Capturing the Cosmic Web for Cosmology

Krishna Naidoo

Under the supervision of
Prof. Ofer Lahav
Dr. Benjamin Joachimi

Submitted for the degree of
Doctor of Philosophy

Department of Physics & Astronomy
University College London
To my Father,
I wish you were still with us,
I wish you could have been a part of this,
I hope you are proud.
I, Krishna Naidoo, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.
Abstract

In this thesis the minimum spanning tree (MST) is developed to infer cosmological parameters from future galaxy redshift surveys. Studies of the distribution of galaxies typically determine constraints on cosmological models by measuring how galaxies are clustered using two-point statistics. However, these techniques do not extract all the information present, especially from the non-linear cosmic web. To incorporate this information the MST is used and is shown to improve constraints on the sum of neutrino masses and parameters from the standard model of cosmology ($\Lambda$CDM).

In Chapter 2 the MST is introduced and developed in a cosmological context. The MST is shown to be sensitive to information additional to that provided by the power spectrum or the bispectrum. In Chapter 3 the Quijote simulations are used to derive Fisher matrix constraints on the MST and power spectrum. The MST is shown to be much more sensitive to the sum of neutrino masses and when combined with the power spectrum the constraints on $\Lambda$CDM parameters improve by a factor $\sim 2$. In Chapter 4 the MST is constructed on BOSS galaxies and shown to be consistent with $\Lambda$CDM mocks.

In the latter Chapters, the effect of large scale structure evolution on light from the cosmic microwave background, known as the integrated Sachs-Wolfe (ISW), is considered. In Chapter 5 voids in $\Lambda$CDM are shown to be correlated with the Cold Spot on the CMB. Therefore the discovery of such voids is not evidence of new physics. In Chapter 6, ISW maps for the MICE and Flagship simulations are constructed. These maps will enable studies of the ISW from a larger parameter space. In Chapter 7 the work presented in this thesis is summarised and future work is discussed.
In this thesis techniques are developed to maximise the information extracted from future galaxy redshift surveys. In Chapters 2, 3 and 4 the minimum spanning tree (MST) is developed to incorporate the cosmic web for cosmological parameter estimations from large scale structure. The technique is part of a growing effort, by the cosmology community as a whole, to extract as much information from the distribution of galaxies. However, much of this effort has typically been concentrated towards artificial intelligence (AI) and machine learning (ML) techniques where an understanding of what is being learnt is often limited. The MST is unique in that its measures remain physically interpretable. This allows the MST to be an intermediate measure between conventional correlation functions and ‘black box’ methods such as AI/ML. The MST has had a broad range of applications in astronomy (cosmic web classification and mass segregation in star clusters) and outside of astronomy (including applications in particle physics, epidemiology, social science and computer science). During the course of these studies I developed the public PYTHON MST package MiSTree to enable future MST studies. In Chapter 6 integrated Sachs-Wolfe (ISW) maps for the MICE and Flagship lightcone simulations are produced. These simulations are core to the creation of mock galaxy catalogues for the current and next generation of galaxy surveys, such as DES, DESI, Euclid, LSST, PFS and 4MOST. These maps can be used for future cross-correlation studies with the cosmic microwave background which could be pivotal in understanding whether dark energy is more than just a cosmological constant. Furthermore, the pipeline developed to construct these maps is significantly faster than the standard technique and will facilitate the study of the ISW from simulations exploring a larger parameter space.
Acknowledgements

First and foremost I would like to thank my supervisors Ofer Lahav and Benjamin Joachimi for their guidance and support during the course of my PhD.

I want to especially thank Ofer Lahav for guiding me during these last six years, first as an MSci student and in the last four years as a PhD student. You have given me the freedom to explore new research ideas, while at the same time helping me to develop and grow as a scientist. In particular, I’ve grown to appreciate your ability to nudge me towards the right answers whenever I’ve been stuck or confused. I’ve always left our meetings feeling very excited to get back to work which I think is down to your very encouraging and motivating nature. I also want to thank you for being so understanding when I was faced with obstacles beyond my control. It has been a pleasure to work with you and to be a part of your research group.

I would like to thank Lorne Whiteway for always being happy to discuss and aid my research, for always being ready to read through my numerous paper drafts and for providing much needed advice when I was making difficult life choices. I often have conversations with other PhD students and often think I’m very lucky I can speak to Lorne when I’m puzzled.

In addition, I would like to thank Aurélien Benoit-Lévy, Andreu Font-Ribera, Steve Fossey, Will Hartley, Donnacha Kirk, Bruno Moraes and Roger Wesson for always being available to discuss bits of science, problems in and out of academia and for giving advice and guidance when I was so often confused.

Of course, I would be remiss if I did not discuss the people that helped me come down this very difficult and strange path. In many ways I’m an intruder, academia is very middle class – how a poor boy without a care for cheese or wine finds himself here is a bit of a mystery. I owe this in part to my family and my own stubborn reluctance to work on anything I find boring.

Firstly I want to thank my mother, despite the hardships she has faced, she has never given up. She fostered my love for space very early on and has always given me the love and support I needed as well as instilled a bullish never-give-up attitude in me. To my sister, since we were running in the gardens in Mauritius, you have always been very supportive of what I wanted to do. I’ve never felt pressure from either of you to ‘cash in’ on my education. You always allowed me to pursue what I was most passionate about.
ACKNOWLEDGEMENTS

Over the years the Pakkiri family have always been very kind to us. You helped us in our darkest days and provided a community which we had sorely been lacking since we left Mauritius. Ricco, you were very important in my early development. Not only did you give me the opportunity for a brighter future, you were also always on hand to try to guide me through difficult decisions. I only really went against your advice once, in deciding to choose Physics instead of Economics at University, but I think you can agree that decision turned out quite well in the end.

To Uncle Ram, Aunty Frances and Jodi, thank you for all the support and guidance you have provided me and my family throughout my life but especially over the last few years. These years have been difficult for us but you gave me the strength to keep going.

To my friends from my undergraduate days at UCL: Yatra Desai, Krishanth “Larry” Kangatharan, Arfa Karani, Kirit Mohane, Gautam Prasad, David Price, Luxshman Ravichandran and Baqar Rizvi, thank you for being there for me and for always making me laugh, even at my own expense.

Finally Jasleen Matharu, thank you for being who you are. I needed you throughout this and cannot imagine doing this without you. As you would say, you’re “(I’m) the best.”
Contents

1 Introduction 25
  1.1 The Standard Model of Cosmology 26
    1.1.1 The Cosmological Principle 26
    1.1.2 The Friedmann-Lemaître-Robertson-Walker Metric 26
    1.1.3 Redshift and Distances 28
    1.1.4 Hubble-Lemaître Law 30
    1.1.5 Einstein Field Equations 31
    1.1.6 Dynamics of a Smooth Universe 32
    1.1.7 Constituents of the Universe 35
    1.1.8 Epochs of Domination and Accelerated Expansion 40
    1.1.9 Inflation 41
    1.1.10 Initial Conditions to Large Scale Structure 43
  1.2 Observational Cosmology 48
    1.2.1 Cosmological Probes 48
    1.2.2 Standard Measures 51
    1.2.3 The Cosmic Web 57
    1.2.4 The Integrated Sachs-Wolfe Effect 58
    1.2.5 Beyond ΛCDM 59
    1.2.6 Surveys: Current and Future 61
    1.2.7 Thesis Roadmap 63

2 Using the Minimum Spanning Tree as a Tool for Cosmology 67
  2.1 Abstract 67
  2.2 Introduction 67
  2.3 Method 69
    2.3.1 Statistics from the Minimum Spanning Tree 71
    2.3.2 Error Estimation 73
    2.3.3 Simulation Summary 73
  2.4 Sensitivity of the MST to Cosmic Web Patterns 75
    2.4.1 Heuristic Argument 75
    2.4.2 Illustris vs. Adjusted Lévy Flight 75
  2.5 Boundary Effects and Redshift Space Distortions 77
## CONTENTS

2.5.1 Boundary Effects ........................................... 77
2.5.2 Scale Cuts .................................................. 78
2.5.3 Redshift Space Distortion on MICE Galaxies .......... 78

2.6 What Does the Minimum Spanning Tree Measure? ............ 79
   2.6.1 $\nu N$-body Simulations ................................ 79
   2.6.2 MST Application to Dark Matter Particles .............. 80
   2.6.3 MST Application to Haloes .............................. 87

2.7 Comparing the Sensitivity to Cosmology of the Power Spectrum, Bispectrum and the Minimum Spanning Tree .......... 88
   2.7.1 COLA Simulation Suites ................................ 88
   2.7.2 Measurements ........................................... 89
   2.7.3 Parameter Estimation ................................... 91
   2.7.4 Comparison ............................................. 93

2.8 Discussion .................................................... 95

3 Cosmology and Neutrinos with the Minimum Spanning Tree 101
   3.1 Abstract ................................................... 101
   3.2 Introduction ............................................... 101
   3.3 Method ...................................................... 103
      3.3.1 Fisher Formalism .................................... 103
      3.3.2 Summary Statistics in Real and Redshift Space .... 104
      3.3.3 Quijote Simulations ................................ 105
   3.4 Result ....................................................... 106
      3.4.1 Covariance Matrix .................................... 107
      3.4.2 Fisher Matrix and Partial Derivatives ............... 112
      3.4.3 Sensitivity to Neutrinos and $\Lambda$CDM ............. 113
   3.5 Discussion ................................................ 120

4 Applying the Minimum Spanning Tree to Test $\Lambda$CDM with BOSS 125
   4.1 Abstract ................................................... 125
   4.2 Introduction ............................................... 126
   4.3 Method ...................................................... 126
      4.3.1 Two-Point Multipoles ............................... 127
      4.3.2 Minimum Spanning Tree ............................. 127
      4.3.3 Weights for Point Processes ....................... 128
   4.4 BOSS Galaxies and Mocks ................................ 129
   4.5 Results .................................................... 130
      4.5.1 Two-Point Multipoles with Stochastic Integer Weights 130
      4.5.2 Stochastic Binary Weights and Density Corrections 131
7 Conclusions and Future Work 175
  7.1 Thesis Summary ............................................. 175
  7.2 Cosmology in the Next Decade ............................. 177
  7.3 Future Work ................................................... 178
    7.3.1 Parameter Constraints from the Minimum Spanning Tree . 178
    7.3.2 Exploring the Integrated Sachs-Wolfe for Different Cosmologies . 179

Appendix 183
  A.1 Gaussian Process Interpolation ................................ 183
    A.1.1 Introduction ............................................. 183
    A.1.2 Kernel .................................................... 184
    A.1.3 Hyperparameter Optimization ................................ 184
    A.1.4 Implementation and Validation ............................ 185
  A.2 MiSTree: a Python Package for Constructing and Analysing Minimum
    Spanning Trees .................................................. 187
    A.2.1 Summary ................................................. 187
    A.2.2 Motivation .............................................. 187
    A.2.3 MiSTree ................................................... 188
    A.2.4 Basic Usage ............................................. 190
    A.2.5 Dependencies ........................................... 191
  A.3 Numerical Derivative Estimators .............................. 192
    A.3.1 Symmetric Derivative Estimator .......................... 192
    A.3.2 Non-Symmetric Derivative Estimators ................... 192

References 195
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Cosmic web large scale structure evolution in $N$-body simulations.</td>
<td>48</td>
</tr>
<tr>
<td>1.2</td>
<td>The minimum spanning tree constructed from galaxies from the IllustrisTNG simulation is shown to trace the underlying cosmic web.</td>
<td>57</td>
</tr>
<tr>
<td>1.3</td>
<td>The sky area for photometric surveys are plotted against the $5\sigma$ magnitude limit for the survey’s photometric bands.</td>
<td>60</td>
</tr>
<tr>
<td>1.4</td>
<td>The sky area for spectroscopic surveys are plotted against the redshift range for different intended targets.</td>
<td>61</td>
</tr>
<tr>
<td>2.1</td>
<td>An example of the minimum spanning tree and the statistics measured.</td>
<td>70</td>
</tr>
<tr>
<td>2.2</td>
<td>A comparison of the 2-point correlation function and minimum spanning tree statistics measured from the Illustris $N$-body simulation and Lévy Flight simulation.</td>
<td>74</td>
</tr>
<tr>
<td>2.3</td>
<td>Minimum spanning tree statistics measured with and without a sky mask.</td>
<td>76</td>
</tr>
<tr>
<td>2.4</td>
<td>The effects of redshift space distortions on the minimum spanning tree.</td>
<td>78</td>
</tr>
<tr>
<td>2.5</td>
<td>Minimum spanning tree statistics as a function of redshift and $\nu N$-body cosmology.</td>
<td>81</td>
</tr>
<tr>
<td>2.6</td>
<td>Minimum spanning tree statistics as a function of density and redshift.</td>
<td>82</td>
</tr>
<tr>
<td>2.7</td>
<td>Power spectra for the $\nu N$-body simulations.</td>
<td>83</td>
</tr>
<tr>
<td>2.8</td>
<td>Measuring the minimum spanning tree sensitivity to simulation resolution.</td>
<td>85</td>
</tr>
<tr>
<td>2.9</td>
<td>Minimum spanning tree statistics measured on haloes from the $\nu N$-body simulations.</td>
<td>86</td>
</tr>
<tr>
<td>2.10</td>
<td>PICOLA power spectrum accuracy.</td>
<td>89</td>
</tr>
<tr>
<td>2.11</td>
<td>PICOLA bispectrum accuracy.</td>
<td>90</td>
</tr>
<tr>
<td>2.12</td>
<td>Posterior constrains from the individual minimum spanning tree statistics.</td>
<td>92</td>
</tr>
<tr>
<td>2.13</td>
<td>Comparison of the constraints from the power spectrum, bispectrum and minimum spanning tree.</td>
<td>93</td>
</tr>
<tr>
<td>2.14</td>
<td>Combined constraints from the power spectrum, bispectrum and minimum spanning tree.</td>
<td>94</td>
</tr>
<tr>
<td>2.15</td>
<td>Marginalised constraints on $A_s$ and $\Omega_m$ for the power spectrum, bispectrum and minimum spanning tree.</td>
<td>94</td>
</tr>
<tr>
<td>3.1</td>
<td>The correlation matrix for the minimum spanning tree in redshift space.</td>
<td>107</td>
</tr>
</tbody>
</table>
3.2 The correlation matrix for the power spectrum multipoles and minimum spanning tree in redshift space. ........................................ 108
3.3 Derivatives for the power spectrum monopole. .......................... 109
3.4 Derivatives for the power spectrum quadrupole. .......................... 109
3.5 Derivatives for the minimum spanning tree degree. ...................... 110
3.6 Derivatives for the minimum spanning tree edge length. ............... 110
3.7 Derivatives for the minimum spanning tree branch length. ........... 111
3.8 Derivatives for the minimum spanning tree branch shape. .......... 111
3.9 Fisher matrix constraints for $\nu\Lambda$CDM parameters from the minimum spanning tree edge and branch length distributions at $z = 0$. .................. 116
3.10 Fisher matrix constraints for $\nu\Lambda$CDM parameters from the minimum spanning tree edge and branch length distributions in combination with the distribution of degree and branch shape at $z = 0$. ......................... 117
3.11 Fisher matrix constraints on $\nu\Lambda$CDM parameters from the power spectrum and minimum spanning tree separately and combined in redshift space at $z = 0$. .................................................... 118
3.12 Fisher matrix constraints on $\nu\Lambda$CDM parameters from the power spectrum and minimum spanning tree separately and combined from redshift $z = 0, 0.5$ and $1$ in redshift space. ........................................ 119

4.1 The 2-point correlation function monopole and quadrupole for BOSS real galaxies and PATCHY mocks. ............................... 131
4.2 The stochastic binary weighted histogram for BOSS real and mock galaxies with respect to redshift. ...................................... 132
4.3 The density-corrected stochastic binary weighted histogram for BOSS real and mock galaxies with respect to redshift. ...................... 132
4.4 The total stochastic binary weights and total nodes for BOSS real and mock galaxies in the final catalogues. ............................. 133
4.5 The minimum spanning tree statistics for BOSS galaxies and PATCHY mocks. ................................................................. 135
4.6 Correlation matrix for the minimum spanning tree statistics measured from PATCHY mocks. .................................................. 136
4.7 The $\chi^2$ for the minimum spanning tree statistics for real and PATCHY galaxies. ............................................................... 137

5.1 The mask and effective mask used for the spherical mexican hat wavelet. .......................... 144
5.2 The relation between the full-sky and cut-sky minima. .............. 147
5.3 Temperature profiles for full- and cut-sky minima. ...................... 148
5.4 The significance of the Cold Spot as a function of mask fraction. 149
5.5 Predicted integrated Sachs-Wolfe effect in the line-of-sight of the Cold Spot. 150
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1</td>
<td>The comoving distances in the MICE lightcone is shown for a range of redshifts in comparison to the MICE simulation box.</td>
</tr>
<tr>
<td>6.2</td>
<td>Displaying the boundary effects from calculations of the spherical Fourier Bessel transform with different $R_{\text{max}}$.</td>
</tr>
<tr>
<td>6.3</td>
<td>A comparison of theoretical auto- and cross-$C_\ell$ for the density and integrate Sachs-Wolfe maps.</td>
</tr>
<tr>
<td>6.4</td>
<td>The impact of periodic boxes on the integrated Sachs-Wolfe angular power spectra.</td>
</tr>
<tr>
<td>6.5</td>
<td>Integrated Sachs-Wolfe maps for the MICE lightcone simulations.</td>
</tr>
<tr>
<td>6.6</td>
<td>Integrated Sachs-Wolfe maps for the Flagship lightcone simulations.</td>
</tr>
<tr>
<td>A.1.1</td>
<td>The position of PICOLA simulations in parameter space.</td>
</tr>
<tr>
<td>A.1.2</td>
<td>Gaussian process residuals for the predicted power spectra, bispectrum and minimum spanning tree data vectors.</td>
</tr>
<tr>
<td>A.2.1</td>
<td>MiSTree's minimum spanning tree construction process.</td>
</tr>
<tr>
<td>A.2.2</td>
<td>An example of the minimum spanning tree histogram plots produced by MiSTree.</td>
</tr>
</tbody>
</table>
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Growing and decaying modes in the Newtonian approximation for matter perturbations as a function of time and the scale factor.</td>
<td>46</td>
</tr>
<tr>
<td>1.2</td>
<td>Relativistic growth of matter perturbations for subhorizon scales and superhorizon scales as a function of comoving time.</td>
<td>47</td>
</tr>
<tr>
<td>2.1</td>
<td>A summary of simulations used in this study.</td>
<td>73</td>
</tr>
<tr>
<td>2.2</td>
<td>Simulation names and cosmological parameters for the $\nu N$-body simulations.</td>
<td>79</td>
</tr>
<tr>
<td>2.3</td>
<td>Haloes counts in the $\nu N$-body simulations.</td>
<td>87</td>
</tr>
<tr>
<td>2.4</td>
<td>A summary of the COLA simulation suites.</td>
<td>89</td>
</tr>
<tr>
<td>3.1</td>
<td>The Quijote simulation parameters and realisations used in this study.</td>
<td>106</td>
</tr>
<tr>
<td>3.2</td>
<td>The sensitivity of the neutrino mass constraint to different partial derivative estimators.</td>
<td>113</td>
</tr>
<tr>
<td>3.3</td>
<td>Separate and combined constraints for parameters from the $\nu\Lambda$CDM model determined from measurements of the power spectrum and minimum spanning tree.</td>
<td>114</td>
</tr>
<tr>
<td>5.1</td>
<td>The probability that the measured cut-sky minima is equivalent to the full-sky minima is given as a function of the angular scale of the spherical Mexican hat wavelet.</td>
<td>151</td>
</tr>
<tr>
<td>6.1</td>
<td>A summary of the cosmological parameters used to generate the MICE and Flagship simulations.</td>
<td>163</td>
</tr>
<tr>
<td>6.2</td>
<td>A summary of the calculated spherical Fourier Bessel coefficients used to generate the integrated Sachs-Wolfe maps.</td>
<td>163</td>
</tr>
<tr>
<td>6.3</td>
<td>A summary of the integrated Sachs-Wolfe maps generated in this analysis.</td>
<td>167</td>
</tr>
</tbody>
</table>
“Anybody can be happy and cozy.
Nothing good in the world happens by being happy and cozy.
No one achieves anything great because they’re happy and cozy.”

– Alex Honnold, Free Solo (2018)
Astronomy is an ancient science, ever since humans have roamed the Earth they have looked up into the heavens and wondered what our role was in the vast expanse of space. It is clear now that astronomy played an important role in the development of civilisation; the location of stars in the sky were critical for the production of agriculture and later for navigation particularly across the oceans. This dependence on the heavens, made the ancients revere the skies which is seen in the role constellations played in religion, myths and legends. While astronomy has detached from its religious connections many of the fundamental questions still remain. How was the Universe created, what laws govern it and how will it end? These are all question which lie at the heart of the subject of cosmology.

Cosmology, as we know it now, is a relatively new science. This is due in part to the way our definition of the Universe has expanded over history – only a century ago the milky way was thought to be the entire Universe, with galaxies in the distance being mistaken for nebulae. When it was realised that some of these nebulae were in fact galaxies like our own the size of the Universe became unimaginably large. On these enormous scales, gravity is the force that governs the dynamics and evolution of the Universe, for which we rely on general relativity (GR; Einstein, 1916). The discovery by Hubble (1929) that galaxies further away from us were receding faster than those nearest to us showed the Universe was expanding. If we rewind the clock this means the Universe was much denser earlier on and would eventually approach a singularity, this phenomenon was derisively called the ‘Big Bang’ by Fred Hoyle but the idea is now widely accepted with the dismissive name sticking. More recently observations of Type Ia supernova showed distant galaxies were further away than expected (Riess et al., 1998; Perlmutter et al., 1999). To reconcile theory with observations a new constituent was required called dark energy which exhibits negative pressure. Coupled with the need for dark matter to explain the rotation velocities of stars in galaxies (Rubin et al., 1980) we are now left with a Universe where almost 95% of it is unknown (in the form of dark matter and dark energy) with only 5% being in the form of ‘normal’ matter.

Fundamental to determining the nature of dark matter, dark energy and the physics of the Universe are cosmological probes that observe aspects of the Universe such as the cosmic microwave background (CMB) or the large scale distribution of galaxies. The latter
is of particular interest to this thesis, since observations of galaxies will be the focus of many new galaxy redshift surveys. Typically in cosmology we rely on measurements of clustering using two-point statistics in real or Fourier space. However, these statistics fails to capture the complicated structure of voids, walls, filaments and clusters collectively known as the cosmic web which dominate the structure of the Universe at late times. The cosmic web is difficult to describe analytically and has so far been unexploited but could be crucial in deriving constraints on cosmological parameters. In this thesis I focus on using the minimum spanning tree, a technique widely used for filament finding, to statistically incorporate cosmic web information for parameter estimation. I later explore the effects of large scale structure, in particular voids, on the CMB via the integrated Sachs-Wolfe effect.

