneville, M. Manera, D. Margala, P. Martini, W. Matthewson, M. Maus, P. McDonald, L. Medina-Varela, A. Meisner, J. Mena-Fernández, R. Miquel, J. Moon, S. Moore, J. Moustakas, N. Mudur, E. Mueller, A. Muñoz-Gutiérrez, A.D. Myers, S. Nadathur, L. Napolitano, R. Neveux, J.A. Newman, N.M. Nguyen, J. Nie, G. Niz, H.E. Noriega, N. Padmanabhan, E. Paillas, N. Palanque-Delabrouille, J. Pan, S. Penmetsa, W.J. Percival, M.M. Pieri, M. Pinon, C. Poppett, A. Porredon, F. Prada, A. Pérez-Fernández, I. Pérez-Ràfols, D. Rabinowitz, A. Raichoor, C. Ramírez-Pérez, S. Ramirez-Solano, M. Rashkovetskyi, C. Ravoux, M. Rezaie, J. Rich, A. Rocher, C. Rockosi, N.A. Roe, A. Rosado-Marin, A.J. Ross, G. Rossi, R. Ruggeri, V. Ruhlmann-Kleider, L. Samushia, E. Sanchez, C. Saulder, E.F. Schlafly, D. Schlegel, M. Schubnell, H. Seo, A. Shafieloo, R. Sharples, J. Silber, A. Slosar, A. Smith, D. Sprayberry, T. Tan, G. Tarlé, P. Taylor, S. Trusov, R. Vaisakh, D. Valcin, F. Valdes, G. Valogiannis, M. Vargas-Magaña, L. Verde, M. Walther, B. Wang, M.S. Wang, B.A. Weaver, N. Weaverdyck, R.H. Wechsler, D.H. Weinberg, M. White, M.J. Wilson, L. Yi, J. Yu, Y. Yu, S. Yuan, C. Yèche, E.A. Zaborowski, P. Zarrouk, H. Zhang, C. Zhao, R. Zhao, R. Zhou, T. Zhuang, and H. and Zou. Desi 2024 vii: cosmological constraints from the full-shape modeling of clustering measurements. *Journal of Cosmology and Astroparticle Physics*, 2025(07):028, 2025.

[6] A.G. Adame, J. Aguilar, S. Ahlen, S. Alam, D.M. Alexander, M. Alvarez, O. Alves, A. Anand, U. Andrade, E. Armengaud, S. Avila, A. Aviles, H. Awan, B. Bahr-Kalus, S. Bailey, C. Baltay, A. Bault, J. Behera, S. BenZvi, A. Bera, F. Beutler, D. Bianchi, C. Blake, R. Blum, S. Brieden, A. Brodzeller, D. Brooks, E. Buckley-Geer, E. Burtin, R. Calderon, R. Canning, A. Carnero Rosell, R. Cereskaite, J.L. Cervantes-Cota, S. Chabanier, E. Chaussidon, J. Chaves-Montero, S. Chen, X. Chen, T. Claybaugh, S. Cole, A. Cuceu, T.M. Davis, K. Dawson, A. de la Macorra, A. de Mattia, N. Deiosso, A. Dey, B. Dey, Z. Ding, P. Doel, J. Edelstein, S. Eftekharzadeh, D.J. Eisenstein, A. Elliott, P. Fagrelius, K. Fanning, S. Ferraro, J. Ereza, N. Findlay, B. Flaugher, A. Font-Ribera, D. Forero-Sánchez, J.E. Forero-Romero, C.S. Frenk, C. Garcia-Quintero, E. Gaztañaga, H. Gil-Marín, S.Gontcho A. Gontcho, A.X. Gonzalez-Morales, V. Gonzalez-Perez, C. Gordon, D. Green, D. Gruen, R. Gsponer, G. Gutierrez, J. Guy, B. Hadzhiyska, C. Hahn, M.M.S. Hanif, H.K. Herrera-Alcantar, K. Honscheid, C. Howlett, D. Huterer, V. Iršič, M. Ishak, S. Juneau, N.G. Karaçaylı, R. Kehoe, S. Kent, D. Kirkby, A. Kremin, A. Krolewski, Y. Lai, T.-W. Lan, M. Landriau, D. Lang, J. Lasker, J.M. Le Goff, L. Le Guillou, A. Leauthaud,

M.E. Levi, T.S. Li, E. Linder, K. Lodha, C. Magneville, M. Manera, D. Margala, P. Martini, M. Maus, P. McDonald, L. Medina-Varela, A. Meisner, J. Mena-Fernández, R. Miquel, J. Moon, S. Moore, J. Moustakas, E. Mueller, A. Muñoz-Gutiérrez, A.D. Myers, S. Nadathur, L. Napolitano, R. Neveux, J.A. Newman, N.M. Nguyen, J. Nie, G. Niz, H.E. Noriega, N. Padmanabhan, E. Paillas, N. Palanque-Delabrouille, J. Pan, S. Penmetsa, W.J. Percival, M.M. Pieri, M. Pinon, C. Poppett, A. Porredon, F. Prada, A. Pérez-Fernández, I. Pérez-Ràfols, D. Rabinowitz, A. Raichoor, C. Ramírez-Pérez, S. Ramirez-Solano, M. Rashkovetskyi, C. Ravoux, M. Rezaie, J. Rich, A. Rocher, C. Rockosi, N.A. Roe, A. Rosado-Marin, A.J. Ross, G. Rossi, R. Ruggeri, V. Ruhlmann-Kleider, L. Samushia, E. Sanchez, C. Saulder, E.F. Schlafly, D. Schlegel, M. Schubnell, H. Seo, A. Shafieloo, R. Sharples, J. Silber, A. Slosar, A. Smith, D. Sprayberry, T. Tan, G. Tarlé, P. Taylor, S. Trusov, L.A. Ureña-López, R. Vaisakh, D. Valcin, F. Valdes, M. Vargas-Magaña, L. Verde, M. Walther, B. Wang, M.S. Wang, B.A. Weaver, N. Weaverdyck, R.H. Wechsler, D.H. Weinberg, M. White, J. Yu, Y. Yu, S. Yuan, C. Yèche, E.A. Zaborowski, P. Zarrouk, H. Zhang, C. Zhao, R. Zhao, R. Zhou, T. Zhuang, and H. Zou. Desi 2024 vi: cosmological constraints from the measurements of baryon acoustic oscillations. *Journal of Cosmology and Astroparticle Physics*, 2025(02):021, February 2025.

[7] A.G. Adame, J. Aguilar, S. Ahlen, S. Alam, D.M. Alexander, M. Alvarez, O. Alves, A. Anand, U. Andrade, E. Armengaud, S. Avila, A. Aviles, H. Awan, S. Bailey, C. Baltay, A. Bault, J. Behera, S. BenZvi, F. Beutler, D. Bianchi, C. Blake, R. Blum, S. Brieden, A. Brodzeller, D. Brooks, Z. Brown, E. Buckley-Geer, E. Burtin, R. Calderon, R. Canning, A. Carnero Rosell, R. Cereskaite, J.L. Cervantes-Cota, S. Chabanier, E. Chaussidon, J. Chaves-Montero, S. Chen, X. Chen, T. Claybaugh, S. Cole, A. Cuceu, T.M. Davis, K. Dawson, A. de la Macorra, A. de Mattia, N. Deiosso, R. Demina, A. Dey, B. Dey, Z. Ding, P. Doel, J. Edelstein, S. Eftekharzadeh, D.J. Eisenstein, A. Elliott, P. Fagrelius, K. Fanning, S. Ferraro, J. Ereza, N. Findlay, B. Flaugher, A. Font-Ribera, D. Forero-Sánchez, J.E. Forero-Romero, C.S. Frenk, C. Garcia-Quintero, E. Gaztañaga, H. Gil-Marín, S.Gontcho A. Gontcho, A.X. Gonzalez-Morales, V. Gonzalez-Perez, C. Gordon, D. Green, D. Gruen, R. Gsponer, G. Gutierrez, J. Guy, B. Hadzhiyska, C. Hahn, M.M.S. Hanif, H.K. Herrera-Alcantar, K. Honscheid, J. Hou, C. Howlett, D. Huterer, V. Iršič, M. Ishak, S. Juneau, N.G. Karaçaylı, R. Kehoe, S. Kent, D. Kirkby, F.-S. Kitaura, H. Kong, A. Kremin, A. Krolewski, Y. Lai, T.-W. Lan, M. Landriau, D. Lang, J. Lasker, J.M. Le Goff, L. Le Guillou, A. Leauthaud, M.E. Levi, T.S. Li, K. Lodha, C. Magneville, M. Manera, D. Margala, P. Martini, M. Maus, P. McDonald, L. Medina-Varela, A. Meisner, J. Mena-Fernández, R. Miquel, J. Moon,

S. Moore, J. Moustakas, N. Mudur, E. Mueller, A. Muñoz-Gutiérrez, A.D. Myers, S. Nadathur, L. Napolitano, R. Neveux, J.A. Newman, N.M. Nguyen, J. Nie, G. Niz, H.E. Noriega, N. Padmanabhan, E. Paillas, N. Palanque-Delabrouille, J. Pan, S. Penmetsa, W.J. Percival, M.M. Pieri, M. Pinon, C. Poppett, A. Porredon, F. Prada, A. Pérez-Fernández, I. Pérez-Ràfols, D. Rabinowitz, A. Raichoor, C. Ramírez-Pérez, S. Ramirez-Solano, M. Rashkovetskyi, C. Ravoux, M. Rezaie, J. Rich, A. Rocher, C. Rockosi, N.A. Roe, A. Rosado-Marin, A.J. Ross, G. Rossi, R. Ruggeri, V. Ruhlmann-Kleider, L. Samushia, E. Sanchez, C. Saulder, E.F. Schlafly, D. Schlegel, D. Scholte, M. Schubnell, H. Seo, R. Sharples, J. Silber, A. Slosar, A. Smith, D. Sprayberry, T. Tan, G. Tarlé, S. Trusov, R. Vaisakh, D. Valcin, F. Valdes, M. Vargas-Magaña, L. Verde, M. Walther, B. Wang, M.S. Wang, B.A. Weaver, N. Weaverdyck, R.H. Wechsler, D.H. Weinberg, M. White, M.J. Wilson, J. Yu, Y. Yu, S. Yuan, C. Yèche, E.A. Zaborowski, P. Zarrouk, H. Zhang, C. Zhao, R. Zhao, R. Zhou, and H. and Zou. Desi 2024 ii: sample definitions, characteristics, and two-point clustering statistics. *Journal of Cosmology and Astroparticle Physics*, 2025(07):017, 2025.

[8] A.G. Adame, J. Aguilar, S. Ahlen, S. Alam, D.M. Alexander, M. Alvarez, O. Alves, A. Anand, U. Andrade, E. Armengaud, S. Avila, A. Aviles, H. Awan, S. Bailey, C. Baltay, A. Bault, J. Behera, S. BenZvi, F. Beutler, D. Bianchi, C. Blake, R. Blum, S. Brieden, A. Brodzeller, D. Brooks, E. Buckley-Geer, E. Burtin, R. Calderon, R. Canning, A. Carnero Rosell, R. Cereskaite, J.L. Cervantes-Cota, S. Chabanier, E. Chaussidon, J. Chaves-Montero, S. Chen, X. Chen, T. Claybaugh, S. Cole, A. Cuceu, T.M. Davis, K. Dawson, A. de la Macorra, A. de Mattia, N. Deiosso, A. Dey, B. Dey, Z. Ding, P. Doel, J. Edelstein, S. Eftekharzadeh, D.J. Eisenstein, A. Elliott, P. Fagrelius, K. Fanning, S. Ferraro, J. Ereza, N. Findlay, B. Flaugher, A. Font-Ribera, D. Forero-Sánchez, J.E. Forero-Romero, C. Garcia-Quintero, L.H. Garrison, E. Gaztañaga, H. Gil-Marín, S.Gontcho A. Gontcho, A.X. Gonzalez-Morales, V. Gonzalez-Perez, C. Gordon, D. Green, D. Gruen, R. Gsponer, G. Gutierrez, J. Guy, B. Hadzhiyska, C. Hahn, M.M.S. Hanif, H.K. Herrera-Alcantar, K. Honscheid, C. Howlett, D. Huterer, V. Iršič, M. Ishak, S. Juneau, N.G. Karaçaylı, R. Kehoe, S. Kent, D. Kirkby, H. Kong, S.E. Koposov, A. Kremin, A. Krolewski, Y. Lai, T.-W. Lan, M. Landriau, D. Lang, J. Lasker, J.M. Le Goff, L. Le Guillou, A. Leauthaud, M.E. Levi, T.S. Li, K. Lodha, C. Magneville, M. Manera, D. Margala, P. Martini, M. Maus, P. McDonald, L. Medina-Varela, A. Meisner, J. Mena-Fernández, R. Miquel, J. Moon, S. Moore, J. Moustakas, E. Mueller, A. Muñoz-Gutiérrez, A.D. Myers, S. Nadathur, L. Napolitano, R. Neveux, J.A. Newman, N.M. Nguyen, J. Nie, G. Niz, H.E. Noriega, N. Padmanabhan, E. Paillas, N. Palanque-Delabrouille,

J. Pan, S. Penmetsa, W.J. Percival, M.M. Pieri, M. Pinon, C. Poppett, A. Porredon, F. Prada, A. Pérez-Fernández, I. Pérez-Ràfols, D. Rabinowitz, A. Raichoor, C. Ramírez-Pérez, S. Ramirez-Solano, M. Rashkovetskyi, C. Ravoux, M. Rezaie, J. Rich, A. Rocher, C. Rockosi, F. Rodríguez-Martínez, N.A. Roe, A. Rosado-Marin, A.J. Ross, G. Rossi, R. Ruggeri, V. Ruhlmann-Kleider, L. Samushia, E. Sanchez, C. Saulder, E.F. Schlafly, D. Schlegel, M. Schubnell, H. Seo, R. Sharples, J. Silber, A. Slosar, A. Smith, D. Sprayberry, T. Tan, G. Tarlé, S. Trusov, R. Vaisakh, D. Valcin, F. Valdes, M. Vargas-Magaña, L. Verde, M. Walther, B. Wang, M.S. Wang, B.A. Weaver, N. Weaverdyck, R.H. Wechsler, D.H. Weinberg, M. White, M.J. Wilson, J. Yu, Y. Yu, S. Yuan, C. Yèche, E.A. Zaborowski, P. Zarrouk, H. Zhang, C. Zhao, R. Zhao, R. Zhou, and H. and Zou. Desi 2024 v: Full-shape galaxy clustering from galaxies and quasars. *Journal of Cosmology and Astroparticle Physics*, 2025(09):008, 2025.

[9] Nishant Agarwal, Vincent Desjacques, Donghui Jeong, and Fabian Schmidt. Information content in the redshift-space galaxy power spectrum and bispectrum. *Journal of Cosmology and Astroparticle Physics*, 2021(03):021, March 2021.

[10] Shankar Agarwal, Filipe B. Abdalla, Hume A. Feldman, Ofer Lahav, and Shaun A. Thomas. PkANN - I. Non-linear matter power spectrum interpolation through artificial neural networks. *Monthly Notices of the Royal Astronomical Society*, 424(2):1409–1418, August 2012.

[11] N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed, J.-P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J.-F. Cardoso, J. Carron, A. Challinor, H. C. Chiang, J. Chluba, L. P. L. Colombo, C. Combet, D. Contreras, B. P. Crill, F. Cuttaia, P. de Bernardis, G. de Zotti, J. Delabrouille, J.-M. Delouis, E. Di Valentino, J. M. Diego, O. Doré, M. Douspis, A. Ducout, X. Dupac, S. Dusini, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, Y. Fantaye, M. Farhang, J. Fergusson, R. Fernandez-Cobos, F. Finelli, F. Forastieri, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, M. Gerbino, T. Ghosh, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gruppuso, J. E. Gudmundsson, J. Hamann, W. Handley, F. K. Hansen, D. Herranz, S. R. Hildebrandt, E. Hivon, Z. Huang, A. H. Jaffe, W. C. Jones, A. Karakci, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Kim, T. S. Kisner, L. Knox, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, G. Lagache, J.-M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, P. Lemos,

J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Lilley, V. Lindholm, M. López-Caniego, P. M. Lubin, Y.-Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M.-A. Miville-Deschênes, D. Molinari, L. Montier, G. Morgante, A. Moss, P. Natoli, H. U. Nørgaard-Nielsen, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, L. Polastri, G. Polenta, J.-L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, E. P. S. Shellard, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A.-S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Toffolatti, M. Tomasi, T. Trombetti, L. Valenziano, J. Valiviita, B. Van Tent, L. Vibert, P. Vielva, F. Villa, N. Vittorio, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Zacchei, and A. Zonca. Planck2018 results: Vi. cosmological parameters. *Astronomy amp; Astrophysics*, 641:A6, September 2020.

[12] Christopher P. Ahn, Rachael Alexandroff, Carlos Allende Prieto, Scott F. Anderson, Timothy Anderton, Brett H. Andrews, Éric Aubourg, Stephen Bailey, Eduardo Balbinot, Rory Barnes, Julian Bautista, Timothy C. Beers, Alessandra Beifiori, Andreas A. Berlind, Vaishali Bhardwaj, Dmitry Bizyaev, Cullen H. Blake, Michael R. Blanton, Michael Blomqvist, John J. Bochanski, Adam S. Bolton, Arnaud Borde, Jo Bovy, W. N. Brandt, J. Brinkmann, Peter J. Brown, Joel R. Brownstein, Kevin Bundy, N. G. Busca, William Carithers, Aurelio R. Carnero, Michael A. Carr, Dana I. Casetti-Dinescu, Yanmei Chen, Cristina Chiappini, Johan Comparat, Natalia Connolly, Justin R. Crepp, Stefano Cristiani, Rupert A. C. Croft, Antonio J. Cuesta, Luiz N. da Costa, James R. A. Davenport, Kyle S. Dawson, Roland de Putter, Nathan De Lee, Timothée Delubac, Saurav Dhital, Anne Ealet, Garrett L. Ebelke, Edward M. Edmondson, Daniel J. Eisenstein, S. Escoffier, Massimiliano Esposito, Michael L. Evans, Xiaohui Fan, Bruno Femenía Castellá, Emma Fernández Alvar, Leticia D. Ferreira, N. Filiz Ak, Hayley Finley, Scott W. Fleming, Andreu Font-Ribera, Peter M. Frinchaboy, D. A. García-Hernández, A. E. García Pérez, Jian Ge, R. Génova-Santos, Bruce A. Gillespie, Léo Girardi, Jonay I. González Hernández, Eva K. Grebel, James E. Gunn, Hong Guo, Daryl Haggard, Jean-Christophe Hamilton, David W. Harris, Suzanne L. Hawley, Frederick R. Hearty, Shirley Ho, David W. Hogg, Jon A. Holtzman, Klaus Honscheid, J. Huehnerhoff, Inese I. Ivans, Željko Ivezić, Heather R. Jacobson, Linhua Jiang, Jonas Johansson, Jennifer A. Johnson, Guinevere Kauffmann, David Kirkby, Jessica A. Kirkpatrick, Mark A. Klaene, Gillian R. Knapp,

Jean-Paul Kneib, Jean-Marc Le Goff, Alexie Leauthaud, Khee-Gan Lee, Young Sun Lee, Daniel C. Long, Craig P. Loomis, Sara Lucatello, Britt Lundgren, Robert H. Lupton, Bo Ma, Zhibo Ma, Nicholas MacDonald, Claude E. Mack, Suvrath Mahadevan, Marcio A. G. Maia, Steven R. Majewski, Martin Makler, Elena Malanushenko, Viktor Malanushenko, A. Manchado, Rachel Mandelbaum, Marc Manera, Claudia Maraston, Daniel Margala, Sarah L. Martell, Cameron K. McBride, Ian D. McGreer, Richard G. McMahon, Brice Ménard, Sz. Meszaros, Jordi Miralda-Escudé, Antonio D. Montero-Dorta, Francesco Montesano, Heather L. Morrison, Demitri Muna, Jeffrey A. Munn, Hitoshi Murayama, Adam D. Myers, A. F. Neto, Duy Cuong Nguyen, Robert C. Nichol, David L. Nidever, Pasquier Noterdaeme, Sebastián E. Nuza, Ricardo L. C. Ogando, Matthew D. Olmstead, Daniel J. Oravetz, Russell Owen, Nikhil Padmanabhan, Nathalie Palanque-Delabrouille, Kaike Pan, John K. Parejko, Prachi Parihar, Isabelle Pâris, Petchara Pattarakijwanich, Joshua Pepper, Will J. Percival, Ismael Pérez-Fournon, Ignasi Pérez-Ràfols, Patrick Petitjean, Janine Pforr, Matthew M. Pieri, Marc H. Pinsonneault, G. F. Porto de Mello, Francisco Prada, Adrian M. Price-Whelan, M. Jordan Raddick, Rafael Rebolo, James Rich, Gordon T. Richards, Annie C. Robin, Helio J. Rocha-Pinto, Constance M. Rockosi, Natalie A. Roe, Ashley J. Ross, Nicholas P. Ross, Graziano Rossi, J. A. Rubiño-Martin, Lado Samushia, J. Sanchez Almeida, Ariel G. Sánchez, Basílio Santiago, Conor Sayres, David J. Schlegel, Katharine J. Schlesinger, Sarah J. Schmidt, Donald P. Schneider, Mathias Schultheis, Axel D. Schwope, C. G. Scóccola, Uros Seljak, Erin Sheldon, Yue Shen, Yiping Shu, Jennifer Simmerer, Audrey E. Simmons, Ramin A. Skibba, M. F. Skrutskie, A. Slosar, Flavia Sobreira, Jennifer S. Sobeck, Keivan G. Stassun, Oliver Steele, and Matthias Steinmetz. The Ninth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Baryon Oscillation Spectroscopic Survey. *The Astrophysical Journal Supplement Series*, 203(2):21, December 2012.

[13] Joël Akeret, Alexandre Refregier, Adam Amara, Sebastian Seehars, and Caspar Hasner. Approximate Bayesian computation for forward modeling in cosmology. *Journal of Cosmology and Astroparticle Physics*, 2015(8):043–043, August 2015.

[14] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework, 2019.

[15] C. Alcock and B. Paczynski. An evolution free test for non-zero cosmological constant. *Nature*, 281:358, October 1979.

[16] R. Alonso, J. Ojeda, N. Hitschfeld, C. Hervías, and L. E. Campusano. Delaunay based algorithm for finding polygonal voids in planar point sets. *Astronomy and Computing*, 22:48–62, January 2018.

[17] R. A. Alpher, H. Bethe, and G. Gamow. The Origin of Chemical Elements. *Physical Review*, 73(7):803–804, April 1948.

[18] Luis A. Anchordoqui, Eleonora Di Valentino, Supriya Pan, and Weiqiang Yang. Dissecting the h0 and s8 tensions with planck + bao + supernova type ia in multi-parameter cosmologies. *Journal of High Energy Astrophysics*, 32:28–64, November 2021.

[19] Y. and Akrami, F. Arroja, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, K. Benabed, J.-P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J.-F. Cardoso, J. Carron, A. Challinor, H. C. Chiang, L. P. L. Colombo, C. Combet, D. Contreras, B. P. Crill, F. Cuttaia, P. de Bernardis, G. de Zotti, J. Delabrouille, J.-M. Delouis, E. Di Valentino, J. M. Diego, S. Donzelli, O. Doré, M. Douspis, A. Ducout, X. Dupac, S. Dusini, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, Y. Fantaye, J. Fergusson, R. Fernandez-Cobos, F. Finelli, F. Forastieri, M. Frailis, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, C. Gauthier, R. T. Génova-Santos, M. Gerbino, T. Ghosh, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gruppuso, J. E. Gudmundsson, J. Hamann, W. Handley, F. K. Hansen, D. Herranz, E. Hivon, D. C. Hooper, Z. Huang, A. H. Jaffe, W. C. Jones, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Kim, T. S. Kisner, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, G. Lagache, J.-M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, V. Lindholm, M. López-Caniego, P. M. Lubin, Y.-Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, M. Maris, P. G. Martin, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. D. Meerburg, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, S. Mitra, M.-A. Miville-Deschênes, D. Molinari, A. Moneti, L. Montier, G. Morgante, A. Moss, M. Münchmeyer, P. Natoli, H. U. Nørgaard-Nielsen, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, L. Polastri, G. Polenta, J.-L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, E. P. S. Shellard, M. Shiraishi, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A.-S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Toffolatti,

M. Tomasi, T. Trombetti, J. Valiviita, B. Van Tent, P. Vielva, F. Villa, N. Vittorio, B. D. Wandelt, I. K. Wehus, S. D. M. White, A. Zacchei, J. P. Zibin, and A. Zonca. ¡i¿planck¡/i¿ 2018 results: X. constraints on inflation. *Astronomy amp; Astrophysics*, 641:A10, 2020.

[20] F. J. and Castander, P. Fosalba, J. Stadel, D. Potter, J. Carretero, P. Talladа-Crespí, L. Pozzetti, M. Bolzonella, G. A. Mamon, L. Blot, K. Hoffmann, M. Huertas-Company, P. Monaco, E. J. Gonzalez, G. De Lucia, C. Scarlata, M.-A. Breton, L. Linke, C. Viglione, S.-S. Li, Z. Zhai, Z. Baghkhani, K. Pardede, C. Neissner, R. Teyssier, M. Crocce, I. Tutusaus, L. Miller, G. Congedo, A. Biviano, M. Hirschmann, A. Pezzotta, H. Aussel, H. Hoekstra, T. Kitching, W. J. Percival, L. Guzzo, Y. Mellier, P. A. Oesch, R. A. A. Bowler, S. Bruton, V. Allevato, V. Gonzalez-Perez, M. Manera, S. Avila, A. Kovács, N. Aghanim, B. Altieri, A. Amara, L. Amendola, S. Andreon, N. Auricchio, C. Baccigalupi, M. Baldi, A. Balestra, S. Bardelli, R. Bender, F. Bernardeau, C. Bodendorf, D. Bonino, E. Branchini, M. Brescia, J. Brinchmann, S. Camera, V. Capobianco, C. Carbone, S. Casas, M. Castellano, G. Castignani, S. Cavuoti, A. Cimatti, C. Colodro-Conde, C. J. Conselice, L. Conversi, Y. Copin, L. Corcione, F. Courbin, H. M. Courtois, A. Da Silva, H. Degaudenzi, A. M. Di Giorgio, J. Dinis, M. Douspis, F. Dubath, C. A. J. Duncan, X. Dupac, S. Dusini, A. Ealet, M. Farina, S. Farrens, S. Ferriol, S. Fotopoulou, N. Fourmanoit, M. Frailis, E. Franceschi, P. Franzetti, S. Galeotta, W. Gillard, B. Gillis, C. Giocoli, P. Gómez-Alvarez, B. R. Granett, A. Grazian, F. Grupp, S. V. H. Haugan, M. S. Holliman, W. Holmes, I. Hook, F. Hormuth, A. Hornstrup, P. Hudelot, S. Ilić, K. Jahnke, M. Jhabvala, B. Joachimi, E. Keihänen, S. Kermiche, A. Kiessling, M. Kilbinger, R. Kohley, B. Kubik, M. Kümmel, M. Kunz, H. Kurki-Suonio, O. Lahav, R. Laureijs, D. Le Mignant, P. Liebing, S. Ligori, P. B. Lilje, V. Lindholm, I. Lloro, D. Maino, E. Maiorano, O. Mansutti, S. Marcin, O. Marggraf, K. Markovic, M. Martinelli, N. Martinet, F. Marulli, R. Massey, D. C. Masters, S. Maurogordato, H. J. McCracken, E. Medinaceli, S. Mei, M. Melchior, M. Meneghetti, E. Merlin, G. Meylan, J. J. Mohr, M. Moresco, L. Moscardini, E. Munari, R. Nakajima, R. C. Nichol, S.-M. Niemi, C. Padilla, K. Paech, S. Paltani, F. Pasian, J. A. Peacock, K. Pedersen, V. Pettorino, S. Pires, G. Polenta, M. Poncet, L. A. Popa, F. Raison, R. Rebolo, A. Renzi, J. Rhodes, G. Riccio, E. Romelli, M. Roncarelli, C. Rosset, E. Rossetti, B. Rusholme, R. Saglia, Z. Sakr, A. G. Sánchez, D. Sapone, J. A. Schewtschenko, M. Schirmer, P. Schneider, T. Schrabback, M. Scodeggio, A. Secroun, E. Sefusatti, G. Seidel, S. Serrano, C. Sirignano, G. Sirri, L. Stanco, J.-L. Starck, J. Steinwagner, A. N. Taylor, H. I. Teplitz, I. Tereno, R. Toledo-Moreo, F. Torradeflot, A. Tsyganov, L. Valenziano,

T. Vassallo, A. Veropalumbo, Y. Wang, J. Weller, A. Zacchei, G. Zamorani, F. M. Zerbi, J. Zoubian, E. Zucca, A. Boucaud, E. Bozzo, C. Burigana, M. Calabrese, P. Casenove, D. Di Ferdinando, J. A. Escartin Vigo, G. Fabbian, F. Finelli, J. Gracia-Carpio, S. Matthew, N. Mauri, M. Pöntinen, C. Porciani, V. Scottez, M. Tenti, M. Viel, M. Wiesmann, Y. Akrami, S. Anselmi, M. Archidiacono, F. Atrio-Barandela, E. Aubourg, A. Balaguera-Antolinez, M. Ballardini, D. Bertacca, M. Bethermin, A. Blanchard, H. Böhringer, S. Borgani, T. Bouvard, R. Cabanac, A. Calabro, B. Camacho Quevedo, G. Canas-Herrera, A. Cappi, F. Caro, C. S. Carvalho, T. Castro, K. C. Chambers, S. Contarini, T. Contini, A. R. Cooray, M. Costanzi, O. Cucciati, S. Davini, B. De Caro, S. de la Torre, G. Desprez, A. Díaz-Sánchez, J. J. Diaz, S. Di Domizio, H. Dole, S. Escoffier, M. Ezziati, A. G. Ferrari, P. G. Ferreira, I. Ferrero, A. Finoguenov, A. Fontana, F. Fornari, L. Gabarra, K. Ganga, J. García-Bellido, T. Gasparetto, E. Gaztanaga, F. Giacomini, F. Gianotti, A. H. Gonzalez, G. Gozaliasl, A. Hall, W. G. Hartley, H. Hildebrandt, J. Hjorth, A. D. Holland, O. Ilbert, S. Joudaki, E. Jullo, J. J. E. Kajava, V. Kansal, D. Karagiannis, C. C. Kirkpatrick, J. Le Graet, L. Legrand, J. Lesgourgues, T. I. Liaudat, A. Loureiro, J. Macias-Perez, M. Magliocchetti, C. Mancini, F. Mannucci, R. Maoli, C. J. A. P. Martins, L. Maurin, R. B. Metcalf, M. Migliaccio, M. Miluzio, A. Mora, C. Moretti, G. Morgante, S. Nadathur, L. Nicastro, Nicholas A. Walton, M. Oguri, L. Patrizii, V. Popa, A. Pourtsidou, P. Reimberg, I. Risso, P.-F. Rocci, R. P. Rollins, M. Sahlén, J. Schaye, A. Schneider, M. Schultheis, M. Sereno, F. Shankar, A. Shulevski, A. Silvestri, P. Simon, A. Spurio Mancini, S. A. Stanford, K. Tanidis, C. Tao, N. Tessore, G. Testera, M. Tewes, S. Toft, S. Tosi, A. Troja, M. Tucci, C. Valieri, J. Valiviita, D. Vergani, F. Vernizzi, G. Verza, P. Vielzeuf, J. R. Weaver, L. Zalesky, P. Dimauro, P.-A. Duc, Y. Fang, A. M. N. Ferguson, C. M. Gutierrez, I. Kovačić, S. Kruk, A. M. C. Le Brun, A. Montoro, C. Murray, L. Pagano, D. Paoletti, E. Sarpa, A. Viitanen, J. Martín-Fleitas, D. Scott, and L. Y. A. Yung. ¡i¿euclid¡/i¿: V. the flagship galaxy mock catalogue: A comprehensive simulation for the ¡i¿euclid¡/i¿ mission. *Astronomy amp; Astrophysics*, 697:A5, April 2025.

[21] Tessa Baker, Joseph Clampitt, Bhuvnesh Jain, and Mark Trodden. Void lensing as a test of gravity. *Physical Review D*, 98(2):023511, July 2018.

[22] Arka Banerjee and Neal Dalal. Simulating nonlinear cosmological structure formation with massive neutrinos. *Journal of Cosmology and Astroparticle Physics*, 2016(11):015, November 2016.

[23] Sudipto Banerjee, Bradley Carlin, and Alan Gelfand. *Hierarchical Modeling and Analysis of Spatial Data*, volume 101. 01 2004.

[24] Alexandre Barreira, Marius Cautun, Baojiu Li, Carlton M. Baugh, and Silvia Pascoli. Weak lensing by voids in modified lensing potentials. *Journal of Cosmology and Astroparticle Physics*, 2015(8):028–028, August 2015.

[25] Daniel Baumann, Alberto Nicolis, Leonardo Senatore, and Matias Zaldarriaga. Cosmological non-linearities as an effective fluid. *Journal of Cosmology and Astroparticle Physics*, 2012(07):051–051, 2012.

[26] M Birkinshaw. The sunyaev–zel'dovich effect. *Physics Reports*, 310(2-3):97–195, March 1999.

[27] Rahul Biswas, Esfandiar Alizadeh, and Benjamin D. Wandelt. Voids as a precision probe of dark energy. *Physical Review D*, 82(2):023002, July 2010.

[28] Michael R. Blanton, Huan Lin, Robert H. Lupton, F. Miller Maley, Neal Young, Idit Zehavi, and Jon Loveday. An Efficient Targeting Strategy for Multiobject Spectrograph Surveys: the Sloan Digital Sky Survey "Tiling" Algorithm. *The Astronomical Journal*, 125(4):2276–2286, April 2003.

[29] G. R. Blumenthal, L. N. da Costa, D. S. Goldwirth, M. Lecar, and T. Piran. The Largest Possible Voids. *The Astrophysical Journal*, 388:234, April 1992.

[30] M. Bonici, C. Carbone, S. Davini, P. Vielzeuf, L. Paganin, V. Cardone, N. Hamaus, A. Pisani, A. J. Hawken, A. Kovacs, S. Nadathur, S. Contarini, G. Verza, I. Tutusaus, F. Marulli, L. Moscardini, M. Aubert, C. Giocoli, A. Pourtsidou, S. Camera, S. Escoffier, A. Caminata, S. Di Domizio, M. Martinelli, M. Pallavicini, V. Pettorino, Z. Sakr, D. Sapone, G. Testera, S. Tosi, V. Yankelevich, A. Amara, N. Auricchio, M. Baldi, D. Bonino, E. Branchini, M. Brescia, J. Brinchmann, V. Capobianco, J. Carretero, M. Castellano, S. Cavuoti, R. Cledassou, G. Congedo, L. Conversi, Y. Copin, L. Corcione, F. Courbin, M. Cropper, A. Da Silva, H. Degaudenzi, M. Douspis, F. Dubath, C. A. J. Duncan, X. Dupac, S. Dusini, A. Ealet, S. Farrens, S. Ferriol, P. Fosalba, M. Frailis, E. Franceschi, M. Fumana, P. Gómez-Alvarez, B. Garilli, B. Gillis, A. Grazian, F. Grupp, L. Guzzo, S. V. H. Haugan, W. Holmes, F. Hormuth, A. Hornstrup, K. Jahnke, M. Kümmel, S. Kermiche, A. Kiessling, M. Kilbinger, M. Kunz, H. Kurki-Suonio, R. Laureijs, S. Ligori, P. B. Lilje, I. Lloro, E. Maiorano, O. Mansutti, O. Marggraf, K. Markovic, R. Massey, E. Medinaceli, M. Melchior, M. Meneghetti, G. Meylan, M. Moresco, E. Munari, S. M. Niemi, C. Padilla, S. Paltani, F. Pasian, K. Pedersen, W. J. Percival, S. Pires, G. Polenta, M. Poncet, L. Popa, F. Raison, R. Rebolo, A. Renzi, J. Rhodes, E. Rossetti, R. Saglia, B. Sartoris, M. Scodeggio, A. Secroun, G. Seidel, C. Sirignano, G. Sirri, L. Stanco,

J.-L. Starck, C. Surace, P. Tallada-Crespí, D. Tavagnacco, A. N. Taylor, I. Tereno, R. Toledo-Moreo, F. Torradeflot, E. A. Valentijn, L. Valenziano, Y. Wang, J. Weller, G. Zamorani, J. Zoubian, and S. Andreon. Euclid: Forecasts from the void-lensing cross-correlation. *Astronomy & Astrophysics*, 670:A47, February 2023.

[31] Supranta S Boruah, Michael J Hudson, and Guilhem Lavaux. Peculiar velocities in the local universe: comparison of different models and the implications for ¡i¿h¡/i¿0 and dark matter. *Monthly Notices of the Royal Astronomical Society*, 507(2):2697–2713, August 2021.

[32] Dillon Brout, Dan Scolnic, Brodie Popovic, Adam G. Riess, Anthony Carr, Joe Zuntz, Rick Kessler, Tamara M. Davis, Samuel Hinton, David Jones, W. D'Arcy Kenworthy, Erik R. Peterson, Khaled Said, Georgie Taylor, Noor Ali, Patrick Armstrong, Pranav Charvu, Arianna Dwomoh, Cole Meldorf, Antonella Palmese, Helen Qu, Benjamin M. Rose, Bruno Sanchez, Christopher W. Stubbs, Maria Vincenzi, Charlotte M. Wood, Peter J. Brown, Rebecca Chen, Ken Chambers, David A. Coulter, Mi Dai, Georgios Dimitriadis, Alexei V. Filippenko, Ryan J. Foley, Saurabh W. Jha, Lisa Kelsey, Robert P. Kirshner, Anais Möller, Jessie Muir, Seshadri Nadathur, Yen-Chen Pan, Armin Rest, Cesar Rojas-Bravo, Masao Sako, Matthew R. Siebert, Mat Smith, Benjamin E. Stahl, and Phil Wiseman. The pantheon+ analysis: Cosmological constraints. *The Astrophysical Journal*, 938(2):110, October 2022.

[33] A. Burden, W. J. Percival, M. Manera, Antonio J. Cuesta, Mariana Vargas Magana, and Shirley Ho. Efficient reconstruction of linear baryon acoustic oscillations in galaxy surveys. *Monthly Notices of the Royal Astronomical Society*, 445(3):3152–3168, 12 2014.

[34] Yan-Chuan Cai, Baojiu Li, Shaun Cole, Carlos S. Frenk, and Mark Neyrinck. The integrated Sachs-Wolfe effect in f(R) gravity. *Monthly Notices of the Royal Astronomical Society*, 439(3):2978–2989, April 2014.

[35] Yan-Chuan Cai, Mark Neyrinck, Qingqing Mao, John A. Peacock, Istvan Szapudi, and Andreas A. Berlind. The lensing and temperature imprints of voids on the cosmic microwave background. *Monthly Notices of the Royal Astronomical Society*, 466(3):3364–3375, April 2017.

[36] Yan-Chuan Cai, Mark C. Neyrinck, István Szapudi, Shaun Cole, and Carlos S. Frenk. A Possible Cold Imprint of Voids on the Microwave Background Radiation. *The Astrophysical Journal*, 786(2):110, May 2014.

[37] Yan-Chuan Cai, Nelson Padilla, and Baojiu Li. Testing gravity using cosmic voids. *Monthly Notices of the Royal Astronomical Society*, 451(1):1036–1055, July 2015.

[38] Yan-Chuan Cai, Andy Taylor, John A. Peacock, and Nelson Padilla. Redshift-space distortions around voids. *Monthly Notices of the Royal Astronomical Society*, 462(3):2465–2477, July 2016.

[39] Erminia Calabrese, Renée A. Hložek, J. Richard Bond, Mark J. Devlin, Joanna Dunkley, Mark Halpern, Adam D. Hincks, Kent D. Irwin, Arthur Kosowsky, Kavilan Moodley, Laura B. Newburgh, Michael D. Niemack, Lyman A. Page, Blake D. Sherwin, Jonathan L. Sievers, David N. Spergel, Suzanne T. Staggs, and Edward J. Wollack. Cosmological parameters from pre-planck cmb measurements: A 2017 update. *Phys. Rev. D*, 95:063525, Mar 2017.

[40] John Joseph M. Carrasco, Mark P. Hertzberg, and Leonardo Senatore. The effective field theory of cosmological large scale structures. *Journal of High Energy Physics*, 2012:82, September 2012.

[41] Marius Cautun, Enrique Paillas, Yan-Chuan Cai, Sownak Bose, Joaquin Armijo, Baojiu Li, and Nelson Padilla. The santiago–harvard–edinburgh–durham void comparison – i. shedding light on chameleon gravity tests. *Monthly Notices of the Royal Astronomical Society*, 476(3):3195–3217, February 2018.

[42] Teeraparb Chantavat, Utane Sawangwit, Paul M. Sutter, and Benjamin D. Wandelt. Cosmological parameter constraints from CMB lensing with cosmic voids. *Physical Review D*, 93(4):043523, February 2016.

[43] Teeraparb Chantavat, Utane Sawangwit, and Benjamin D. Wandelt. Void Profile from Planck Lensing Potential Map. *The Astrophysical Journal*, 836(2):156, February 2017.

[44] Michael J. Chapman, Faizan G. Mohammad, Zhongxu Zhai, Will J. Percival, Jeremy L. Tinker, Julian E. Bautista, Joel R. Brownstein, Etienne Burtin, Kyle S. Dawson, Héctor Gil-Marín, Axel de la Macorra, Ashley J. Ross, Graziano Rossi, Donald P. Schneider, and Gong-Bo Zhao. The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the growth rate of structure from the small-scale clustering of the luminous red galaxy sample. *Monthly Notices of the Royal Astronomical Society*, 516(1):617–635, October 2022.

[45] Zhao Chen and Yu Yu. Extending csst emulator to post-desi era. *Journal of Cosmology and Astroparticle Physics*, 2026(06):086, 2026.

[46] Chia-Hsun Chuang, Francisco-Shu Kitaura, Francisco Prada, Cheng Zhao, and Gustavo Yepes. Ezmocks: extending the zel'dovich approximation to generate mock galaxy catalogues with accurate clustering statistics. *Monthly Notices of the Royal Astronomical Society*, 446(3):2621–2628, November 2014.

[47] Joseph Clampitt, Yan-Chuan Cai, and Baojiu Li. Voids in modified gravity: excursion set predictions. *MNRAS*, 431(1):749–766, May 2013.

[48] Joseph Clampitt and Bhuvnesh Jain. Lensing measurements of the mass distribution in SDSS voids. *Monthly Notices of the Royal Astronomical Society*, 454(4):3357–3365, December 2015.

[49] Jörg M. Colberg, Frazer Pearce, Caroline Foster, Erwin Platen, Riccardo Brunino, Mark Neyrinck, Spyros Basilakos, Anthony Fairall, Hume Feldman, Stefan Gottlöber, Oliver Hahn, Fiona Hoyle, Volker Müller, Lorne Nelson, Manolis Plionis, Cristiano Porciani, Sergei Shandarin, Michael S. Vogeley, and Rien van de Weygaert. The Aspen-Amsterdam void finder comparison project. *MNRAS*, 387(2):933–944, June 2008.

[50] Shaun Cole and Nick Kaiser. Biased clustering in the cold dark matter cosmogony. *Monthly Notices of the Royal Astronomical Society*, 237:1127–1146, April 1989.

[51] DES Collaboration, T. M. C. Abbott, M. Acevedo, M. Aguena, A. Alarcon, S. Allam, O. Alves, A. Amon, F. Andrade-Oliveira, J. Annis, P. Armstrong, J. Asorey, S. Avila, D. Bacon, B. A. Bassett, K. Bechtol, P. H. Bernardinelli, G. M. Bernstein, E. Bertin, J. Blazek, S. Bocquet, D. Brooks, D. Brout, E. Buckley-Geer, D. L. Burke, H. Camacho, R. Camilleri, A. Campos, A. Carnero Rosell, D. Carollo, A. Carr, J. Carretero, F. J. Castander, R. Cawthon, C. Chang, R. Chen, A. Choi, C. Conselice, M. Costanzi, L. N. da Costa, M. Crocce, T. M. Davis, D. L. DePoy, S. Desai, H. T. Diehl, M. Dixon, S. Dodelson, P. Doel, C. Doux, A. Drlica-Wagner, J. Elvin-Poole, S. Everett, I. Ferrero, A. Ferté, B. Flaugher, R. J. Foley, P. Fosalba, D. Friedel, J. Frieman, C. Frohmaier, L. Galbany, J. García-Bellido, M. Gatti, E. Gaztanaga, G. Giannini, K. Glazebrook, O. Graur, D. Gruen, R. A. Gruendl, G. Gutierrez, W. G. Hartley, K. Herner, S. R. Hinton, D. L. Hollowood, K. Honscheid, D. Huterer, B. Jain, D. J. James, N. Jeffrey, E. Kasai, L. Kelsey, S. Kent, R. Kessler, A. G. Kim, R. P. Kirshner, E. Kovacs, K. Kuehn, O. Lahav, J. Lee, S. Lee, G. F. Lewis, T. S. Li, C. Lidman, H. Lin, U. Malik, J. L. Marshall, P. Martini, J. Mena-Fernández, F. Menanteau, R. Miquel, J. J. Mohr, J. Mould, J. Muir, A. Möller, E. Neilsen, R. C. Nichol, P. Nugent, R. L. C. Ogando, A. Palmese, Y. C.

Pan, M. Paterno, W. J. Percival, M. E. S. Pereira, A. Pieres, A. A. Plazas Malagón, B. Popovic, A. Porredon, J. Prat, H. Qu, M. Raveri, M. Rodríguez-Monroy, A. K. Romer, A. Roodman, B. Rose, M. Sako, E. Sanchez, D. Sanchez Cid, M. Schubnell, D. Scolnic, I. Sevilla-Noarbe, P. Shah, J. Allyn. Smith, M. Smith, M. Soares-Santos, E. Suchyta, M. Sullivan, N. Suntzeff, M. E. C. Swanson, B. O. Sánchez, G. Tarle, G. Taylor, D. Thomas, C. To, M. Toy, M. A. Troxel, B. E. Tucker, D. L. Tucker, S. A. Uddin, M. Vincenzi, A. R. Walker, N. Weaverdyck, R. H. Wechsler, J. Weller, W. Wester, P. Wiseman, M. Yamamoto, F. Yuan, B. Zhang, and Y. Zhang. The dark energy survey: Cosmology results with 1500 new high-redshift type ia supernovae using the full 5-year dataset, 2025.