The introduction is organised as follows: in Section 1.1 I will explain the fundamental concepts and constituents of the standard model of cosmology and the evolution of large scale structure and in Section 1.2 I will describe the probes, methods and problems in observational cosmology which are addressed in this thesis.

1.1 THE STANDARD MODEL OF COSMOLOGY

1.1.1 The Cosmological Principle

The Cosmological Principle is based on the Copernican Principle which simply states “the Universe looks the same whoever and wherever you are” (Liddle, 2015). In other words there are no special observers and the part of the Universe we live in is a fair representation of the whole. The consequence of this principle is the Universe must be both homogeneous and isotropic. This is consistent with observations made over large scales (\(\sim 100 h^{-1}\)Mpc). It is important to note that this is different to the Perfect Cosmological Principle where the Universe is homogeneous and isotropic both in space and time. This symmetry is philosophically appealing and underpins the core ideas behind steady state theories. Unfortunately, observations of the Universe show clear time evolution, and therefore in order for the Cosmological Principle to be consistent with observations it can only be true in space and not in time. Lemaitre (1927) showed that the equations for a static universe in GR were unstable and would result in a universe that was either contracting or expanding. In other words, GR appeared to naturally predict the time evolution we see in observations.

1.1.2 The Friedmann-Lemaître-Robertson-Walker Metric

GR underpins the foundations of cosmology. In order to describe the fundamental properties of the Universe we need to first define the metric, which describes the geometry of the spacetime fabric. This is often described by the length of an infinitesimal distance on
the metric using Einstein’s summation convention
\[ ds^2 = g_{\mu\nu} dx^\mu dx^\nu \equiv \sum_{\mu=0}^{3} \sum_{\nu=0}^{3} g_{\mu\nu} dx^\mu dx^\nu, \] (1.1)

where \( g_{\mu\nu} \) is the metric tensor and \( x \) is a position four-vector. For a completely flat and static Universe we can define the geometry with the Minkowski metric tensor \( \eta_{\mu\nu} = \text{diag}(c^2, -1, -1, -1) \) using the (+, −, −, −) signature,
\[ ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu = c^2 dt^2 - (dx^2 + dy^2 + dz^2), \] (1.2)

where \( c \) is the speed of light, \( t \) is time and \( x, y \) and \( z \) are the conventional 3D Cartesian coordinates. In cosmology we need to be able to describe an expanding, homogeneous and isotropic Universe. So we introduce a term \( a \), called the scale factor, which relates the distance between objects at the present cosmological time to the distance between objects in the past. This coordinate system is called comoving coordinates. The metric can be defined as
\[ ds^2 = c^2 dt^2 - a^2 (dx^2 + dy^2 + dz^2), \] (1.3)
in comoving Cartesian coordinates. This metric is known as the Friedmann-Lemaître-Robertson-Walker (FLRW) metric in Euclidean space and can be defined in spherical co-moving coordinates as
\[ ds^2 = c^2 dt^2 - a^2 [d\chi^2 + S_\kappa^2(\chi) (d\theta^2 + \sin^2 \theta d\phi^2)], \] (1.4)

where \( \chi \) is the comoving radial coordinate, \( \kappa \) defines the curvature of the universe and \( \phi \) and \( \theta \) are the longitudinal and latitudinal angles on the sphere. The function \( S_\kappa(\chi) \) is defined as
\[ S_\kappa(\chi) = \begin{cases} \frac{1}{\sqrt{\kappa}} \sin \left( \chi \sqrt{\kappa} \right), & \text{for } \kappa > 0, \\ \chi, & \text{for } \kappa = 0, \\ \frac{1}{\sqrt{|\kappa|}} \sinh \left( \chi \sqrt{|\kappa|} \right), & \text{for } \kappa < 0. \end{cases} \] (1.5)

Positive curvature (\( \kappa > 0 \)) is associated with a closed universe with spherical geometry, zero curvature (\( \kappa = 0 \)) is associated with an open universe with Euclidean geometry and negative curvature (\( \kappa < 0 \)) is associated with an open universe with hyperbolic geometry.
This is sometimes written in reduced-circumference polar coordinates as
\[ ds^2 = c^2 dt^2 - a^2 \left[ \frac{dr^2}{1 - \kappa r^2} + r^2 \left( d\theta^2 + \sin^2 \theta d\phi^2 \right) \right], \] (1.6)
which has the benefit of not requiring conditions for curvature but the drawback is \( r \) is no longer the radial axis in spherical coordinates but the radial axis in polar coordinates embedded in three dimensions.

### 1.1.3 Redshift and Distances

#### 1.1.3.1 Redshift

Due to the expansion of the Universe, light from distant sources are Doppler shifted to redder wavelengths. The redshift \( z \) is defined as the ratio between the shift in wavelength \( \Delta \lambda \) and the emitted wavelength \( \lambda_e \),
\[ z = \frac{\Delta \lambda}{\lambda_e} = \frac{\lambda_o - \lambda_e}{\lambda_e}, \] (1.7)
where \( \lambda_o \) is the observed redshifted wavelength. The scale factor can be used to relate the observed and emitted wavelengths, \( \lambda_e = a \lambda_o \). This allows us to derive the relation between the redshift and the scale factor,
\[ a = \frac{1}{1 + z}. \] (1.8)

The relativistic Doppler shift relates the ratio of the observed and emitted wavelengths by
\[ \frac{\lambda_o}{\lambda_e} = \frac{1 + \frac{v}{c}}{\sqrt{1 - \frac{v^2}{c^2}}}, \] (1.9)
where \( v \) is the velocity of the source moving away from the observer. When \( v \ll c \) this implies \( v^2/c^2 \ll v/c \) and therefore this can be approximated as
\[ \frac{\lambda_o}{\lambda_e} \simeq 1 + \frac{v}{c} \] (1.10)
which is the normal Doppler shift equation. The redshift can therefore be approximated when \( v \ll c \) to
\[ z \simeq \frac{v}{c}. \] (1.11)
1.1.3.2 Distances

The comoving distance $\chi$ can be simply calculated by taking the path of a photon (a null geodesic) in the FLRW metric (equation 1.4). This means setting $ds = 0$ and to simplify the calculation further, we will only consider the radial component (i.e. setting $d\theta = d\phi = 0$). The comoving distance is found to be

$$\chi = c \int_{t_{e}}^{t_{o}} \frac{dt}{a},$$  \hfill  (1.12)

where $t_{o}$ is the time for the observer and $t_{e}$ is the time the light was emitted. This can be rewritten as a function of the scale factor,

$$\chi = \frac{c}{H_{0}} \int_{a}^{1} \frac{1}{a'^{2}E(a')} da',$$  \hfill  (1.13)

where the Hubble expansion is defined as

$$H(a) = H_{0}E(a) = \frac{\dot{a}}{a},$$  \hfill  (1.14)

where $H_{0}$ is the Hubble constant, $E(a)$ is the normalised Hubble parameter and the dot represents a derivative with respect to time. The transverse comoving distance $\chi_{T}$ is found to be

$$\chi_{T} = S_{e}(\chi),$$  \hfill  (1.15)

for a small angle $\delta \theta$ this is given by $\chi_{T}\delta \theta$. Similarly, the particle horizon $\chi_{H}$ is defined as the conformal time $\eta$ that has elapsed since the Big Bang multiplied by the speed of light. This is identical to equation 1.12 except we integrate from the Big Bang (i.e. $t_{e} = 0$) to the current time $t$,

$$\chi_{H} = c\eta = c \int_{0}^{t} \frac{dt}{a} \implies \chi_{H} = \frac{c}{H_{0}} \int_{0}^{a} \frac{1}{a'^{2}E(a')} da'.$$  \hfill  (1.16)

The angular diameter distance $D_{A}$ is defined as the ratio between the proper size of an object $X$ and the angle it subtends on the sky,

$$D_{A} = \frac{X}{\theta}.$$  \hfill  (1.17)

Since objects are observed at a time $t$ they are observed with a proper distance equal to $a(t)\chi_{T}$ and thus the angular diameter distance is defined as

$$D_{A} = a\chi_{T} \implies D_{A} = \frac{\chi_{T}}{1 + z}.$$  \hfill  (1.18)
INTRODUCTION

The luminosity distance \( D_L \) is defined by the relationship between the bolometric flux \( F \) (integrated over all frequencies) and bolometric luminosity \( L \),

\[
D_L = \sqrt{\frac{L}{4\pi F}}. \tag{1.19}
\]

We recall that the expansion of the universe means the transverse comoving distance increases by \( a\chi_T \) which means for the above relation to hold,

\[
D_L = \frac{\chi_T}{a} = \frac{D_A}{a^2} \implies D_L = (1 + z)\chi_T = (1 + z)^2 D_A \tag{1.20}
\]

1.1.4 Hubble-Lemaître Law

The Hubble-Lemaître law describes, to first approximation, the relation between the recession velocity of galaxies and their distances to an observer caused by the expansion of the Universe. Taking the Taylor series expansion of \( a \) at a time \( t_0 \) (i.e. the present day) we can find \( a \) at \( t \) using

\[
a(t) = \frac{1}{1 + z} = 1 + H_0(t - t_0) - \frac{q_0H_0^2}{2}(t - t_0)^2 + ..., \tag{1.21}
\]

where the subscripts 0 indicate quantities evaluated at the present time. Using equation 1.14 the Hubble constant is defined as

\[
H_0 = \frac{\dot{a}_0}{a_0}, \tag{1.22}
\]

where \( a_0 = 1 \) and the deceleration parameter \( q(t) \) is defined as

\[
q(t) = -\frac{\ddot{a}(t)a(t)}{\dot{a}(t)^2} \implies q_0 = -\frac{\ddot{a}_0}{H_0^2}. \tag{1.23}
\]

For small distances the luminosity distance can be defined as a power law using the Taylor series expansion defined in equation 1.21 as

\[
D_L = \frac{c}{H_0} \left( z + \frac{1}{2}(1 - q_0)z^2 + ... \right). \tag{1.24}
\]

In the linear approximate this gives the Hubble-Lemaître law (Lemaitre, 1927; Hubble, 1929),

\[
z \simeq \frac{H_0}{c} D_L, \tag{1.25}
\]

where for small distances \( D_L \simeq \chi \simeq D_A \). The exact relation for the Hubble parameter are described later in Section 1.1.6 (equation 1.39).
1.1.5 Einstein Field Equations

The dynamics of the Universe are dominated by gravity and defined according to Einstein’s theory of GR. The dynamics of any gravitational state is defined by the Einstein field equations,

\[ R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} - \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}, \]  

(1.26)

where \( R_{\mu\nu} \) is the Ricci curvature tensor, \( R \) is the Ricci scalar, \( \Lambda \) is the cosmological constant, \( T_{\mu\nu} \) is the energy-momentum (or stress-energy) tensor and \( G \) is Newton’s gravitational constant. On the Right Hand Side (RHS) of equation (1.26) the energy/matter is described and on the Left Hand Side (LHS) the spacetime curvature is described. In other words “Spacetime tells matter how to move; matter tells spacetime how to curve” (Wheeler & Ford, 2000). In this particular form we choose to include the cosmological constant on the LHS as it was introduced by Einstein (1917) although in principle, as we will show later, we can choose to describe it as a rather strange fluid with negative pressure as part of the energy-momentum tensor on the RHS (Zel’dovich, 1967).

The energy-momentum tensor is a generalised tensor representation of the momentum four-vector. While a particle can be completely described by the momentum four-vector, in cosmology we need to describe extended systems such as fluids that are described by macroscopic properties such as density, pressure, entropy and viscosity. For a perfect fluid \( T_{\mu\nu} \) can be defined generally as

\[ T_{\mu\nu} = \left( \rho + \frac{P}{c^2} \right) U_{\mu} U_{\nu} - P g_{\mu\nu}, \]  

(1.27)

where \( \rho \) is the energy density, \( P \) is the pressure and \( U_{\mu} \) is the velocity four-vector of an observer \( U_{\mu} = \partial X_{\mu}/\partial \tau \) where \( \tau \) is proper time. The components \( T_{\mu\nu} \) describe the following,

- \( T_{00} \): energy density in time,
- \( T_{0i} = T_{i0} \): momentum density,
- \( T_{ij} = T_{ji} \): momentum flux or stress between neighbouring infinitesimal elements (diagonal elements represent the pressure exerted in the \( i \)th direction and off diagonal represent the shear terms due to viscosity).

We can express this in a metric-independent form by raising an index (i.e. by multiplying by the metric contravariant),

\[ T^{\mu}_{\nu} \equiv g^{\mu\gamma} T_{\gamma\nu} = \left( \rho + \frac{P}{c^2} \right) U^{\mu} U_{\nu} - P \delta^{\mu}_{\nu}. \]  

(1.28)
INTRODUCTION

For a comoving observer the velocity four-vector is \( U_\mu = (c, 0, 0, 0) \) and therefore the metric independent energy-momentum tensor is

\[
T^\mu_\nu = \begin{pmatrix}
\rho c^2 & 0 & 0 & 0 \\
0 & -P & 0 & 0 \\
0 & 0 & -P & 0 \\
0 & 0 & 0 & -P \\
\end{pmatrix}.
\]

1.1.6 Dynamics of a Smooth Universe

The expansion and evolution of the Universe is best expressed in terms of the Friedmann and continuity equations.

Continuity Equation

For a fluid with no gravity and negligible velocities pressure and energy are conserved by the continuity equation \( \partial \rho / \partial t = 0 \) and the Euler equation \( \partial P / \partial x^i = 0 \). The four component conservation equation for the energy momentum tensor becomes

\[
\frac{\partial T^\mu_\nu}{\partial x^\mu} = 0,
\]

but for an expanding Universe this has to be modified. Conservation implies the vanishing of the covariant derivative (which replaces regular partial derivatives to carry the laws of physics from flat spacetime to the curved spacetime of GR) and therefore the conservation becomes

\[
T^\mu_\nu,\mu \equiv \frac{\partial T^\mu_\nu}{\partial x^\mu} + \Gamma^\mu_{\alpha\nu} T^\alpha_\nu - \Gamma^\alpha_{\nu\mu} T^\mu_\alpha = 0,
\]

where \( \Gamma^\alpha_{\beta\gamma} \) are Christoffel symbols. If we consider \( \nu = 0 \) we arrive at the continuity equation

\[
\dot{\rho} + 3 \frac{\dot{a}}{a} \left( \rho + \frac{P}{c^2} \right) = 0.
\]

Taking the equation of state \( w = P/(\rho c^2) \) the continuity equation allows us to define the energy density as a function of the scale factor,

\[
\rho \propto a^{-3(1+w)}.
\]

The equation of state for the different constituents of the Universe are summarised below:

- Matter (dust): Pressureless \( \implies P = 0 \implies w = 0 \implies \rho \propto a^{-3} \).
• Radiation: \( P = \frac{\rho c^2}{3} \implies w = \frac{1}{3} \implies \rho \propto a^{-4} \),

• Vacuum energy: Negative pressure \( \implies P = -\rho c^2 \implies w = -1 \implies \rho \propto 1 \).

These relationships will be useful for defining the dimensionless energy density parameter \( \Omega \) described later.

**Friedmann Equations**

The components of the Einstein field equation (equation 1.26) can be defined by using the FLRW metric. The Ricci tensor is a special case of the Riemann curvature tensor defined as

\[
R^\alpha_{\beta \gamma \delta} = \frac{\partial \Gamma^\alpha_{\delta \beta}}{\partial x^\gamma} - \frac{\partial \Gamma^\alpha_{\gamma \delta}}{\partial x^\beta} + \Gamma^\alpha_{\gamma \epsilon} \Gamma^\epsilon_{\beta \delta} - \Gamma^\alpha_{\delta \epsilon} \Gamma^\epsilon_{\beta \gamma},
\]

where \( \alpha = \gamma \),

\[
R_{\mu \nu} = R^\alpha_{\mu \alpha \nu}.
\]

For the FLRW metric, only the diagonal elements are non-zero. The temporal component (when \( \mu = \nu = 0 \)) is given by

\[
R_{00} = -3 \frac{\dot{a}}{a},
\]

the spatial components (when \( \mu = \nu = 1, 2 \) or 3) are given by

\[
R_{ii} = -\frac{g_{ii}}{a^2 c^2} \left[ \ddot{a} a + 2 \dot{a}^2 + 2 \kappa c^2 \right],
\]

and the Ricci scalar defined as \( \mathcal{R} = g^{\mu \nu} R_{\mu \nu} \) is determined to be

\[
\mathcal{R} = -\frac{6}{c^2} \left[ \frac{\ddot{a}}{a} + \left( \frac{\dot{a}}{a} \right)^2 + \frac{\kappa c^2}{a^2} \right].
\]

Taking the temporal component of the Einstein Field equation (\( \mu = \nu = 0 \)) we eventually arrive at the first Friedmann equation,

\[
\left( \frac{\dot{a}}{a} \right)^2 = \frac{8\pi G}{3} \rho + \frac{\Lambda c^2}{3} - \frac{\kappa c^2}{a^2},
\]

which is also known as the Hubble parameter \( H(t) = \dot{a}/a \). The second Friedmann equation can be derived by taking any of the spatial components (\( \mu = \nu = 1, 2 \) or 3),

\[
\frac{\ddot{a}}{a} = -\frac{4\pi G}{3} \left( \rho + \frac{3P}{c^2} \right) + \frac{\Lambda c^2}{3}.
\]
The Friedmann equations can be redefined with the dimensionless energy density

\[
\Omega = \frac{8\pi G}{3H^2} \frac{\rho}{\rho_c},
\]

(1.41)

where \(\rho_c\) is the critical density. This allows us to derive how the energy density is related to the overall curvature of the Universe. We start by rewriting the first Friedmann equation (equation 1.39) as

\[
1 = \Omega + \frac{\Lambda c^2}{3H^2} - \frac{\kappa c^2}{a^2 H^2}.
\]

(1.42)

In our derivation of equation 1.39 and 1.40 we have taken the cosmological constant on the LHS of the Einstein field equation. If we instead take this to be a perfect fluid with negative pressure we can take it to be a component of the energy-momentum tensor (on the RHS of the Einstein field equation). This means

\[
\Omega + \Omega_\Lambda = \frac{8\pi G}{3H^2} \frac{\Lambda c^2}{3H^2},
\]

(1.43)

and therefore

\[
\Lambda = \frac{8\pi G}{c^2} \rho_\Lambda.
\]

(1.44)

This can then be absorbed in \(\Omega\) which is the sum of energy densities for all the constituents in the Universe. Additionally it is convenient to also express curvature in this form by substituting

\[
\Omega_\kappa = -\frac{\kappa c^2}{a^2 H^2}.
\]

(1.45)

This means

\[
\Omega + \Omega_\kappa = 1,
\]

(1.46)

and since \(\Omega = \rho/\rho_c\), the curvature of the Universe is simply defined by \(\rho\) in relation to \(\rho_c\):

- \(\rho < \rho_c \Rightarrow \Omega < 1 \Rightarrow \kappa < 0 \Rightarrow \text{open Universe},\)
- \(\rho = \rho_c \Rightarrow \Omega = 1 \Rightarrow \kappa = 0 \Rightarrow \text{flat Universe},\)
- \(\rho > \rho_c \Rightarrow \Omega > 1 \Rightarrow \kappa > 0 \Rightarrow \text{closed Universe}.\)

It is more common to describe the Friedmann equations in terms of the current Hubble expansion \(H_0\), current critical density \(\rho_{c,0} = 8\pi G/3H_0^2\) and current dimensionless energy
density,
\[ \Omega_{i,0} = \frac{8\pi G}{3H_0^2} \rho_{i,0} = \frac{\rho_{i,0}}{\rho_{c,0}}. \] (1.47)

Therefore, equation 1.39 can be rewritten as
\[ H^2(a) = H_0^2 \left( \frac{\Omega_{r,0}}{a^4} + \frac{\Omega_{m,0}}{a^3} + \frac{\Omega_{\Lambda,0}}{a^2} + \frac{\Omega_{\kappa,0}}{a^2} \right), \] (1.48)

where the dimensionless energy density for curvature is defined as \( \Omega_{\kappa,0} = -\kappa c^2 / H_0^2 \) (similar to equation 1.45). Similarly the second Friedmann equation (equation 1.39) can be rewritten as
\[ \frac{\ddot{a}}{a} = \frac{H_0^2}{2} \left( -2\frac{\Omega_{i,0}}{a^4} - \frac{\Omega_{m,0}}{a^3} + 2\Omega_{\Lambda,0} \right). \] (1.49)

The subscripts in equations 1.48 and 1.49 are used to define the different constituents in the \( \Lambda \)CDM model: \( r \) is used for radiation, \( m \) for matter, \( \Lambda \) for the cosmological constant and \( \kappa \) for curvature. The dimensionless energy density, equation 1.41, can be rewritten using equations 1.14 and 1.33 as
\[ \Omega_i = \frac{\Omega_{i,0} a^{-3(1+w_i)}}{E(a)^2}. \] (1.50)

1.1.7 Constituents of the Universe

Photons

The largest contribution to the number density of photons in the Universe is from the CMB. Early on the density and temperature of the Universe is so high that ordinary matter is in the form of a hot ionized plasma and photons continuously interact with free electrons through Compton scattering. This ionized plasma eventually cools due to the expansion of the Universe allowing the electrons and protons to form helium and neutral hydrogen. This period known as last scattering allows photons to escape and propagate through the Universe producing the relic radiation known as the CMB. The CMB has been observed to have one of the most precise black body spectrum’s in nature with a temperature of \( T = 2.7255 \pm 0.0006 \) K (Fixsen, 2009) with small anisotropies of order \( \Delta T/T \sim 10^{-5} \). The energy density for photons \( \rho_\gamma \) can be defined as
\[ \frac{\rho_{\gamma}}{\rho_{c,0}} = \frac{2.46 \times 10^{-5}}{h^2 a^4} = \frac{\Omega_{\gamma,0}}{a^4} \implies \Omega_{\gamma,0} = \frac{2.46 \times 10^{-5}}{h^2}, \] (1.51)

where \( \rho_{c,0} \) is the critical density today, \( \Omega_{\gamma,0} \) is the current radiation energy density (Dodelson, 2003) and \( h \) is the reduced Hubble constant following the relation \( H_0 = 100h \).
INTRODUCTION

Notice that the energy density scales by $a^{-4}$ which follows from the fact that the number density of any particle will scale by $a^{-3}$ (i.e. scaling according to the volume). The expansion of the Universe which causes the photon’s wavelength to redshift and stretch results in the additional $a^{-1}$ reduction in energy.