[52] DESI Collaboration, B. Abareshi, J. Aguilar, S. Ahlen, Shadab Alam, David M. Alexander, R. Alfarsy, L. Allen, C. Allende Prieto, O. Alves, J. Ameel, E. Armengaud, J. Asorey, Alejandro Aviles, S. Bailey, A. Balaguera-Antolínez, O. Ballester, C. Baltay, A. Bault, S. F. Beltran, B. Benavides, S. BenZvi, A. Berti, R. Besuner, Florian Beutler, D. Bianchi, C. Blake, P. Blanc, R. Blum, A. Bolton, S. Bose, D. Bramall, S. Brieden, A. Brodzeller, D. Brooks, C. Brownewell, E. Buckley-Geer, R. N. Cahn, Z. Cai, R. Canning, R. Capasso, A. Carnero Rosell, P. Carton, R. Casas, F. J. Castander, J. L. Cervantes-Cota, S. Chabanier, E. Chaussidon, C. Chuang, C. Circosta, S. Cole, A. P. Cooper, L. da Costa, M.-C. Cousinou, A. Cuceu, T. M. Davis, K. Dawson, R. de la Cruz-Noriega, A. de la Macorra, A. de Mattia, J. Della Costa, P. Demmer, M. Derwent, A. Dey, B. Dey, G. Dhungana, Z. Ding, C. Dobson, P. Doel, J. Donald-McCann, J. Donaldson, K. Douglass, Y. Duan, P. Dunlop, J. Edelstein, S. Eftekharzadeh, D. J. Eisenstein, M. Enriquez-Vargas, S. Escoffier, M. Evatt, P. Fagrelius, X. Fan, K. Fanning, V. A. Fawcett, S. Ferraro, J. Ereza, B. Flaugher, A. Font-Ribera, J. E. Forero-Romero, C. S. Frenk, S. Fromenteau, B. T. Gänsicke, C. Garcia-Quintero, L. Garrison, E. Gaztañaga, F. Gerardi, H. Gil-Marín, S. Gontcho A Gontcho, Alma X. Gonzalez-Morales, G. Gonzalez-de Rivera, V. Gonzalez-Perez, C. Gordon, O. Graur, D. Green, C. Grove, D. Gruen, G. Gutierrez, J. Guy, C. Hahn, S. Harris, D. Herrera, Hiram K. Herrera-Alcantar, K. Honscheid, C. Howlett, D. Huterer, V. Iršič, M. Ishak, P. Jelinsky, L. Jiang, J. Jimenez, Y. P. Jing, R. Joyce, E. Jullo, S. Juneau, N. G. Karaçaylı, M. Karamanis, A. Karcher, T. Karim, R. Kehoe, S. Kent, D. Kirkby, T. Kisner, F. Kitaura, S. E. Koposov, A. Kovács, A. Kremin, Alex Krolewski, B. L'Huillier, O. Lahav, A. Lambert, C. Lamman, Ting-Wen Lan, M. Landriau, S. Lane, D. Lang, J. U. Lange, J. Lasker, L. Le Guillou, A. Leauthaud, A. Le Van Suu, Michael E. Levi, T. S. Li, C. Magneville, M. Manera, Christopher J. Manser, B. Marshall, Paul Martini, W. McCollam, P. McDonald, Aaron M. Meisner, J. Mena-Fernández, J. Meneses-

Rizo, M. Mezcua, T. Miller, R. Miquel, P. Montero-Camacho, J. Moon, J. Moustakas, E. Mueller, Andrea Muñoz-Gutiérrez, Adam D. Myers, S. Nadathur, J. Najita, L. Napolitano, E. Neilsen, Jeffrey A. Newman, J. D. Nie, Y. Ning, G. Niz, P. Norberg, Hernán E. Noriega, T. O'Brien, A. Obuljen, N. Palanque-Delabrouille, A. Palmese, P. Zhiwei, D. Pappalardo, X. PENG, W. J. Percival, S. Perruchot, R. Pogge, C. Poppett, A. Porredon, F. Prada, J. Prochaska, R. Pucha, A. Pérez-Fernández, I. Pérez-Ràfols, D. Rabinowitz, A. Raichoor, S. Ramirez-Solano, César Ramírez-Pérez, C. Ravoux, K. Reil, M. Rezaie, A. Rocher, C. Rockosi, N. A. Roe, A. Roodman, A. J. Ross, G. Rossi, R. Ruggeri, V. Ruhlmann-Kleider, C. G. Sabiu, S. Gaines, K. Said, A. Saintonge, Javier Salas Catonga, L. Samushia, E. Sanchez, C. Saulder, E. Schaan, E. Schlafly, D. Schlegel, J. Schmoll, D. Scholte, M. Schubnell, A. Secroun, H. Seo, S. Serrano, Ray M. Sharples, Michael J. Sholl, Joseph Harry Silber, D. R. Silva, M. Sirk, M. Siudek, A. Smith, D. Sprayberry, R. Staten, B. Stupak, T. Tan, Gregory Tarlé, Suk Sien Tie, R. Tojeiro, L. A. Ureña-López, F. Valdes, O. Valenzuela, M. Valluri, M. Vargas-Magaña, L. Verde, M. Walther, B. Wang, M. S. Wang, B. A. Weaver, C. Weaverdyck, R. Wechsler, Michael J. Wilson, J. Yang, Y. Yu, S. Yuan, Christophe Yèche, H. Zhang, K. Zhang, Cheng Zhao, Rongpu Zhou, Zhimin Zhou, H. Zou, J. Zou, S. Zou, and Y. Zu. Overview of the instrumentation for the dark energy spectroscopic instrument. *The Astronomical Journal*, 164(5):207, October 2022.

[53] S. Contarini, G. Verza, A. Pisani, N. Hamaus, M. Sahlén, C. Carbone, S. Dusini, F. Marulli, L. Moscardini, A. Renzi, C. Sirignano, L. Stanco, M. Aubert, M. Bonici, G. Castignani, H. M. Courtois, S. Escoffier, D. Guinet, A. Kovacs, G. Lavaux, E. Massara, S. Nadathur, G. Pollina, T. Ronconi, F. Ruppin, Z. Sakr, A. Veropalumbo, B. D. Wandelt, A. Amara, N. Auricchio, M. Baldi, D. Bonino, E. Branchini, M. Brescia, J. Brinchmann, S. Camera, V. Capobianco, J. Carretero, M. Castellano, S. Cavuoti, R. Cledassou, G. Congedo, C. J. Conselice, L. Conversi, Y. Copin, L. Corcione, F. Courbin, M. Cropper, A. Da Silva, H. Degaudenzi, F. Dubath, C. A. J. Duncan, X. Dupac, A. Ealet, S. Farrens, S. Ferriol, P. Fosalba, M. Frailis, E. Franceschi, B. Garilli, W. Gillard, B. Gillis, C. Giocoli, A. Grazian, F. Grupp, L. Guzzo, S. Haugan, W. Holmes, F. Hormuth, K. Jahnke, M. Kümmel, S. Kermiche, A. Kiessling, M. Kilbinger, M. Kunz, H. Kurki-Suonio, R. Laureijs, S. Ligori, P. B. Lilje, I. Lloro, E. Maiorano, O. Mansutti, O. Marggraf, K. Markovic, R. Massey, M. Melchior, M. Meneghetti, G. Meylan, M. Moresco, E. Munari, S. M. Niemi, C. Padilla, S. Paltani, F. Pasian, K. Pedersen, W. J. Percival, V. Pettorino, S. Pires, G. Polenta, M. Poncet, L. Popa, L. Pozzetti, F. Raison, J. Rhodes, E. Rossetti, R. Saglia, B. Sartoris, P. Schneider, A. Secroun, G. Seidel, G. Sirri, C. Surace, P. Tallada-

Crespí, A. N. Taylor, I. Tereno, R. Toledo-Moreo, F. Torradeflot, E. A. Valentijn, L. Valenziano, Y. Wang, J. Weller, G. Zamorani, J. Zoubian, S. Andreon, D. Maino, and S. Mei. Euclid: Cosmological forecasts from the void size function. *Astronomy amp; Astrophysics*, 667:A162, November 2022.

[54] Sofia Contarini, Federico Marulli, Lauro Moscardini, Alfonso Veropalumbo, Carlo Giocoli, and Marco Baldi. Cosmic voids in modified gravity models with massive neutrinos. *Monthly Notices of the Royal Astronomical Society*, 504(4):5021–5038, July 2021.

[55] Sofia Contarini, Alice Pisani, Nico Hamaus, Federico Marulli, Lauro Moscardini, and Marco Baldi. Cosmological Constraints from the BOSS DR12 Void Size Function. *The Astrophysical Journal*, 953(1):46, August 2023.

[56] Ryan J. Cooke, Max Pettini, Kenneth M. Nollett, and Regina Jorgenson. The primordial deuterium abundance of the most metal-poor damped ly system. *The Astrophysical Journal*, 830(2):148, October 2016.

[57] Carlos M. Correa, Dante J. Paz, Ariel G. Sánchez, Andrés N. Ruiz, Nelson D. Padilla, and Raúl E. Angulo. Redshift-space effects in voids and their impact on cosmological tests. Part I: the void size function. *Monthly Notices of the Royal Astronomical Society*, 500(1):911–925, January 2021.

[58] Carlos Mauricio Correa. Cosmic voids as cosmological laboratories, 2022.

[59] M. C. Cousinou, A. Pisani, A. Tilquin, N. Hamaus, A. J. Hawken, and S. Escoffier. Multivariate analysis of cosmic void characteristics. *Astronomy and Computing*, 27:53, April 2019.

[60] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Science*, 117(48):30055–30062, December 2020.

[61] SM Croom, RJ Smith, BJ Boyle, T Shanks, NS Loaring, L Miller, and IJ Lewis. The 2df qso redshift survey—v. the 10k catalogue. *Monthly Notices of the Royal Astronomical Society*, 322(4):L29–L36, 2001.

[62] Carolina Cuesta-Lazaro, Enrique Paillas, Sihan Yuan, Yan-Chuan Cai, Seshadri Nadathur, Will J. Percival, Florian Beutler, Arnaud de Mattia, Daniel Eisenstein, Daniel Forero-Sanchez, Nelson Padilla, Mathilde Pinon, Vanina Ruhlmann-Kleider,

Ariel G. Sánchez, Georgios Valogiannis, and Pauline Zarrouk. Sunbird: A simulation-based model for full-shape density-split clustering, 2023.

[63] Olivia Curtis, Bryanne McDonough, and Tereasa G. Brainerd. Void Galaxies and Active Galactic Nucleus Activity in ZOBOV-identified TNG300 Voids from z = 3 to z = 0. *The Astrophysical Journal*, 998(2):325, February 2026.

[64] Cihan H Dagli. *Artificial neural networks for intelligent manufacturing*. Springer Science & Business Media, 2012.

[65] Christopher T. Davies, Marius Cautun, Benjamin Giblin, Baojiu Li, Joachim Harnois-Déraps, and Yan-Chuan Cai. Constraining cosmology with weak lensing voids. *Monthly Notices of the Royal Astronomical Society*, 507(2):2267–2282, October 2021.

[66] M. Davis and P. J. E. Peebles. A survey of galaxy redshifts. V. The two-point position and velocity correlations. *The Astrophysical Journal*, 267:465–482, April 1983.

[67] Kyle S. Dawson, David J. Schlegel, Christopher P. Ahn, Scott F. Anderson, Éric Aubourg, Stephen Bailey, Robert H. Barkhouser, Julian E. Bautista, Alessandra Beifiori, Andreas A. Berlind, Vaishali Bhardwaj, Dmitry Bizyaev, Cullen H. Blake, Michael R. Blanton, Michael Blomqvist, Adam S. Bolton, Arnaud Borde, Jo Bovy, W. N. Brandt, Howard Brewington, Jon Brinkmann, Peter J. Brown, Joel R. Brownstein, Kevin Bundy, N. G. Busca, William Carithers, Aurelio R. Carnero, Michael A. Carr, Yanmei Chen, Johan Comparat, Natalia Connolly, Frances Cope, Rupert A. C. Croft, Antonio J. Cuesta, Luiz N. da Costa, James R. A. Davenport, Timothée Delubac, Roland de Putter, Saurav Dhital, Anne Ealet, Garrett L. Ebelke, Daniel J. Eisenstein, S. Escoffier, Xiaohui Fan, N. Filiz Ak, Hayley Finley, Andreu Font-Ribera, R. Génova-Santos, James E. Gunn, Hong Guo, Daryl Haggard, Patrick B. Hall, Jean-Christophe Hamilton, Ben Harris, David W. Harris, Shirley Ho, David W. Hogg, Diana Holder, Klaus Honscheid, Joe Huehnerhoff, Beatrice Jordan, Wendell P. Jordan, Guinevere Kauffmann, Eyal A. Kazin, David Kirkby, Mark A. Klaene, Jean-Paul Kneib, Jean-Marc Le Goff, Khee-Gan Lee, Daniel C. Long, Craig P. Loomis, Britt Lundgren, Robert H. Lupton, Marcio A. G. Maia, Martin Makler, Elena Malanushenko, Viktor Malanushenko, Rachel Mandelbaum, Marc Manera, Claudia Maraston, Daniel Margala, Karen L. Masters, Cameron K. McBride, Patrick McDonald, Ian D. McGreer, Richard G. McMahon, Olga Mena, Jordi Miralda-Escudé, Antonio D. Montero-Dorta, Francesco Montesano, Demitri Muna, Adam D. Myers, Tracy Naugle, Robert C. Nichol, Pasquier Noterdaeme, Sebastián E. Nuza,

Matthew D. Olmstead, Audrey Oravetz, Daniel J. Oravetz, Russell Owen, Nikhil Padmanabhan, Nathalie Palanque-Delabrouille, Kaike Pan, John K. Parejko, Isabelle Pâris, Will J. Percival, Ismael Pérez-Fournon, Ignasi Pérez-Ràfols, Patrick Petitjean, Robert Pfaffenberger, Janine Pforr, Matthew M. Pieri, Francisco Prada, Adrian M. Price-Whelan, M. Jordan Raddick, Rafael Rebolo, James Rich, Gordon T. Richards, Constance M. Rockosi, Natalie A. Roe, Ashley J. Ross, Nicholas P. Ross, Graziano Rossi, J. A. Rubiño-Martin, Lado Samushia, Ariel G. Sánchez, Conor Sayres, Sarah J. Schmidt, Donald P. Schneider, C. G. Scóccola, Hee-Jong Seo, Alaina Shelden, Erin Sheldon, Yue Shen, Yiping Shu, Anže Slosar, Stephen A. Smee, Stephanie A. Snedden, Fritz Stauffer, Oliver Steele, Michael A. Strauss, Alina Streblyanska, Nao Suzuki, Molly E. C. Swanson, Tomer Tal, Masayuki Tanaka, Daniel Thomas, Jeremy L. Tinker, Rita Tojeiro, Christy A. Tremonti, M. Vargas Magaña, Licia Verde, Matteo Viel, David A. Wake, Mike Watson, Benjamin A. Weaver, David H. Weinberg, Benjamin J. Weiner, Andrew A. West, Martin White, W. M. Wood-Vasey, Christophe Yeche, Idit Zehavi, Gong-Bo Zhao, and Zheng Zheng. The Baryon Oscillation Spectroscopic Survey of SDSS-III. *The Astronomical Journal*, 145(1):10, January 2013.

[68] G. Degni, E. Sarpa, M. Aubert, E. Branchini, A. Pisani, and H. M. Courtois. Extracting the alcock-paczyński signal from voids: A novel approach via reconstruction. *Astronomy amp; Astrophysics*, 706:A286, February 2026.

[69] DESI Collaboration, Amir Aghamousa, Jessica Aguilar, Steve Ahlen, Shadab Alam, Lori E. Allen, Carlos Allende Prieto, James Annis, Stephen Bailey, Christophe Balland, Otger Ballester, Charles Baltay, Lucas Beaufore, Chris Bebek, Timothy C. Beers, Eric F. Bell, José Luis Bernal, Robert Besuner, Florian Beutler, Chris Blake, Hannes Bleuler, Michael Blomqvist, Robert Blum, Adam S. Bolton, Cesar Briceno, David Brooks, Joel R. Brownstein, Elizabeth Buckley-Geer, Angela Burden, Etienne Burtin, Nicolas G. Busca, Robert N. Cahn, Yan-Chuan Cai, Laia Cardiel-Sas, Raymond G. Carlberg, Pierre-Henri Carton, Ricard Casas, Francisco J. Castander, Jorge L. Cervantes-Cota, Todd M. Claybaugh, Madeline Close, Carl T. Coker, Shaun Cole, Johan Comparat, Andrew P. Cooper, M. C. Cousinou, Martin Crocce, Jean-Gabriel Cuby, Daniel P. Cunningham, Tamara M. Davis, Kyle S. Dawson, Axel de la Macorra, Juan De Vicente, Timothée Delubac, Mark Derwent, Arjun Dey, Govinda Dhungana, Zhejie Ding, Peter Doel, Yutong T. Duan, Anne Ealet, Jerry Edelstein, Sarah Eftekharzadeh, Daniel J. Eisenstein, Ann Elliott, Stéphanie Escoffier, Matthew Evatt, Parker Fagrelius, Xiaohui Fan, Kevin Fanning, Arya Farahi, Jay Farihi, Ginevra Favole, Yu Feng, Enrique Fernandez, Joseph R. Findlay, Douglas P. Finkbeiner, Michael J. Fitzpatrick, Brenna Flaugher, Samuel Flender, An-

dreu Font-Ribera, Jaime E. Forero-Romero, Pablo Fosalba, Carlos S. Frenk, Michele Fumagalli, Boris T. Gaensicke, Giuseppe Gallo, Juan Garcia-Bellido, Enrique Gaztanaga, Nicola Pietro Gentile Fusillo, Terry Gerard, Irena Gershkovich, Tommaso Giannantonio, Denis Gillet, Guillermo Gonzalez-de-Rivera, Violeta Gonzalez-Perez, Shelby Gott, Or Graur, Gaston Gutierrez, Julien Guy, Salman Habib, Henry Heetderks, Ian Heetderks, Katrin Heitmann, Wojciech A. Hellwing, David A. Herrera, Shirley Ho, Stephen Holland, Klaus Honscheid, Eric Huff, Timothy A. Hutchinson, Dragan Huterer, Ho Seong Hwang, Joseph Maria Illa Laguna, Yuzo Ishikawa, Dianna Jacobs, Niall Jeffrey, Patrick Jelinsky, Elise Jennings, Linhua Jiang, Jorge Jimenez, Jennifer Johnson, Richard Joyce, Eric Jullo, Stéphanie Juneau, Sami Kama, Armin Karcher, Sonia Karkar, Robert Kehoe, Noble Kennamer, Stephen Kent, Martin Kilbinger, Alex G. Kim, David Kirkby, Theodore Kisner, Ellie Kitanidis, Jean-Paul Kneib, Sergey Koposov, Eve Kovacs, Kazuya Koyama, Anthony Kremin, Richard Kron, Luzius Kronig, Andrea Kueter-Young, Cedric G. Lacey, Robin Lafever, Ofer Lahav, Andrew Lambert, Michael Lampton, Martin Landriau, Dustin Lang, Tod R. Lauer, Jean-Marc Le Goff, Laurent Le Guillou, Auguste Le Van Suu, Jae Hyeon Lee, Su-Jeong Lee, Daniela Leitner, Michael Lesser, Michael E. Levi, Benjamin L'Huillier, Baojiu Li, Ming Liang, Huan Lin, Eric Linder, Sarah R. Loebman, Zarija Lukić, Jun Ma, Niall MacCrann, Christophe Magneville, Laleh Makarem, Marc Manera, Christopher J. Manser, Robert Marshall, Paul Martini, Richard Massey, Thomas Matheson, Jeremy McCauley, Patrick McDonald, Ian D. McGreer, Aaron Meisner, Nigel Metcalfe, Timothy N. Miller, Ramon Miquel, John Moustakas, Adam Myers, Milind Naik, Jeffrey A. Newman, Robert C. Nichol, Andrina Nicola, Luiz Nicolati da Costa, Jundan Nie, Gustavo Niz, Peder Norberg, Brian Nord, Dara Norman, Peter Nugent, Thomas O'Brien, Minji Oh, Knut A. G. Olsen, Cristobal Padilla, Hamsa Padmanabhan, Nikhil Padmanabhan, Nathalie Palanque-Delabrouille, Antonella Palmese, Daniel Pappalardo, Isabelle Pâris, Changbom Park, Anna Patej, John A. Peacock, Hiranya V. Peiris, Xiyan Peng, Will J. Percival, Sandrine Perruchot, Matthew M. Pieri, Richard Pogge, Jennifer E. Pollack, Claire Poppett, Francisco Prada, Abhishek Prakash, Ronald G. Probst, David Rabinowitz, Anand Raichoor, Chang Hee Ree, Alexandre Refregier, Xavier Regal, Beth Reid, Kevin Reil, Mehdi Rezaie, Constance M. Rockosi, Natalie Roe, Samuel Ronayette, Aaron Roodman, Ashley J. Ross, Nicholas P. Ross, Graziano Rossi, Eduardo Rozo, Vanina Ruhlmann-Kleider, Eli S. Rykoff, Cristiano Sabiu, Lado Samushia, Eusebio Sanchez, Javier Sanchez, David J. Schlegel, Michael Schneider, Michael Schubnell, Aurélia Secroun, Uros Seljak, Hee-Jong Seo, Santiago Serrano, Arman Shafieloo, Huanyuan Shan, Ray Sharples, Michael J. Sholl, William V. Shourt, Joseph H. Silber, David R. Silva, Martin M. Sirk, Anze Slosar, Alex Smith, George F. Smoot, Debopam Som,

Yong-Seon Song, David Sprayberry, Ryan Staten, Andy Stefanik, Gregory Tarle, Suk Sien Tie, Jeremy L. Tinker, Rita Tojeiro, Francisco Valdes, Octavio Valenzuela, Monica Valluri, Mariana Vargas-Magana, Licia Verde, Alistair R. Walker, Jiali Wang, Yuting Wang, Benjamin A. Weaver, Curtis Weaverdyck, Risa H. Wechsler, David H. Weinberg, Martin White, Qian Yang, Christophe Yeche, Tianmeng Zhang, Gong-Bo Zhao, Yi Zheng, Xu Zhou, Zhimin Zhou, Yaling Zhu, Hu Zou, and Ying Zu. The DESI Experiment Part I: Science,Targeting, and Survey Design. *arXiv e-prints*, page arXiv:1611.00036, October 2016.