**Baryonic Matter**

Baryonic matter is what we consider to be ‘normal’ matter which forms the structures we are most familiar with including stars, planets and everything on Earth. Baryonic matter is mostly found in the form of gas and the energy density for baryonic matter $\rho_b$ can be defined as

$$\frac{\rho_b}{\rho_c,0} = \frac{\Omega_{b,0}}{a^3}. \quad (1.52)$$

where $\Omega_{b,0}$ is the current baryonic matter energy density which has been measured to be $\Omega_{b,0} \simeq 0.049$ (Planck Collaboration et al., 2018), only $\sim 5\%$ of the critical density required for the Universe to be flat. Since baryonic matter are particles the number density scales by $a^{-3}$.

**Dark Matter**

Observations of the radial velocities of stars in galaxies (Rubin et al., 1980) provided the first evidence for dark matter. Subsequent studies of the strength of the baryonic acoustic oscillation peaks in the cosmic microwave background angular power spectra and measurements of weak lensing have further strengthened the evidence for dark matter while simulations have shown it is a vital component of the Universe required to form large scale structure and galaxies. Studies sensitive to the ratio of baryonic matter and the total matter have shown that baryonic matter forms only $\sim 20\%$ of all the matter in the Universe. While dark matter’s gravitational effects can be felt, whatever it is, must interact very weakly with the electromagnetic, strong and weak forces of nature. For this reason the word ‘dark’ is used to represent our ignorance of this form of matter but perhaps a better descriptive term would be to call it invisible matter. Like normal baryonic matter the energy density $\rho_{dm}$ can be defined as

$$\frac{\rho_{dm}}{\rho_c,0} = \frac{\Omega_{dm,0}}{a^3}. \quad (1.53)$$

where $\Omega_{dm,0}$ is the current dark matter energy density. While the particle responsible for dark matter remains unknown its mass will determine its velocity and influence on large scale structure formation. Hot or warm dark matter will move rather quickly erasing small structures while cold dark matter will move slowly allowing for small structures to form
Dark Energy

For much of the last century it was predicted that the Universe, which was thought to be matter dominated, would be experiencing a deceleration in its expansion. However, observations of Type Ia supernova showed the Universe’s expansion was actually acceleration (Riess et al., 1998; Perlmutter et al., 1999). To reconcile theory with observation we needed to add a new constituent to the Universe, which we call dark energy, that has negative pressure and would on the very largest scales push things apart.

The simplest solution was to reintroduce the cosmological constant $\Lambda$ originally introduced, by Einstein (1917), to enable GR to be consistent with a steady state Universe. Einstein later called this introduction his “greatest blunder” since Lemaître (1927) had shown that this solution was unstable and the Universe would need to expand or contract. While there was no reason not to include this term in the field equations it was assumed that $\Lambda = 0$. The cosmological constant or vacuum energy can be reinterpreted as a scalar meaning it remains constant regardless of the expansion of the Universe. The energy density can be defined as

$$\frac{\rho_\Lambda}{\rho_c,0} = \frac{\Lambda c^2}{3H^2} = \Omega_{\Lambda,0}. \quad (1.54)$$

While observations remain consistent with a cosmological constant dark energy may actually be something else and may evolve over time. To describe dynamical dark energy (denoted with a subscript DE) we redefine the energy density as

$$\frac{\rho_{\text{DE}}}{\rho_c,0} = \Omega_{\text{DE},0} a^{-3(1+w_{\text{DE}})} \quad (1.55)$$

as a function of its equation of state $w_{\text{DE}}$ often parametrised as a function of $a$,

$$w_{\text{DE}}(a) = w_0 + (1-a)w_a, \quad (1.56)$$

where $w_0$ is the equation of state today and $w_a$ is its variation with the scale factor. Current observations find $\Omega_{\Lambda,0} \simeq 0.689$ (Planck Collaboration et al., 2018) which means dark energy is currently the dominant constituent of the Universe.

Neutrinos

Neutrinos are chargeless fermions that only interact with gravity and the weak nuclear forces. For this reason neutrinos interact very weakly with other forms of matter and are
notoriously difficult to measure. Until very recently it was thought that neutrinos were massless particles but the discovery of neutrino oscillation (Fukuda et al., 1998; Ahmad et al., 2001) showed neutrinos must have a small mass. Unlike photons, neutrinos are not reheated through electron-positron annihilation, since this occurs after neutrinos have decoupled from the primordial plasma. However, prior to this neutrinos remained in thermal equilibrium with photons. Assuming electron-positron annihilation conserves entropy we can use the relation,

\[ S \propto g_{\text{eff}} T^3, \quad (1.57) \]

where \( S \) is the entropy, \( g_{\text{eff}} \) is the effective degrees of freedom of relativistic particles at temperature \( T \) given by

\[ g_{\text{eff}} = \sum_{i=\text{bosons}} g_i + \frac{7}{8} \sum_{j=\text{fermions}} g_j, \quad (1.58) \]

where \( g_\gamma = 2 \) is the degrees of freedom for photons and \( g_{e^-/+} = 2 \) are the degrees of freedom for electrons/positrons. The ratio between the temperature of photons before electron-positron annihilation \( T_0 \) (but after neutrino decoupling) and after electron-positron annihilation \( T_1 \) is given by,

\[ \frac{T_0}{T_1} = \left( \frac{g_{\text{eff},1}}{g_{\text{eff},0}} \right)^{1/3}. \quad (1.59) \]

This ratio can be recast in terms of the neutrino temperature \( T_\nu = T_0 \) and the photon temperature \( T_\gamma = T_1 \) (also referred to as \( T \) in discussions of the CMB). The temperature ratio between neutrinos and photons is given by

\[ \frac{T_\nu}{T_\gamma} = \left[ \frac{g_\gamma}{g_\gamma + \frac{7}{8} (g_{e^-} + g_{e^+})} \right]^{1/3} = \left( \frac{4}{11} \right)^{1/3}. \quad (1.60) \]

From measurements of the current CMB temperature it follows that neutrinos have a current temperature \( T_\nu \simeq 1.9 \) K. Due to the neutrinos small mass they are the only particles, that we know of, that behave both as radiation at early times and as matter at late times. The era in which neutrinos behave as radiation is determined by when neutrinos are relativistic. During this period the dimensionless neutrino density is related to the radiation dimensionless density (Tanabashi et al., 2018) by

\[ \Omega_\nu = \frac{7}{8} N_{\text{eff}} \left( \frac{T_\nu}{T_\gamma} \right)^{4/3} \Omega_\gamma \simeq 0.69 \Omega_\gamma, \quad (1.61) \]
where \( N_{\text{eff}} = 3.045 \) is the effective number of neutrino species (taking into consideration the effect of neutrino oscillation with the present observed mixing parameters and collision terms). The dimensionless neutrino density in the present day \( \Omega_{\nu,0} \) is related to the sum of the three species \( \sum m_\nu \) (also referred to as \( M_\nu \) in this thesis) by

\[
\Omega_{\nu,0} = \frac{\sum m_\nu}{93.14 \, h^2 \, \text{eV}}. \tag{1.62}
\]

Like matter when neutrinos becomes non-relativistic they scale by \( a^{-3} \) meaning for any given species the energy density \( \rho_\nu^i \) for species \( i \) scales by

\[
\frac{\rho_\nu^i}{\rho_{c,0}} = \begin{cases} 
\frac{1}{3} \left[ \frac{7}{8} N_{\text{eff}} \left( \frac{T_\nu}{T_\gamma} \right)^{4/3} \right] \frac{\Omega_{\gamma,0}}{a^4}, & \text{for a relativistic neutrino,} \\
\frac{m_i}{93.14 \, a^3 \, h^2 \, \text{eV}}, & \text{for a non-relativistic neutrino,}
\end{cases} \tag{1.63}
\]

where the relativistic dimensionless energy density, given by equation 1.61, is divided by 3 for the 3 neutrino species. Neutrinos become non-relativistic when the mass equals the average momentum. The redshift at which this occurs \( z_{\text{NR}} \) can be calculated by equating the two condition above giving

\[
z_{\text{NR}} \simeq \frac{m_i}{0.53 \times 10^{-3} \, \text{eV}} - 1, \tag{1.64}
\]

for a neutrino species with \( m = 0.06 \, \text{eV} \) they become non-relativistic at \( z_{\text{NR}} \simeq 110 \).

When neutrinos become non-relativistic they have velocities \( v_i \) defined as

\[
v_i = \frac{0.53 \times 10^{-3} \left( 1 + z \right) \, \text{eV}}{m_i}, \tag{1.65}
\]

which is several orders of magnitude larger than the velocities for the dominant cold dark matter. This characteristic diffusion scale, known as the free-streaming length, affects the growth of structure and clustering by washing out small structures. This effect can be determined for example by measuring the suppression of the power spectrum below the free-streaming scale. The strength of this effect is dependent on the mass of the neutrinos.

**Total Radiation**

As discussed earlier neutrinos behave as radiation while they are relativistic. This means the density of radiation is a sum of the density of photons and relativistic neutrinos,

\[
\frac{\rho_\nu}{\rho_{c,0}} = \frac{\rho_\gamma}{\rho_{c,0}} + \left[ \frac{\rho_\nu}{\rho_{c,0}} \right]_{\text{relativistic}}. \tag{1.66}
\]
Since the density of relativistic neutrinos is related to the photon density the above relative can be written as

\[
\frac{\rho_r}{\rho_{c,0}} = \frac{\Omega_{r,0}}{a^4} = \left\{ 1 + \frac{N_{\nu}^R}{3} \left[ \frac{7}{8} N_{\text{eff}} \left( \frac{T_{\nu}}{T_{\gamma}} \right)^{4/3} \right] \right\} \frac{\Omega_{\gamma,0}}{a^4},
\]

(1.67)

where \(N_{\nu}^R\) is the number of relativistic neutrino species.

### Total Matter

The total matter density is given by the sum of the density of baryons, cold dark matter and non-relativistic neutrinos

\[
\frac{\rho_m}{\rho_{c,0}} = \frac{\rho_b}{\rho_{c,0}} + \frac{\rho_{dm}}{\rho_{c,0}} + \left[ \frac{\rho_{\nu}}{\rho_{c,0}} \right]_{\text{non-relativistic}},
\]

which can be rewritten as

\[
\frac{\rho_m}{\rho_{c,0}} \equiv \frac{\Omega_{m,0}}{a^3} = \frac{1}{a^3} \left( \Omega_{b,0} + \Omega_{dm,0} + \Omega_{\nu,0}^{NR} \right),
\]

(1.69)

where \(\Omega_{\nu,0}^{NR}\) is the density of non-relativistic neutrinos today for neutrinos that remain non-relativistic at a scale factor \(a\).

### 1.1.8 Epochs of Domination and Accelerated Expansion

#### Matter-Radiation Equality

The early Universe is dominated by radiation and later transitions to matter domination. The scale factor at matter-radiation equality \(a_{eq}\), which occurs while neutrinos are still relativistic, can be determined by

\[
\frac{\rho_r(a_{eq})}{\rho_{c,0}} = \frac{\rho_m(a_{eq})}{\rho_{c,0}} \implies a_{eq} \simeq \frac{1.69 \Omega_{\gamma,0}}{\Omega_{m,0}}.
\]

(1.70)

From current constraints \(a_{eq} \simeq 2.93 \times 10^{-4}\) which corresponds to a redshift of \(z_{eq} \simeq 3410\). This means radiation domination (when \(z > z_{eq}\)) occurs far before the moment of last scattering \(z_{LS} \simeq 1100\) and the majority of the Universe which we observe takes place during matter domination (when \(z < z_{eq}\)).
Λ-Domination

At later times matter domination transitions to a period of Λ-domination. Assuming dark energy is a cosmological constant this occurs at a scale factor $a_{\Lambda}$ which is determined by

$$\frac{\rho_m(a_{\Lambda})}{\rho_{c,0}} = \frac{\rho_{\Lambda}(a_{\Lambda})}{\rho_{c,0}} \implies a_{\Lambda} = \left(\frac{\Omega_{m,0}}{\Omega_{\Lambda,0}}\right)^{1/3}.$$  \hspace{1cm} (1.71)

From current constraints Λ-domination occurs when $a_{\Lambda} \gtrsim 0.766$ which corresponds to a redshift $z_{\Lambda} \lesssim 0.305$, fairly recent in cosmic time.

Accelerated Expansion

The Universe is currently undergoing a period of accelerated expansion. This period can be calculated by determining when the deceleration parameter given by equation 1.23, is $q(a_{\text{acc}}) < 0$. To calculate when this occurs it is useful to first rewrite equation 1.23 as

$$q(a) = -\frac{\ddot{a}}{a} \frac{a^2}{\dot{a}^2}. \hspace{1cm} (1.72)$$

Since $\ddot{a}/a > 0$, accelerated expansion is determined by $-\ddot{a}/a < 0$. Taking the approximation for $\ddot{a}/a$ during matter-Λ equality this gives the relation

$$-\frac{H_0^2}{2} \left( -\frac{\Omega_{m,0}}{a_{\text{acc}}^3} + 2\Omega_{\Lambda,0} \right) \lesssim 0 \implies a_{\text{acc}} \gtrsim \left(\frac{\Omega_{m,0}}{2\Omega_{\Lambda,0}}\right)^{1/3}.$$  \hspace{1cm} (1.73)

From current constraints the Universe undergoes accelerated expansion when $a_{\text{acc}} \gtrsim 0.608$ which corresponds to a redshift $z_{\text{acc}} \lesssim 0.645$.

1.1.9 Inflation

Inflation introduces a scalar field, that very early on after the Big Bang causes an exponential accelerated expansion of spacetime (Guth, 1981; Linde, 1982). This early expansion solved several open problems in cosmology: the flatness problem, the relic problem and the horizon problem.

The Flatness Problem

The flatness problems is based on the observation that the curvature of the Universe shows $\Omega_{\kappa,0} \simeq 0$. Since $\Omega_{\kappa}$ is defined as $\Omega_{\kappa} = -kc^2/a^2H^2$ it can be redefined as

$$|\Omega_{\kappa}| = \frac{|k|c^2}{a^2}.$$  \hspace{1cm} (1.74)
INTRODUCTION

Until very recently in cosmic time, the Universe has been dominated by matter and prior to this dominated by radiation. In both of these eras $\Omega_\kappa$ is increasing as a function of the scale factor,

$$|\Omega_\kappa| \approx \begin{cases} \frac{|\kappa|c^2}{H_0^2\Omega_m,0}a, & \text{for matter domination,} \\ \frac{|\kappa|c^2}{H_0^2\Omega_r,0}a^2, & \text{for radiation domination.} \end{cases}$$

(1.75)

This means $\Omega_\kappa$ is growing during matter domination and is growing even faster during radiation domination. For $\Omega_{\kappa,0}$ to be so small now $\Omega_\kappa$ must have been extremely small early on. This becomes a problem of fine tuning: why is the early Universe so flat? However, we can remove this problem if we assume the Universe is exactly flat (i.e. $\kappa = 0$) and has always been that way. While curvature is allowed by the equations we consider it may just be a mathematical curiosity and not a phenomenon that is allowed by nature.

The Relic Problem

The relic problem is based on the idea that at very early times the Universe would be so dense and hot the electromagnetic, weak and strong forces of nature would merge into a single force. Many of these theories which work under the assumption of General Unification predict the production of relic particles, such as the monopole, which have not been observed in the Universe. Some have argued that the lack of these particles is a problem for the standard model of cosmology but sceptics have rightfully pointed out the lack of particles from theories which are yet to be proven should not be an argument for new physics.

The Horizon Problem

The horizon problem is based on observations of the CMB. During recombination, when protons and electrons combine and the Universe first becomes transparent, the particle horizon at the last scattering surface $z_{LS} \approx 1100$ would subtend an angle $\sim 1.6^\circ$ on the sky. Any perturbations larger than this would not have had time to smooth out. However, the CMB is extraordinarily uniform and exhibits a black body spectrum meaning the Universe was previously in a state of equilibrium with very small perturbations.

The Solution

To explain this, Guth (1981) introduced a period of inflation where the Universe undergoes exponential accelerated expansion. This resolves the three problems in the following way:

- The curvature problem: A universe with arbitrary levels of curvature would be stretched and expanded to such an extent that it would appear locally flat as we observe today.
• The relic problem: Relic particles would be spread across vast distances and would be so rare that it would be difficult to observe them.

• The horizon problem: In order for the observable universe to be in equilibrium the universe would initially need to be in causal contact. Regions of space would then be exponentially expanded to such an extent that patches of space originally in causal contact now appear beyond their particle horizon. This would explain why patches of space not in causal contact are in equilibrium.

The most serious of these problems, the horizon problem, requires the largest period of inflation and therefore the less serious problems of curvature and relic particles are solved for free.

Inflation additionally provides an explanation for the seeds of structure. Since inflation is caused by a scalar field it would on very small scales be quantum mechanical. As a result quantum fluctuations in the field would grow and be embedded into the distribution of matter and would later grow to form large scale structure.

1.1.10 Initial Conditions to Large Scale Structure

The FLRW metric allows us to describe the evolution of a smooth homogeneous and isotropic Universe, however the Universe we live in is certainly not smooth. Determining how small perturbations eventually grow to form clusters, galaxies, stars and eventually us is an important step for observational cosmology. It is believed that inflation seeds small perturbations in the distribution of matter, over time these small perturbations will grow due to gravity. Schematically the growth of structure is dependent on a competition between gravity and pressure (described later in equation 1.93).

The local overdensity $\delta(x)$ at a point $x$ in Cartesian coordinates is defined as

$$\delta(x) = \frac{\rho(x) - \bar{\rho}}{\bar{\rho}},$$ (1.76)

where $\bar{\rho}$ is the average matter density. The variance for an isotropic and homogeneous overdensity field $\delta(x)$ can be represented by the power spectrum $P(k)$ in Fourier space,

$$\langle \delta(k)\delta(k')^* \rangle = (2\pi)^3 \delta_D(k - k') P(k),$$ (1.77)

where $k = |k|$ and since homogeneous fields are invariant, the Fourier modes are uncorrelated which is characterised by the Dirac delta function $\delta_D(k - k')$.

The growth of perturbations in Fourier space $\delta(k, t)$ at time $t_0$ can be related to the initial perturbations $\delta(k, t_i)$ by the transfer function $T(k)$,

$$\delta(k, t_0) = T(k, t_0)\delta(k, t_0).$$ (1.78)
INTRODUCTION

Using the transfer function we can relate the primordial power spectrum $P_s(k)$ to the power spectrum measured at a time $t_0$,

$$ P(k, t_0) = T^2(k, t_0)P_s(k). $$

(1.79)

Slow roll inflationary models, which are preferred by observations, predict scalar fluctuations with a primordial power spectrum that obeys a power law defined as

$$ P_s(k) = A_s \left(\frac{k}{k_*}\right)^{n_s}, $$

(1.80)

where $A_s$ is the amplitude for scalar fluctuation, $n_s$ is the scalar spectral index and $k_*$ is an arbitrary pivot scale set to $k_* = 0.05h\text{Mpc}^{-1}$.

1.1.10.1 Linear Perturbation Theory

Perturbations in the matter distribution $\delta$ can be described by linear perturbation theory when perturbations are small (i.e. $\delta \ll 1$). This can be extremely useful to describe the evolution of the matter distribution from primordial perturbations to large scale structure.

Newtonian Approximation

We can describe the distribution of matter in the Newtonian approximation by approximating the matter as a fluid described by the three equations of motion: continuity equation (equation 1.81), Euler equation (equation 1.82) and Poisson equation (equation 1.83):

$$ \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0, $$

(1.81)

$$ \frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -\frac{1}{\rho} \nabla P - \nabla \Phi, $$

(1.82)

$$ \nabla^2 \Phi = 4\pi G \rho - \Lambda, $$

(1.83)

given in real coordinates where $\rho$ is the density, $\mathbf{u}$ is the velocity, $P$ is the pressure, $\Phi$ is the gravitational potential and $\nabla$ is the conventional $\nabla$ operator carried out in real coordinates. Taking this into account we can arrive at the non-perturbed equations of motion (assuming $\Phi = 0$ at $r = 0$) which when combined allow us to construct the Newtonian approximation to the Friedmann equations:

$$ H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3} \rho + \frac{\Lambda}{3} - \frac{\kappa}{a^2}, $$

(1.84)

$$ \frac{\ddot{a}}{a} = -\frac{4\pi G}{3} \rho + \frac{\Lambda}{3}, $$

(1.85)
where in the relativistic case $\rho$ in equation 1.85 is replaced with $\rho + 3P$. We can translate this into comoving coordinates $\mathbf{x}$ by noting that $\mathbf{r} = a\mathbf{x}$, $\nabla_r = \nabla_r/a$, $\mathbf{u} = a\dot{H}\mathbf{x}$ and the time derivative in comoving coordinates changes to $(\partial/\partial t)_r = (\partial/\partial t)_x - H\mathbf{x} \cdot \nabla$. The properties of the Universe can be perturbed such that we replace

$$\rho \mapsto \bar{\rho} + \delta \rho \equiv \bar{\rho}(1 + \delta),$$

$$P \mapsto \bar{P} + \delta P,$$

$$\mathbf{u} \mapsto \bar{\mathbf{u}} + \mathbf{v} \equiv a\dot{H}\mathbf{x} + \mathbf{v},$$

$$\Phi \mapsto \bar{\Phi} + \phi.$$  

Where the bar denotes the background evolution of these quantities in a non-perturbed Universe. The equation of motions in the linear regime becomes

$$\ddot{\delta} + \frac{1}{a} \nabla \cdot \mathbf{v} = 0,$$  

$$\dot{\mathbf{v}} + \frac{\dot{a}}{a} \mathbf{v} = -\frac{1}{a\rho} \nabla \delta P - \frac{1}{a} \nabla \phi,$$  

$$\nabla^2 \phi = 4\pi G a^2 \bar{\rho} \delta.$$  

If we combine these three equations we come to the Newtonian description of matter evolution in Fourier space (taking $\nabla^2 = -k^2$),

$$\ddot{\delta} + 2\frac{\dot{a}}{a} \dot{\delta} + \left( \frac{c_s^2 k^2}{a^2} - 4\pi G \bar{\rho} \right) \delta = 0,$$  

where we assume a barotropic fluid, i.e. pressure is a function of density. The sound speed is defined to be $c_s^2 = dP/d\rho$ and therefore the relation $\delta P = c_s^2 \bar{\rho} \delta$ follows. If $\dot{a} = 0$ and $a = 1$ we can show the solution to this equation is

$$\delta \propto \exp \left( \pm \sqrt{4\pi G \bar{\rho} - c_s^2 k^2 t} \right),$$

which for $k > \sqrt{4\pi G \bar{\rho}}/c_s$ gives oscillating solutions due to pressure support and for $k < \sqrt{4\pi G \bar{\rho}}/c_s$ gives exponential growth. This transition is often characterised by the Jean’s length $\lambda_J = 2\pi/k_J$ and therefore,

$$\lambda_J = c_s \sqrt{\frac{\pi}{G \bar{\rho}}},$$

meaning if $\lambda > \lambda_J$ density fluctuations grow exponentially but if $\lambda < \lambda_J$ density fluctuations oscillate. This gives rise to the baryonic acoustic oscillations (BAO) features which occur in the primordial plasma before recombination. It is important to note the oscillating solutions are removed once $c_s = 0$ since this means the Jean’s wavelength is $\lambda_J = 0$. 

45


**INTRODUCTION**

Table 1.1: Growing and decaying modes in the Newtonian approximation for matter perturbations as a function of time and the scale factor.

<table>
<thead>
<tr>
<th>Era</th>
<th>Growing Terms</th>
<th>Decaying Term</th>
<th>Potential Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiation</td>
<td>$\delta \propto \ln t \propto \ln a$ &amp; $\delta \propto 1$ &amp; $\phi \propto a^{-1}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matter</td>
<td>$\delta \propto t^{2/3} \propto a$ &amp; $\delta \propto a^{-1/2}$ &amp; $\phi \propto 1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dark Energy</td>
<td>$\delta \propto 1$ &amp; $\delta \propto e^{-2\sqrt{\Lambda/3t}} \propto a^{-2}$ &amp; $\phi \propto a^{-1}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It can be shown in the Newtonian approximation that perturbations in the matter distribution will grow in the different eras according to the relations shown in Table 1.1 which follows from solving equation 1.93 in the radiation, matter and dark energy eras. During matter-radiation equality it can be shown that $\delta \propto 2/3 + \rho_M/\rho_R$; during the radiation era $\rho_M \ll \rho_R$ and $\delta \propto 1$ and during the matter era $\delta \propto a$ as shown in Table 1.1.

**Relativistic Approach**

In the relativistic approach, the perturbed metric in the linear regime is defined with a line element in conformal time as

$$ds^2 = a^2(\eta_{\mu\nu} + h_{\mu\nu})dx^\mu dx^\nu,$$

where the flat FLRW metric is defined as $\bar{g}_{\mu\nu} = a^2\eta_{\mu\nu}$ and perturbations to the metric $|h_{\mu\nu}| \ll 1$ are small. In the synchronous gauge $h_{00} = h_{0i} = 0$ and

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + \delta g_{\mu\nu} = a^2(\eta_{\mu\nu} + h_{\mu\nu})$$
$$g^{\mu\nu} = \bar{g}^{\mu\nu} + \delta g^{\mu\nu} = a^{-2}(\eta^{\mu\nu} - h^{\mu\nu})$$
$$\therefore g_{\mu\alpha}g^{\alpha\nu} = \delta_{\mu}^{\nu} + O(h^2).$$

This is analogous to equation 1.93 and gives rise to the relativistic equation of motion,

$$\delta'' + \frac{a'}{a} \delta' = \frac{3\omega_r \delta_r}{a^2} + \frac{3\omega_m \delta}{2a},$$

where $' = \partial/\partial \eta$ is a derivative with respect to comoving time, $\delta_r$ are density perturbations in radiation, $\omega_r = \Omega_r H_0^2$ and $\omega_m = \Omega_m H_0^2$.

The relativistic approach gives rise to the Newtonian solutions for matter density growth for scales below the horizon scale, i.e. $k\eta > 1$. Scales with $k\eta \gg 1$ are defined as subhorizon scales and scales with $k\eta \ll 1$ are defined as superhorizon scales. During the matter and dark energy eras the solutions can be shown to be scale dependent and therefore follow the solutions shown in the Newtonian approximation. However, during the radiation era superhorizon perturbations can be shown to have growing modes where $\delta \propto \eta^2$ and
Table 1.2: Relativistic growth of matter perturbations for subhorizon scales and superhorizon scales as a function of comoving time.

<table>
<thead>
<tr>
<th>Era</th>
<th>Subhorizon</th>
<th>Superhorizon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiation</td>
<td>$\delta \propto 1$</td>
<td>$\delta \propto \eta^2$</td>
</tr>
<tr>
<td>Matter</td>
<td>$\delta \propto \eta^2$</td>
<td>$\delta \propto \eta^2$</td>
</tr>
<tr>
<td>Dark Energy</td>
<td>$\delta \propto 1$</td>
<td>$\delta \propto 1$</td>
</tr>
</tbody>
</table>

decaying modes where $\delta \propto \eta^{-2}$. A summary of the growing modes for subhorizon and superhorizon scales are provided in Table 1.2.

Fourier modes that enter the horizon at matter-radiation equality are called $k_{eq}$. Modes larger than this ($k > k_{eq}$) would have entered the horizon during radiation domination while modes smaller than this ($k < k_{eq}$) would have entered during matter domination. Therefore the transfer function can be shown to be

$$T(k) \propto \begin{cases} 1, & \text{for } k < k_{eq}, \\ k^{-2}, & \text{for } k > k_{eq}, \end{cases}$$

and therefore the power spectrum can be described as

$$P(k) \propto \begin{cases} k^{n_s}, & \text{for } k < k_{eq}, \\ k^{n_s-4}, & \text{for } k > k_{eq}, \end{cases}$$

which gives the characteristic shape of the matter power spectrum.

1.1.10.2 Non-Linear Structures and the Cosmic Web

Linear perturbation theory is able to describe the Universe when $\delta \ll 1$, meaning it can be used for much of the early Universe including features in the CMB. At late times these density perturbations grow and can no longer be described by linear perturbation theory and we need to start to include higher-order terms. This can be done by including second-order perturbation theory and so on. Some progress can also be made using analytical methods such as applying the spherical collapse model (Gunn & Gott, 1972), the Zel'dovich approximation (Zel'dovich, 1970) and the Press-Schechter formalism (Press & Schechter, 1974) but eventually $N$-body simulations are required.

Large $N$-body simulations have revealed highly non-Gaussian features in the distribution of matter. Over time density perturbations grow from Gaussian fluctuations and begin to exhibit a web like structure called the cosmic web shown in Figure 1.1. This is characterised by a complicated structure of voids, walls, filaments and clusters. This highly non-linear structure is difficult to describe analytically but is a manifestation of many effects. The clustering of matter on large scales is dominated by the clustering of dark matter,
with baryonic matter tracing this background clustering. On smaller scales the clustering of dark matter is still dominant but baryonic feedback from astrophysical effects from galaxies, and the stars and supermassive black holes that reside in them, start to become important. Furthermore non-relativistic neutrinos start to behave like warm dark matter and depending on their free-streaming scale will begin to wash out clustering on small scales.

1.2 OBSERVATIONAL COSMOLOGY

In this Section I will outline the different probes and methods used to measure the observable Universe and how these are used to test the standard model of cosmology. I will then discuss the current limitations in these techniques and the hints of new physics and the frontiers for future advancements in cosmology.
1.2.1 Cosmological Probes

Cosmological probes are measurements of the observable universe that can be tested against theoretical predictions from the standard model of cosmology and beyond.

**Cosmic Microwave Background**

The CMB has been the most powerful probe for cosmology to date. The CMB was discovered serendipitously by Penzias & Wilson (1965) and was quickly realised to be the relic radiation from the Big Bang (Dicke et al., 1965). The CMB has since been the focus of many observational experiments including the space based missions COBE\(^1\) (Cosmic Background Explorer), WMAP\(^2\) (Wilkinson Microwave Anisotropy Probe) and Planck\(^3\).

The CMB is incredibly smooth, exhibiting a black body spectrum with a temperature \(T \approx 2.7255\) K (Fixsen, 2009) with anisotropies on the scale of \(\Delta T/T \sim 10^{-5}\) which are measured by CMB measurements. These can be quantified by measuring the CMB's angular power spectra and compared to theoretical models. Measurements by Planck show the CMB is consistent with the standard model of cosmology \(\Lambda\)CDM (Planck Collaboration et al., 2018) and show a preference for simple slow roll inflationary models.

**The Distance Ladder and Standard Candles**

Standard candles are astrophysical sources for which we know the intrinsic luminosity. Measurements of a source’s luminosity can then be used to determine its luminosity distance.

The earliest standard candle Cepheids, variable stars with oscillating luminosities, were shown to have pulsation periods that were directly correlated to the stars luminosity (Leavitt, 1912). This allowed astronomers to determine the Cepheid’s distance. We can use this for cosmology by comparing this to the recession velocity caused by the Hubble expansion. In principle Cepheids can be used for this study but current telescopes can only resolve these large stars for galaxies relatively close by.

At larger distances cosmologist tend to use a different standard candle, Type Ia supernovae, which are supernovae explosions caused by the accretion of matter onto a white dwarf. A significant fraction of the star is consumed by a runaway reaction which is triggered by the white dwarf accumulating enough matter for the core to reach a high enough temperature for carbon fusion. Since this process is virtually the same for every Type Ia supernovae, and since it is so luminous, it can be used as a standard candle for cosmology. The luminosity distance determined by the light profile of the supernova is compared to the host galaxies redshift through spectroscopic observations. This allows us to measure

\[^1\]https://lambda.gsfc.nasa.gov/product/cobe/
\[^2\]https://wmap.gsfc.nasa.gov
\[^3\]https://sci.esa.int/web/planck
the luminosity distance as a function of redshift. This analysis led to the discovery of the accelerated expansion of the Universe and the discovery of dark energy (Riess et al., 1998; Perlmutter et al., 1999).

The discovery of gravitational waves (Abbott et al., 2016) by the Laser Interferometer Gravitational-Wave Observatory (LIGO) has introduced a new way to measure the distance ladder. Since gravitational waves give us the luminosity distance directly (rather than requiring standard candle relations) they can be used to measure the Hubble expansion. In the case of neutron star binary mergers an optical counterpart can be measured. As such they are often referred to as ‘standard sirens’ or ‘bright sirens’. This can be used to relate the luminosity distance to redshift (Abbott et al., 2017). From standard sirens measured by LIGO we can expect a 2% determination of the Hubble constant in ∼ 5 years (Chen et al., 2018). However, the majority of mergers observed by LIGO will be from black hole mergers known as ‘dark sirens’ since no optical counterpart is expected. For this reason techniques have been developed to use dark sirens to infer the Hubble constant. Without an optical counterpart the host galaxy of the merger cannot be determined. The Hubble constant is instead determined by correlating the localisation of the source on the sky to galaxies from large scale structure surveys (see Soares-Santos et al., 2019, for a measurement of the Hubble constant from a dark siren).

Large Scale Structure

Observations of large scale structure are carried out by surveying the distribution of galaxies in the Universe through imaging surveys. The redshifts for the galaxies are then obtained either through photometric estimations or from follow-up spectroscopic observations. This allows cosmologist to build a three dimensional picture of the galaxy distribution. From this several properties can be measured such as: clustering (two-point correlation functions or power spectrum), redshift space distortions (RSD; Kaiser, 1987) and BAO (Eisenstein & Hu, 1998). Observations of the distribution of galaxies will be the focus of many future cosmological surveys and will be the focus of the techniques and studies described in this thesis.

Weak Lensing

Einstein showed that clumps of matter could lens background light (Einstein, 1936). It was later realised that weak lensing (Bartelmann & Schneider, 2001), a subtle version of this effect, could be used to measure the amount of intervening matter. Unlike galaxies which trace the underlying dark matter distribution this would directly measure the matter distribution and would allow to test models of cosmology. This is carried out by measuring the shapes of distant galaxies and correlating this with intervening galaxies. This measurement is complicated by the fact that galaxies have varied shapes due to astrophysics,
observational effects and correlations with their position in the cosmic web. While this effect is small the measurement can be made by using large sample statistics and is at the forefront of many of the current and future galaxy redshift surveys.

\textit{Lyman-α Forest}

Distant quasars were observed to have a ‘forest’ of absorption features below the quasar’s rest frame Lyman-α emission line (Lynds, 1971). This feature is caused by clouds of neutral hydrogen absorbing the quasar light via the Lyman-α transition. Since the light is redshifted when it arrives in the cloud these absorption lines appear for shorter wavelengths than the quasars rest frame Lyman-α emission. Cosmologist can use the Lyman-α forest to trace the clouds of neutral hydrogen. This feature can not be measured at low redshifts since neutral hydrogen becomes fully ionized. This is a particularly useful measurement since quasars are very luminous and more plentiful at high redshifts giving a measurement of the clustering of matter at high redshift and at smaller scales, typically not accessible in large scale structure measurements due to systematics.

\textit{Big Bang Nucleosynthesis}

The light elements, in particular Deuterium, Helium and Lithium, were produced during the first few minutes of the Universe. The relative abundance of these elements tells us about the conditions of the early Universe, such as the duration of the conditions for efficient production of these light elements. As a result Big Bang nucleosynthesis (BBN) places strong constraints on the baryon density $\Omega_b$ and the effective number of neutrino species $N_{\text{eff}}$ (Walker et al., 1991; Cyburt et al., 2005, 2016; Cooke et al., 2018).

1.2.2 Standard Measures

In observational cosmology we typically measure the level of clustering in a continuous field in Fourier space or from discrete points in real space. These methods are used for most of the cosmic probes discussed prior and are outlined below.

1.2.2.1 \textit{Two-Point Statistics}

The variance of a density distribution can be determined by measuring its variance around the mean. This is measured as a function of scale, Fourier modes in Fourier space and the distances between points in real space. The two point correlation function (2PCF) $\xi(r)$ can be calculated from the power spectrum $P(k)$ using

$$\xi(r) = \frac{1}{2\pi^2} \int k^2 P(k) \text{sinc}(kr) dk,$$

(1.103)
where \( \text{sinc}(x) = \frac{\sin(x)}{x} \). This relation shows \( \xi(r) \) and \( P(k) \) are equivalent measurements.

**Fourier Space: Power Spectrum**

In Section 1.1.10.1 we demonstrate the evolution of Fourier modes \( k \) in the linear regime, showing the matter power spectrum is a natural prediction of linear perturbation theory. For this reason it is typically the main summary statistics measured from cosmological datasets. The power spectrum is calculated using equation 1.77. For simulations we often calculate statistics in Fourier space since this allows us to take advantage of the Fast Fourier Transform (FFT). For this reason throughout this thesis we will use the power spectrum for our two-point statistics from simulations and the 2PCF for our two-point statistics from real data. In real space the 2PCF is chosen because it is more difficult to incorporate the complicated survey selection function in Fourier space.

To calculate the power spectrum in simulations we must first calculate the density in a 3D grid in the simulation box. To do this we use a mass assignment scheme where a particle with a position \( x_p \) along one axis is assigned a weight \( w \) for a grid point at \( x_g \) on the same Cartesian axis. There are several widely used distribution schemes (see Cui et al., 2008, for a discussion of these different mass assignment schemes) used:

\[
W_{\text{NGP}}(x) = \begin{cases} 
1, & \text{for } |x| < \delta L/2, \\
\frac{1}{2}, & \text{for } |x| = \delta L/2, \\
0, & \text{otherwise,}
\end{cases} 
\]

\[
W_{\text{CIC}}(x) = \begin{cases} 
1 - \frac{|x|}{\delta L}, & \text{for } |x| < \delta L/2, \\
0, & \text{otherwise,}
\end{cases} 
\]

\[
W_{\text{TSC}}(x) = \begin{cases} 
\frac{3}{4} - \left( \frac{x}{\delta L} \right)^2, & \text{for } |x| \leq \delta L/2, \\
\frac{1}{2} \left( \frac{3}{2} - \frac{|x|}{\delta L} \right)^2, & \text{for } \delta L/2 < |x| \leq 3\delta L/2, \\
0, & \text{otherwise,}
\end{cases} 
\]

where \( x = x_p - x_g \). \( W_{\text{NGP}} \) is the nearest-grid-point (NGP) weight, \( W_{\text{CIC}} \) is the cloud-in-cell (CIC) weight, \( W_{\text{TSC}} \) is the triangular-shaped-cloud (TSC) weight, \( \delta L = L/N_g \) is the length of each grid cell, \( L \) is the simulation box size and \( N_g \) is the number of grids across each axis. The number density \( n \) at a cell with coordinates \( x_g = (x_g, y_g, z_g) \) is given by,

\[
n(x_g) = \sum_{i=0}^{N_p-1} W(x_g - x_i) W(y_g - y_i) W(z_g - z_i),
\]

summed over the number of particles. In practice there is no need to loop through each
particle for each point in the grid, instead we can loop through the particles and determine the grid points which have non-zero weights. The density contrast is then calculated from

$$\delta(x_g) = \frac{n(x_g)}{\langle n \rangle} - 1.$$  \hspace{1cm} (1.108)

We then apply an FFT to $\delta(x)$ to calculate its Fourier space coefficients $\tilde{\delta}(k)$,

$$\delta(x) \xrightarrow{\text{FFT}} \tilde{\delta}(k).$$  \hspace{1cm} (1.109)

However, this is not the Fourier transform of the field but rather the binned field, to remove the effect of the binning scheme (mass assignment scheme) we apply the following deconvolution

$$\delta(k) = \tilde{\delta}(k) \left[ \text{sinc} \left( \frac{k_x L}{2N_g} \right) \cdot \text{sinc} \left( \frac{k_y L}{2N_g} \right) \cdot \text{sinc} \left( \frac{k_z L}{2N_g} \right) \right]^{-p},$$  \hspace{1cm} (1.110)

where $p = 1$ for NGP, $p = 2$ for CIC, $p = 3$ for TSC and $k = (k_x, k_y, k_z)$ (Jing, 2005).

The one dimensional power spectrum is then obtained by calculating

$$\bar{P}(k) = \frac{L^3}{N_g^3} \sum_{i=0}^{N_g^3-1} \left\{ \left| \delta(k_i) \right|^2, \quad \text{for } k - \Delta k/2 \leq |k - |k_i|| < k + \Delta k/2 \right\},$$  \hspace{1cm} (1.111)

Where $\Delta k$ is a defined $k$-bin. Notice, the power spectrum takes the sum of the amplitudes of the Fourier modes and discards all phase information. Finally we subtract shot noise,

$$P(k) = \bar{P}(k) - \frac{L^3}{N_p}.$$  \hspace{1cm} (1.112)

Another cosmological observable of interest is the amplitude of the linear matter power spectrum at $R = 8 \, h^{-1}\text{Mpc}$ (defined by Peebles, 1980, to be roughly of order unity) defined as

$$\sigma^2_R = \int_0^\infty \frac{1}{k} w(kR)^2 \Delta^2(k)dk,$$  \hspace{1cm} (1.113)

where

$$w(kR) = \frac{3}{kR} \left( \text{sinc}(kR) - \cos(kR) \right)$$  \hspace{1cm} (1.114)

and $\Delta^2(k) \equiv k^3 P(k)/(2\pi^2)$ is the dimensionless power spectrum. While $\sigma_8$ is not a parameter like others discussed before, in that it is not an input parameter, it is a useful measure used frequently to characterise the power spectrum amplitude for low redshift observations.
**INTRODUCTION**

**Real Space: Two-Point Correlation Function**

In real space the 2PCF gives the level of extra clustering at a scale \( r \) in comparison to a homogenous background distribution. The joint probability of finding a point in a sphere centred at \( \mathbf{x}_1 \) with volume \( dV_1 \) and at a point \( \mathbf{x}_2 \) with volumes \( dV_2 \) is given by

\[
dP_{12} = \bar{n}^2 [1 + \xi(r)] dV_1 dV_2,
\]  

(1.115)

where \( \bar{n} \) is the average density, \( r \) is the distance between \( \mathbf{x}_1 \) and \( \mathbf{x}_2 \) and \( \xi(r) \) is the 2PCF. Several estimators exist for calculating the 2PCF which are shown below,

\[
\xi_{DP}(r) = \frac{N_R^2 DD(r)}{N_D^2} - 1, 
\]  

(1.116)

\[
\xi_{HAM}(r) = \frac{DD(r)RR(r)}{DR(r)^2} - 1, 
\]  

(1.117)

\[
\xi_{LS}(r) = 1 + \frac{N_R^2 DD(r)}{N_D^2} RR(r) - 2\frac{N_R}{N_D} DR(r), 
\]  

(1.118)

where \( \xi_{DP}(r) \) is the Davis & Peebles (1983) estimator, \( \xi_{HAM}(r) \) is the Hamilton (1993) estimator and \( \xi_{LS}(r) \) is the Landy & Szalay (1993) estimator. Here the distribution of data points (with \( N_D \) points) is compared to the distribution of randoms points (with \( N_R \) points) which are used to represent a homogeneous distribution. In the estimators above we count pairs with separations between \( r \) and \( r + dr \) for the data-to-data pairs \( DD(r) \), for random-to-random pairs \( RR(r) \) and for data-to-random pairs \( DR(r) \). Throughout this thesis we will use the Landy & Szalay (1993) estimator since it has been shown to be an unbiased estimator that converges to the true 2PCF faster than other estimators (Szapudi & Szalay, 1998). Note that the estimators above can be used in any spatial coordinates. The separation \( r \) can be taken to be the real two or three dimensional separation (for Cartesian coordinate systems) or the angle between points on a sphere (for tomographic analysis).

**1.2.2.2 Biased Tracers**

Perturbation theory gives us an idea of how matter is clustered in the Universe, however observations of the Universe do not allow us to probe directly the distribution of dark matter (the dominant source of matter) in the Universe. So we instead rely on baryonic tracers of matter, usually in the form of galaxies. Galaxies will not trace the matter exactly, since galaxies form from haloes which occur at peaks in the matter distribution. To account for this a biasing scheme is often used. A typical biasing scheme is to assume that matter and galaxies are linearly biased. Kaiser (1984) and Bardeen et al. (1986) showed by the statistics of peaks that the power spectrum of a tracer distribution \( P_T(k) \) is related to the power spectrum of the underlying dark matter distribution \( P(k) \) by a linear bias term \( b_T \),
\[ P_T(k) = b_T^2 P(k). \]  

(1.119)

It follows that the density contrast of a tracer distribution \( \delta_T(x) \) is related to the underlying dark matter density contrast \( \delta(x) \) by

\[ \delta_T(x) = b_T \delta(x). \]  

(1.120)

### 1.2.2.3 Baryonic Acoustic Oscillations

BAO are set in the primordial plasma where the sound speed \( c_s > 0 \), which creates acoustic density waves that are frozen in the distribution of matter at the moment of last scattering. This length scale is used as a standard ruler in cosmology and is one of the main features in the CMB angular power spectrum. At later times this can be measured by looking at the ‘wiggle’ in the matter power spectrum or looking for the BAO feature in the 2PCF at \( \sim 100 \ h^{-1}\text{Mpc} \) (Eisenstein & Hu, 1998; Eisenstein et al., 1998). At later times due to non-linear evolution the distribution of matter becomes more non-Gaussian and the BAO feature can be smeared and difficult to detect. To counter the effect reconstruction techniques are used to revert the distribution of matter to the linear regime where the BAO feature is easier to detect.