[70] DESI Collaboration, Amir Aghamousa, Jessica Aguilar, Steve Ahlen, Shadab Alam, Lori E. Allen, Carlos Allende Prieto, James Annis, Stephen Bailey, Christophe Balland, Otger Ballester, Charles Baltay, Lucas Beaufore, Chris Bebek, Timothy C. Beers, Eric F. Bell, José Luis Bernal, Robert Besuner, Florian Beutler, Chris Blake, Hannes Bleuler, Michael Blomqvist, Robert Blum, Adam S. Bolton, Cesar Briceno, David Brooks, Joel R. Brownstein, Elizabeth Buckley-Geer, Angela Burden, Etienne Burtin, Nicolas G. Busca, Robert N. Cahn, Yan-Chuan Cai, Laia Cardiel-Sas, Raymond G. Carlberg, Pierre-Henri Carton, Ricard Casas, Francisco J. Castander, Jorge L. Cervantes-Cota, Todd M. Claybaugh, Madeline Close, Carl T. Coker, Shaun Cole, Johan Comparat, Andrew P. Cooper, M. C. Cousinou, Martin Crocce, Jean-Gabriel Cuby, Daniel P. Cunningham, Tamara M. Davis, Kyle S. Dawson, Axel de la Macorra, Juan De Vicente, Timothée Delubac, Mark Derwent, Arjun Dey, Govinda Dhungana, Zhejie Ding, Peter Doel, Yutong T. Duan, Anne Ealet, Jerry Edelstein, Sarah Eftekharzadeh, Daniel J. Eisenstein, Ann Elliott, Stéphanie Escoffier, Matthew Evatt, Parker Fagrelius, Xiaohui Fan, Kevin Fanning, Arya Farahi, Jay Farihi, Ginevra Favole, Yu Feng, Enrique Fernandez, Joseph R. Findlay, Douglas P. Finkbeiner, Michael J. Fitzpatrick, Brenna Flaugher, Samuel Flender, Andreu Font-Ribera, Jaime E. Forero-Romero, Pablo Fosalba, Carlos S. Frenk, Michele Fumagalli, Boris T. Gaensicke, Giuseppe Gallo, Juan Garcia-Bellido, Enrique Gaztanaga, Nicola Pietro Gentile Fusillo, Terry Gerard, Irena Gershkovich, Tommaso Giannantonio, Denis Gillet, Guillermo Gonzalez-de-Rivera, Violeta Gonzalez-Perez, Shelby Gott, Or Graur, Gaston Gutierrez, Julien Guy, Salman Habib, Henry Heetderks, Ian Heetderks, Katrin Heitmann, Wojciech A. Hellwing, David A. Herrera, Shirley Ho, Stephen Holland, Klaus Honscheid, Eric Huff, Timothy A. Hutchinson, Dragan Huterer, Ho Seong Hwang, Joseph Maria Illa Laguna, Yuzo Ishikawa, Dianna Jacobs, Niall Jeffrey, Patrick Jelinsky, Elise Jennings, Linhua Jiang, Jorge Jimenez, Jennifer Johnson, Richard Joyce, Eric Jullo, Stéphanie Juneau, Sami Kama, Armin Karcher, Sonia Karkar, Robert Kehoe, Noble Kennamer, Stephen Kent, Martin Kilbinger, Alex G. Kim, David Kirkby, Theodore Kisner, Ellie Kitanidis, Jean-Paul

Kneib, Sergey Koposov, Eve Kovacs, Kazuya Koyama, Anthony Kremin, Richard Kron, Luzius Kronig, Andrea Kueter-Young, Cedric G. Lacey, Robin Lafever, Ofer Lahav, Andrew Lambert, Michael Lampton, Martin Landriau, Dustin Lang, Tod R. Lauer, Jean-Marc Le Goff, Laurent Le Guillou, Auguste Le Van Suu, Jae Hyeon Lee, Su-Jeong Lee, Daniela Leitner, Michael Lesser, Michael E. Levi, Benjamin L'Huillier, Baojiu Li, Ming Liang, Huan Lin, Eric Linder, Sarah R. Loebman, Zarija Lukić, Jun Ma, Niall MacCrann, Christophe Magneville, Laleh Makarem, Marc Manera, Christopher J. Manser, Robert Marshall, Paul Martini, Richard Massey, Thomas Matheson, Jeremy McCauley, Patrick McDonald, Ian D. McGreer, Aaron Meisner, Nigel Metcalfe, Timothy N. Miller, Ramon Miquel, John Moustakas, Adam Myers, Milind Naik, Jeffrey A. Newman, Robert C. Nichol, Andrina Nicola, Luiz Nicolati da Costa, Jundan Nie, Gustavo Niz, Peder Norberg, Brian Nord, Dara Norman, Peter Nugent, Thomas O'Brien, Minji Oh, Knut A. G. Olsen, Cristobal Padilla, Hamsa Padmanabhan, Nikhil Padmanabhan, Nathalie Palanque-Delabrouille, Antonella Palmese, Daniel Pappalardo, Isabelle Pâris, Changbom Park, Anna Patej, John A. Peacock, Hiranya V. Peiris, Xiyan Peng, Will J. Percival, Sandrine Perruchot, Matthew M. Pieri, Richard Pogge, Jennifer E. Pollack, Claire Poppett, Francisco Prada, Abhishek Prakash, Ronald G. Probst, David Rabinowitz, Anand Raichoor, Chang Hee Ree, Alexandre Refregier, Xavier Regal, Beth Reid, Kevin Reil, Mehdi Rezaie, Constance M. Rockosi, Natalie Roe, Samuel Ronayette, Aaron Roodman, Ashley J. Ross, Nicholas P. Ross, Graziano Rossi, Eduardo Rozo, Vanina Ruhlmann-Kleider, Eli S. Rykoff, Cristiano Sabiu, Lado Samushia, Eusebio Sanchez, Javier Sanchez, David J. Schlegel, Michael Schneider, Michael Schubnell, Aurélia Secroun, Uros Seljak, Hee-Jong Seo, Santiago Serrano, Arman Shafieloo, Huanyuan Shan, Ray Sharples, Michael J. Sholl, William V. Shourt, Joseph H. Silber, David R. Silva, Martin M. Sirk, Anze Slosar, Alex Smith, George F. Smoot, Debopam Som, Yong-Seon Song, David Sprayberry, Ryan Staten, Andy Stefanik, Gregory Tarle, Suk Sien Tie, Jeremy L. Tinker, Rita Tojeiro, Francisco Valdes, Octavio Valenzuela, Monica Valluri, Mariana Vargas-Magana, Licia Verde, Alistair R. Walker, Jiali Wang, Yuting Wang, Benjamin A. Weaver, Curtis Weaverdyck, Risa H. Wechsler, David H. Weinberg, Martin White, Qian Yang, Christophe Yeche, Tianmeng Zhang, Gong-Bo Zhao, Yi Zheng, Xu Zhou, Zhimin Zhou, Yaling Zhu, Hu Zou, and Ying Zu. The DESI Experiment Part II: Instrument Design. *arXiv e-prints*, page arXiv:1611.00037, October 2016.

[71] Arjun Dey, David J. Schlegel, Dustin Lang, Robert Blum, Kaylan Burleigh, Xiaohui Fan, Joseph R. Findlay, Doug Finkbeiner, David Herrera, Stéphanie Juneau, Martin Landriau, Michael Levi, Ian McGreer, Aaron Meisner, Adam D. Myers, John

Moustakas, Peter Nugent, Anna Patej, Edward F. Schlafly, Alistair R. Walker, Francisco Valdes, Benjamin A. Weaver, Christophe Yèche, Hu Zou, Xu Zhou, Behzad Abareshi, T. M. C. Abbott, Bela Abolfathi, C. Aguilera, Shadab Alam, Lori Allen, A. Alvarez, James Annis, Behzad Ansarinejad, Marie Aubert, Jacqueline Beechert, Eric F. Bell, Segev Y. BenZvi, Florian Beutler, Richard M. Bielby, Adam S. Bolton, César Briceño, Elizabeth J. Buckley-Geer, Karen Butler, Annalisa Calamida, Raymond G. Carlberg, Paul Carter, Ricard Casas, Francisco J. Castander, Yumi Choi, Johan Comparat, Elena Cukanovaite, Timothée Delubac, Kaitlin DeVries, Sharmila Dey, Govinda Dhungana, Mark Dickinson, Zhejie Ding, John B. Donaldson, Yutong Duan, Christopher J. Duckworth, Sarah Eftekharzadeh, Daniel J. Eisenstein, Thomas Etourneau, Parker A. Fagrelius, Jay Farihi, Mike Fitzpatrick, Andreu Font-Ribera, Leah Fulmer, Boris T. Gänsicke, Enrique Gaztanaga, Koshy George, David W. Gerdes, Satya Gontcho A. Gontcho, Claudio Gorgoni, Gregory Green, Julien Guy, Diane Harmer, M. Hernandez, Klaus Honscheid, Lijuan Wendy Huang, David J. James, Buell T. Jannuzi, Linhua Jiang, Richard Joyce, Armin Karcher, Sonia Karkar, Robert Kehoe, Jean-Paul Kneib, Andrea Kueter-Young, Ting-Wen Lan, Tod R. Lauer, Laurent Le Guillou, Auguste Le Van Suu, Jae Hyeon Lee, Michael Lesser, Laurence Perreault Levasseur, Ting S. Li, Justin L. Mann, Robert Marshall, C. E. Martínez-Vázquez, Paul Martini, Hélion du Mas des Bourboux, Sean McManus, Tobias Gabriel Meier, Brice Ménard, Nigel Metcalfe, Andrea Muñoz-Gutiérrez, Joan Najita, Kevin Napier, Gautham Narayan, Jeffrey A. Newman, Jundan Nie, Brian Nord, Dara J. Norman, Knut A. G. Olsen, Anthony Paat, Nathalie Palanque-Delabrouille, Xiyan Peng, Claire L. Poppett, Megan R. Poremba, Abhishek Prakash, David Rabinowitz, Anand Raichoor, Mehdi Rezaie, A. N. Robertson, Natalie A. Roe, Ashley J. Ross, Nicholas P. Ross, Gregory Rudnick, Sasha Safonova, Abhijit Saha, F. Javier Sánchez, Elodie Savary, Heidi Schweiker, Adam Scott, Hee-Jong Seo, Huanyuan Shan, David R. Silva, Zachary Slepian, Christian Soto, David Sprayberry, Ryan Staten, Coley M. Stillman, Robert J. Stupak, David L. Summers, Suk Sien Tie, H. Tirado, Mariana Vargas-Magaña, A. Katherina Vivas, Risa H. Wechsler, Doug Williams, Jinyi Yang, Qian Yang, Tolga Yapici, Dennis Zaritsky, A. Zenteno, Kai Zhang, Tianmeng Zhang, Rongpu Zhou, and Zhimin Zhou. Overview of the DESI Legacy Imaging Surveys. *The Astronomical Journal*, 157(5):168, May 2019.

[72] Eleonora Di Valentino, Olga Mena, Supriya Pan, Luca Visinelli, Weiqiang Yang, Alessandro Melchiorri, David F Mota, Adam G Riess, and Joseph Silk. In the realm of the hubble tension—a review of solutions *. *Classical and Quantum Gravity*, 38(15):153001, July 2021.

[73] Scott Dodelson and Michael D. Schneider. The effect of covariance estimator error on cosmological parameter constraints. *Physical Review D*, 88(6):063537, September 2013.

[74] Fuyu Dong, Yu Yu, Jun Zhang, Xiaohu Yang, and Pengjie Zhang. Measuring the integrated Sachs-Wolfe effect from the low-density regions of the universe. *Monthly Notices of the Royal Astronomical Society*, 500(3):3838–3853, January 2021.

[75] Kelly Douglass, Dahlia Veyrat, Stephen O'Neill, Segev BenZvi, Fatima Zaidouni, and Michaela Guzzetti. VAST: the Void Analysis Software Toolkit. *The Journal of Open Source Software*, 7(77):4033, September 2022.

[76] Kelly A. Douglass, Dahlia Veyrat, and Segev BenZvi. Updated void catalogs of the sdss dr7 main sample. *The Astrophysical Journal Supplement Series*, 265(1):7, feb 2023.

[77] Shiv Ram Dubey, Satish Kumar Singh, and Bidyut Baran Chaudhuri. Activation functions in deep learning: A comprehensive survey and benchmark, 2022.

[78] Dumerchat, Tyann and Bautista, Julian. Galaxy clustering multi-scale emulation. *AA*, 684:A57, 2024.

[79] A. Einstein. The foundation of the general theory of relativity. 1916.

[80] Daniel J. Eisenstein, James Annis, James E. Gunn, Alexander S. Szalay, Andrew J. Connolly, R. C. Nichol, Neta A. Bahcall, Mariangela Bernardi, Scott Burles, Francisco J. Castander, Masataka Fukugita, David W. Hogg, Željko Ivezić, G. R. Knapp, Robert H. Lupton, Vijay Narayanan, Marc Postman, Daniel E. Reichart, Michael Richmond, Donald P. Schneider, David J. Schlegel, Michael A. Strauss, Mark SubbaRao, Douglas L. Tucker, Daniel Vanden Berk, Michael S. Vogeley, David H. Weinberg, and Brian Yanny. Spectroscopic Target Selection for the Sloan Digital Sky Survey: The Luminous Red Galaxy Sample. *The Astronomical Journal*, 122(5):2267–2280, November 2001.

[81] Daniel J. Eisenstein, David H. Weinberg, Eric Agol, Hiroaki Aihara, Carlos Allende Prieto, Scott F. Anderson, James A. Arns, Éric Aubourg, Stephen Bailey, Eduardo Balbinot, Robert Barkhouser, Timothy C. Beers, Andreas A. Berlind, Steven J. Bickerton, Dmitry Bizyaev, Michael R. Blanton, John J. Bochanski, Adam S. Bolton, Casey T. Bosman, Jo Bovy, W. N. Brandt, Ben Breslauer, Howard J. Brewington, J. Brinkmann, Peter J. Brown, Joel R. Brownstein, Dan Burger, Nicolas G. Busca,

Heather Campbell, Phillip A. Cargile, William C. Carithers, Joleen K. Carlberg, Michael A. Carr, Liang Chang, Yanmei Chen, Cristina Chiappini, Johan Comparat, Natalia Connolly, Marina Cortes, Rupert A. C. Croft, Katia Cunha, Luiz N. da Costa, James R. A. Davenport, Kyle Dawson, Nathan De Lee, Gustavo F. Porto de Mello, Fernando de Simoni, Janice Dean, Saurav Dhital, Anne Ealet, Garrett L. Ebelke, Edward M. Edmondson, Jacob M. Eiting, Stephanie Escoffier, Massimiliano Esposito, Michael L. Evans, Xiaohui Fan, Bruno Femenía Castellá, Leticia Dutra Ferreira, Greg Fitzgerald, Scott W. Fleming, Andreu Font-Ribera, Eric B. Ford, Peter M. Frinchaboy, Ana Elia García Pérez, B. Scott Gaudi, Jian Ge, Luan Ghezzi, Bruce A. Gillespie, G. Gilmore, Léo Girardi, J. Richard Gott, Andrew Gould, Eva K. Grebel, James E. Gunn, Jean-Christophe Hamilton, Paul Harding, David W. Harris, Suzanne L. Hawley, Frederick R. Hearty, Joseph F. Hennawi, Jonay I. González Hernández, Shirley Ho, David W. Hogg, Jon A. Holtzman, Klaus Honscheid, Naohisa Inada, Inese I. Ivans, Linhua Jiang, Peng Jiang, Jennifer A. Johnson, Cathy Jordan, Wendell P. Jordan, Guinevere Kauffmann, Eyal Kazin, David Kirkby, Mark A. Klaene, G. R. Knapp, Jean-Paul Kneib, C. S. Kochanek, Lars Koesterke, Juna A. Kollmeier, Richard G. Kron, Hubert Lampeitl, Dustin Lang, James E. Lawler, Jean-Marc Le Goff, Brian L. Lee, Young Sun Lee, Jarron M. Leisenring, Yen-Ting Lin, Jian Liu, Daniel C. Long, Craig P. Loomis, Sara Lucatello, Britt Lundgren, Robert H. Lupton, Bo Ma, Zhibo Ma, Nicholas MacDonald, Claude Mack, Suvrath Mahadevan, Marcio A. G. Maia, Steven R. Majewski, Martin Makler, Elena Malanushenko, Viktor Malanushenko, Rachel Mandelbaum, Claudia Maraston, Daniel Margala, Paul Maseman, Karen L. Masters, Cameron K. McBride, Patrick McDonald, Ian D. McGreer, Richard G. McMahon, Olga Mena Requejo, Brice Ménard, Jordi Miralda-Escudé, Heather L. Morrison, Fergal Mullally, Demitri Muna, Hitoshi Murayama, Adam D. Myers, Tracy Naugle, Angelo Fausti Neto, Duy Cuong Nguyen, Robert C. Nichol, David L. Nidever, Robert W. O'Connell, Ricardo L. C. Ogando, Matthew D. Olmstead, Daniel J. Oravetz, Nikhil Padmanabhan, Martin Paegert, Nathalie Palanque-Delabrouille, Kaike Pan, Parul Pandey, John K. Parejko, Isabelle Pâris, Paulo Pellegrini, Joshua Pepper, Will J. Percival, Patrick Petitjean, Robert Pfaffenberger, Janine Pforr, Stefanie Phleps, Christophe Pichon, Matthew M. Pieri, Francisco Prada, Adrian M. Price-Whelan, M. Jordan Raddick, Beatriz H. F. Ramos, I. Neill Reid, Celine Reyle, James Rich, Gordon T. Richards, George H. Rieke, Marcia J. Rieke, Hans-Walter Rix, Annie C. Robin, Helio J. Rocha-Pinto, Constance M. Rockosi, Natalie A. Roe, Emmanuel Rollinde, Ashley J. Ross, Nicholas P. Ross, Bruno Rossetto, Ariel G. Sánchez, Basilio Santiago, Conor Sayres, Ricardo Schiavon, David J. Schlegel, Katharine J. Schlesinger, Sarah J. Schmidt, Donald P. Schneider, Kris Sellgren, Alaina Shelden, Erin Sheldon, and Matthew Shetrone. SDSS-III: Massive Spec-

troscopic Surveys of the Distant Universe, the Milky Way, and Extra-Solar Planetary Systems. *The Astronomical Journal*, 142(3):72, September 2011.

[82] Hagai El-Ad and Tsvi Piran. Voids in the large-scale structure. *The Astrophysical Journal*, 491(2):421–435, December 1997.

[83] Bridget Falck, Kazuya Koyama, Gong-Bo Zhao, and Marius Cautun. Using voids to unscreen modified gravity. *Monthly Notices of the Royal Astronomical Society*, 475(3):3262–3272, April 2018.

[84] Nathan Findlay and Seshadri Nadathur. Versus: An excursion-set-inspired void-finder for the stage-iv era, 2026.

[85] D. J. Fixsen. The temperature of the cosmic microwave background. *The Astrophysical Journal*, 707(2):916–920, November 2009.

[86] Tristan S. Fraser, Enrique Paillas, Will J. Percival, Seshadri Nadathur, Slaana Radinović, and Hans A. Winther. Modelling the boss void-galaxy cross-correlation function using a neural-network emulator, 2024.

[87] A. Friedmann. Über die Krümmung des Raumes. *Zeitschrift fur Physik*, 10:377–386, January 1922.

[88] A. Friedmann. Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes. *Zeitschrift fur Physik*, 21(1):326–332, December 1924.

[89] J. N. Fry. The Galaxy correlation hierarchy in perturbation theory. *The Astrophysical Journal*, 279:499–510, April 1984.

[90] J. N. Fry and P. J. E. Peebles. Statistical analysis of catalogs of extragalactic objects. IX. The four-point galaxy correlation function. *The Astrophysical Journal*, 221:19–33, April 1978.

[91] G. Gamow. Expanding Universe and the Origin of Elements. *Physical Review*, 70(7-8):572–573, October 1946.

[92] Lehman H. Garrison, Daniel J. Eisenstein, Douglas Ferrer, Nina A. Maksimova, and Philip A. Pinto. The ABACUS cosmological N-body code. *Monthly Notices of the Royal Astronomical Society*, 508(1):575–596, November 2021.

[93] Payam Ghafari, Mahdi Najafi, Mina Ghodsi Yengejeh, Emre Özülker, Eleonora Di Valentino, and Javad T. Firouzjaee. A Multi-Probe ISW Study of Dark Energy Models with Negative Energy Density: Galaxy Correlations, Lensing Bispectrum, and Planck ISW-Lensing Likelihood. *arXiv e-prints*, page arXiv:2512.07060, December 2025.

[94] Tommaso Giannantonio, Ryan Scranton, Robert G. Crittenden, Robert C. Nichol, Stephen P. Boughn, Adam D. Myers, and Gordon T. Richards. Combined analysis of the integrated Sachs-Wolfe effect and cosmological implications. *Physical Review D*, 77(12):123520, June 2008.

[95] Héctor Gil-Marín, Julián E. Bautista, Romain Paviot, Mariana Vargas-Magaña, Sylvain de la Torre, Sebastien Fromenteau, Shadab Alam, Santiago Ávila, Etienne Burtin, Chia-Hsun Chuang, Kyle S. Dawson, Jiamin Hou, Arnaud de Mattia, Faizan G. Mohammad, Eva-Maria Müller, Seshadri Nadathur, Richard Neveux, Will J. Percival, Anand Raichoor, Mehdi Rezaie, Ashley J. Ross, Graziano Rossi, Vanina Ruhlmann-Kleider, Alex Smith, Amélie Tamone, Jeremy L. Tinker, Rita Tojeiro, Yuting Wang, Gong-Bo Zhao, Cheng Zhao, Jonathan Brinkmann, Joel R. Brownstein, Peter D. Choi, Stephanie Escoffier, Axel de la Macorra, Jeongin Moon, Jeffrey A. Newman, Donald P. Schneider, Hee-Jong Seo, and Mariappan Vivek. The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the BAO and growth rate of structure of the luminous red galaxy sample from the anisotropic power spectrum between redshifts 0.6 and 1.0. *Monthly Notices of the Royal Astronomical Society*, 498(2):2492–2531, October 2020.

[96] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[97] Jochen Görtler, Rebecca Kehlbeck, and Oliver Deussen. A visual exploration of gaussian processes. *Distill*, 2019.

[98] S. A. Gregory and L. A. Thompson. The Coma/A1367 supercluster and its environs. *The Astrophysical Journal*, 222:784–799, June 1978.

[99] Alternative Clustering Methods Topical Group. https://github.com/epaillas/acm, title = Alterative Clustering Methods (ACM) Pipeline, year = 2024-2026.

[100] D. Gruen, O. Friedrich, A. Amara, D. Bacon, C. Bonnett, W. Hartley, B. Jain, M. Jarvis, T. Kacprzak, E. Krause, A. Mana, E. Rozo, E. S. Rykoff, S. Seitz, E. Sheldon, M. A. Troxel, V. Vikram, T. M. C. Abbott, F. B. Abdalla, S. Allam, R. Armstrong, M. Banerji, A. H. Bauer, M. R. Becker, A. Benoit-Lévy, G. M. Bernstein,

R. A. Bernstein, E. Bertin, S. L. Bridle, D. Brooks, E. Buckley-Geer, D. L. Burke, D. Capozzi, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, M. Crocce, C. E. Cunha, C. B. D'Andrea, L. N. da Costa, D. L. DePoy, S. Desai, H. T. Diehl, J. P. Dietrich, P. Doel, T. F. Eifler, A. Fausti Neto, E. Fernandez, B. Flaugher, P. Fosalba, J. Frieman, D. W. Gerdes, R. A. Gruendl, G. Gutierrez, K. Honscheid, D. J. James, K. Kuehn, N. Kuropatkin, O. Lahav, T. S. Li, M. Lima, M. A. G. Maia, M. March, P. Martini, P. Melchior, C. J. Miller, R. Miquel, J. J. Mohr, B. Nord, R. Ogando, A. A. Plazas, K. Reil, A. K. Romer, A. Roodman, M. Sako, E. Sanchez, V. Scarpine, M. Schubnell, I. Sevilla-Noarbe, R. C. Smith, M. Soares-Santos, F. Sobreira, E. Suchyta, M. E. C. Swanson, G. Tarle, J. Thaler, D. Thomas, A. R. Walker, R. H. Wechsler, J. Weller, Y. Zhang, and J. Zuntz. Weak lensing by galaxy troughs in des science verification data. *Monthly Notices of the Royal Astronomical Society*, 455(3):3367–3380, 01 2016.

[101] Davide Gualdi, Héctor Gil-Marín, Robert L Schuhmann, Marc Manera, Benjamin Joachimi, and Ofer Lahav. Enhancing boss bispectrum cosmological constraints with maximal compression. *Monthly Notices of the Royal Astronomical Society*, 484(3):3713–3730, January 2019.

[102] Giulia Gubitosi, Federico Piazza, and Filippo Vernizzi. The effective field theory of dark energy. *Journal of Cosmology and Astroparticle Physics*, 2013(2):032, February 2013.

[103] Alan H. Guth. Inflationary universe: A possible solution to the horizon and flatness problems. *Physical Review D*, 23(2):347–356, January 1981.

[104] L. Guzzo, M. Pierleoni, B. Meneux, E. Branchini, O. Le Fèvre, C. Marinoni, B. Garilli, J. Blaizot, G. De Lucia, A. Pollo, H. J. McCracken, D. Bottini, V. Le Brun, D. Maccagni, J. P. Picat, R. Scaramella, M. Scodeggio, L. Tresse, G. Vettolani, A. Zanichelli, C. Adami, S. Arnouts, S. Bardelli, M. Bolzonella, A. Bongiorno, A. Cappi, S. Charlot, P. Ciliegi, T. Contini, O. Cucciati, S. de la Torre, K. Dolag, S. Foucaud, P. Franzetti, I. Gavignaud, O. Ilbert, A. Iovino, F. Lamareille, B. Marano, A. Mazure, P. Memeo, R. Merighi, L. Moscardini, S. Paltani, R. Pellò, E. Perez-Montero, L. Pozzetti, M. Radovich, D. Vergani, G. Zamorani, and E. Zucca. A test of the nature of cosmic acceleration using galaxy redshift distortions. *Nature*, 451(7178):541–544, January 2008.

[105] Salman Habib, Katrin Heitmann, David Higdon, Charles Nakhleh, and Brian Williams. Cosmic calibration: Constraints from the matter power spectrum and the cosmic microwave background. *Physical Review D*, 76(8), October 2007.

[106] Boryana Hadzhiyska, Daniel Eisenstein, Sownak Bose, Lehman H. Garrison, and Nina Maksimova. compaso: A new halo finder for competitive assignment to spherical overdensities. *Mon. Not. Roy. Astron. Soc.*, 509(1):501–521, 2021.

[107] Boryana Hadzhiyska, Lehman H. Garrison, Daniel Eisenstein, and Sownak Bose. The halo light-cone catalogues of ABACUSSUMMIT. *Monthly Notices of the Royal Astronomical Society*, 509(2):2194–2208, January 2022.

[108] ChangHoon Hahn, Roman Scoccimarro, Michael R. Blanton, Jeremy L. Tinker, and Sergio A. Rodríguez-Torres. The effect of fibre collisions on the galaxy power spectrum multipoles. *Mon. Not. Roy. Astron. Soc.*, 467(2):1940–1956, 2017.