### 1.2.2.4 Redshift Space Distortions

Galaxy peculiar velocities along the line-of-sight causes a distortion in the galaxy distribution in redshift space along the radial axis – an effect known as RSD (Kaiser, 1987). This effect can be measured by additionally binning the 2PCF (or power spectrum) by the cosine of the angle \( \cos \theta = \mu \) where \( \theta \) is the angle between \( x \) (the vector between two points) and the line-of-sight vector \( r \),

\[ \mu = \frac{x \cdot r}{|x||r|}. \]  

(1.121)

This is useful because the clustering along the line-of-sight has RSD due to peculiar velocities. To capture this difference we can calculate the multipoles of the 2PCF,

\[ \xi_\ell(r) = (2\ell + 1) \int_0^1 \xi(r, \mu) \mathcal{L}_\ell(\mu) d\mu, \]  

(1.122)
\( \ell \) is the Legendre polynomial and \( \ell \) defines the multipole. This can be extended to the power spectrum,

\[
P_\ell(k) = (2\ell + 1) \int_0^1 P(k, \mu) \ell_\ell(\mu) d\mu. \quad (1.123)
\]

Typically the monopole (where \( \ell = 0 \) and equivalent to not binning in \( \mu \)) and the quadrupole (where \( \ell = 2 \)) are measured. This measurement can be similarly made in redshift space by measuring the angle between the \( \mathbf{k} \) mode vectors and the line-of-sight \( \mathbf{k}_r \) mode. Theoretically the RSD power spectrum for a tracer \( T \) can be determined from

\[
P_T(k, \mu) = \left[ b_T + \mu^2 f(z) \right]^2 P(k), \quad (1.124)
\]

where \( f(z) = d \ln D(z)/d \ln a \) (which can be approximated as \( f(z) \simeq \Omega_M^0 \); Peebles, 1980; Lahav et al., 1991) and \( D(z) \) is the linear growth factor (i.e. the solution to equation 1.93 when \( c_s = 0 \)).

1.2.2.5 Three-Point and Beyond

At late times we know the Universe exhibits a highly non-linear cosmic web structure. Since two-point statistics simply look at the variance in the density distribution it cannot capture this complex non-Gaussian structure.

To capture this cosmologist have often resorted to measuring the three-point correlation function (3PCF) in real space or the bispectrum (its Fourier space equivalent) but this leads to technical challenges as will be explained below.

A general estimator used for the \( N \)-point correlation function (NPCF) following Szapudi & Szalay (1998) is given by,

\[
\zeta_N = \frac{(D - R)^N}{R^N}. \quad (1.125)
\]

This reduces to the Landy & Szalay (1993) estimator for the 2PCF and gives the following 3PCF estimator,

\[
\zeta_3(s_{12}, s_{23}, s_{31}) = \frac{DDD - 3DDR + 3DRR - RRR}{RRR}, \quad (1.126)
\]

where \( s_{12}, s_{23} \) and \( s_{31} \) are the sides of a triangle where \( s_{12} \leq s_{23} \leq s_{32} \). The counts for triangles between the data points and a random catalogue are given by the normalised terms \( DDD, DDR, DRR \) and \( RRR \); where \( DDD \) are the counts for triangles between data points, \( DDR \) are the counts for triangles between two data points and one random point, \( DRR \) are the counts for triangles between one data point and two random points and \( RRR \) are the counts for triangles between random points.
We can quickly see that the computation of even the 3PCF scales naively by $O(N_D^3)$ which for modern cosmological surveys can be intractable. This is further complicated by the fact that the random catalogues are usually 2 or 3 orders larger in size than the galaxy catalogues and will provide the main bottleneck to this calculation. While there exist techniques to reduce this calculation, see Scoccimarro (2015) and Slepian & Eisenstein (2016) which can reduce these calculations to $O(N_D^2)$, these do not exist for the NPCF which scales by $O(N_D^N)$.

Another approach is to look at the Fourier space versions. For the 3PCF this means looking at the bispectrum $B(k_1, k_2, k_3)$ which is defined as

$$\langle \delta(k_1)\delta(k_2)\delta(k_3) \rangle = (2\pi)^3 \delta_D(k_1+k_2+k_3) B(k_1, k_2, k_3).$$

Using FFTs this can be calculated fairly quickly. However, on real data FFT routines are more difficult to apply, although not impossible. So while it may be possible to determine the Fourier space NPCF it is difficult to transfer these routines to real data.

Furthermore, in order to maximise cosmological constraints from the bispectrum and higher-order Fourier space correlation functions it is important to consider as many triangle configurations (or $N$-dimensional polygons) as possible which requires making the binning widths $\Delta k$ small. This leads to very large data vectors which require many more mock simulations to accurately calculate covariance matrices. One way around this is to use the Karhunen and Loève technique (Tegmark et al., 1997; Heavens et al., 2017; Gualdi et al., 2018) which reduces the dimensionality of the data vector removing the requirement for large mock galaxy catalogues but requires analytical predictions for the initial data vector.

### 1.2.3 The Cosmic Web

The distribution of matter at late times exhibits a cosmic web like structure (Bond et al., 1996). This complex structure is easily distinguished from simulations but is challenging to capture in real data.

From simulations there exist numerous tools for cosmic web classification; including looking at the Hessian matrix of the gravitational potential (Forero-Romero et al., 2009) or defining structures according to caustic mathematics (Feldbrugge et al., 2018). In either case this requires an accurate knowledge of the density and velocity field making it difficult to apply to a tracer distribution such as galaxies in real data. The density field can then be classified to clusters, filaments, walls or voids. Cosmic voids and clusters are the simplest of these structures; cluster are peaks in the density distribution and voids the troughs. From spherical collapse we know clusters are usually spherical while voids can be fairly elliptical. Walls are more subtle to describe and can be thought of as a weak overdensity which forms between voids. While filaments are long overdensities that form when two walls collide and are usually stretched between clusters. This has a natural connection to
Figure 1.2: A 10 $h^{-1}$Mpc slice through the IllustrisTNG simulation (Pillepich et al., 2018) is shown in both plots with higher densities, specifically $\log_{10}(\delta+1)$ of the dark matter particles, represented by darker shades of blue. On the left the distribution of galaxies is shown in red and on the right the MST graph (constructed from the galaxies) is shown in red. The MST can be seen to trace the cosmic web structure of the underlying dark matter distribution.

galaxy evolution, since early type galaxies are more likely to be found in clusters while younger types in filaments, walls and voids.

Looking for the cosmic web in the distribution of galaxies is more difficult since galaxies are biased tracers of the density distribution. Large galaxy surveys tend to search for the most luminous galaxies meaning the catalogues of galaxies produced are biased towards bright and massive galaxies typically found in clusters. This means we can resolve the peaks and troughs of the distribution well but have difficulty resolving filaments and walls which are not very dense. Alpaslan et al. (2014) used the minimum spanning tree (MST), a graph based algorithm used to search for filaments, to classify galaxies from the GAMA survey into filaments, voids and clusters. The study showed good agreement with data and has been used in other studies to search for filaments. The MST is very good at finding structures in point distributions and in Figure 1.2 is shown to be able to trace the underlying cosmic web structure from galaxies alone.

A useful review of cosmic web classification tools from the density fields and halo catalogues is presented in Libeskind et al. (2018).

1.2.4 The Integrated Sachs-Wolfe Effect

The CMB distribution is incredibly uniform with the temperature of the photons being described very precisely by a black body emitter. However due to small perturbations in the primordial plasma on the surface of last scattering the CMB is not completely uniform and has small anisotropies of the order of $\Delta T/T \sim 10^{-5}$. These anisotropies come from
a range of different processes that occur on the surface and processes that alter the light as it travels from the last scattering surface to an observer.

The effect we will concentrate on is the integrated Sachs-Wolfe (ISW; Sachs & Wolfe, 1967) effect which is caused by the linear decay of gravitational potentials as photons travel through them. This decay is caused by the accelerated expansion caused by dark energy and is usually written as an integral,

$$\left( \frac{\Delta T(\hat{n})}{T} \right)_{\text{ISW}} = \frac{2}{c^2} \int_{t_{LS}}^{t_0} \dot{\Phi}(\hat{n}, t) dt,$$

where $\hat{n}$ is a direction in the sky, $t_{LS}$ is the time at last scattering, $t_0$ is the time for the observer and $\dot{\Phi}$ is the time derivative of the gravitational potential.

A similar effect called the Rees-Sciama effect (Rees & Sciama, 1968) is caused by the gravitational collapse of potentials as photons travel across them. This effect is quite small, $\sim 10^{-2}$ of the magnitude of the ISW and occurs on characteristically smaller scales. The strength of this effect is determined by the size of gravitational potentials, therefore is strongest for a Universe which is matter dominated.

During matter domination, potentials in the linear regime do not grow and therefore we do not expect to see any ISW except from non-linear gravitational collapse (i.e. from the Rees-Sciama effect) however during $\Lambda$-domination potentials decays by $\propto a^{-1}$ (see Table 1.1) and gives rise to the dominant linear ISW effect. For this reason the linear ISW is particularly sensitive to dark energy.

Measurements of the ISW have typically been carried out either by cross-correlating large scale structure with the CMB (for example Giannantonio et al., 2008) or by stacking the temperature profiles of voids or clusters (for example Flender et al., 2013). Cross-correlation studies have typically shown agreement with the standard $\Lambda$CDM model but studies from stacked voids have found much larger decrements than is predicted by $\Lambda$CDM (Granett et al., 2008; Cai et al., 2014; Kovács, 2018). This has motivated studies linking the CMB Cold Spot anomaly to voids (which was first theoretically proposed by Inoue & Silk, 2006, 2007); Nadathur et al. (2014) showed that no void could produce the correct ISW profile or amplitude and Naidoo et al. (2016) showed no set of multiple voids could produce the Cold Spot even in the most extreme scenario. In both studies a $\Lambda$CDM model was assumed. The excess signal from voids has been debated with other studies such as Nadathur & Crittenden (2016) finding stacked voids to be in agreement with $\Lambda$CDM. Voids exhibiting excess signals were found to be highly elliptical (Kovács & García-Bellido, 2016), with the longest side aligned with the line-of-sight direction, since they were observed from photometric surveys. However studies that took this into account (Kovács et al., 2017) were unable to remove the discrepancy.

A common criticism of these studies is that they have failed to explore alternative cosmologies, either by allowing parameters from the $\Lambda$CDM models to vary or by choosing
INTRODUCTION

alternative models with for example dynamical dark energy. Although there exist analytical expressions for void temperature profiles (see Nadathur et al., 2014), in order to forward model the effects of observing with a biased tracer (such as galaxies), studies often make use of the Jubilee simulations (a large $N$-body simulation with galaxies and a full-sky ISW map; Watson et al., 2014). However the dependence on this single simulations means these studies can only compare to the fiducial WMAP cosmology and not alternative models.

1.2.5 Beyond $\Lambda$CDM

$\Lambda$CDM has been remarkably effective at explaining multiple cosmological probes, in some sense this has become a cause of frustration since the model is certainly incomplete; the particle or particles that make up dark matter still remain undetected and the nature of dark energy still remains a mystery – is it more than just a cosmological constant?

In recent years a few tensions have emerged from the inference of cosmological parameters from measurements of the CMB and measurements of the distance ladder and weak lensing.

Measurements of the CMB by Planck constrain the Hubble constant to $H_0 = 67.36 \pm 0.54 \, \text{km s}^{-1} \, \text{Mpc}^{-1}$ (Planck Collaboration et al., 2018) while measurements of the luminosity distance from Type Ia supernovae constrain the Hubble constant to $H_0 = 74.03 \pm 1.42 \, \text{km s}^{-1} \, \text{Mpc}^{-1}$, a 4.4$\sigma$ tension with CMB measurements (Riess et al., 2019). The exact cause of this tension remains a point of contentious debate and will probably not be fully resolved until another experiment using a different cosmological probe is able to validate or invalidate the results from Type Ia supernova.

Another tension is the so called $\sigma_8$ tension which has arisen from measurements of weak lensing from large scale structure by the Kilo Degree Survey (KiDS) which have measured smaller values of $\sigma_8$ (Hildebrandt et al., 2017) than is predicted by measurements of the CMB from Planck (Planck Collaboration et al., 2018). This tension is less severe than the Hubble tension, at 2.3$\sigma$, but results from the Dark Energy Survey (DES) although currently not in tension with Planck appear to be going in the same direction (see Figure 10 of Abbott et al., 2018).

Finally, the discovery of neutrino oscillation (Fukuda et al., 1998; Ahmad et al., 2001) revealed that neutrinos are not massless. This means they will have an observable effect on clustering on cosmological scales. This effect remains undetected and measurements of the sum of the neutrino masses remains consistent with zero, however in future as measurements of clustering become more precise we will be able to measure this effect and determine the sum of neutrino masses. This is a departure from $\Lambda$CDM that we expect to take place in the next decade.
1.2.6 Surveys: Current and Future

Galaxy redshift surveys will provide unparalleled three dimensional maps of the distribution of galaxies in the late Universe. Redshift surveys obtain redshifts for objects by either using photometric redshifts which calculate redshifts based on the objects magnitude in different photometric filters or by follow-up spectroscopic observations.

The quality of redshift estimations is determined by the spectroscopic resolution. Finer resolution will allow for better redshift determinations. Similarly, since the light for a galaxy in a photometric survey is binned across the survey’s photometric filters (typically ~ 5), we can think of photometric observations as spectroscopic observation with poor resolution. Since photometric surveys tend to have large errors in their determination of redshift, cosmological analysis of these surveys tend to be carried out in redshift slices. This means the analysis is performed on the surface of a sphere often using the same statistics used for the analysis of the CMB. For spectroscopic observations the analysis is usually carried out in three dimensions however this requires careful consideration of RSD.

In Figure 1.3 the magnitude limits are plotted against the survey area for the photometric surveys: Dark Energy Survey⁴ (DES), Euclid,⁵ the Nancy Grace Roman Space Telescope.

---

⁴http://www.darkenergysurvey.org
⁵http://www.euclid-ec.org/
The spectroscopic survey’s sky area is plotted against the observed distribution’s redshift range. BOSS is shown in blue for the galaxy distribution; DESI is shown with a brown line for the galaxy distribution and dotted brown for the Lyman-α distribution; eBOSS is shown in green for the galaxy and quasar distribution; Euclid is shown in yellow for the galaxy distribution; HETDEX is shown in pink for the galaxy distribution; the PFS distribution is shown with a dashed dotted purple line for the galaxy distribution, dotted purple line for the Lyman-α distribution and full purple line for the deep galaxy distribution and the Roman Space Telescope is shown in pale blue for the galaxy distribution.

The next generation of galaxy surveys will look to answer several fundamental cosmological question, in particular:

6 https://roman.gsfc.nasa.gov
7 https://www.lsst.org/
8 http://sumire.ipmu.jp
9 http://www.sdss3.org/surveys/boss.php
10 https://www.sdss.org/surveys/eboss/
11 http://desi.lbl.gov/
12 http://hetdex.org
13 https://pfs.ipmu.jp
1. What is the sum of neutrino masses and can we determine their mass hierarchy?

Neutrino oscillations have provided a lower limit on $\sum m_\nu \geq 0.06 \text{eV}$ while the best constraints from CMB measurements by Planck and constraints from BOSS (using BAO, clustering and Lyman-$\alpha$ forest) provide an upper limits of $\sum m_\nu \leq 0.12 \text{eV}$ (Planck Collaboration et al., 2018; Palanque-Delabrouille et al., 2015; Alam et al., 2017; Loureiro et al., 2019). On the other hand, the strictest constraints on the effective number of neutrino species is obtained from BBN, $N_\nu = 2.86 \pm 0.15$ (Fields et al., 2020). Experiments such as DESI are expected to be sensitive to neutrino masses below the lower limit provided by neutrino oscillations and therefore should provide the first determination of the mass of neutrinos. A rather more difficult task will be to determine whether neutrinos have a normal mass hierarchy or an inverted mass hierarchy. Revealing the nature of neutrinos will be pivotal for the standard model of particle physics and determining what questions need to be asked to move forward.

2. Is dark energy more than a cosmological constant?

Future redshift surveys will aim to provide better constraints on the dark energy equation of state. In particular they will determine, using the figure of merit, whether the data is still consistent with $w = -1$, as expected for the cosmological constant, or whether the equation of state varies with time, as expected for dynamical dark energy models often characterised by an equation of state $w = w_0 + (1 - a)w_a$.

1.2.7 Thesis Roadmap

In this thesis I develop methods for incorporating the cosmic web for cosmological parameter estimation and study the imprints of large scale structure on the CMB from real data and for simulations. Of particular interest is to understand the role these methods can play in determining new physics, such as determining the sum of neutrino masses.

In Chapters 2, 3 and 4 I develop the MST to incorporate information from the cosmic web into cosmological analysis. In Chapter 2 the method is introduced and explored in a limited parameter space to determine whether the MST adds new information. In Chapter 3 the MST is explored in a much larger parameter space using the Quijote (Villaescusa-Navarro et al., 2019) simulations. In this Chapter we determine the relative sensitivities of the MST and power spectrum to the $\nu\Lambda$CDM parameters. Finally, in Chapter 4 techniques are developed for incorporating galaxy weights (used to remove survey systematics) to the MST. Using these techniques the MST is constructed on BOSS galaxies and $\Lambda$CDM mocks to test whether they are consistent.

In Chapter 5 I explore the role of the ISW on the CMB Cold Spot anomaly and dis-
cuss the impact of masking on its significance. In Chapter 6 I construct a pipeline for computing the ISW from the density distribution of lightcone simulations. This is used to construct ISW maps for the MICE (Crocce et al., 2015) and Flagship lightcone simulations used heavily in the analysis of current and future redshift surveys such as DESI, Euclid and LSST.

Some of the Chapters or Sections have previously been published. This includes: Chapter 2 and Appendix A.1 which was published in Naidoo et al. (2020), Chapter 5 which was published in Naidoo et al. (2017) and Appendix A.2 which was published in Naidoo (2019). The other Chapters have been written with the intention of being published as they appear and for this reason they are written as self contained pieces of work.
“

Shifu: Master! Master!
Oogway: Hmm?
Shifu: I... I... have... it’s... it’s very bad news.
Oogway: Aah, Shifu. There's just news. There’s no good or bad.
Shifu: Master, your vision... your vision was right. Tai Lung has broken out of jail! He’s on his way!
Oogway: That is bad news... if you do not believe that the Dragon Warrior can stop him.
Shifu: The panda? Master, that panda is not the Dragon Warrior. He wasn’t even meant to be here! It was an accident!
Oogway: There are no accidents.
Shifu: [sighs] Yes, I know. You said that already... twice.
Oogway: Well, that was no accident either.
Shifu: ... Thrice.
Oogway: My friend, the panda will never fulfill his destiny, nor you yours, until you let go of the illusion of control.
Shifu: Illusion?
Oogway: Yes. Look at this tree, Shifu. I cannot make it blossom when it suits me, nor make it bear fruit before it’s time.
Shifu: But there are things we can control. I can control when the fruit will fall. And I can control where to plant the seed. That is no illusion, Master.
Oogway: Ah, yes. But no matter what you do, that seed will grow to be a peach tree. You may wish for an apple or an orange, but you will get a peach.
Shifu: But a peach cannot defeat Tai Lung!
Oogway: Maybe it can... if you are willing to guide it, to nurture it. To believe in it.”

– Kung Fu Panda (2008)
CHAPTER 2

USING THE MINIMUM SPANNING TREE AS A TOOL FOR COSMOLOGY

2.1 ABSTRACT

Cosmological studies of large-scale structure have relied on two-point statistics, not fully exploiting the rich structure of the cosmic web. In this Chapter we show how to capture some of this cosmic web information by using the minimum spanning tree (MST), for the first time using it to estimate cosmological parameters in simulations. Discrete tracers of dark matter such as galaxies, N-body particles or haloes are used as nodes to construct a unique graph, the MST, that traces skeletal structure. We study the dependence of the MST on cosmological parameters using haloes from a suite of COLA simulations with a box size of $250 \, h^{-1} \text{Mpc}$, varying the amplitude of scalar fluctuations ($A_s$), matter density ($\Omega_m$), and neutrino mass ($\sum m_\nu$). The power spectrum $P$ and bispectrum $B$ are measured for wavenumbers between $0.125$ and $0.5 \, h \text{Mpc}^{-1}$, while a corresponding lower cut of $\sim 12.6 \, h^{-1} \text{Mpc}$ is applied to the MST. The constraints from the individual methods are fairly similar but when combined we see improved $1\sigma$ constraints of $\sim 17\%$ ($\sim 12\%$) on $\Omega_m$ and $\sim 12\%$ ($\sim 10\%$) on $A_s$ with respect to $P (P + B)$ thus showing the MST is providing additional information. The MST can be applied to current and future spectroscopic surveys (BOSS, DESI, Euclid, PSF, Roman and MOST) in 3D and photometric surveys (DES and LSST) in tomographic shells to constrain parameters and/or test systematics.

2.2 INTRODUCTION

Over the years, a series of probes have emerged as standard tools for cosmological parameter inference. Surveys of the cosmic microwave background (CMB), large-scale structure (LSS), weak lensing (WL), and distance ladder have dominated our knowledge of cosmological parameters through measurements of the CMB angular power spectra (e.g. Planck Collaboration et al., 2018), galaxy clustering (e.g. Loureiro et al., 2019), weak lensing (e.g. Abbott et al., 2018; Hildebrandt et al., 2017), baryonic acoustic oscillations (BAO) from galaxies (e.g. Alam et al., 2017) and Lyman alpha (e.g. de Sainte Agathe et al., 2019), standard candles (e.g. Riess et al., 2016) and, more recently, standard sirens (e.g. Abbott et al.,
These techniques are relatively mature, well understood and most importantly, reliable and trusted.

However, many of these techniques (but not all) rely on measuring the two-point correlation function (2PCF) or its Fourier space equivalent, the power spectrum. Studies that include higher order statistics, such as the three-point correlation function (e.g. Slepian et al., 2017) or bispectrum (e.g. Gil-Marín et al., 2017), have already provided interesting constraints on cosmological parameters, demonstrating the need to go beyond the 2PCF. Despite solutions to improve the speed of 2PCF and 3PCF estimators (see Scoccimarro, 2015; Slepian & Eisenstein, 2016), going beyond the 3PCF is currently computationally intractable. The computational cost of current N-point correlation function (NPCF) estimators scale by $O(n^N)$; for this reason this information remains to be exploited.