[109] ChangHoon Hahn, Mohammadjavad Vakili, Kilian Walsh, Andrew P. Hearin, David W. Hogg, and Duncan Campbell. Approximate Bayesian computation in large-scale structure: constraining the galaxy-halo connection. *Monthly Notices of the Royal Astronomical Society*, 469(3):2791–2805, August 2017.

[110] N. Hamaus, M. Aubert, A. Pisani, S. Contarini, G. Verza, M.-C. Cousinou, S. Escoffier, A. Hawken, G. Lavaux, G. Pollina, B. D. Wandelt, J. Weller, M. Bonici, C. Carbone, L. Guzzo, A. Kovacs, F. Marulli, E. Massara, L. Moscardini, P. Ntelis, W. J. Percival, S. Radinović, M. Sahlén, Z. Sakr, A. G. Sánchez, H. A. Winther, N. Auricchio, S. Awan, R. Bender, C. Bodendorf, D. Bonino, E. Branchini, M. Brescia, J. Brinchmann, V. Capobianco, J. Carretero, F. J. Castander, M. Castellano, S. Cavuoti, A. Cimatti, R. Cledassou, G. Congedo, L. Conversi, Y. Copin, L. Corcione, M. Cropper, A. Da Silva, H. Degaudenzi, M. Douspis, F. Dubath, C. A. J. Duncan, X. Dupac, S. Dusini, A. Ealet, S. Ferriol, P. Fosalba, M. Frailis, E. Franceschi, P. Franzetti, M. Fumana, B. Garilli, B. Gillis, C. Giocoli, A. Grazian, F. Grupp, S. V. H. Haugan, W. Holmes, F. Hormuth, K. Jahnke, S. Kermiche, A. Kiessling, M. Kilbinger, T. Kitching, M. Kümmel, M. Kunz, H. Kurki-Suonio, S. Ligori, P. B. Lilje, I. Lloro, E. Maiorano, O. Marggraf, K. Markovic, R. Massey, S. Maurogordato, M. Melchior, M. Meneghetti, G. Meylan, M. Moresco, E. Munari, S. M. Niemi, C. Padilla, S. Paltani, F. Pasian, K. Pedersen, V. Pettorino, S. Pires, M. Poncet, L. Popa, L. Pozzetti, R. Rebolo, J. Rhodes, H. Rix, M. Roncarelli, E. Rossetti, R. Saglia, P. Schneider, A. Secroun, G. Seidel, S. Serrano, C. Sirignano, G. Sirri, J.-L. Starck, P. Tallada-Crespí, D. Tavagnacco, A. N. Taylor, I. Tereno, R. Toledo-Moreo, F. Torradeflot, E. A. Valentijn, L. Valenziano, Y. Wang, N. Welikala, G. Zamorani, J. Zoubian, S. Andreon, M. Baldi, S. Camera, S. Mei, C. Neissner, and E. Romelli. Euclid: Forecasts from redshift-space distortions and the alcock–paczynski test with cosmic voids. *Astronomy amp; Astrophysics*, 658:A20, January 2022.

[111] Nico Hamaus, Marie-Claude Cousinou, Alice Pisani, Marie Aubert, Stéphanie Escoffier, and Jochen Weller. Multipole analysis of redshift-space distortions around cosmic voids. *Journal of Cosmology and Astroparticle Physics*, 2017(07):014–014, 2017.

[112] Nico Hamaus, Alice Pisani, Jin-Ah Choi, Guilhem Lavaux, Benjamin D. Wandelt, and Jochen Weller. Precision cosmology with voids in the final boss data. *Journal of Cosmology and Astroparticle Physics*, 2020(12):023–023, December 2020.

[113] Nico Hamaus, Alice Pisani, Paul M Sutter, Guilhem Lavaux, Stéphanie Escoffier, Benjamin D Wandelt, and Jochen Weller. Constraints on cosmology and gravity from the dynamics of voids. *Physical Review Letters*, 117(9):091302, 2016.

[114] Nico Hamaus, P. M. Sutter, and Benjamin D. Wandelt. Universal density profile for cosmic voids. *Phys. Rev. Lett.*, 112:251302, Jun 2014.

[115] Nico Hamaus, P.M. Sutter, Guilhem Lavaux, and Benjamin D. Wandelt. Probing cosmology and gravity with redshift-space distortions around voids. *Journal of Cosmology and Astroparticle Physics*, 2015(11):036–036, November 2015.

[116] Nico Hamaus, Benjamin D. Wandelt, P. M. Sutter, Guilhem Lavaux, and Michael S. Warren. Cosmology with Void-Galaxy Correlations. *Physical Review Letters*, 112(4):041304, January 2014.

[117] A. J. S. Hamilton. Toward Better Ways to Measure the Galaxy Correlation Function. *The Astrophysical Journal*, 417:19, November 1993.

[118] A. J. S. Hamilton. *Linear Redshift Distortions: A Review*, pages 185–275. Springer Netherlands, 1998.

[119] Katrin Heitmann, David Higdon, Martin White, Salman Habib, Brian J. Williams, Earl Lawrence, and Christian Wagner. The coyote universe. ii. cosmological models and precision emulation of the nonlinear matter power spectrum. *The Astrophysical Journal*, 705(1):156–174, October 2009.

[120] Shirley Ho, Christopher Hirata, Nikhil Padmanabhan, Uros Seljak, and Neta Bahcall. Correlation of CMB with large-scale structure. I. Integrated Sachs-Wolfe tomography and cosmological implications. *Physical Review D*, 78(4):043519, August 2008.

[121] David W. Hogg. Distance measures in cosmology, 2000.

[122] Z. Hou, C. L. Reichardt, K. T. Story, B. Follin, R. Keisler, K. A. Aird, B. A. Benson, L. E. Bleem, J. E. Carlstrom, C. L. Chang, H.-M. Cho, T. M. Crawford, A. T. Crites, T. de Haan, R. de Putter, M. A. Dobbs, S. Dodelson, J. Dudley, E. M. George, N. W. Halverson, G. P. Holder, W. L. Holzapfel, S. Hoover, J. D. Hrubes, M. Joy, L. Knox, A. T. Lee, E. M. Leitch, M. Lueker, D. Luong-Van, J. J. McMahon, J. Mehl, S. S. Meyer, M. Millea, J. J. Mohr, T. E. Montroy, S. Padin, T. Plagge, C. Pryke, J. E. Ruhl, J. T. Sayre, K. K. Schaffer, L. Shaw, E. Shirokoff, H. G. Spieler, Z. Staniszewski, A. A. Stark, A. van Engelen, K. Vanderlinde, J. D. Vieira, R. Williamson, and O. Zahn. Constraints on Cosmology from the Cosmic Microwave Background Power Spectrum of the 2500 deg^2 SPT-SZ Survey. *The Astrophysical Journal*, 782(2):74, February 2014.

[123] Cullan Howlett and Will J. Percival. Galaxy two-point covariance matrix estimation for next generation surveys. *Monthly Notices of the Royal Astronomical Society*, 472(4):4935–4952, December 2017.

[124] Fiona Hoyle and Michael S. Vogeley. Voids in the point source catalogue survey and the updated zwicky catalog. *The Astrophysical Journal*, 566(2):641–651, February 2002.

[125] Edwin Hubble. A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15(3):168–173, March 1929.

[126] Mike Hudson. The s8 tension in cosmology, may 2023. PIRSA:23050129 see, https://pirsa.org.

[127] Dragan Huterer, Carlos E. Cunha, and Wenjuan Fang. Calibration errors unleashed: effects on cosmological parameters and requirements for large-scale structure surveys. *Monthly Notices of the Royal Astronomical Society*, 432(4):2945–2961, 07 2013.

[128] E. E. O. Ishida, S. D. P. Vitenti, M. Penna-Lima, J. Cisewski, R. S. de Souza, A. M. M. Trindade, E. Cameron, V. C. Busti, and COIN Collaboration. COSMOABC: Likelihood-free inference via Population Monte Carlo Approximate Bayesian Computation. *Astronomy and Computing*, 13:1–11, November 2015.

[129] Mikhel Jõeveer, Jaan Einasto, and Erik Tago. Spatial distribution of galaxies and of clusters of galaxies in the southern galactic hemisphere. *Monthly Notices of the Royal Astronomical Society*, 185:357–370, November 1978.

[130] D. Heath Jones, Will Saunders, Matthew Colless, Mike A. Read, Quentin A. Parker, Fred G. Watson, Lachlan A. Campbell, Daniel Burkey, Thomas Mauch, Lesa Moore, Malcolm Hartley, Paul Cass, Dionne James, Ken Russell, Kristin Fiegert, John Dawe, John Huchra, Tom Jarrett, Ofer Lahav, John Lucey, Gary A. Mamon, Dominique Proust, Elaine M. Sadler, and Ken-ichi Wakamatsu. The 6dF Galaxy Survey: samples, observational techniques and the first data release. *Monthly Notices of the Royal Astronomical Society*, 355(3):747–763, December 2004.

[131] Nick Kaiser. Clustering in real space and in redshift space. *Monthly Notices of the Royal Astronomical Society*, 227:1–21, July 1987.

[132] N. Katz, T. Quinn, and J. M. Gelb. Galaxy Formation and the Peaks Formalism. *Monthly Notices of the Royal Astronomical Society*, 265:689, December 1993.

[133] F. S. Kitaura, G. Yepes, and F. Prada. Modelling baryon acoustic oscillations with perturbation theory and stochastic halo biasing. *Monthly Notices of the Royal Astronomical Society*, 439:L21–L25, March 2014.

[134] Francisco-Shu Kitaura, Sergio Rodríguez-Torres, Chia-Hsun Chuang, Cheng Zhao, Francisco Prada, Héctor Gil-Marín, Hong Guo, Gustavo Yepes, Anatoly Klypin, Claudia G. Scóccola, Jeremy Tinker, Cameron McBride, Beth Reid, Ariel G. Sánchez, Salvador Salazar-Albornoz, Jan Niklas Grieb, Mariana Vargas-Magana, Antonio J. Cuesta, Mark Neyrinck, Florian Beutler, Johan Comparat, Will J. Percival, and Ashley Ross. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: mock galaxy catalogues for the BOSS Final Data Release. *Monthly Notices of the Royal Astronomical Society*, 456(4):4156–4173, March 2016.

[135] Arthur Kosowsky and Michael S. Turner. Cbr anisotropy and the running of the scalar spectral index. *Phys. Rev. D*, 52:R1739(R)–R1743(R), Aug 1995.

[136] András Kovács. The part and the whole: voids, supervoids, and their ISW imprint. *Monthly Notices of the Royal Astronomical Society*, 475(2):1777–1790, April 2018.

[137] Christina D. Kreisch, Alice Pisani, Carmelita Carbone, Jia Liu, Adam J. Hawken, Elena Massara, David N. Spergel, and Benjamin D. Wandelt. Massive neutrinos leave fingerprints on cosmic voids. *Monthly Notices of the Royal Astronomical Society*, 488(3):4413–4426, September 2019.

[138] Vladimír Kunc and Jiří Kléma. Three decades of activations: A comprehensive survey of 400 activation functions for neural networks, 2024.

[139] Juliana Kwan, Katrin Heitmann, Salman Habib, Nikhil Padmanabhan, Earl Lawrence, Hal Finkel, Nicholas Frontiere, Pope, and Adrian. Cosmic emulation: Fast predictions for the galaxy power spectrum. *The Astrophysical Journal*, 810(1):35, aug 2015.

[140] Y. Lai, C. Howlett, J. Aguilar, S. Ahlen, A.J. Amsellem, J. Bautista, S. BenZvi, D. Bianchi, C. Blake, D. Brooks, A. Carr, T. Claybaugh, T.M. Davis, A. de la Macorra, P. Doel, K. Douglass, S. Ferraro, A. Font-Ribera, J.E. Forero-Romero, E. Gaztañaga, G. Gutierrez, J. Guy, H.K. Herrera-Alcantar, D. Huterer, M. Ishak, R. Joyce, A. Kim, D. Kirkby, T. Kisner, A. Kremin, O. Lahav, C. Lamman, M. Landriau, L. Le Guillou, A. Leauthaud, M.E. Levi, M. Manera, P. Martini, A. Meisner, R. Miquel, J. Moustakas, A. Muñoz-Gutiérrez, S. Nadathur, W.J. Percival, C. Poppett, F. Prada, I. Pérez-Ràfols, F. Qin, C. Ross, G. Rossi, K. Said, E. Sanchez, D. Schlegel, M. Schubnell, H. Seo, J. Silber, D. Sprayberry, G. Tarlé, R. Turner, B.A. Weaver, P. Zarrouk, R. Zhou, and H. Zou. The desi dr1 peculiar velocity survey: growth rate measurements from the maximum likelihood fields method. *Journal of Cosmology and Astroparticle Physics*, 2026(04):026, April 2026.

[141] Stephen D. Landy and Alexander S. Szalay. Bias and Variance of Angular Correlation Functions. *The Astrophysical Journal*, 412:64, July 1993.

[142] R. Laureijs, J. Amiaux, S. Arduini, J. L. Augueres, J. Brinchmann, R. Cole, M. Cropper, C. Dabin, L. Duvet, A. Ealet, B. Garilli, P. Gondoin, L. Guzzo, J. Hoar, H. Hoekstra, R. Holmes, T. Kitching, T. Maciaszek, Y. Mellier, F. Pasian, W. Percival, J. Rhodes, G. Saavedra Criado, M. Sauvage, R. Scaramella, L. Valenziano, S. Warren, R. Bender, F. Castander, A. Cimatti, O. Le Fèvre, H. Kurki-Suonio, M. Levi, P. Lilje, G. Meylan, R. Nichol, K. Pedersen, V. Popa, R. Rebolo Lopez, H. W. Rix, H. Rottgering, W. Zeilinger, F. Grupp, P. Hudelot, R. Massey, M. Meneghetti, L. Miller, S. Paltani, S. Paulin-Henriksson, S. Pires, C. Saxton, T. Schrabback, G. Seidel, J. Walsh, N. Aghanim, L. Amendola, J. Bartlett, C. Baccigalupi, J. P. Beaulieu, K. Benabed, J. G. Cuby, D. Elbaz, P. Fosalba, G. Gavazzi, A. Helmi, I. Hook, M. Irwin, J. P. Kneib, M. Kunz, F. Mannucci, L. Moscardini, C. Tao, R. Teyssier, J. Weller, G. Zamorani, M. R. Zapatero Osorio, O. Boulade, J. J. Foumond, A. Di Giorgio, P. Guttridge, A. James, M. Kemp, J. Martignac, A. Spencer, D. Walton, T. Blümchen, C. Bonoli, F. Bortoletto, C. Cerna, L. Corcione, C. Fabron, K. Jahnke, S. Ligori, F. Madrid, L. Martin, G. Morgante, T. Pamplona, E. Prieto, M. Riva, R. Toledo, M. Trifoglio, F. Zerbi, F. Abdalla, M. Douspis, C. Grenet, S. Borgani, R. Bouwens, F. Courbin, J. M. Delouis, P. Dubath, A. Fontana, M. Frailis, A. Grazian, J. Koppenhöfer, O. Mansutti, M. Melchior, M. Mignoli, J. Mohr,

C. Neissner, K. Noddle, M. Poncet, M. Scodeggio, S. Serrano, N. Shane, J. L. Starck, C. Surace, A. Taylor, G. Verdoes-Kleijn, C. Vuerli, O. R. Williams, A. Zacchei, B. Altieri, I. Escudero Sanz, R. Kohley, T. Oosterbroek, P. Astier, D. Bacon, S. Bardelli, C. Baugh, F. Bellagamba, C. Benoist, D. Bianchi, A. Biviano, E. Branchini, C. Carbone, V. Cardone, D. Clements, S. Colombi, C. Conselice, G. Cresci, N. Deacon, J. Dunlop, C. Fedeli, F. Fontanot, P. Franzetti, C. Giocoli, J. Garcia-Bellido, J. Gow, A. Heavens, P. Hewett, C. Heymans, A. Holland, Z. Huang, O. Ilbert, B. Joachimi, E. Jennins, E. Kerins, A. Kiessling, D. Kirk, R. Kotak, O. Krause, O. Lahav, F. van Leeuwen, J. Lesgourgues, M. Lombardi, M. Magliocchetti, K. Maguire, E. Majerotto, R. Maoli, F. Marulli, S. Maurogordato, H. McCracken, R. McLure, A. Melchiorri, A. Merson, M. Moresco, M. Nonino, P. Norberg, J. Peacock, R. Pello, M. Penny, V. Pettorino, C. Di Porto, L. Pozzetti, C. Quercellini, M. Radovich, A. Rassat, N. Roche, S. Ronayette, E. Rossetti, B. Sartoris, P. Schneider, E. Semboloni, S. Serjeant, F. Simpson, C. Skordis, G. Smadja, S. Smartt, P. Spano, S. Spiro, M. Sullivan, A. Tilquin, R. Trotta, L. Verde, Y. Wang, G. Williger, G. Zhao, J. Zoubian, and E. Zucca. Euclid Definition Study Report. *arXiv e-prints*, page arXiv:1110.3193, October 2011.

[143] Guilhem Lavaux and Benjamin D. Wandelt. Precision cosmology with voids: definition, methods, dynamics. *Monthly Notices of the Royal Astronomical Society*, 403(3):1392–1408, April 2010.

[144] Guilhem Lavaux and Benjamin D. Wandelt. Precision cosmography with stacked voids. *The Astrophysical Journal*, 754(2):109, July 2012.

[145] Jounghun Lee and Daeseong Park. Constraining the Dark Energy Equation of State with Cosmic Voids. *The Astrophysical Journal Letters*, 696(1):L10–L12, May 2009.

[146] Georges Lemaitre. A Homogeneous Universe of Constant Mass and Growing Radius Accounting for the Radial Velocity of Extragalactic Nebulae. *Annales Soc. Sci. Bruxelles A*, 47:49–59, 1927.

[147] M. Francis. "b" is for baryon acoustic oscillations (alphabet of cosmology), 2013. [Online; accessed November 1, 2026].

[148] David JC MacKay. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

[149] Nina A. Maksimova, Lehman H. Garrison, Daniel J. Eisenstein, Boryana Hadzhiyska, Sownak Bose, and Thomas P. Satterthwaite. ABACUSSUMMIT: a massive set of

high-accuracy, high-resolution N-body simulations. *Monthly Notices of the Royal Astronomical Society*, 508(3):4017–4037, December 2021.

[150] Elena Massara and Ravi K. Sheth. Density and velocity profiles around cosmic voids. *arXiv e-prints*, page arXiv:1811.03132, November 2018.

[151] Peter Melchior, P. M. Sutter, Erin S. Sheldon, Elisabeth Krause, and Benjamin D. Wandelt. First measurement of gravitational lensing by cosmic voids in SDSS. *Monthly Notices of the Royal Astronomical Society*, 440(4):2922–2927, June 2014.

[152] Christopher B. Morrison and Michael D. Schneider. On estimating cosmology-dependent covariance matrices. *Journal of Cosmology and Astroparticle Physics*, 2013(11):009, November 2013.

[153] Adam D. Myers, John Moustakas, Stephen Bailey, Benjamin A. Weaver, Andrew P. Cooper, Jaime E. Forero-Romero, Bela Abolfathi, David M. Alexander, David Brooks, Edmond Chaussidon, Chia-Hsun Chuang, Kyle Dawson, Arjun Dey, Biprateep Dey, Govinda Dhungana, Peter Doel, Kevin Fanning, Enrique Gaztañaga, Satya Gontcho A Gontcho, Alma X. Gonzalez-Morales, ChangHoon Hahn, Hiram K. Herrera-Alcantar, Klaus Honscheid, Mustapha Ishak, Tanveer Karim, David Kirkby, Theodore Kisner, Sergey E. Koposov, Anthony Kremin, Ting-Wen Lan, Martin Landriau, Dustin Lang, Michael E. Levi, Christophe Magneville, Lucas Napolitano, Paul Martini, Aaron Meisner, Jeffrey A. Newman, Nathalie Palanque-Delabrouille, Will Percival, Claire Poppett, Francisco Prada, Anand Raichoor, Ashley J. Ross, Edward F. Schlafly, David Schlegel, Michael Schubnell, Ting Tan, Gregory Tarle, Michael J. Wilson, Christophe Yèche, Rongpu Zhou, Zhimin Zhou, and Hu Zou. The target-selection pipeline for the dark energy spectroscopic instrument. *The Astronomical Journal*, 165(2):50, January 2023.

[154] Seshadri Nadathur, Paul Carter, and Will J Percival. A zeldovich reconstruction method for measuring redshift space distortions using cosmic voids. *Monthly Notices of the Royal Astronomical Society*, 482(2):2459–2470, 01 2019.

[155] Seshadri Nadathur, Paul M. Carter, Will J. Percival, Hans A. Winther, and Julian E. Bautista. Beyond bao: Improving cosmological constraints from boss data with measurement of the void-galaxy cross-correlation. *Physical Review D*, 100(2), July 2019.

[156] Seshadri Nadathur, Shaun Hotchkiss, and Robert Crittenden. Tracing the gravitational potential using cosmic voids. *Monthly Notices of the Royal Astronomical Society*, 467(4):4067–4079, 06 2017.

[157] Seshadri Nadathur and Will J Percival. An accurate linear model for redshift space distortions in the void–galaxy correlation function. *Monthly Notices of the Royal Astronomical Society*, 483(3):3472–3487, 03 2019.

[158] Seshadri Nadathur, Alex Woodfinden, Will J. Percival, Marie Aubert, Julian Bautista, Kyle Dawson, Stéphanie Escoffier, Sebastien Fromenteau, Héctor Gil-Marín, James Rich, Ashley J. Ross, Graziano Rossi, Mariana Vargas Magaña, Joel R. Brownstein, and Donald P. Schneider. The completed SDSS-IV extended baryon oscillation spectroscopic survey: geometry and growth from the anisotropic void-galaxy correlation function in the luminous red galaxy sample. *Monthly Notices of the Royal Astronomical Society*, 499(3):4140–4157, December 2020.

[159] Mark C. Neyrinck. zobov: a parameter-free void-finding algorithm. *Monthly Notices of the Royal Astronomical Society*, 386(4):2101–2109, June 2008.

[160] Luis E. Padilla, Luis O. Tellez, Luis A. Escamilla, and Jose Alberto Vazquez. Cosmological parameter inference with bayesian statistics. *Universe*, 7(7):213, 2021.

[161] N. D. Padilla, L. Ceccarelli, and D. G. Lambas. Spatial and dynamical properties of voids in a cold dark matter universe. *Monthly Notices of the Royal Astronomical Society*, 363(3):977–990, 2005.

[162] Enrique Paillas, Marius Cautun, Baojiu Li, Yan-Chuan Cai, Nelson Padilla, Joaquín Armijo, and Sownak Bose. The santiago–harvard–edinburgh–durham void comparison ii: unveiling the vainshtein screening using weak lensing. *Monthly Notices of the Royal Astronomical Society*, 484(1):1149–1165, January 2019.

[163] Enrique Paillas, Carolina Cuesta-Lazaro, Will J. Percival, Seshadri Nadathur, Yan-Chuan Cai, Sihan Yuan, Florian Beutler, Arnaud de Mattia, Daniel Eisenstein, Daniel Forero-Sanchez, Nelson Padilla, Mathilde Pinon, Vanina Ruhlmann-Kleider, Ariel G. Sánchez, Georgios Valogiannis, and Pauline Zarrouk. Cosmological constraints from density-split clustering in the boss cmass galaxy sample, 2023.

[164] Enrique Paillas, Carolina Cuesta-Lazaro, Will J. Percival, Seshadri Nadathur, Yan-Chuan Cai, Sihan Yuan, Florian Beutler, Arnaud de Mattia, Daniel J. Eisenstein, Daniel Forero-Sanchez, Nelson Padilla, Mathilde Pinon, Vanina Ruhlmann-Kleider, Ariel G. Sánchez, Georgios Valogiannis, and Pauline Zarrouk. Cosmological constraints from density-split clustering in the BOSS CMASS galaxy sample. *Monthly Notices of the Royal Astronomical Society*, 531(1):898–918, June 2024.

[165] Daeseong Park and Jounghun Lee. Void ellipticity distribution as a probe of cosmology. *Physical Review Letters*, 98(8), February 2007.