The most attractive reason to explore methods that incorporate higher order statistics is their potential to break existing parameter degeneracies, to provide tighter constraints and to test systematics. Of growing interest to cosmologists is the total mass of the three neutrino species, $\sum m_\nu$. Neutrinos are massless in the standard model of particle physics; however this cannot be the case since neutrinos oscillate (Fukuda et al., 1998; Ahmad et al., 2001). Fortunately, LSS is sensitive to the mass of these elusive particles. As neutrinos are very light they possess high thermal velocities and dampen structure formation at scales below the free streaming scale (set by when they become non-relativistic). This effect is dependent on $\sum m_\nu$ and although it can be measured, the effect is small and highly degenerate with the matter density ($\Omega_m$) and the variance of density perturbations (e.g. as measured at $8 h^{-1}$Mpc ($\sigma_8$)). Currently, upper bounds of $\sum m_\nu \lesssim 0.12 - 0.23$ eV (95% confidence limits) (Palanque-Delabrouille et al., 2015; Planck Collaboration et al., 2016a; Alam et al., 2017; Loureiro et al., 2019) have been established from cosmology (specifically CMB and galaxy surveys) whilst the lower bound of $\gtrsim 0.06$ eV is given by neutrino oscillation experiments. Future experiments will be able to go further; in particular experiments such as the Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration et al., 2016) are expected to probe below the lower bound of $\sim 0.06$ eV, and are expected to make a detection of the neutrino mass (see Font-Ribera et al., 2014). However, this is to be achieved purely by a more precise measurement of the 2PCF, not by the inclusion of extra information.

We know from $N$-body simulations that the universe at late times appears as a cosmic web (Bond et al., 1996). Currently this cosmic web structure is not fully incorporated into the inference of cosmological parameters. In this Chapter we turn to graph theory, looking specifically at the minimum spanning tree (MST), to try to capture some of this rich information. The MST was first introduced to astronomy by Barrow et al. (1985). It has been typically used in cosmology for LSS classification, for example to search for cosmic web features such as filaments (see Bhavsar & Ling, 1988; Pearson & Coles, 1995; Krzewina & Saslaw, 1996; Ueda & Itoh, 1997; Coles et al., 1998; Adami & Mazure, 1999; Doroshke-
vich et al., 1999, 2001; Colberg, 2007; Balázs et al., 2008; Park & Lee, 2009; Adami et al., 2010; Demianínski et al., 2011; Durret et al., 2011; Cybulski et al., 2014; Alpaslan et al., 2014; Shim & Lee, 2013; Shim et al., 2014, 2015; Beuret et al., 2017; Campana et al., 2018a,b; Libeskind et al., 2018; Clarke et al., 2019). It has also been used in other contexts such as determining mass segregation in star clusters (Allison et al., 2009) and the generalized dimensionality of data points, fractals and percolation analysis (see Martinez & Jones, 1990; van de Weygaert et al., 1992; Bhavsar & Splinter, 1996). More recently, the MST was used in particle physics to distinguish between different classes of events in collider experiments (Rainbolt & Schmitt, 2017). The MST’s strength is in its ability to extract patterns; this is precisely why it has been used to extract cosmic web features (the type of information currently missing from most cosmological studies). The MST’s weakness lies in the fact that the statistics cannot be described analytically and depend heavily on the density of the tracer. This means any comparison of models via the MST will be dependent on simulations. While this makes parameter inference more challenging, the reliance on simulations is not new; in fact parameter inference through artificial intelligence (AI) and machine learning (ML) will be similarly reliant. Here, the MST may provide a bridge between the traditional 2PCF and AI/ML, allowing us to understand the information being extracted by these AI/ML algorithms.

Our goal in this Chapter is to understand whether the MST could be a useful tool for cosmological parameter inference for current or future photometric and spectroscopic galaxy redshift surveys. These include the Baryon Oscillation Spectroscopic Survey, the Dark Energy Survey, DESI, the Nancy Grace Roman Space Telescope, the Prime Focus Spectrograph, the Rubin Observatory Legacy Survey of Space and Time and 4-metre Multi-Object Spectroscopic Telescope. With this in mind, the Chapter is organized as follows. In Section 2.3, we describe the MST construction and statistics and we summarize the suites of simulations used in later Sections. In Section 2.4, we demonstrate that the MST is sensitive to higher order statistics (i.e. beyond two-point). In Section 2.5, we explore relevant sources of systematics and methods to mitigate them. In addition, we test the sensitivity to redshift space distortions (RSD). In Section 2.6, we explore the MST statistics on an unbiased tracer, and try to determine what the MST is actually measuring about the underlining density distribution. Lastly, in Section 2.7, we compare the MST’s constraining power to that of the more traditional power spectrum and bispectrum measurements.

1http://www.sdss3.org/surveys/boss.php
2http://www.darkenergysurvey.org
3http://desi.lbl.gov/
4http://www.euclid-ec.org/
5https://roman.gsfc.nasa.gov
6https://pfs.ipmu.jp/index.html
7https://www.lsst.org/
8https://www.4most.eu/cms/
In mathematics, a graph is a set of nodes (points) together with a set of edges, where each edge joins two distinct nodes; given any two distinct nodes, there will be either zero or one edge between them. In this Chapter, all graphs are undirected and weighted i.e. an edge does not have an orientation, but it does have a (positive) weight (which in this Chapter will be the distance (defined below) between the nodes that it connects). A path is a sequence of nodes in which each consecutive pair of nodes is connected by an edge (and no edge is used twice); a path that returns to its starting point is a cycle. If there is an edge (respectively path) between any two distinct nodes then the graph is complete (respectively connected). Given a connected graph $G$ (not necessarily complete), one can discard edges to obtain the MST of $G$. By definition this new graph is spanning (i.e. contains all the nodes of $G$), is a tree (i.e. is connected and contains no cycles) and is minimal in that the sum of the
edge weights is minimal among all spanning trees. Every connected graph has a unique MST (Abraham, 1962; Gower & Ross, 1969; Zahn, 1971), which can be demonstrated by the deterministic algorithms used to construct them (such as Prim 1957 or Kruskal 1956). Since a global minima is sought for the entire tree, the MST is stable to relatively small perturbations and such changes will only occur for edges local to any perturbation.

In this work we consider sets of points in various spaces, with distance between points defined to be:

- In two and three dimensions: Euclidean distance;
- On the sphere (i.e. RA, Dec.): subtended angle;
- Using RA, Dec., redshift: convert redshift to comoving distance (using the fiducial cosmology), then use Euclidean distance.

Given a set of points $S$ we wish to investigate the MST of the complete graph on these points (i.e. there is an edge between every pair of points and all these edges are candidates for inclusion in the MST); we refer to this as the MST of $S$. See Figure 2.1 for an example of such an MST. Now Kruskal’s algorithm (Kruskal, 1956) (described below) takes as input a connected graph (not necessarily complete) and discards certain edges so as to find its MST. In theory, we should input to this algorithm the complete graph on $S$. However this is inefficient as the complete graph contains many edges (e.g. between widely separated points) that are very unlikely to appear in the output MST; it is sufficient to input to Kruskal’s algorithm a pruned graph that retains only shorter edges.

To this end, we use as input to Kruskal’s algorithm the $k$ nearest neighbours graph ($k$NN), i.e. the graph in which each point has an edge to its $k$ nearest neighbours. Here $k$ is a free parameter (and should not be confused with the wavenumber used in harmonic analysis). We calculate this graph using the kneighbors_graph function from scikit-learn.9 Note that if $k$ is too small then the $k$NN graph need not be connected (it might consist of several isolated islands); in most cases considered, $k > 10$ ensures that $k$NN will be connected (but when applying scale cuts (see Section 2.5.2) a larger $k$ is needed).

We then apply the scipy minimum spanning tree10 function, which implements Kruskal’s algorithm. This algorithm removes all the edges from the graph, sorts these removed edges by length (shortest to longest), and then sequentially re-embeds them, omitting an edge if its inclusion would create a cycle. This continues until all points are connected into a single tree. The Kruskal algorithm can be shown to scale as $O(N_E \log N_V)$ (see Cormen et al., 2009, Section on Kruskal’s algorithm) where $N_E$ is the number of edges in the supplied spanning graph and $N_V$ is the number of nodes. At most $N_E \simeq N_V^2$ but this can

---

9 http://www.scikit-learn.org
10 https://scipy.org/
be greatly reduced by using the \( k \text{NN} \) graph, which changes the scaling from \( \mathcal{O}(n^2 \log n) \), where \( n \) is the number of nodes, to \( \mathcal{O}(kn \log n) \). Since usually \( k \ll n \) this greatly reduces computation time.

We tested the sensitivity to the choice of \( k \) by using a graph with \( 256^3 \) points (HZ = High \( \sigma_8 \) and zero \( \sum m_\nu \) simulations at \( z = 0 \) explained later in Section 2.6). We compared the total length of the MST when \( k = 50 \) (a proxy for \( k = \infty \)) and found a fractional difference of \( \sim 2 \times 10^{-6} \) for \( k = 20 \), \( \sim 2 \times 10^{-7} \) for \( k = 30 \), and \( \sim 3 \times 10^{-8} \) for \( k = 40 \). It appears that \( k = 20 \) gives a good balance between computation time and an accurate estimation of the MST, so we use this as a fiducial value unless stated otherwise.

2.3.1 Statistics from the Minimum Spanning Tree

Any given MST is a complex structure with many interesting features. In this study, we are not interested in these individual features but rather the overall properties and their relation to cosmological parameters. Taking inspiration from Rainbolt & Schmitt (2017) and Krzewina & Saslaw (1996) we measure the probability distribution (i.e. histograms) of the following:

- Degree \( (d) \): the number of edges attached to each node.
- Edge lengths \( (l) \): the length of edges.
- From branches, which are chains of edges connected with intermediary nodes of \( d = 2 \), we measure:
  - Branch lengths \( (b) \): the sum of edges that make up the branch.
  - Branch shape \( (s) \): the straight line distance between the branch ends divided by the branch length.

These statistics are displayed in Figure 2.1. Of course one could consider other statistics to extract from the MST (see Alpaslan et al., 2014) but we choose to explore these as they have been shown to successfully aid in the classification of particle physics interactions (see Rainbolt & Schmitt, 2017). The MST will have a total of \( n - 1 \) edges (Kruskal, 1956), where \( n \) is the number of nodes. Since each edge has a node on either end, each edge contributes twice to the total degree of the MST. Hence the expectation value for \( d \) will be:

\[
(d) = \frac{2(n - 1)}{n} \simeq 2.
\]

(2.1)

By definition the branch shapes satisfies \( 0 \leq s \leq 1 \). Often \( s \) is near 1, so to facilitate visual comparison we frequently plot \( \sqrt{1 - s} \) instead of \( s \). Straighter branches correspond to \( \sqrt{1 - s} \) closer to zero.
Table 2.1: A summary of the simulation suites used in this study. For each simulation suite we list its name, the method used to produce it, the point distribution used and the use to which it is put.

<table>
<thead>
<tr>
<th>Name</th>
<th>Method</th>
<th>Points</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illustris</td>
<td>Hydrodynamic</td>
<td>Subhaloes</td>
<td>Testing the sensitivity of the MST to higher order statistics</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(i.e. beyond two-point).</td>
</tr>
<tr>
<td>MICE</td>
<td>N-body</td>
<td>Galaxies</td>
<td>Exploring the sensitivity to RSD.</td>
</tr>
<tr>
<td>νN-body</td>
<td>N-body</td>
<td>Dark matter particles and haloes</td>
<td>Using an unbiased tracer we look to find what the MST is actually measuring.</td>
</tr>
<tr>
<td>PICOLA</td>
<td>COLA</td>
<td>Haloes</td>
<td>Comparing sensitivity of the MST to traditional methods.</td>
</tr>
</tbody>
</table>

Additionally it is useful in certain circumstances, particularly when comparing MSTs that contain different number of nodes, to look at the dimensionless parameters of:

- \( \ln(\bar{l}) \), where \( \bar{l} = l/\langle l \rangle \) and \( \langle l \rangle \) is the average edge length.
- \( \ln(\bar{b}) \), where \( \bar{b} = b/\langle b \rangle \) and \( \langle b \rangle \) is the average branch length.

Comparing the distribution of these dimensionless parameters is only appropriate if the distribution of points is scale-independent. In cosmology this is not necessarily the case for higher order statistics, so these should be used sparingly.

### 2.3.1 Computational Issues for Finding Branches

Once the MST is constructed, we know the edge lengths \( l \) and the indices of the nodes at either end of the edges. These can be trivially used to find the degree \( d \) of each node and edge end. To find branches, we search for edges joining a \( d = 2 \) node to a \( d \neq 2 \) node (i.e. ‘branch ends’) and edges joining two \( d = 2 \) nodes (such edges, which form the middle parts of branches, are referred to as ‘branch mids’). To find the branches we begin with a branch end, search for a branch mid that is connected to it, and continue to grow the branch until no more branch mids can be added. At this point we then search for the branch end that finishes it. This is a computationally expensive procedure but can be trivially made faster by dividing the entire tree into Sections and running the algorithm on the Sections independently. Branches straddling the boundaries will be left incomplete, but can be completed by matching any remaining incomplete branches.

MSTree (Naidoo, 2019), the Python package to construct the MST and derive its statistics, is made publicly available.\(^\text{11}\)

### 2.3.2 Error Estimation

Uncertainties for the MST statistics are generated in two ways.

- In the cases where many realizations of a data set can be generated easily we will estimate the mean and standard deviation from an ensemble of realizations.

\(^{11}\text{https://github.com/knaidoo/two.lf/nine.lf/mistree} \)
• If only a single realization is available we will use jackknife errors. Here, we divide up our data set into $n$ regions and run the analysis $n$ times, each time removing a single different region from the analysis yielding an output $\theta_i$. The errors, $\Delta \theta_{\text{jack}}$, are estimated using

$$
\Delta \theta_{\text{jack}} = \left[ \frac{n-1}{n} \sum_{i=1}^{n} (\theta_i - \bar{\theta})^2 \right]^{1/2},
$$

where $\bar{\theta}$ is the average of $\theta_i$.

### 2.3.3 Simulation Summary

We use several simulations suites; these are summarised in Table 2.1. We discuss these simulations in greater detail in the relevant Sections of the Chapter where they are used.

### 2.4 SENSITIVITY OF THE MST TO COSMIC WEB PATTERNS

#### 2.4.1 Heuristic Argument

There are compelling reasons to believe the MST should be sensitive to cosmic web patterns. Consider how the Kruskal algorithm constructs the MST (see Section 2.3). An edge is added only if this does not create a cycle; this means that the very construction of the MST requires an awareness of neighbouring edges or more generally the environment each edge inhabits. More generally this means the inclusion of a single edge is not defined solely by the 2PCF but by its local environment. Therefore, we should expect the MST to contain more information than is present in the 2PCF.

#### 2.4.2 Illustris vs. Adjusted Lévy Flight

Testing whether the MST is sensitive to higher order statistics is rather challenging since at present there are no analytical descriptions of the MST statistics.

To go around this theoretical limitation we instead carry out an analysis similar to that of Hong et al. (2016), comparing the Illustris\textsuperscript{12} (Nelson et al., 2015; Vogelsberger et al., 2014) simulations (see Section 2.4.2.1) to an adjusted Lévy flight (ALF) simulation that is tuned to have almost identical 2PCF but different higher order information.

Lévy flights (Mandelbrot, 1982) are random walk simulations where the step size (the distance between one point and the next) is given by a fat-tailed power-law probability distribution function (PDF). This ensures that its 2PCF will follow a power law (see Mandelbrot, 1982) similar to that found for galaxies. However, although a standard Lévy flight scheme may be able to replicate the 2PCF at large scales, at small scales, the 2PCF eventually

\textsuperscript{12}http://www.illustris-project.org
Figure 2.2: Top panels: the left shows the Illustris galaxy sample and the middle panel shows one realization of the ALF. Visually these two simulations are different in their distribution of galaxies. However they have virtually identical 2PCF by construction (right-hand panel). Illustris measurements are shown in blue and the mean for 100 realizations of the ALF is shown by the green dashed line and green envelopes show the $1\sigma$ (darker) and $2\sigma$ regions. Bottom panels: the histogram distributions of the MST statistics (from left to right): degree ($d$), edge length ($l$), branch length ($b$), and branch shape ($s$; note we plot the $\sqrt{T - \hat{s}}$ value instead because the distribution peaks towards 1 and it is easier to see the difference in this projection). The difference between the PDF is displayed in the bottom subplots where zero on the $y$-axis corresponds to the mean counts for the ALF PDF. The measurements from the MST are significantly different for each of these simulations. In particular the distributions of edge lengths and branches show some bimodality for the Illustris sample which is not present in the ALF. This demonstrates the sensitivity of the MST to patterns in the cosmic web as the bimodal distribution appears to be driven by void and cluster environments (explored in Section 2.6.2.2).
plateaus (see Hong et al., 2016). Since the MST is sensitive to small scales, it is important that the Lévy flight simulation match that of the Illustris sample at small scales. We are able to match the 2PCF of the Illustris sample at all scales using an adjusted Lévy flight (ALF) simulation explained below.

### 2.4.2.1 Illustris Galaxy Sample

We use the subhalo catalogue of the Illustris-I snap 100 sample and follow Hong et al. (2016) to include only subhaloes which are large and dark-matter-dominated:

\[
M_* \geq 10^8 M_\odot, \\
M_* < 0.63M_{DM},
\]

where \(M_*\) and \(M_{DM}\) are the stellar and dark matter mass of the subhaloes respectively. We will refer to this as the Illustris galaxy sample.

### 2.4.2.2 Adjusted Lévy Flight

We generate an ALF simulation with the same number of ‘galaxies’ as our Illustris sample and (almost) the same 2PCF. For comparison with Illustris we enforce periodic boundary conditions. The standard Lévy flight has step sizes \(t\) with cumulative distribution function (CDF),

\[
\text{CDF}(t) = \begin{cases} 
0 & \text{for } t < t_0, \\
1 - \left(\frac{t}{t_0}\right)^{-\alpha} & \text{for } t \geq t_0,
\end{cases}
\]

where \(t_0\) and \(\alpha\) are free parameters. This yields a simulation with a power-law 2PCF of the form \(C(t_0, \alpha)t^{3-\alpha}\) at scales larger than \(t_0\) (where \(C(t_0, \alpha)\) is a constant determined by the free parameters), below this scale the 2PCF plateaus (see Hong et al., 2016). To have control of the 2PCF below scales of \(t_0\) we introduce an ALF model with the following CDF:

\[
\text{CDF}(t) = \begin{cases} 
0 & \text{for } t < t_s, \\
\beta \left(\frac{t-t_s}{t_0-t_s}\right)^\gamma & \text{for } t_s \leq t < t_0, \\
(1 - \beta) \left[1 - \left(\frac{t}{t_0}\right)^{-\alpha}\right] + \beta & \text{for } t \geq t_0.
\end{cases}
\]

This introduces three new parameters: \(\beta, \gamma,\) and \(t_s\). Rather than having a step size probability distribution function (PDF) that jumps from zero to a maximum at \(t_0\), the ALF is constructed to have a slow rise to the maximum at \(t_0\). The second piece of the CDF describes a transfer function that operates between \(t_s\) and \(t_0\) (where by definition \(t_s < t_0\)). Here \(\gamma\) allows us to control the gradient of this rise and \(\beta\) allows us to define the fraction of step sizes below \(t_0\).
CHAPTER 2

Figure 2.3: The MST statistics, calculated tomographically on random points placed in the BOSS CMASS North footprint (placed with the same density as the BOSS CMASS galaxies), with (red) and without (blue) using the CMASS mask. We see a significant shift towards longer edges in the MST performed with the mask, with a similar effect seen in the distribution of branch lengths. For the degree and branch shape the masking has no statistically significant effect.

2.4.2.3 Comparison

The Illustris sample contains 63,453 galaxies. We create a sample of the same size using an ALF model with parameters $\alpha = 1.5$, $t_0 = 0.325$, $t_s = 0.015$, $\beta = 0.45$, and $\gamma = 1.3$ (where length-scales $t_0$ and $t_s$ are given in $h^{-1}$ Mpc). The two samples have approximately equal 2PCFs down to scales of 0.01 $h^{-1}$ Mpc by construction. The 2PCF was calculated on a single realization of the ALF model with varying $\beta$, $\gamma$, $t_s$ and $t_0$ ($\alpha = 1.5$ was kept constant, see Hong et al. 2016). We then chose the parameters that produced the closest match, i.e. by minimizing the sum of difference between the 2PCF in log space. The Illustris and ALF sample show widely different MST statistics (see Figure 2.2), thereby demonstrating the sensitivity of the MST to higher order statistics. The bimodal distribution of edge and branch lengths shown in Figure 2.2 occurs in over- and underdensities (explored in more detail in Section 2.6). Note also that we see differences in the shape of branches and the distribution of degrees to a statistically significant level, although these differences are not as striking as the difference in edge and branch length distributions.

2.5 BOUNDARY EFFECTS AND REDSHIFT SPACE DISTORTIONS

We study possible sources of systematic errors that could affect the MST. In particular we would like to establish to what extent simulations need to replicate survey properties.

2.5.1 Boundary Effects

Galaxy surveys often contain complex survey footprints with regions masked due to stars and varying completeness and it is important to understand how such footprints will affect the MST. Imposing a mask on the data set results in two effects:

1. Additional edges are included to join nodes near the boundaries. These would have otherwise been joined by nodes outside the boundary in a larger MST.
2. New edges are located near the centre whose purpose appears to be to unify the structure as a single spanning tree. In a larger spanning tree, these separated regions would be connected through routes that extend beyond the boundary.

The net result of these effects is to create a slight bias towards longer edges and slightly longer branches. Interestingly, all edges in the larger MST (within the boundary) are present in the smaller MST. This property always holds, as can easily be proven using the ‘cycle property’ of the MST (see Katriel et al., 2003).

We investigate the effects of a realistic mask by using the BOSS CMASS MD-Patchy mocks North mask (Rodríguez-Torres et al., 2016), which includes masking for bright stars, bad fields, centrepost and collision priority.11 In Figure 2.3 we demonstrate the effects of this mask on random points placed within the CMASS footprint (with the same density as the CMASS galaxies) with and without a mask. The MST is then calculated on 1000 realizations tomographically (i.e. on the sphere). The degree and branch shape show little change but the distribution of edge lengths show a significant tendency towards longer edges when a mask is used. This is mirrored by a similar effect in the distribution of branch lengths. This is because the mask eliminates shorter paths, forcing the MST to include longer edges that would (without the mask) have been excluded. This demonstrates that realistic masks with holes do have an impact on the MST and must be included in any future analysis.

2.5.2 Scale Cuts

In cosmology there is often a need to apply scale cuts in real space. This can occur for a variety of reasons: theoretical uncertainty at small scales both from simulation and from analytic formulae and also practically from fibre collisions in spectroscopic surveys. For the 2PCF this is rather simple to mitigate; you simply restrict the domain of the 2PCF to exclude separations below the scale cut. With the MST this is more complicated. Unfortunately there does not appear to be a way to deal with this after the MST has been constructed; this is because the problematic smallest edges will by construction be incorporated in the graph. To ensure that problematic small scales are removed from the MST we alter the $k$NN graph that is the input to the Kruskal algorithm by removing edges whose length is below the desired scale cut.