[166] Dante J Paz, Carlos M Correa, Sebastiín R Gualpa, Andres N Ruiz, Carlos S Bederián, R Dario Graña, and Nelson D Padilla. Guess the cheese flavour by the size of its holes: a cosmological test using the abundance of popcorn voids. *Monthly Notices of the Royal Astronomical Society*, 522(2):2553–2569, April 2023.

[167] J. A. Peacock and S. J. Dodds. Non-linear evolution of cosmological power spectra. *Monthly Notices of the Royal Astronomical Society*, 280(3):L19–L26, June 1996.

[168] P. J. E. Peebles. Primordial Helium Abundance and the Primordial Fireball. II. *The Astrophysical Journal*, 146:542, November 1966.

[169] P. J. E. Peebles. *The large-scale structure of the universe*. 1980.

[170] P. J. E. Peebles. *Principles of Physical Cosmology*. 1993.

[171] P. J. E. Peebles and Edward Harrison. Principles of Physical Cosmology. *American Journal of Physics*, 62(4):381–381, April 1994.

[172] A. A. Penzias and R. W. Wilson. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *The Astrophysical Journal*, 142:419–421, July 1965.

[173] Will J Percival, Oliver Friedrich, Elena Sellentin, and Alan Heavens. Matching bayesian and frequentist coverage probabilities when using an approximate data covariance matrix. *Monthly Notices of the Royal Astronomical Society*, 510(3):3207–3221, December 2021.

[174] Will J. Percival, Will Sutherland, John A. Peacock, Carlton M. Baugh, Joss Bland-Hawthorn, Terry Bridges, Russell Cannon, Shaun Cole, Matthew Colless, Chris Collins, Warrick Couch, Gavin Dalton, Roberto De Propris, Simon P. Driver, George Efstathiou, Richard S. Ellis, Carlos S. Frenk, Karl Glazebrook, Carole Jackson, Ofer Lahav, Ian Lewis, Stuart Lumsden, Steve Maddox, Stephen Moody, Peder Norberg, Bruce A. Peterson, and Keith Taylor. Parameter constraints for flat cosmologies from cosmic microwave background and 2dFGRS power spectra. *Monthly Notices of the Royal Astronomical Society*, 337(3):1068–1080, December 2002.

[175] Will J. Percival and Martin White. Testing cosmological structure formation using redshift-space distortions. *Monthly Notices of the Royal Astronomical Society*, 393(1):297–308, February 2009.

[176] Eder L. D. Perico, Rodrigo Voivodic, Marcos Lima, and David F. Mota. Cosmic voids in modified gravity scenarios. *Astronomy & Astrophysics*, 632:A52, December 2019.

[177] Oliver H E Philcox, Daniel J Eisenstein, Ross O'Connell, and Alexander Wiegand. ¡scp¿rascalc¡/scp¿: a jackknife approach to estimating single- and multitracer galaxy covariance matrices. *Monthly Notices of the Royal Astronomical Society*, 491(3):3290–3317, November 2019.

[178] Oliver H. E. Philcox, Jiamin Hou, and Zachary Slepian. A first detection of the connected 4-point correlation function of galaxies using the boss cmass sample, 2021.

[179] Oliver H.E. Philcox and Mikhail M. Ivanov. Boss dr12 full-shape cosmology: ¡mml:math xmlns:mml="http://www.w3.org/1998/math/mathml" display="inline"¿¡mml:mrow¿¡mml:mi mathvariant="normal"¿¡/mml:mi¿¡mml:mi¿cdm¡/mml:mi¿¡/m constraints from the large-scale galaxy power spectrum and bispectrum monopole. *Physical Review D*, 105(4), February 2022.

[180] Alice Pisani, Elena Massara, David N. Spergel, David Alonso, Tessa Baker, Yan-Chuan Cai, Marius Cautun, Christopher Davies, Vasiliy Demchenko, Olivier Doré, Andy Goulding, Mélanie Habouzit, Nico Hamaus, Adam Hawken, Christopher M. Hirata, Shirley Ho, Bhuvnesh Jain, Christina D. Kreisch, Federico Marulli, Nelson Padilla, Giorgia Pollina, Martin Sahlén, Ravi K. Sheth, Rachel Somerville, Istvan Szapudi, Rien van de Weygaert, Francisco Villaescusa-Navarro, Benjamin D. Wandelt, and Yun Wang. Cosmic voids: a novel probe to shed light on our universe, 2019.

[181] Alice Pisani, P. M. Sutter, Nico Hamaus, Esfandiar Alizadeh, Rahul Biswas, Benjamin D. Wandelt, and Christopher M. Hirata. Counting voids to probe dark energy. *Phys. Rev. D*, 92:083531, Oct 2015.

[182] Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed, J.-P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J.-F. Cardoso, J. Carron, A. Challinor, H. C. Chiang, J. Chluba, L. P. L. Colombo, C. Combet, D. Contreras, B. P. Crill, F. Cuttaia, P. de Bernardis, G. de Zotti, J. Delabrouille, J.-M. Delouis, E. Di Valentino, J. M. Diego, O. Doré, M. Douspis, A. Ducout, X. Dupac, S. Dusini, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, Y. Fantaye, M. Farhang, J. Fergusson, R. Fernandez-Cobos, F. Finelli,

F. Forastieri, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, M. Gerbino, T. Ghosh, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gruppuso, J. E. Gudmundsson, J. Hamann, W. Handley, F. K. Hansen, D. Herranz, S. R. Hildebrandt, E. Hivon, Z. Huang, A. H. Jaffe, W. C. Jones, A. Karakci, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Kim, T. S. Kisner, L. Knox, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, G. Lagache, J.-M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, P. Lemos, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Lilley, V. Lindholm, M. López-Caniego, P. M. Lubin, Y.-Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M.-A. Miville-Deschênes, D. Molinari, L. Montier, G. Morgante, A. Moss, P. Natoli, H. U. Nørgaard-Nielsen, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, L. Polastri, G. Polenta, J.-L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, E. P. S. Shellard, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A.-S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Toffolatti, M. Tomasi, T. Trombetti, L. Valenziano, J. Valiviita, B. Van Tent, L. Vibert, P. Vielva, F. Villa, N. Vittorio, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Zacchei, and A. Zonca. Planck 2018 results. VI. Cosmological parameters (Corrigendum). *Astronomy & Astrophysics*, 652:C4, August 2021.

[183] Erwin Platen, Rien van de Weygaert, and Bernard J. T. Jones. A Cosmic Watershed: The WVF Void Detection Technique. *Mon. Not. Roy. Astron. Soc.*, 380:551–570, 2007.

[184] Erwin Platen, Rien van de Weygaert, and Bernard J. T. Jones. Alignment of voids in the cosmic web. *Monthly Notices of the Royal Astronomical Society*, 387(1):128–136, June 2008.

[185] Alvise Raccanelli, Daniele Bertacca, Davide Pietrobon, Fabian Schmidt, Lado Samushia, Nicola Bartolo, Olivier Doré, Sabino Matarrese, and Will J. Percival. Testing gravity using large-scale redshift-space distortions. *Monthly Notices of the Royal Astronomical Society*, 436(1):89–100, November 2013.

[186] S. Radinović, S. Nadathur, H.-A. Winther, W. J. Percival, A. Woodfinden, E. Massara, E. Paillas, S. Contarini, N. Hamaus, A. Kovacs, A. Pisani, G. Verza, M. Aubert,

A. Amara, N. Auricchio, M. Baldi, D. Bonino, E. Branchini, M. Brescia, S. Camera, V. Capobianco, C. Carbone, V. F. Cardone, J. Carretero, M. Castellano, S. Cavuoti, A. Cimatti, R. Cledassou, G. Congedo, L. Conversi, Y. Copin, L. Corcione, F. Courbin, A. Da Silva, M. Douspis, F. Dubath, X. Dupac, S. Farrens, S. Ferriol, P. Fosalba, M. Frailis, E. Franceschi, M. Fumana, S. Galeotta, B. Garilli, W. Gillard, B. Gillis, C. Giocoli, A. Grazian, F. Grupp, S. V. H. Haugan, W. Holmes, A. Hornstrup, K. Jahnke, M. Kümmel, A. Kiessling, M. Kilbinger, T. Kitching, H. Kurki-Suonio, S. Ligori, P. B. Lilje, I. Lloro, E. Maiorano, O. Mansutti, O. Marggraf, K. Markovic, F. Marulli, R. Massey, S. Mei, M. Melchior, Y. Mellier, M. Meneghetti, E. Merlin, G. Meylan, M. Moresco, L. Moscardini, S.-M. Niemi, J. W. Nightingale, T. Nutma, C. Padilla, S. Paltani, F. Pasian, K. Pedersen, V. Pettorino, S. Pires, G. Polenta, M. Poncet, L. A. Popa, L. Pozzetti, F. Raison, A. Renzi, J. Rhodes, G. Riccio, E. Romelli, M. Roncarelli, C. Rosset, R. Saglia, D. Sapone, B. Sartoris, P. Schneider, A. Secroun, G. Seidel, S. Serrano, C. Sirignano, G. Sirri, L. Stanco, J.-L. Starck, C. Surace, P. Tallada-Crespí, I. Tereno, R. Toledo-Moreo, F. Torradeflot, I. Tutusaus, E. A. Valentijn, L. Valenziano, T. Vassallo, Y. Wang, J. Weller, G. Zamorani, J. Zoubian, and V. Scottez. Euclid: Cosmology forecasts from the void-galaxy cross-correlation function with reconstruction. *Astronomy amp; Astrophysics*, 677:A78, September 2023.

[187] Slaana Radinović, Hans A. Winther, Seshadri Nadathur, Will J. Percival, Enrique Paillas, Tristan Sohrab Fraser, Elena Massara, and Alex Woodfinden. Alcock–paczyński effect on void-finding: Implications for void-galaxy cross-correlation modelling. *Astronomy amp; Astrophysics*, 691:A39, October 2024.

[188] Carl Edward Rasmussen. *Gaussian Processes in Machine Learning*, pages 63–71. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[189] Beth Reid, Shirley Ho, Nikhil Padmanabhan, Will J. Percival, Jeremy Tinker, Rita Tojeiro, Martin White, Daniel J. Eisenstein, Claudia Maraston, Ashley J. Ross, Ariel G. Sánchez, David Schlegel, Erin Sheldon, Michael A. Strauss, Daniel Thomas, David Wake, Florian Beutler, Dmitry Bizyaev, Adam S. Bolton, Joel R. Brownstein, Chia-Hsun Chuang, Kyle Dawson, Paul Harding, Francisco-Shu Kitaura, Alexie Leauthaud, Karen Masters, Cameron K. McBride, Surhud More, Matthew D. Olmstead, Daniel Oravetz, Sebastián E. Nuza, Kaike Pan, John Parejko, Janine Pforr, Francisco Prada, Sergio Rodríguez-Torres, Salvador Salazar-Albornoz, Lado Samushia, Donald P. Schneider, Claudia G. Scóccola, Audrey Simmons, and Mariana Vargas-Magana. Sdss-iii baryon oscillation spectroscopic survey data release 12:

galaxy target selection and large-scale structure catalogues. *Monthly Notices of the Royal Astronomical Society*, 455(2):1553–1573, 01 2016.

[190] Beth Reid, Shirley Ho, Nikhil Padmanabhan, Will J. Percival, Jeremy Tinker, Rita Tojeiro, Martin White, Daniel J. Eisenstein, Claudia Maraston, Ashley J. Ross, Ariel G. Sanchez, David Schlegel, Erin Sheldon, Michael A. Strauss, Daniel Thomas, David Wake, Florian Beutler, Dmitry Bizyaev, Adam S. Bolton, Joel R. Brownstein, Chia-Hsun Chuang, Kyle Dawson, Paul Harding, Francisco-Shu Kitaura, Alexie Leauthaud, Karen Masters, Cameron K. McBride, Surhud More, Matthew D. Olmstead, Daniel Oravetz, Sebastian E. Nuza, Kaike Pan, John Parejko, Janine Pforr, Francisco Prada, Sergio Rodriguez-Torres, Salvador Salazar-Albornoz, Lado Samushia, Donald P. Schneider, Claudia G. Scoccola, Audrey Simmons, and Mariana Vargas-Magana. Sdss-iii baryon oscillation spectroscopic survey data release 12: galaxy target selection and large scale structure catalogues, 2015.

[191] Beth A. Reid, Lado Samushia, Martin White, Will J. Percival, Marc Manera, Nikhil Padmanabhan, Ashley J. Ross, Ariel G. Sánchez, Stephen Bailey, Dmitry Bizyaev, Adam S. Bolton, Howard Brewington, J. Brinkmann, Joel R. Brownstein, Antonio J. Cuesta, Daniel J. Eisenstein, James E. Gunn, Klaus Honscheid, Elena Malanushenko, Viktor Malanushenko, Claudia Maraston, Cameron K. McBride, Demitri Muna, Robert C. Nichol, Daniel Oravetz, Kaike Pan, Roland de Putter, N. A. Roe, Nicholas P. Ross, David J. Schlegel, Donald P. Schneider, Hee-Jong Seo, Alaina Shelden, Erin S. Sheldon, Audrey Simmons, Ramin A. Skibba, Stephanie Snedden, Molly E. C. Swanson, Daniel Thomas, Jeremy Tinker, Rita Tojeiro, Licia Verde, David A. Wake, Benjamin A. Weaver, David H. Weinberg, Idit Zehavi, and Gong-Bo Zhao. The clustering of galaxies in the sdss-iii baryon oscillation spectroscopic survey: measurements of the growth of structure and expansion rate at¡i¿z¡/i¿= 0.57 from anisotropic clustering: Anisotropic clustering in cmass galaxies. *Monthly Notices of the Royal Astronomical Society*, 426(4):2719–2737, October 2012.

[192] Beth A. Reid and Martin White. Towards an accurate model of the redshift-space clustering of haloes in the quasi-linear regime: Redshift-space distortions. *Monthly Notices of the Royal Astronomical Society*, 417(3):1913–1927, August 2011.

[193] Adam G. Riess. The expansion of the Universe is faster than expected. *Nature Reviews Physics*, 2(1):10–12, January 2020.

[194] Adam G. Riess, Wenlong Yuan, Lucas M. Macri, Dan Scolnic, Dillon Brout, Stefano Casertano, David O. Jones, Yukei Murakami, Gagandeep S. Anand, Louise Breuval,

Thomas G. Brink, Alexei V. Filippenko, Samantha Hoffmann, Saurabh W. Jha, W. D'arcy Kenworthy, John Mackenty, Benjamin E. Stahl, and WeiKang Zheng. A comprehensive measurement of the local value of the hubble constant with 1 km s ¡sup¿1¡/sup¿ mpc ¡sup¿1¡/sup¿ uncertainty from the hubble space telescope and the sh0es team. *The Astrophysical Journal Letters*, 934(1):L7, 2022.

[195] H. P. Robertson. Kinematics and World-Structure. *The Astrophysical Journal*, 82:284, November 1935.

[196] H. P. Robertson. Kinematics and World-Structure II. *The Astrophysical Journal*, 83:187, April 1936.

[197] H. P. Robertson. Kinematics and World-Structure III. *The Astrophysical Journal*, 83:257, May 1936.

[198] Sergio A. Rodríguez-Torres, Chia-Hsun Chuang, Francisco Prada, Hong Guo, Anatoly Klypin, Peter Behroozi, Chang Hoon Hahn, Johan Comparat, Gustavo Yepes, Antonio D. Montero-Dorta, Joel R. Brownstein, Claudia Maraston, Cameron K. McBride, Jeremy Tinker, Stefan Gottlöber, Ginevra Favole, Yiping Shu, Francisco-Shu Kitaura, Adam Bolton, Román Scoccimarro, Lado Samushia, David Schlegel, Donald P. Schneider, and Daniel Thomas. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: modelling the clustering and halo occupation distribution of BOSS CMASS galaxies in the Final Data Release. *Monthly Notices of the Royal Astronomical Society*, 460(2):1173–1187, August 2016.

[199] A. J. Ross, J. Aguilar, S. Ahlen, S. Alam, A. Anand, S. Bailey, D. Bianchi, S. Brieden, D. Brooks, E. Burtin, A. Carnero Rosell, E. Chaussidon, T. Claybaugh, S. Cole, K. Dawson, A. de la Macorra, A. de Mattia, Arjun Dey, Biprateep Dey, P. Doel, K. Fanning, S. Ferraro, J. Ereza, A. Font-Ribera, J. E. Forero-Romero, E. Gaztañaga, H. Gil-Marín, S. Gontcho A Gontcho, A. X. Gonzalez-Morales, J. Guy, C. Hahn, S. Heydenreich, K. Honscheid, C. Howlett, M. Ishak, T. Karim, D. Kirkby, T. Kisner, H. Kong, A. Kremin, A. Krolewski, A. Lambert, M. Landriau, J. Lasker, L. Le Guillou, M. E. Levi, M. Manera, P. Martini, P. McDonald, A. Meisner, R. Miquel, J. Moon, J. Moustakas, A. Muñoz-Gutiérrez, A. D. Myers, S. Nadathur, L. Napolitano, J. A. Newman, J. Nie, G. Niz, N. Palanque-Delabrouille, W. J. Percival, C. Poppett, F. Prada, A. Raichoor, C. Ravoux, M. Rezaie, A. Rosado-Marin, G. Rossi, L. Samushia, E. Sanchez, E. F. Schlafly, D. Schlegel, H. Seo, A. Smith, D. Sprayberry, G. Tarlé, D. Valcin, M. Vargas-Magaña, B. A. Weaver, M. Wilson, J. Yu, P. Zarrouk, C. Zhao, R. Zhou, and H. Zou. The construction of large-scale structure catalogs for the dark energy spectroscopic instrument, 2024.

[200] Cheng-Zong Ruan, Carolina Cuesta-Lazaro, Alexander Eggemeier, Baojiu Li, Carlton M. Baugh, Christian Arnold, Sownak Bose, César Hernández-Aguayo, Pauline Zarrouk, and Christopher T. Davies. An emulator-based halo model in modified gravity – i. the halo concentration-mass relation and density profile, 2023.

[201] R. K. Sachs and A. M. Wolfe. Perturbations of a Cosmological Model and Angular Variations of the Microwave Background. *The Astrophysical Journal*, 147:73, January 1967.

[202] Lado Samushia, Zachary Slepian, and Francisco Villaescusa-Navarro. Information content of higher order galaxy correlation functions. *Monthly Notices of the Royal Astronomical Society*, 505(1):628–641, July 2021.

[203] C. Sánchez, J. Clampitt, A. Kovacs, B. Jain, J. García-Bellido, S. Nadathur, D. Gruen, N. Hamaus, D. Huterer, P. Vielzeuf, A. Amara, C. Bonnett, J. DeRose, W. G. Hartley, M. Jarvis, O. Lahav, R. Miquel, E. Rozo, E. S. Rykoff, E. Sheldon, R. H. Wechsler, J. Zuntz, T. M. C. Abbott, F. B. Abdalla, J. Annis, A. Benoit-Lévy, G. M. Bernstein, R. A. Bernstein, E. Bertin, D. Brooks, E. Buckley-Geer, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, M. Crocce, C. E. Cunha, C. B. D'Andrea, L. N. da Costa, S. Desai, H. T. Diehl, J. P. Dietrich, P. Doel, A. E. Evrard, A. Fausti Neto, B. Flaugher, P. Fosalba, J. Frieman, E. Gaztanaga, R. A. Gruendl, G. Gutierrez, K. Honscheid, D. J. James, E. Krause, K. Kuehn, M. Lima, M. A. G. Maia, J. L. Marshall, P. Melchior, A. A. Plazas, K. Reil, A. K. Romer, E. Sanchez, M. Schubnell, I. Sevilla-Noarbe, R. C. Smith, M. Soares-Santos, F. Sobreira, E. Suchyta, G. Tarle, D. Thomas, A. R. Walker, and J. Weller. Cosmic voids and void lensing in the dark energy survey science verification data. *Monthly Notices of the Royal Astronomical Society*, 465(1):746–759, October 2016.

[204] Michael D. Schneider, Lloyd Knox, Salman Habib, Katrin Heitmann, David Higdon, and Charles Nakhleh. Simulations and cosmological inference: A statistical model for power spectra means and covariances. *Physical Review D*, 78(6):063529, September 2008.

[205] Roman Scoccimarro. The bispectrum: From theory to observations. *The Astrophysical Journal*, 544(2):597–615, December 2000.

[206] Roman Scoccimarro, Stephane Colombi, J. N. Fry, Joshua A. Frieman, Eric Hivon, and Adrian Melott. Nonlinear evolution of the bispectrum of cosmological perturbations. *The Astrophysical Journal*, 496(2):586–604, April 1998.

[207] Roman Scoccimarro, H. M. P. Couchman, and Joshua A. Frieman. The bispectrum as a signature of gravitational instability in redshift space. *The Astrophysical Journal*, 517(2):531–540, 1999.

[208] Dan Scolnic, Dillon Brout, Anthony Carr, Adam G. Riess, Tamara M. Davis, Arianna Dwomoh, David O. Jones, Noor Ali, Pranav Charvu, Rebecca Chen, Erik R. Peterson, Brodie Popovic, Benjamin M. Rose, Charlotte M. Wood, Peter J. Brown, Ken Chambers, David A. Coulter, Kyle G. Dettman, Georgios Dimitriadis, Alexei V. Filippenko, Ryan J. Foley, Saurabh W. Jha, Charles D. Kilpatrick, Robert P. Kirshner, Yen-Chen Pan, Armin Rest, Cesar Rojas-Bravo, Matthew R. Siebert, Benjamin E. Stahl, and WeiKang Zheng. The pantheon+ analysis: The full data set and light-curve release. *The Astrophysical Journal*, 938(2):113, October 2022.

[209] Hee-Jong Seo, Florian Beutler, Ashley J. Ross, and Shun Saito. Modeling the reconstructed BAO in Fourier space. *Monthly Notices of the Royal Astronomical Society*, 460(3):2453–2471, August 2016.

[210] Ravi K. Sheth and Giuseppe Tormen. Large-scale bias and the peak background split. *Monthly Notices of the Royal Astronomical Society*, 308(1):119–126, 1999.

[211] Jonathan Shlomi, Peter Battaglia, and Jean-Roch Vlimant. Graph neural networks in particle physics. *Machine Learning: Science and Technology*, 2(2):021001, 2021.

[212] Joseph H. Silber, David J. Schlegel, Ricardo Araujo, Charles Baltay, Robert W. Besuner, Emily Farr, Julien Guy, Jean-Paul Kneib, Claire Poppett, Travis A. Mandeville, Michael Schubnell, Markus Thurneysen, and Sarah Tuttle. 25,000 optical fiber positioning robots for next-generation cosmology, 2022.

[213] Manodeep Sinha and Lehman H. Garrison. CORRFUNC - a suite of blazing fast correlation functions on the CPU. *Monthly Notices of the Royal Astronomical Society*, 491(2):3022–3041, Jan 2020.

[214] Ramin Skibba, Ravi K. Sheth, Andrew J. Connolly, and Ryan Scranton. The luminosity-weighted or 'marked' correlation function. *Monthly Notices of the Royal Astronomical Society*, 369(1):68–76, May 2006.

[215] Alex Smith, Arnaud de Mattia, Etienne Burtin, Chia-Hsun Chuang, and Cheng Zhao. Reducing the variance of redshift space distortion measurements from mock galaxy catalogues with different lines of sight. *Monthly Notices of the Royal Astronomical Society*, 500(1):259–271, October 2020.

[216] Yong-Seon Song and Will J. Percival. Reconstructing the history of structure formation using redshift distortions. *Journal of Cosmology and Astroparticle Physics*, 2009(10):004, oct 2009.

[217] Kate Storey-Fisher, Jeremy L. Tinker, Zhongxu Zhai, Joseph DeRose, Risa H. Wechsler, and Arka Banerjee. The Aemulus Project. VI. Emulation of Beyond-standard Galaxy Clustering Statistics to Improve Cosmological Constraints. *The Astrophysical Journal*, 961(2):208, February 2024.

[218] R. A. Sunyaev and Ia. B. Zeldovich. Microwave background radiation as a probe of the contemporary structure and history of the universe. *Annual Review of Astronomy and Astrophysics*, 18:537–560, January 1980.

[219] R. A. Sunyaev and Ya. B. Zeldovich. Small-Scale Fluctuations of Relic Radiation. *Astrophysics and Space Science*, 7(1):3–19, April 1970.

[220] P. M. Sutter, Guilhem Lavaux, Nico Hamaus, Alice Pisani, Benjamin D. Wandelt, Michael S. Warren, Francisco Villaescusa-Navarro, Paul Zivick, Qingqing Mao, and Benjamin B. Thompson. Vide: The void identification and examination toolkit, 2014.

[221] P. M. Sutter, Guilhem Lavaux, Benjamin D. Wandelt, and David H. Weinberg. A public void catalog from the sdss dr7 galaxy redshift surveys based on the watershed transform. *The Astrophysical Journal*, 761(1):44, November 2012.

[222] Boxin Tang. Orthogonal array-based latin hypercubes. *Journal of the American Statistical Association*, 88(424):1392–1397, 1993.

[223] Max Tegmark, Michael R. Blanton, Michael A. Strauss, Fiona Hoyle, David Schlegel, Roman Scoccimarro, Michael S. Vogeley, David H. Weinberg, Idit Zehavi, Andreas Berlind, Tamas Budavari, Andrew Connolly, Daniel J. Eisenstein, Douglas Finkbeiner, Joshua A. Frieman, James E. Gunn, Andrew J. S. Hamilton, Lam Hui, Bhuvnesh Jain, David Johnston, Stephen Kent, Huan Lin, Reiko Nakajima, Robert C. Nichol, Jeremiah P. Ostriker, Adrian Pope, Ryan Scranton, Uroš Seljak, Ravi K. Sheth, Albert Stebbins, Alexander S. Szalay, Istvan Szapudi, Licia Verde, Yongzhong Xu, James Annis, Neta A. Bahcall, J. Brinkmann, Scott Burles, Francisco J. Castander, Istvan Csabai, Jon Loveday, Mamoru Doi, Masataka Fukugita, J. Richard Gott III, Greg Hennessy, David W. Hogg, Željko Ivezić, Gillian R. Knapp, Don Q. Lamb, Brian C. Lee, Robert H. Lupton, Timothy A. McKay, Peter Kunszt, Jeffrey A. Munn, Liam O'Connell, John Peoples, Jeffrey R. Pier, Michael Richmond,

Constance Rockosi, Donald P. Schneider, Christopher Stoughton, Douglas L. Tucker, Daniel E. Vanden Berk, Brian Yanny, and Donald G. York. The three-dimensional power spectrum of galaxies from the sloan digital sky survey. *The Astrophysical Journal*, 606(2):702–740, May 2004.

[224] Leander Thiele, Elena Massara, Alice Pisani, ChangHoon Hahn, David N. Spergel, Shirley Ho, and Benjamin Wandelt. Neutrino mass constraint from an implicit likelihood analysis of boss voids, 2023.

[225] Shaun A Thomas, Filipe B Abdalla, and Ofer Lahav. Excess clustering on large scales in the megaz dr7 photometric redshift survey. *Physical review letters*, 106(24):241301, 2011.

[226] W. G. Tifft and S. A. Gregory. Direct observations of the large-scale distribution of galaxies. *The Astrophysical Journal*, 205:696–708, May 1976.

[227] Sunny Vagnozzi. Consistency tests of Λ CDM from the early integrated Sachs-Wolfe effect: Implications for early-time new physics and the Hubble tension. *Physical Review D*, 104(6):063524, September 2021.

[228] Georgios Valogiannis, Sihan Yuan, and Cora Dvorkin. Precise cosmological constraints from boss galaxy clustering with a simulation-based emulator of the wavelet scattering transform. *Physical Review D*, 109(10), May 2024.

[229] Giovanni Verza, Carmelita Carbone, and Alessandro Renzi. The Halo Bias inside Cosmic Voids. *The Astrophysical Journal Letters*, 940(1):L16, November 2022.

[230] Pablo Villanueva-Domingo and Francisco Villaescusa-Navarro. Learning cosmology and clustering with cosmic graphs. *The Astrophysical Journal*, 937(2):115, 2022.

[231] Lile Wang, Beth Reid, and Martin White. An analytic model for redshift-space distortions. *Monthly Notices of the Royal Astronomical Society*, 437(1):588–599, November 2013.

[232] Yuting Wang, Ruiyang Zhao, Zhongxu Zhai, Kazuya Koyama, Will J. Percival, Hong Guo, Yin Li, Gong-Bo Zhao, Takahiro Nishimichi, Héctor Gil-Marín, Yonghao Feng, Hanyu Zhang, and Yi Wu. Emulating Power Spectra for Prereconstructed and Postreconstructed Galaxy Samples. *The Astrophysical Journal*, 966(1):35, May 2024.

[233] Martin White. A marked correlation function for constraining modified gravity models. *Journal of Cosmology and Astroparticle Physics*, 2016(11):057–057, November 2016.

[234] Alex Woodfinden, Seshadri Nadathur, Will J Percival, Sladana Radinovic, Elena Massara, and Hans A Winther. Measurements of cosmic expansion and growth rate of structure from voids in the sloan digital sky survey between redshift 0.07 and 1.0. *Monthly Notices of the Royal Astronomical Society*, 516(3):4307–4323, 2022.

[235] Alex Woodfinden, Will J Percival, Seshadri Nadathur, Hans A Winther, T S Fraser, Elena Massara, Enrique Paillas, and Slaana Radinović. Cosmological measurements from void-galaxy and galaxy-galaxy clustering in the sloan digital sky survey. *Monthly Notices of the Royal Astronomical Society*, 523(4):6360–6370, 2023.

[236] Sihan Yuan, Lehman H. Garrison, Daniel J. Eisenstein, and Risa H. Wechsler. Stringent σ_8 constraints from small-scale galaxy clustering using a hybrid MCMC + emulator framework. *Monthly Notices of the Royal Astronomical Society*, 515(1):871–896, September 2022.

[237] Sihan Yuan, Lehman H Garrison, Boryana Hadzhiyska, Sownak Bose, and Daniel J Eisenstein. ¡scp¿abacushod¡/scp¿ : a highly efficient extended multitracer hod framework and its application to boss and eboss data. *Monthly Notices of the Royal Astronomical Society*, 510(3):3301–3320, November 2021.

[238] Sihan Yuan, Hanyu Zhang, Ashley J. Ross, Jamie Donald-McCann, Boryana Hadzhiyska, Risa H. Wechsler, Zheng Zheng, Shadab Alam, Violeta Gonzalez-Perez, Jessica Nicole Aguilar, Steven Ahlen, Davide Bianchi, David Brooks, Axel de la Macorra, Kevin Fanning, Jaime E. Forero-Romero, Klaus Honscheid, Mustapha Ishak, Robert Kehoe, James Lasker, Martin Landriau, Marc Manera, Paul Martini, Aaron Meisner, Ramon Miquel, John Moustakas, Seshadri Nadathur, Jeffrey A. Newman, Jundan Nie, Will Percival, Claire Poppett, Antoine Rocher, Graziano Rossi, Eusebio Sanchez, Lado Samushia, Michael Schubnell, Hee-Jong Seo, Gregory Tarle, Benjamin Alan Weaver, Jiaxi Yu, Zhimin Zhou, and Hu Zou. The desi one-percent survey: Exploring the halo occupation distribution of luminous red galaxies and quasi-stellar objects with abacussummit, 2023.

[239] Ya. B. Zel'dovich. Gravitational instability: An approximate theory for large density perturbations. *Astronomy & Astrophysics*, 5:84–89, March 1970.

[240] Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J. Smola. Dive into deep learning, 2023.

[241] Cheng Zhao, Chia-Hsun Chuang, Julian Bautista, Arnaud de Mattia, Anand Raichoor, Ashley J Ross, Jiamin Hou, Richard Neveux, Charling Tao, Etienne Burtin,

Kyle S Dawson, Sylvain de la Torre, Héctor Gil-Marín, Jean-Paul Kneib, Will J Percival, Graziano Rossi, Amélie Tamone, Jeremy L Tinker, Gong-Bo Zhao, Shadab Alam, and Eva-Maria Mueller. The completed sdss-iv extended baryon oscillation spectroscopic survey: 1000 multi-tracer mock catalogues with redshift evolution and systematics for galaxies and quasars of the final data release. *Monthly Notices of the Royal Astronomical Society*, 503(1):1149–1173, February 2021.

[242] Zheng Zheng, Andreas A. Berlind, David H. Weinberg, Andrew J. Benson, Carlton M. Baugh, Shaun Cole, Romeel Davé, Carlos S. Frenk, Neal Katz, and Cedric G. Lacey. Theoretical Models of the Halo Occupation Distribution: Separating Central and Satellite Galaxies. *The Astrophysical Journal*, 633(2):791–809, November 2005.

[243] Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph neural networks: A review of methods and applications, 2021.

[244] Rongpu Zhou, Biprateep Dey, Jeffrey A. Newman, Daniel J. Eisenstein, K. Dawson, S. Bailey, A. Berti, J. Guy, Ting-Wen Lan, H. Zou, J. Aguilar, S. Ahlen, Shadab Alam, D. Brooks, A. de la Macorra, A. Dey, G. Dhungana, K. Fanning, A. Font-Ribera, S. Gontcho A. Gontcho, K. Honscheid, Mustapha Ishak, T. Kisner, A. Kovács, A. Kremin, M. Landriau, Michael E. Levi, C. Magneville, Marc Manera, P. Martini, Aaron M. Meisner, R. Miquel, J. Moustakas, Adam D. Myers, Jundan Nie, N. Palanque-Delabrouille, W. J. Percival, C. Poppett, F. Prada, A. Raichoor, A. J. Ross, E. Schlafly, D. Schlegel, M. Schubnell, Gregory Tarlé, B. A. Weaver, R. H. Wechsler, Christophe Yéche, and Zhimin Zhou. Target Selection and Validation of DESI Luminous Red Galaxies. *The Astronomical Journal*, 165(2):58, February 2023.

[245] Rongpu Zhou, Jeffrey A. Newman, Kyle S. Dawson, Daniel J. Eisenstein, David D. Brooks, Arjun Dey, Biprateep Dey, Yutong Duan, Sarah Eftekharzadeh, Enrique Gaztañaga, Robert Kehoe, Martin Landriau, Michael E. Levi, Timothy C. Licquia, Aaron M. Meisner, John Moustakas, Adam D. Myers, Nathalie Palanque-Delabrouille, Claire Poppett, Francisco Prada, Anand Raichoor, David J. Schlegel, Michael Schubnell, Ryan Staten, Gregory Tarlé, and Christophe Yèche. Preliminary Target Selection for the DESI Luminous Red Galaxy (LRG) Sample. *Research Notes of the American Astronomical Society*, 4(10):181, October 2020.

[246] Paul Zivick, P. M. Sutter, Benjamin D. Wandelt, Baojiu Li, and Tsz Yan Lam. Using cosmic voids to distinguish f(R) gravity in future galaxy surveys. *Monthly Notices of the Royal Astronomical Society*, 451(4):4215–4222, August 2015.

Appendix

Appendix A

Joint likelihood analysis of void emulator and DESI DR1

Our joint likelihood analysis, discussed in Chapter 5, as a proof-of-concept, explored the potential complementary constraints that could be obtained from a joint void-emulator + BAO (+SNIa dataset) analysis. We explored combining these datasets for flat wCDM and w_0w_aCDM cosmologies, and utilized a shared parameter space of either $\Omega_{\rm m}$ or $\omega_{\rm cdm}$ alongside the dark energy equation of state parameters. We found that these joint-likelihood "constraints" saw a consistent reduction in errors for both $\Omega_{\rm m}$ or $\omega_{\rm cdm}$ when the void-emulator's chain was incorporated. Future work could extend the results discussed in Chapter 5 into a thorough exploration of the joint constraining power, leveraging the added information of cosmic voids from multiple redshift bins in DESI DR1 and DESI DR2.

A.1 Additional Joint Analysis for w_0CDM

We utilised a joint likelihood analysis of the DESI BAO+SNIa chains, which utilised SNIa data from PantheonPlus (Figures), Union3 (Figures), and DES (Figures).

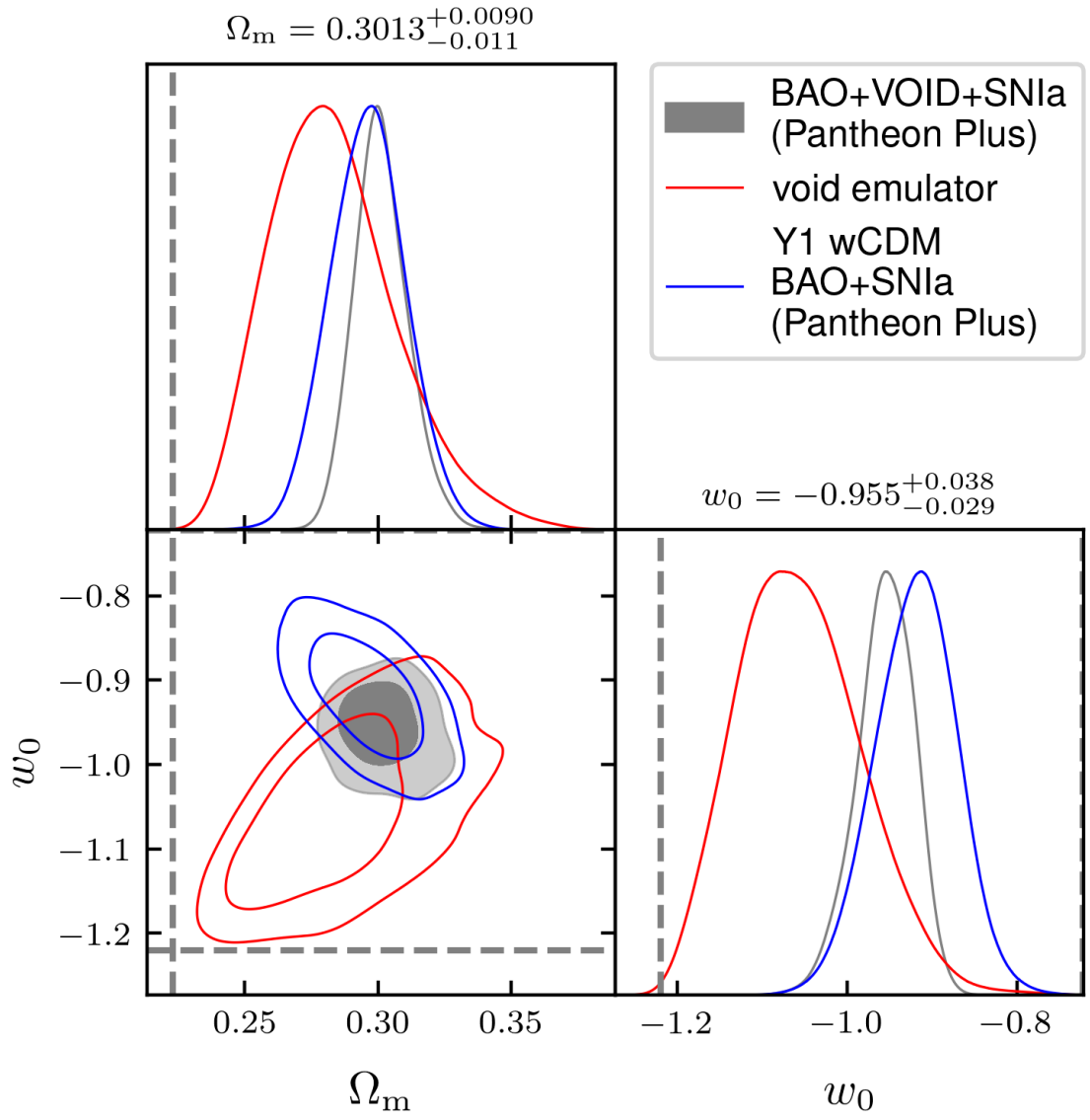

Figure A.1: Joint likelihood analysis combining the BAO+DESY5 result (blue) with the void emulator (red). The grey dashes demarcate the prior walls for the void emulator, and the grey contour denotes the joint likelihood. The joint likelihoods help produce dark energy constraints that are competitive with the BAO+SN+CMB results.

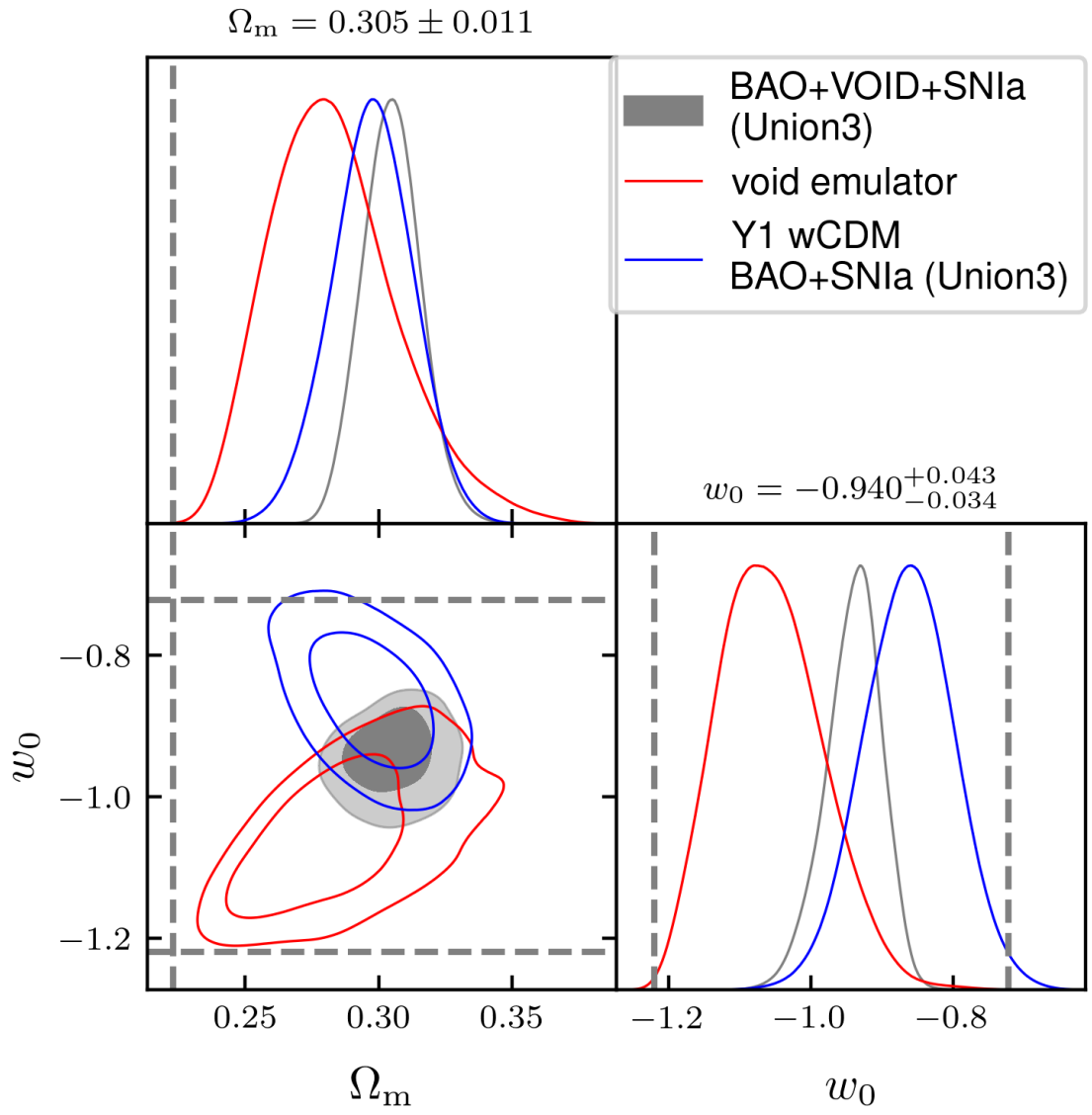

Figure A.2: Joint likelihood analysis combining the BAO+Union3 result (blue) with the void emulator (red). The grey dashes demarcate the prior walls for the void emulator, and the grey contour denotes the joint likelihood. The joint likelihoods help produce dark energy constraints that are competitive with the BAO+SN+CMB results.

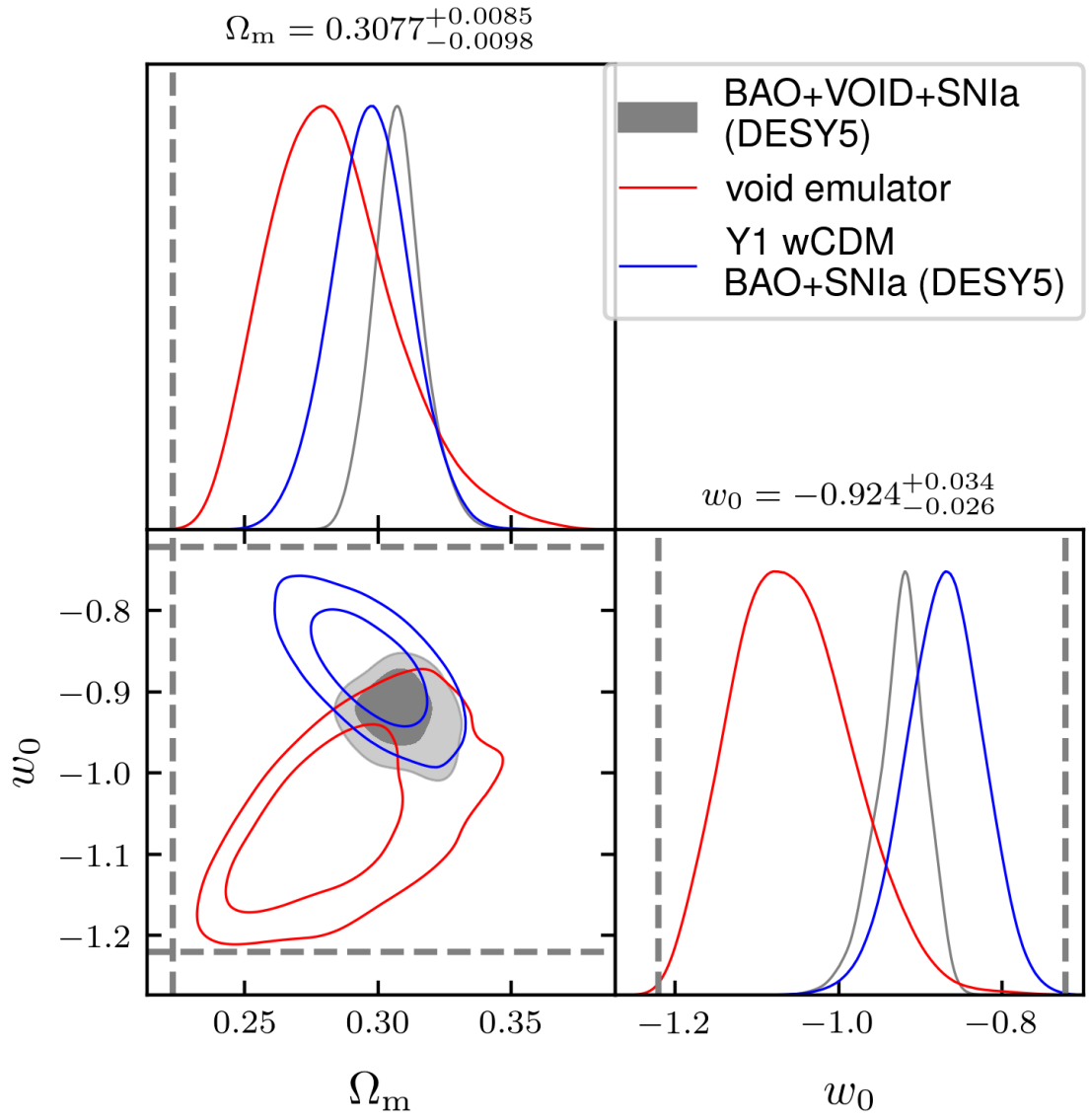

Figure A.3: Joint likelihood analysis combining the BAO+DESY5 result (blue) with the void emulator (red). The grey dashes demarcate the prior walls for the void emulator, and the grey contour denotes the joint likelihood. The joint likelihoods help produce dark energy constraints that are competitive with the BAO+SN+CMB results.