2.5.3 Redshift Space Distortion on MICE Galaxies

RSDs (Kaiser, 1987), caused by the Kaiser and Fingers of God effects, will distort the measured redshift of galaxies and thus will impact the inferred comoving distance. Since this effect alters the 3D distribution of galaxies, it will inevitably affect the MST statistics.

11See http://www.sdss3.org/dr9/algorithms/boss_tiling.php#veto_masks
Figure 2.4: The effects of RSDs on the MST statistics. From left to right: the MST statistics degree ($d$), edge length ($l$), branch length ($b$), and branch shape ($s$). Bottom panels show the differences. Ten realizations of 500 000 MICE galaxies were generated and the MST were constructed on their true positions (grey) and then the measured positions (red), i.e. the inferred positions based on their redshifts including RSD. The envelopes correspond to 1σ uncertainties. Significant differences between the MST statistics show that the MST is sensitive to the RSD effect.

We explore this effect by comparing the MST performed on a subset of the MICE galaxy catalogue (Crocce et al., 2015) in real and redshift space (i.e. with RSD). Here, we randomly draw 10 realizations of 500 000 galaxies with real comoving distances between 1000 to 1500 $h^{-1}$Mpc. We ensure that the density of galaxies is constant so that the number of galaxies $\propto D_c^3$, where $D_c$ is the radial comoving distance from the observer.

Figure 2.4 shows the MST statistics with and without the RSD effect. We see significant results in all the MST statistics demonstrating the importance of including this effect in any future MST study.

2.6 WHAT DOES THE MINIMUM SPANNING TREE MEASURE?

This Section considers the following questions:

1. What do the MST statistics look like on an unbiased tracer (i.e. $N$-body dark matter particles)?
2. What does the MST statistics tell us about the underlining density distribution?
3. What is the relation of MST statistics to 2PCF?
4. What happens when we change simulation resolution?
5. How do the MST statistics change when measured on haloes (i.e. a more galaxy-like tracer)?

2.6.1 $\nu N$-body Simulations

Five $N$-body simulations (see Massara et al., 2015) were made by running the TreePM code GADGET-III (Springel, 2005). The following cosmological parameters were common to
Table 2.2: Simulation names and cosmological parameters for the \(N\)-body simulations. Massara et al. (2015) uses different names, which we list here.

<table>
<thead>
<tr>
<th>Name</th>
<th>Reason for name</th>
<th>Massara et al. (2015)</th>
<th>(L_{\text{Box}}) ((h^{-1}\text{Mpc}))</th>
<th>(N_{\text{cdm}})</th>
<th>(N_\nu)</th>
<th>(\sum m_\nu) ((\text{eV}))</th>
<th>(\sigma_8)</th>
<th>(10^9 A_{\text{s}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>HZ</td>
<td>High (\sigma_8), zero (\sum m_\nu)</td>
<td>L0</td>
<td>1000</td>
<td>256(^3)</td>
<td>0</td>
<td>0.834</td>
<td>2.13</td>
<td></td>
</tr>
<tr>
<td>LZ</td>
<td>Low (\sigma_8), zero (\sum m_\nu)</td>
<td>L0s8</td>
<td>1000</td>
<td>256(^3)</td>
<td>0</td>
<td>0.693</td>
<td>1.473</td>
<td></td>
</tr>
<tr>
<td>LN</td>
<td>Low (\sigma_8), non-zero (\sum m_\nu)</td>
<td>L60</td>
<td>1000</td>
<td>256(^3)</td>
<td>0.6</td>
<td>0.693</td>
<td>2.13</td>
<td></td>
</tr>
<tr>
<td>HZHR</td>
<td>High (\sigma_8), zero (\sum m_\nu), high res.</td>
<td>H00</td>
<td>500</td>
<td>512(^3)</td>
<td>0</td>
<td>0.834</td>
<td>2.13</td>
<td></td>
</tr>
<tr>
<td>LNHR</td>
<td>Low (\sigma_8), zero (\sum m_\nu), high res.</td>
<td>H160</td>
<td>500</td>
<td>512(^3)</td>
<td>0.6</td>
<td>0.693</td>
<td>1.473</td>
<td></td>
</tr>
</tbody>
</table>

all simulations: \(\Omega_m = 0.3175\), \(\Omega_b = 0.049\), \(\Omega_\Lambda = 0.6825\), \(h = 0.6711\), and \(n_s = 0.9624\). See Table 2.2 for a list of the simulations used and their respective cosmological parameters, particle numbers and box sizes. The cold dark matter energy density is set to \(\Omega_c = \Omega_m - \Omega_b - \Omega_\nu\), where \(\Omega_\nu h^2 \approx \sum m_\nu/(94.1 \text{ eV})\). Cold dark matter and neutrinos are both treated as collisionless particles. They differ in their masses and in their initial conditions, where the initial conditions for neutrinos receive an extra thermal velocity obtained by randomly sampling the neutrino Fermi–Dirac momentum distribution (Viel et al., 2010). These are evolved from an initial redshift of \(z = 100\). The simulations used are summarised in Table 2.2.

### 2.6.2 MST Application to Dark Matter Particles

An MST was constructed on the dark matter particles from the HZ, LZ, and LN simulations (see Table 2.2), where errors were calculated using the jackknife method (Section 2.3.2). Figures 2.5, 2.6, 2.7, and 2.9 use the same colour scheme: HZ in blue, LZ in orange and LN in green. We boost the speed of the MST calculation by allowing this to be done in parallel, breaking the \(N\)-body snapshots into 64 cubes. We then implement the scale cut strategy discussed in Section 2.5.2 and partition the data set into four groups (to dilute the sample to look at larger sales) and apply a scale cut of \(l_{\text{min}} = 2 h^{-1}\text{Mpc}\).

#### 2.6.2.1 Features in the Minimum Spanning Tree Statistics

In Figure 2.5 we plot the MST statistics for these different simulations at redshifts \(z = 2\), 1, 0.5 and 0. The plots display how the MST statistics evolve over cosmological time, as discussed below:

- **Degree:** the distribution of degree remains relatively similar in all simulations and does not appear to evolve greatly over redshift, although differences between the simulations become more pronounced at lower redshifts.

- **Edge length:** overall we see that the distribution shows a high sensitivity to redshift, evolving from a single distribution into a bimodal one at smaller redshift.
Figure 2.5: From left to right: the distribution of degree \((d)\), edge length \((l)\), branch length \((b)\) and branch shape \((s)\). These are obtained by dividing the full \(1 (h^{-1}\text{Gpc})^3\) box into \(250 (h^{-1}\text{Mpc})^3\) cubes for speed. These are then partitioned into four groups to minimise the effect of applying a scale cut of \(2 h^{-1}\text{Mpc}\). From top to bottom: distributions are shown with respect to redshift \(2, 1, 0.5\) and \(0\). These are further subdivided into a top subplot of the distributions and a bottom subplot of the differences. Simulations shown are HZ (blue), LZ (orange) and LN (green). See Section 2.6.2.1 for a detailed explanation of the distribution features, differences, and evolution.
Figure 2.6: Contour plots of the average density contrast ($\delta$) is plotted against the MST statistics [from left to right: the average degree ($\langle d \rangle$), edge length ($\langle l \rangle$), branch length ($\langle b \rangle$) and branch shape ($\langle s \rangle$)] in 25 $h^{-1}$Mpc cubes. The 1$\sigma$ and 2$\sigma$ contours are indicated by solid and dashed lines, respectively. The relation for HZ is in blue, LZ in orange and LN in green. See Section 2.6.2.2 for a detailed explanation of the relation and their evolution.
○ $l \geq 3\, h^{-1}\text{Mpc}$: a broad peak is seen in the distribution at $l \simeq 4\, h^{-1}\text{Mpc}$. This feature dampens at lower redshift with the peak consistently highest for LN, followed by LZ and then HZ.

○ $l < 3\, h^{-1}\text{Mpc}$: a secondary peak emerges and dominates at lower redshift, which rises against the scale cut limit of $l_{\text{min}} = 2$.

○ $l \sim 3\, h^{-1}\text{Mpc}$: between the two peak features is a region where seemingly all three distributions appear to converge and the orderings of the peaks above and below this point switch.

- Branch length: the evolution appears virtually identical to the edge length distribution except at larger scales.

- Branch shape:
  ○ A broad peak at $\sqrt{1-s} = 0.6$ which is present in all simulations. This peak is always highest for LN followed by LZ and HZ.
  ○ A subpeak at $\sqrt{1-s} \sim 0.05$ which dampens at lower redshift. This suggests that some branches at low redshift are fairly straight. Since the simulation we use are fairly low in resolution we suspect that this feature is more an indication that the particles have not undergone much mixing and are still very close to their initial perturbed grid layout. This could be used as a diagnostic to test whether $N$-body simulations have moved from their perturbed gridded initial conditions.
  ○ Lastly we see the emergence of two bumps between $\sqrt{1-s} \sim 0.7 - 1$ at low redshift. Comparison of the branch shape statistics with and without a scale cut show this is caused by the introduction of the scale cut, which forces some branches to be more curved. Branch shapes without a scale cut rarely see $\sqrt{1-s} > 0.8$.

2.6.2.2 Exploring the Minimum Spanning Tree Relation to Density

To gain a greater physical intuition of what these statistics are telling us about cosmology, we subdivide the $1\, h^{-1}\text{Gpc}$ cube into smaller $25\, h^{-1}\text{Mpc}$ cubes. In these cubes we calculate the density contrast $\delta$,

$$\delta = \frac{N_{\text{DM}}}{\langle N_{\text{DM}} \rangle} - 1,$$

where $N_{\text{DM}}$ is the number of dark matter particles in a particular cube and and $\langle N_{\text{DM}} \rangle$ is the average across all cubes. Figure 2.6 illustrates the relationship between the average
Figure 2.7: In the top panels the matter power spectra, $P(k)$, are plotted for redshift (from left to right) 2, 1, 0.5 and 0 for simulations HZ (blue), LZ (orange) and LN (green). In the bottom subplots we plot the ratio with respect to the HZ power spectra. Solid lines correspond to the measured $P(k)$ from the respective simulations, while dashed and dotted lines correspond to the theoretical linear and non-linear $P(k)$, respectively. The dashed grey lines shows the level at which the measured $P(k)$ will be affected by the shot noise of the simulation and the regions in red show the scales for which we apply a scale cut in the construction of the MST. Here, we see (that at all redshifts) the power at all $k$ is highest for HZ, and then LZ and lastly LN. Note that LZ is close to HZ at high $k$ and close to LN at low $k$.

degree ($\langle d \rangle$), edge length ($\langle l \rangle$), branch length ($\langle b \rangle$) and branch shape ($\langle s \rangle$) and the density contrast inside these cubes.

- $d$ vs $\delta$: we see that the mean of the degree, $d$, is relatively constant at $d \simeq 2$ as a function of density. The variance shows a strong dependence on density, with over densities having very low variance, i.e. predominantly $d = 2$, and under densities showing a much larger variance and a slight tilt towards $d = 1$. Of course we should expect high-density environment to form the main ‘backbone’ of the MST, since these are the areas where the edges are shortest.

- $l$ and $b$ vs $\delta$: both the edge and branch length distribution show a very similar relation to density. Shorter edges and branches are mostly associated with overdensities and vice versa. Furthermore as the simulations evolve in redshift this relation becomes more pronounced. In both these statistics, we see that HZ appears consistently to have more overdense and underdense regions than the other two simulations. We also see that LN appears to have marginally but consistently higher overdense and underdense regions than LZ.

- $s$ vs $\delta$: the mean of the branch shape appears centred at 0.75 and shifts slightly to a mean of 0.7 for higher densities. Furthermore, as with the degree, the biggest relation to density is with the variance, which increases as the density lowers.
This analysis demonstrates a clear relation between MST statistics and environment (i.e. the local density).

2.6.2.3 Relation to the Matter Power Spectrum

In Figure 2.7, we calculate the matter power spectra, $P(k)$, measured from these simulations. The dependence on redshift can be characterized by a simple shift in amplitude. We see that (at all $k$) HZ has more power, followed by LZ and then LN. At small $k$, LZ converges to HZ while at large $k$, LZ converges to LN. Notice, that the strength in $P(k)$ at large $k$ is matched by a tendency for shorter edges in the MST, demonstrating the MST expected dependence on clustering.

2.6.2.4 Simulation Resolution

The MST of $N$-body simulations will be affected by the resolutions used. To measure the sensitivity of the MST statistics to the simulation resolution we calculate the MST on higher resolution versions of HZ and LN called HZHR and LNHR (see Table 2.2 for details of simulation properties). The resulting distributions of the MST statistics are shown in Figure 2.8. For comparison we additionally subsample these two simulation boxes by randomly selecting particles in the simulation with equal number of particles. In the more sparsely sampled version of HZHR and LNHR the more resolved extreme high- and low-density environments are still imprinted. This can be seen by the fact that in the bottom panels of Figure 2.8 there appears to be more features at high and low values of $l$. This illustrates the importance of high-resolution simulations on the MST profiles inferred. We could also use high-resolution simulations to calibrate the scale cut for low-resolution simulations by allowing the scale cut to vary until the MST statistics reach agreement between the high- and low-resolution simulations. We additionally measure the MST on a completely random set of points (shown in grey) illustrating how the more sparsely subsampled data set appears to be asymptotically approaching these profiles.

2.6.3 MST Application to Haloes

Halo catalogues were derived from the HZ, LZ, and LN simulation snapshots. We study these to get a sense of what the MST statistics will look like when performed on a biased tracer, such as galaxies. We dropped the $z = 2$ snapshots as they contained too few haloes to be meaningful. Unlike the $N$-body simulation, we do not apply a scale cut since the density of haloes is quite low and the fraction of edges below $l_{\text{min}} = 2 \, h^{-1}\text{Mpc}$ is very low. The MST statistics derived from the haloes is shown in Figure 2.9. The number of haloes varies both across simulations and across redshift snapshots (see Table 2.3) – this is different from dark matter particles whose number count is constant across redshift and simulations.
Figure 2.8: The distribution of degree ($d$), normalized edge and branch lengths ($\ln(\bar{l})$ and $\ln(\bar{b})$) and branch shape ($s$) are displayed from left to right. Top panel: the distributions of the high-resolution versions of HZ (dashed line) and LN (dotted line) (i.e. HZHR and LNHR) simulation are shown in red and subsequently subsampled versions are shown in green and blue with dark matter particle densities ($\rho_{DM}$) $256^3$, $128^3$ and $64^3$ per ($h^{-1}$Gpc)$^3$ respectively. Middle panel: the distribution for the HZ (dashed line) and LN (dotted line) simulation is shown. Bottom panel: the differences between the high resolution (top panel) and low-resolution (middle panel) simulations are shown. We additionally illustrate the distribution for random points (dashed grey).
Figure 2.9: The MST constructed on halo catalogues derived from the HZ (blue), LZ (orange), and LN (green) N-body simulations. From left to right are the MST statistics: degree (d), edge length (l), branch length (b) and branch shape (s). They are plotted from top to bottom according to snapshots at redshift 1, 0.5, and 0. Corresponding shaded areas show the jackknife uncertainties in the measurements. The distribution of the MST statistics are indistinguishable from each other at all redshifts, demonstrating that we should expect to see similar lines of degeneracy as power spectrum. Note $N = 10^3$. 

\[ N^* = 10^3 N \]
Table 2.3: The number of haloes found in each simulation (HZ, LN and LZ) for each redshift (z) snapshot. The number of haloes at $z = 2$ was far too little for a meaningful MST and presumably would be uninformative.

<table>
<thead>
<tr>
<th>Redshift</th>
<th>HZ</th>
<th>LN</th>
<th>LZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1791</td>
<td>11168</td>
<td>9892</td>
</tr>
<tr>
<td>0.5</td>
<td>6717</td>
<td>3017</td>
<td>2392</td>
</tr>
<tr>
<td>1</td>
<td>1585</td>
<td>458</td>
<td>262</td>
</tr>
<tr>
<td>2</td>
<td>16</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

To mitigate this issue, for each redshift we match the number of haloes to the lowest number found in the simulations (thus always matching the number of haloes found in the LZ simulations). For those with more haloes, we simply select the most massive haloes. In Figure 2.9, we find no real noticeable difference in the statistics suggesting the degeneracies of the MST may be similar to that found for $P(k)$.

### 2.7 COMPARING THE SENSITIVITY TO COSMOLOGY OF THE POWER SPECTRUM, BISPECTRUM AND THE MINIMUM SPANNING TREE

In this Section we compare the sensitivities to cosmological parameters of the power spectrum $P(k)$, bispectrum $B(k_1, k_2, k_3)$ and MST, measured on the same halo catalogues, to establish whether the MST can improve parameter constraints. Specifically, we compare the constraints on $A_s$, $\Omega_m$ and $\sum m_\nu$ for 10 sets of mock simulations. To obtain reliable posterior distributions for the three methods and their joint constraints, we would normally run an Markov Chain Monte Carlo (MCMC) using an analytic expression for the data vector. However, the MST statistics cannot be obtained analytically and hence have to be obtained from simulations. $P(k)$, $B(k_1, k_2, k_3)$, and MST are therefore estimated from a grid of simulations in parameter space. To limit the noise in the estimates of the theory we take the mean of five simulations rather than just one at each point in parameter space. Additionally, since our simulation grid is rather sparse we use Gaussian process (GP) regression to interpolate the data vector. Finally we use a corrected likelihood function (see Sellentin & Heavens, 2016; Jeffrey & Abdalla, 2019) which accounts for the use of an estimated covariance matrix.

#### 2.7.1 COLA Simulation Suites

A suite of COLA (Tassev et al., 2013) simulations were constructed using the MG-PICOLA software (Winther et al., 2017, an extension to L-PICOLA by Howlett et al. 2015) which, among other things, can model the effects of massive neutrinos (Wright et al., 2017). This allowed us to generate $N$-body-like simulations relatively cheaply (in terms of computation...
Table 2.4: Properties of the simulations suites are shown above; including the reference names, cosmological parameters, realisations \( N_{\text{sims}} \) and information on their eventual uses.

<table>
<thead>
<tr>
<th>Name</th>
<th>(10^9A_s)</th>
<th>(\Omega_m)</th>
<th>(\sum m_\nu) [eV]</th>
<th>(N_{\text{sims}})</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grid</td>
<td>[1, 3.5]</td>
<td>[0.2, 0.5]</td>
<td>[0, 0.6]</td>
<td>5</td>
<td>Simulations carried out at 216 points defined across a 6 ( \times ) 6 ( \times ) 6 grid in parameter space.</td>
</tr>
<tr>
<td>Fiducial</td>
<td>2</td>
<td>0.3</td>
<td>0</td>
<td>500</td>
<td>Used to calculate covariance matrices.</td>
</tr>
<tr>
<td>Mock</td>
<td>2.13</td>
<td>0.3175</td>
<td>0.06</td>
<td>10</td>
<td>Treated as real data.</td>
</tr>
</tbody>
</table>

time), albeit by sacrificing accuracy at small scales. All simulations are run in boxes of lengths 250 \( h^{-1}\)Mpc, with \( 256^3 \) dark matter particles and a discrete Fourier transform (DFT) density grid of \( (3 \times 256)^3 \). The latter is set to satisfy a requirement to produce accurate haloes from COLA simulations (Izard et al., 2016). The dependence on \( A_s, \Omega_m, \) and \( \sum m_\nu \) are explored, while \( h = 0.6711, \Omega_b = 0.049, \) and \( n_s = 0.9624 \) are constant in all simulations. Haloes and particles are outputted at redshift \( z = 0.5 \), using 20 steps from an initial redshift \( z = 10 \). Further details on the simulation suites are summarised in Table 2.4.

The reliability of these simulations is evaluated by comparing the power spectrum, calculated on the dark matter particles from the fiducial suite, to the non-linear power spectrum calculated from CAMB. We plot the 1\( \sigma \) difference variation in the power spectrum in Figure 2.10. Although this test shows the simulations can be trusted up to \( k < 0.7 \) \( h\)Mpc\(^{-1} \), we apply a conservative scale cut of \( k_{\text{max}} < 0.5 \) \( h\)Mpc\(^{-1} \) in Fourier space and \( l_{\text{min}} > 4\pi \) \( h^{-1}\)Mpc in real space.

2.7.2 Measurements

We use haloes from MG-PICOLA as a proxy for galaxies. These are found using the friends-of-friends halo finder MATCHMAKER\(^{14} \) which was found to be consistent (for the heaviest haloes) to the phase space halo finder ROCKSTAR (Behroozi et al., 2013). Unlike \( P(k) \) and \( B(k_1, k_2, k_3) \) which are unaffected by the density of tracers, the MST will exhibit different profiles purely based on the different halo counts. Since different number of haloes are produced from simulations with different cosmologies we mitigate this issue by performing our measurements on only the heaviest 5000 haloes. In practice such a restriction would not be imposed on \( P(k) \) or \( B(k_1, k_2, k_3) \) measurements, but here we wish to simply establish whether the MST improves on the constraints of \( P(k) + B(k_1, k_2, k_3) \).

We will explore replicating realistic survey properties in later work but in practice if we were simulating a galaxy catalogue, we would have to use a halo occupation distribution (HOD) model where we would tune the parameters of the HOD to have the same galaxy density as the actual survey. What we do here is a simplified version of that. The simulations constructed used haloes with masses between \( 10^{12} \) and \( 10^{15} \)\( \odot \). The number density (≈

\(^{14}\)https://github.com/damonge/MatchMaker
Figure 2.10: In the top panel we compare the mean (blue) and 1σ distributions (blue envelopes) of the power spectra calculated on dark matter particles from our fiducial suite of simulations to the linear and non-linear CAMB power spectra. In the bottom panels we show the difference between the measured and non-linear CAMB power spectra. The power spectra from MG-PICOLA appears to be accurately reproduced up to about $k = 0.7$, but we conservatively apply a scale cut of $k < k_{\text{max}}$ where $k_{\text{max}} = 0.5$.