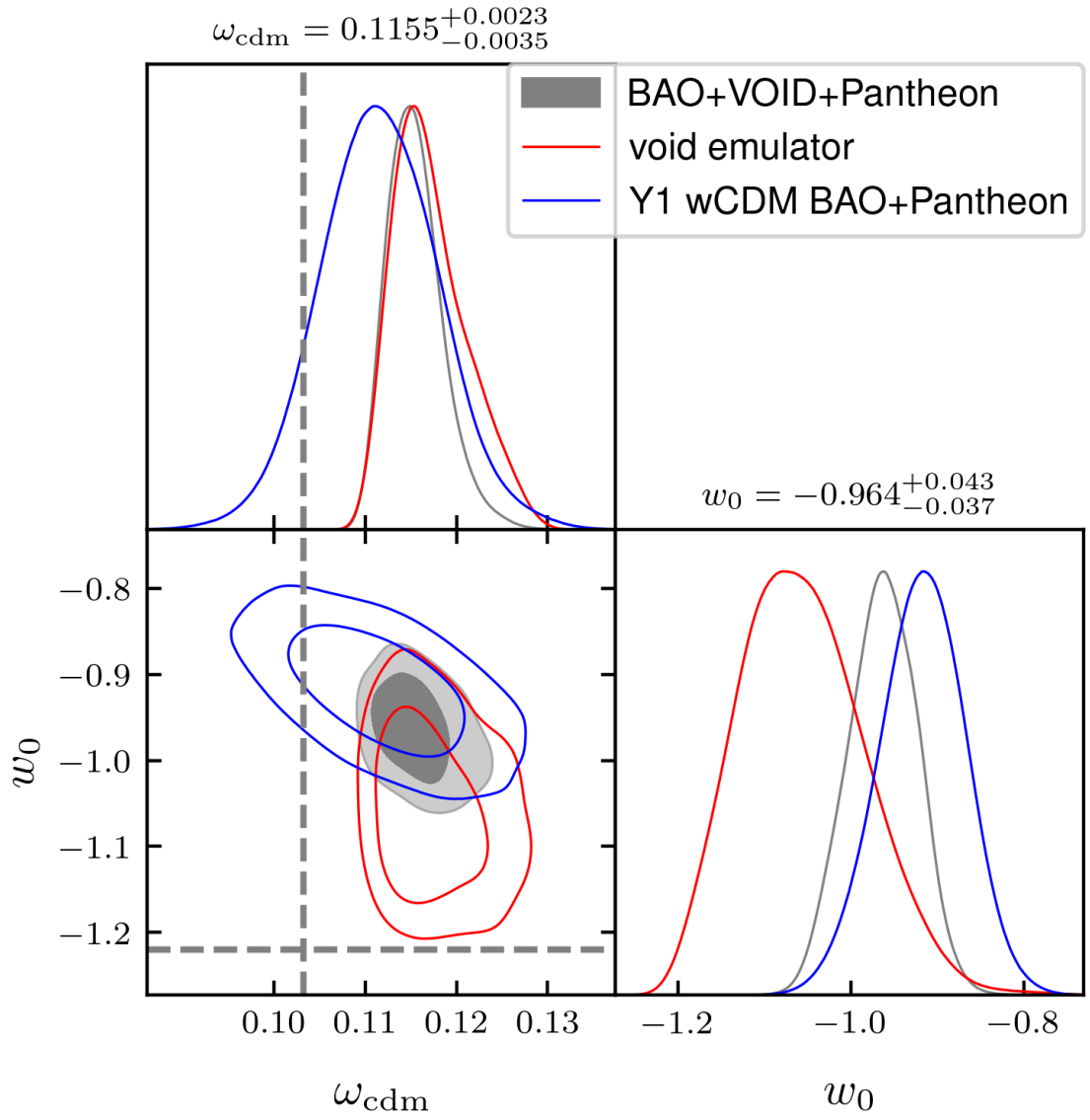

Figure A.4: Joint likelihood analysis combining the BAO+Pantheon plus result (blue) but unlike Figure A.1, we explored the constraints for $\omega_{\rm cdm}$ instead of $\Omega_{\rm m}$.

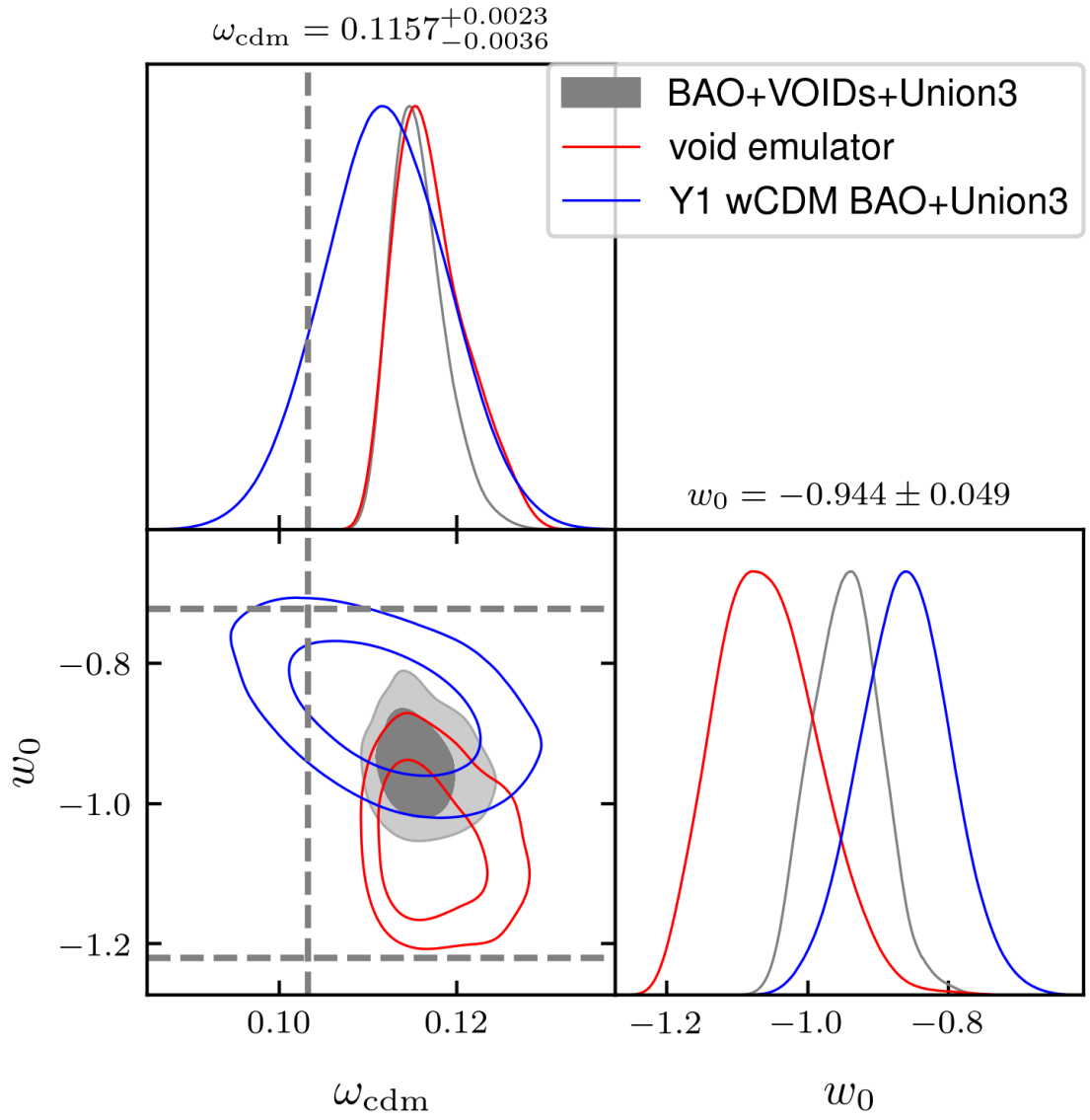

Figure A.5: Joint likelihood analysis combining the BAO+Pantheon plus result (blue) but unlike Figure A.2, we explored the constraints for $\omega_{\rm cdm}$ instead of $\Omega_{\rm m}$.

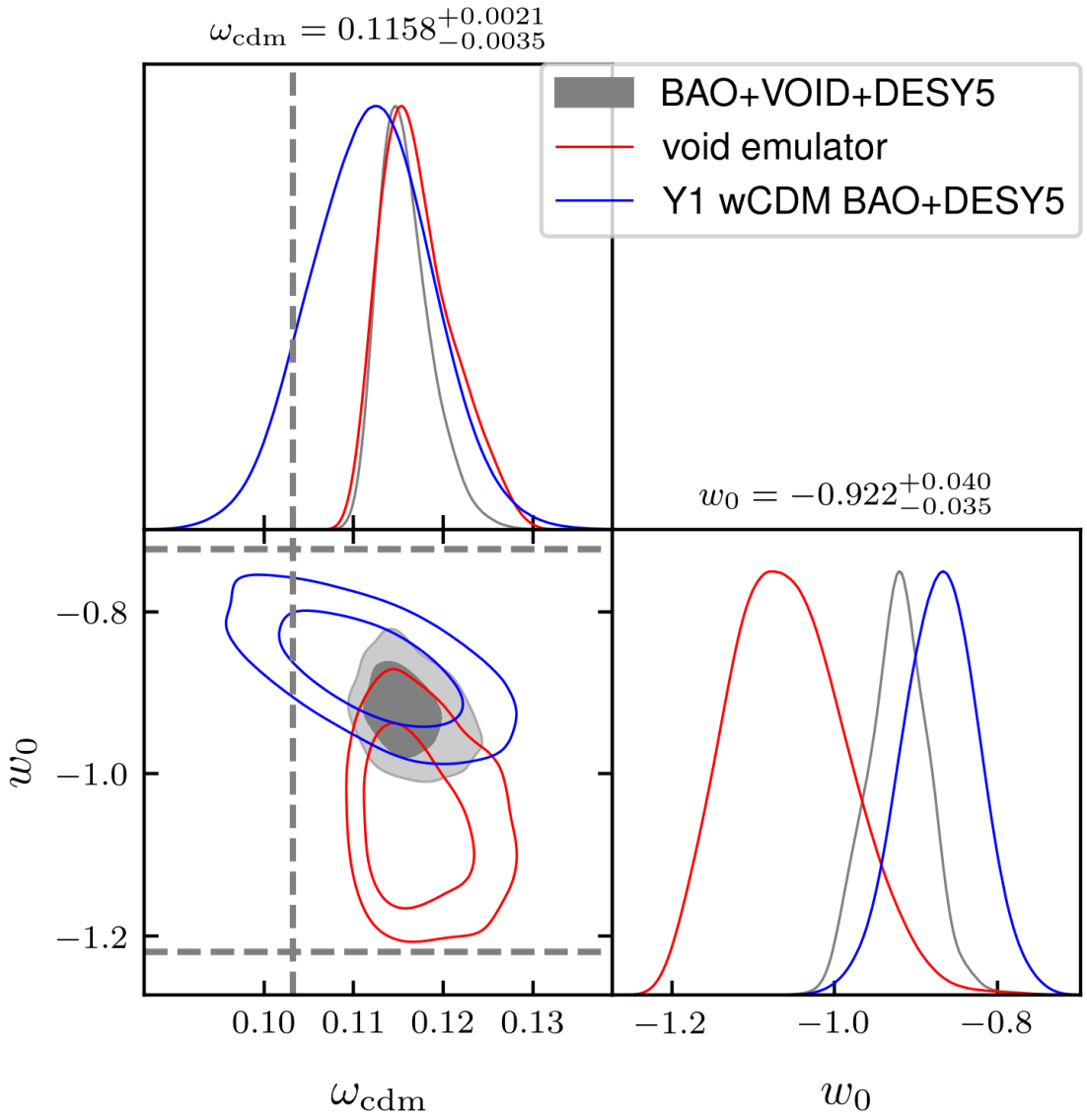

Figure A.6: Joint likelihood analysis combining the BAO+Pantheon plus result (blue) but unlike Figure A.3, we explored the constraints for $\omega_{\rm cdm}$ instead of $\Omega_{\rm m}$.

Depending on the chosen SNIa dataset, we see the mean values of w_0 range from

-0.955 to -0.922, approximately 1-2 σ away from the fiducial value for w_0. Re-visiting this analysis, armed with an emulator that contained more redshift evolution from the VG-CCF would be interesting, and could allow the complementary constraining power to be thoroughly explored.

A.2 Additional Joint Analysis for w_0w_aCDM

The same analysis was extended to w_0w_aCDM cosmologies. We again note that this is a proof-of-concept as discussed above.

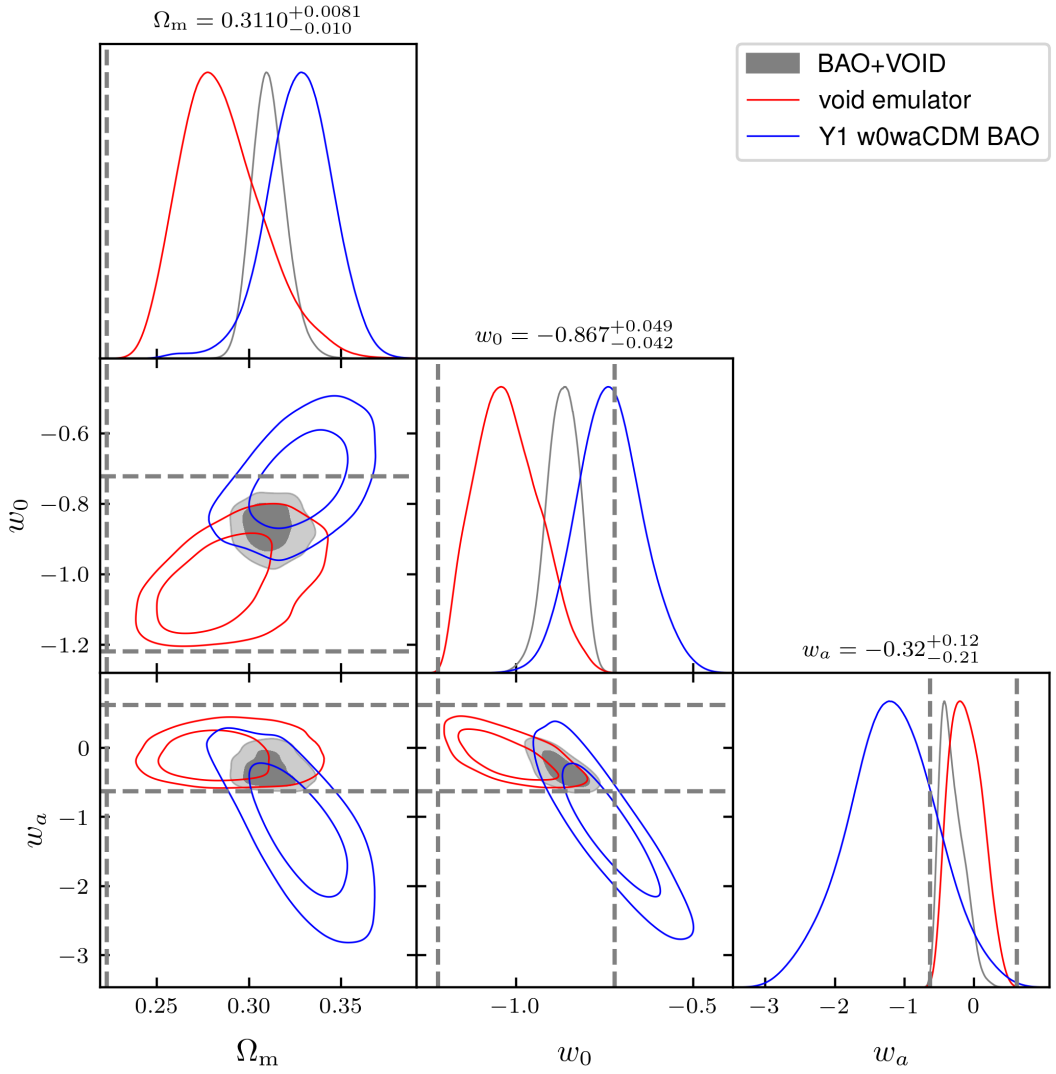

Figure A.7: Joint likelihood analysis combining the BAO+DESY5 result (blue) with the void emulator (red). The grey dashes demarcate the prior walls for the void emulator, and the grey contour denotes the joint likelihood. The joint likelihoods help produce dark energy constraints that are competitive with the BAO+SN+CMB results.

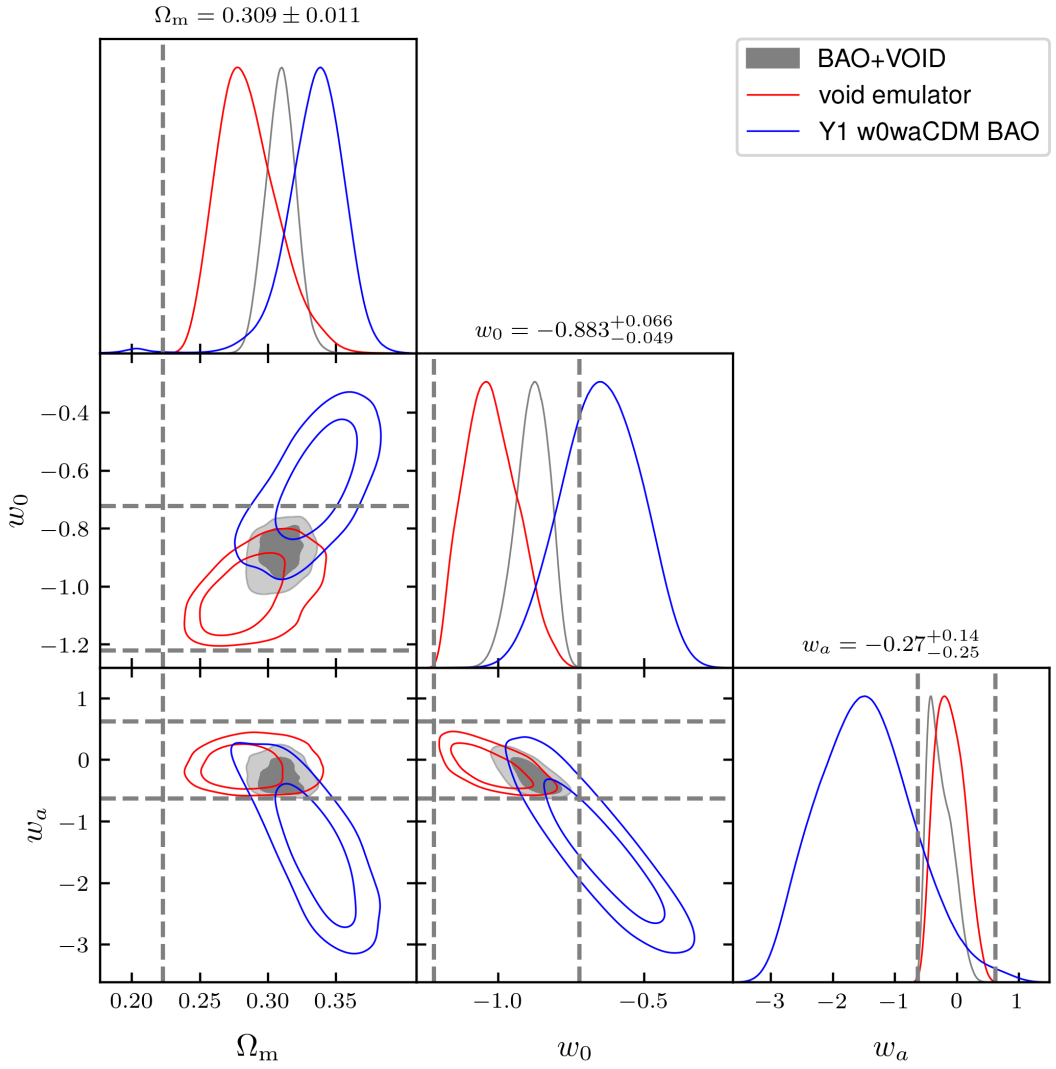

Figure A.8: Joint likelihood analysis combining the BAO+Union3 result (blue) with the void emulator (red). The grey dashes demarcate the prior walls for the void emulator, and the grey contour denotes the joint likelihood.

Our analysis in this subsection points to a similar conclusion: the potential constraining power, if the void-emulator could obtain w_0, w_a constraints once it is trained with the VG-CCF at multiple redshift bins could offer an interesting avenue to explore the complementarity of its information compared to the BAO and BAO+SNIa datasets.

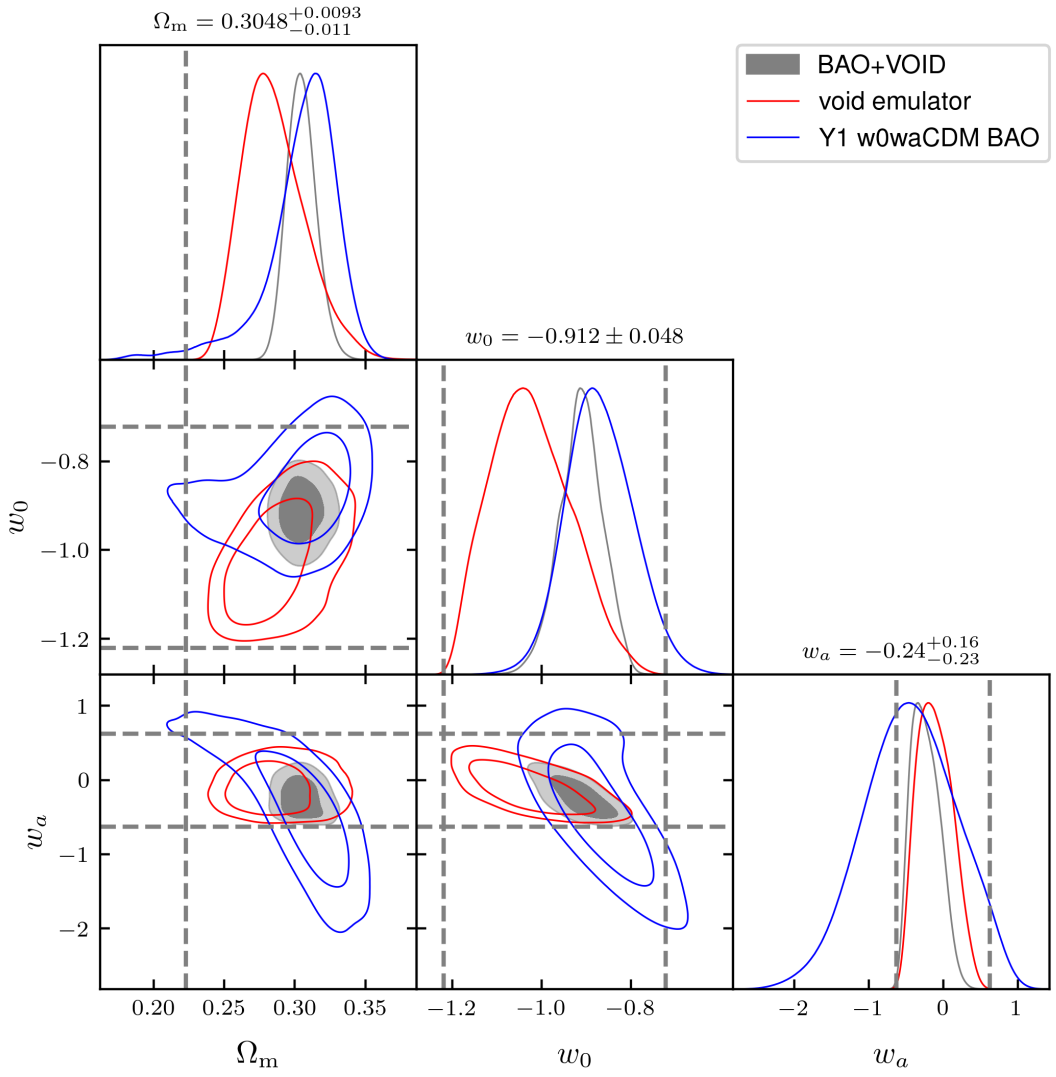

Figure A.9: Joint likelihood analysis combining the BAO+PantheonPlus result (blue) with the void emulator (red). The grey dashes demarcate the prior walls for the void emulator, and the grey contour denotes the joint likelihood.