Figure 2.11: In the top panel we compare the mean (blue) and 1σ and 2σ distributions (blue envelopes) against theoretical bispectra calculated using the linear and non-linear CAMB power spectra. The $x$-axis displays triangle index (generated by listing triangles in lexicographic order based on sides $k_1, k_2$ and $k_3$ where all elements are below $k_{\text{max}}$). In the bottom panel we show the significance between the measured and theoretical values. The theoretical bispectrum measurements are made using Gualdi et al. (2018) and will only be accurate up to the quasi-linear regime; since we are pushing to more non-linear scales the discrepancy for smaller triangles is expected. Using the non-linear $P(k)$ for the bispectrum is an approximation that only helps in partially reducing the discrepancy between the tree-level model and the measurements by using loop corrections for the power spectrum. A better model would be given by using one-loop corrections to the bispectrum.
3.2 \times 10^{-4} \ h^{-3}\text{Mpc}^3) \text{ is similar to the BOSS LOWZ sample between redshift 0.3 – 0.4 and to the CMASS sample between redshift 0.5 – 0.6 (see Figure 1 of Tojeiro et al., 2014). Assuming a linear bias of } b^2 = \frac{P_{\text{haloes}}(k)}{P(k)} \text{ we found the fiducial simulations to have a bias of } b \sim 1.3; \text{ this is more similar to the bias observed in eBOSS for emission line galaxies } (b \sim 1.4) \text{ than in BOSS for luminous red galaxies } (b \sim 2).

2.7.2.1 Power Spectrum and Bispectrum

Power spectrum and bispectrum measurements are performed through DFT algorithms as implemented by fftw3.\textsuperscript{15} We use the cloud-in-cell (CIC) mass assignment scheme using 64\textsuperscript{3} Cartesian grid cells to define a discrete overdensity field in configuration space, later transformed into Fourier space. The size of the simulation box is \( L_{\text{box}} = 250 \ h^{-1}\text{Mpc} \) and therefore, the mass resolution of the discrete over-density field is \( \sim 3.9 \ h^{-1}\text{Mpc} \). We compute the power spectrum between the fundamental frequency, \( k_f = 2\pi/L_{\text{box}} \), and a maximum frequency, \( k_{\text{max}} = 0.5 \ h\text{Mpc}^{-1} \), in bins of width \( k_f \).

The power spectrum and bispectrum measurements are performed using the code and estimator described in Gil-Marín et al. (2017). For the bispectrum we initially perform the measurements in bin sizes of \( k_f \). In this case we ensure that the three \( k \)-vectors of the bispectrum form closed triangles, and without loss of generality we define \( k_1 \leq k_2 \leq k_3 \). We include all the closed triangles with \( k_3 < k_{\text{max}} \). The bispectrum data vector, \( B(k_1, k_2, k_3) \), contains around 700 elements. In Figure 2.11, the bispectra measured on dark matter particles from the fiducial simulations are compared to theoretical values, showing good agreement until we reach non-linear regimes where the theory can no longer be trusted.

Using measurements of the power spectrum and bispectrum on the haloes of the fiducial suite, we were able to determine the skewness and kurtosis of the individual elements of the data vector. We found that elements with \( k < 0.125 \ h\text{Mpc}^{-1} \) contained much higher than expected skewness and kurtosis (i.e. exceeded the expected skewness and excess kurtosis of a Gaussian data set by 2\( \sigma \)) and as such we limit the power spectrum and bispectrum measurements to \( k > 0.125 \ h\text{Mpc}^{-1} \). This reduced the bispectrum data vector from \( \sim 700 \) to \( \sim 500 \). We then use a maximal compression technique (based on the work of Tegmark et al. 1997 and Heavens et al. 2017) to compress the bispectrum data vector to three elements (following Gualdi et al., 2018, 2019). Such a compression allows us to estimate the covariance matrix for a number of triangle configurations much larger than the number of available simulations.

\textsuperscript{15}Fastest Fourier Transform in the West, http://www.fftw.org
Figure 2.12: Posterior distributions on cosmological parameters as constrained by the individual components of the MST. On the left, we show those from the degree and branch shape and on the right from edge and branch lengths. Branch shapes are the least sensitive, whilst the degree gives broad constraints but rules out parts of the parameter space. Edge and branch length show similar posterior distributions with tighter constraints coming from edges.

2.7.2.2 Minimum Spanning Tree

The MST measurements are made with a scale cut of $l_{\text{min}} > 4\pi h^{-1}\text{Mpc}$, which corresponds to the wavelength ($\lambda = 2\pi / k$) of the largest $k$ modes ($k_{\text{max}}$) probed by $P(k)$ and $B(k_1, k_2, k_3)$. The MST statistics are then binned, which presents a problem as counts are discrete. For large counts, the distribution can be approximated by a Gaussian and as such we only select bins which we found the mean of our fiducial data vectors to have counts of greater than 50.

2.7.3 Parameter Estimation

Using the noisy estimates of the theory $d_{\text{Grid}}$ (the mean of five grid simulations at each point in parameter space) we can interpolate using GPs (see Appendix A.1) from a $6 \times 6 \times 6$ to a $20 \times 20 \times 20$ grid with theoretical data vectors $\mu_{\text{GP}}$ and uncertainty $\sigma_{\text{GP}}$ which is used instead of an MCMC due to the low dimensionality of the parameters. The sample covariance matrix, $S$, is estimated from 400 fiducial simulations (the other 100 fiducial simulations are used to apply a coverage correction, Sellentin & Starck 2019). The posterior for each of our ten mocks is evaluated using the following likelihood function,

$$L(d | \theta) \propto \det(C)^{-1/2} \left(1 + \frac{[d - \mu_{\text{GP}}(\theta)]^T \cdot C^{-1} \cdot [d - \mu_{\text{GP}}(\theta)]}{N - 1}\right)^{-\frac{N}{2}}. \quad (2.7)$$

This assumes the underlying data vector and statistics are Gaussian but accounts for noise in both the data vector and the sample covariance matrix with a multivariate T-distribution (see Sellentin & Heavens 2016; Jeffrey & Abdalla 2019). The data vector $d$ contains either
Figure 2.13: The posterior distributions are shown for power spectrum \( P(k) \), shown in grey), bispectrum \( B(k_1, k_2, k_3) \), shown in blue) and MST (shown in red). The tightest constraints on \( A_s \) and \( \Omega_m \) are given by the MST whilst \( B(k_1, k_2, k_3) \) provides better constraints on \( \sum m_\nu \).

one (or combinations) of the following: power spectrum, bispectrum and/or the binned PDFs of the MST statistics \( d, l, b \) and \( s \). The uncertainty in the GP regression is added to the sample covariance, i.e. \( C = S + S_{GP} \), where elements of \( (S_{GP})_{ij} = \sigma_{GP,i} \delta_k(v_i, v_j) \), \( \delta_k \) is the Kronecker delta function and \( v_i = v_j \) if the same GP hyperparameters were used to construct these elements of the data vector (following Bird et al., 2019; Rogers et al., 2019, which assume maximal dependency between elements of the data vector constructed from the same GP hyperparameters).

Finally we apply a coverage correction (Sellentin & Starck, 2019) using 100 fiducial simulations not included in the calculation of the covariance matrix. This accounts for unrecognized sources of biases. We found that all methods exhibited overconfident confidence contours. For \( P(k) \) and \( B(k_1, k_2, k_3) \) this is believed to have arisen due to non-Gaussian features in the data set. Although we attempted to limit this by selecting regions of the data vector that had fairly low skewness and kurtosis, we found that the skewness for \( P(k) \) tended to be consistently positive, whilst the excess kurtosis for the maximally compressed \( B(k_1, k_2, k_3) \) was always > 1\( \sigma \) than expected if the data were Gaussian. For the MST, this effect is larger which we suspect occurs due to two reasons: (1) similar to \( P(k) \) and \( B(k_1, k_2, k_3) \) the data vector is non-Gaussian and (2) the scale cut adds an extra stochasticity to the data vector that is not fully captured by the covariance matrix.
USING THE MST AS A TOOL FOR COSMOLOGY

Figure 2.14: The posterior distributions for cosmological parameters as constrained by (a) power spectrum \( P(k) \) (shown in dark grey) (b) power spectrum and bispectrum \( P(k) + B(k_1, k_2, k_3) \), shown in blue) and (c) power spectrum, bispectrum, and MST \( P(k) + B(k_1, k_2, k_3) + \text{MST} \), shown in purple).

\[ P(k) \]
\[ P(k) + B(k_1, k_2, k_3) \]
\[ P(k) + B(k_1, k_2, k_3) + \text{MST} \]

Figure 2.15: The 1σ constraints on \( A_s \) and \( \Omega_m \) are shown for \( P(k) \) (dark gray), \( P(k) + B(k_1, k_2, k_3) \) (blue) and \( P(k) + B(k_1, k_2, k_3) + \text{MST} \) (purple). This plot shows how including the MST improves constraints on \( A_s \) by \( \sim 12\% \) (\( \sim 10\% \)) and on \( \Omega_m \) by \( \sim 17\% \) (\( \sim 12\% \)) with respect to \( P(k) \) \( P(k) + B(k_1, k_2, k_3) \).

2.7.4 Comparison

The posterior distributions are measured for the three statistics and their combinations. Correlations between each statistic are accounted for by using a covariance matrix that is not block diagonal. In Figures 2.12, 2.13 and 2.14 we show the posterior distributions measured on the mean of the data vectors from 10 mocks allowing for better visual comparison of the errors whilst improvement in parameter constraints are stated according to the average improvement when measured on the mocks independently.

2.7.4.1 Components of the minimum spanning tree

We compare the constraints from the four individual components of the MST. The elements of the MST statistics are counts, and as such they follow a Poisson distribution. We
apply a cut on the data vector based on where the mean of the fiducial MST statistics had counts > 50, where we expect the Poisson distribution to be approximately characterized by a Gaussian. In Figure 2.12, we display the constraints from the individual components of the MST. Of the four statistics $s$ is the least constraining and provides very little information; this is followed by $d$ which, although it has very broad posteriors, appears at least to rule out parts of the parameter space (low $A_s$, $\Omega_m$ and high $\sum m_\nu$). The MST statistics $l$ and $b$ provide constraints having similar degeneracies with $l$ providing somewhat tighter constraints.

2.7.4.2 $P(k)$, $B(k_1, k_2, k_3)$, and MST

In Figure 2.13 we compare the constraints from $P(k)$, $B(k_1, k_2, k_3)$ and MST. All three appear to have similar degeneracies and as such are unable to establish meaningful constraints on $A_s$ and $\sum m_\nu$. The constraints on $\Omega_m$ are more conclusive but are fairly similar. The constraints on $\sum m_\nu$ tend to show a broad peak towards the centre of the prior range. Since the constraints on neutrino mass are poor the kernel-length scale for $\sum m_\nu$ of the GPs is quite broad and as such the estimates of the theory vector are smoother in the centre. This creates a slight bias towards the centre of the parameter space. This effect is also seen in Figure 2.14.

2.7.4.3 Combining $P(k)$, $B(k_1, k_2, k_3)$ and MST

In Figure 2.14, we combine the statistics and compare their relative constraints which is more clearly shown in Figure 2.15. In combining $P(k)$ and $B(k_1, k_2, k_3)$, we find an improvement of $\sim 6\%$ in the constraints of $\Omega_m$ and $\sim 3\%$ for $A_s$. When combined with the MST the constraints on $\Omega_m$ improve by $\sim 17\%$ and on $A_s$ improve by $\sim 12\%$ with respect to $P(k)$ ($\sim 12\%$ for $\Omega_m$ and $\sim 10\%$ for $A_s$ with respect to $P(k) + B(k_1, k_2, k_3)$). Since we have ensured the same scale cuts, i.e. $k_{\text{max}} = 0.5 \, h\text{Mpc}^{-1}$ for $P(k)$ and $B(k_1, k_2, k_3)$ and $l_{\text{min}} = 4\pi \, h^{-1}\text{Mpc}$, we can be fairly certain that the additional information is not coming from the MST having access to smaller scales. Furthermore, the maximally compressed $B(k_1, k_2, k_3)$ has been shown by Gualdi et al. (2018) to improve parameter constraints by allowing the inclusion of many more triangle configurations than standard bispectrum analysis. Therefore, we can be fairly certain that the additional information is coming from the MST’s detection of patterns in the cosmic web, information which would be present in higher order functions such as the trispectrum, thus confirming the heuristic arguments made in Section 2.4.1.
2.8 DISCUSSION

In this Chapter, we have sought to understand whether the MST can be used for parameter inference in cosmology. Until now, the MST has been predominantly used to search for large-scale features. This type of information has largely been overlooked as traditionally two-point statistics are completely insensitive to phase information whereas two-point statistics only measure the Fourier mode amplitudes. In constructing the MST we hope to pick up patterns in the cosmic web and use this to improve parameter constraints.

In Section 2.4, we argue heuristically why the MST should be sensitive to higher order statistics (i.e. three-point and beyond). This is demonstrated using simulated galaxies (from the Illustris N-body simulation) and a random walk simulation (produced using an adjusted Lévy Flight algorithm) with virtually identical 2PCF by design but different higher order statistics.

In Section 2.5, we look at the effects of boundaries and masks, RSD and scale cuts. Boundaries and masks\(^{16}\) tended to produce longer edge lengths, whilst the degree and branch shape appeared to be unaffected. RSD is shown to have a significant impact on the MST statistics and thus should be incorporated in any future study. Lastly, we develop a strategy to impose a scale cut on the MST. This is done by removing edges below a set length in the \(k\)NN graph and then constructing the MST from this. Unfortunately this creates some artefacts in the degree and branch shape distributions. It is also believed that this method distorts some of the information we are trying to learn. As such alternatives or improvements to this method should be explored.

In Section 2.6, we look to determine what the MST actually measures, finding the MST to be highly sensitive to its local density. This is demonstrated by the fact that nodes in overdensities tended to have a degree of 2.

Lastly in Section 2.7, we determine whether the MST provides information not present in power spectrum and bispectrum. We do this by obtaining parameter constraints on \(A_s\), \(\Omega_m\), and \(\sum m_\nu\) for 10 halo mock catalogues. To keep the density of haloes the same in all our simulations we use only the most massive 5000 haloes and measure the power spectrum \(P(k)\), bispectrum \(B(k_1, k_2, k_3)\) and MST statistics. The individual methods provided similar constraints although due to the degeneracies with \(\Omega_m\) we were unable to obtain meaningful constraints on \(\sum m_\nu\). We found that combining the three methods narrows the 1\(\sigma\) constraints on \(\Omega_m\) by \(\sim 17\%\) and on \(A_s\) by \(\sim 12\%\) with respect to \(P(k)\) and \(\sim 12\%\) on \(\Omega_m\) and \(\sim 10\%\) on \(A_s\) with respect to \(P(k) + B(k_1, k_2, k_3)\), thus showing that the MST is providing information not present in the power spectrum or bispectrum. We expect this to improve with improved implementation of scale cuts and greater statistical power from larger samples.

\(^{16}\)Boundaries can be thought of as a survey’s footprint, whilst the mask would also include holes and varying completeness levels.
The MST provides several advantages over existing methods but has some important limitations. The main advantages are: (1) it is sensitive to patterns in the cosmic web and (2) the algorithm is computationally inexpensive. The naive brute force implementation of $N$-point statistics for $n$ points is an $O(n^N)$ process. While there exist faster implementations of the 2PCF and 3PCF (see Scoccimarro, 2015; Slepian & Eisenstein, 2016) there are no such methods for higher order statistics. On the other hand, the MST is sensitive to higher order statistics and the Kruskal algorithm used here is approximately an $O(n \log n)$ process. In the MST, we have a window into these higher order statistics but at a fraction of the computational cost. The main limitations of the MST: (1) we need simulations to estimate the statistics and (2) the statistic is dependent on the density of the tracer. This means we will need to create simulations that both match the survey properties as well as the density of the tracers used.

In future work we look to apply the MST to current and future galaxy redshift surveys. In doing so we hope to better understand how to implement scale cuts and mitigate any of the resulting effects that occur as a result. One thing we have not studied in this Chapter is the effect of galaxy bias which should be explored in future. This could be achieved by varying HOD parameters. Lastly, ML algorithms and AI are powerful new tools to cosmology (see Ravanbakhsh et al., 2017; Fluri et al., 2018), however it is difficult to gain an intuition into what these algorithms are learning. Since the MST is relatively simple this could be used to gain insight into this work, providing a bridge between the traditional two-point and a full ML/AI approach.

Finally, the MST statistics presented in this Chapter have been produced by the PYTHON module MiSTree (Naidoo, 2019), which implements the procedures detailed in Section 2.3. The module is made publicly available (see https://github.com/knaidoo/mistree for documentation) and can handle data sets provided in 2D and 3D Cartesian coordinates, spherical polar coordinates and coordinates on a sphere (either celestial RA, Dec. or simply longitude and latitude).
“It’s the questions we can’t answer that teach us the most. They teach us how to think. If you give a man an answer, all he gains is a little fact. But give him a question and he’ll look for his own answers.”

– Patrick Rothfuss, The Wise Man’s Fear (2011)
3.1 ABSTRACT

The information content of the minimum spanning tree (MST) statistics are measured for a $\nu\Lambda\text{CDM}$ model using the Quijote simulations. The power spectrum (monopole and quadrupole in redshift space) and MST are measured in real and redshift space for haloes of mass $\geq 10^{13.5} \, h^{-1}\text{M}_\odot$ in simulations of length $L_{\text{box}} = 1 \, h^{-1}\text{Gpc}$. The power spectrum is measured for Fourier modes between the range $k_{\min} < k < k_{\max}$, where $k_{\min} = k_F$ the fundamental frequency and $k_{\max} = 0.5 \, h\text{Mpc}^{-1}$. The MST is measured with a minimum length scale of $l_{\min} = 4\pi \simeq 13 \, h^{-1}\text{Mpc}$, which correspond to the maximum Fourier mode $k_{\max}$. Like previous results we show that most of the information from the MST statistics is contained in the distribution of edge and branch lengths. In redshift space we find the MST provides $\sim 50\%$ tighter constraints on the sum of neutrino masses $M_\nu$ ($\Delta M_\nu = 0.16 \, \text{eV at} \, z = 0$) than the power spectrum ($\Delta M_\nu = 0.34 \, \text{eV at} \, z = 0$) and dominates the combined constraints ($\Delta M_\nu = 0.14 \, \text{eV at} \, z = 0$). This is due to the MST’s sensitivity to small scale clustering where the effect of neutrino free-streaming becomes important. On the other hand, the power spectrum provides much tighter constraints on $\Omega_m$ and $\sigma_8$. For the other parameters we find combining the power spectrum and MST improves constraints on the Hubble constant $h$, spectral tilt $n_s$ and baryon energy density $\Omega_b$ by a factor $\sim 2$. The MST’s sensitivity to neutrino mass and its different degeneracies mean it can be a powerful tool for constraining cosmological parameters, determining $M_\nu$ and testing $\Lambda\text{CDM}$ from future galaxy redshift surveys (such as DES, DESI, Euclid and LSST).

3.2 INTRODUCTION

The minimum spanning tree (MST) was first introduced to astronomy by Barrow et al. (1985) and has successfully been used as a filament finder for cosmic web studies (Bhavsar & Ling, 1988; van de Weygaert et al., 1992; Bhavsar & Splinter, 1996; Krzewina & Saslaw, 1996; Ueda & Itoh, 1997; Coles et al., 1998; Adami & Mazure, 1999; Colberg, 2007; Alpaslan et al., 2014; Beuret et al., 2017; Libeskind et al., 2018). The MST is the minimum weighted graph
COSMOLOGY AND NEUTRINOS WITH THE MST

that connects a set of points without forming loops. More recently Naidoo et al. (2020) (Chapter 2) investigated how the MST could be used to incorporate the cosmic web for constraining cosmological parameter. However, unlike conventional two-point analysis, performed by most galaxy redshift surveys, the MST cannot be calculated analytically and instead needs to be calculated from simulations. Fortunately, this problem is not unique to the MST – conventional statistics such as the power spectrum and bispectrum cannot be analytically described in the non-linear regime (for Fourier modes $k \gtrsim 0.3$) and artificial intelligence (AI) and machine learning (ML) algorithms (as well as other algorithms used to measure non-linear features in the cosmic web) will all require simulations. This has created a growing demand for large suites of cosmological simulations and the development of accurate emulators as cosmologists push to extract more information from the distribution of galaxies.

The Quijote simulations (Villaescusa-Navarro et al., 2019) were designed precisely for this use (i.e. to test new summary statistics like the MST and AI/ML and to push conventional statistics to smaller scales). In this Chapter we will use these simulations to measure the information content of the MST. The simulations have previously been used to conduct Fisher matrix analysis for the power spectrum (Villaescusa-Navarro et al., 2019), bispectrum (Hahn et al., 2019), 1D probability distribution function (Uhlemann et al., 2019) and marked power spectrum (Massara et al., 2020). In this Chapter we extend this analysis to the MST; in Naidoo et al. (2020) (Chapter 2) this was tested against measurements of the power spectrum and bispectrum for a few parameters (matter density $\Omega_m$, amplitude of scalar fluctuations $A_s$ and neutrino mass $M_\nu$) to test whether the MST adds new information. Fisher matrix analysis are useful to cosmologists as it allows lower bounds to be placed on the constraints of cosmological parameters from a given statistic. If the posterior distribution can be defined by a multivariate Gaussian then the Fisher matrix constraints are true however in the case that this is not true the constraints on any given parameter will be larger (this is known as the Cramer-Rao bound; Rao, 1945; Cramer, 1946). This analysis is explored for parameters of the $\nu$CDM model (i.e. the standard model of cosmology $\Lambda$CDM + massive neutrinos $M_\nu$). This will help determine the role the MST can play in constraining parameter from the current and next generation of galaxy surveys (such as the Baryon Oscillation Spectroscopic Survey, the Dark Energy Survey, the Dark Energy Spectroscopic Instrument, Euclid, the Nancy Grace Roman Space Telescope, the Prime Focus Spectrograph, the Rubin Observatory Legacy Survey of Space and Time, and the

---

1http://www.sdss3.org/surveys/boss.php
2http://www.darkenergysurvey.org
3http://desi.lbl.gov/
4http://www.euclid-ec.org/
5https://roman.gsfc.nasa.gov
6https://pfs.ipmu.jp/index.html
7https://www.lsst.org/
The Chapter is organised as follows. In section 3.3 we discuss the methodology and Quijote simulations used. In section 3.4 we present the constraints from components of the MST and demonstrate how including it with measurements of the power spectrum improve parameter constraints in a $\nu \Lambda$CDM model. Finally in section 3.5 we discuss the main results and their relevance for cosmology and future surveys.

3.3 METHOD

In this section we explain the Fisher matrix formalism used to measure the information content of several summary statistics and then explain how we measure the following in real and redshift space: power spectrum, power spectrum multipoles and the MST statistics. Lastly, we describe properties of the Quijote simulations used in this analysis.

3.3.1 Fisher Formalism

The Fisher matrix $\mathbf{F}$ is defined with elements

\[
F_{ij} = \sum_{\alpha,\beta} \frac{\partial S_\alpha}{\partial \theta_i} C^{-1}_{\alpha,\beta} \frac{\partial S_\beta}{\partial \theta_j},
\]

where $S_\alpha$ and $S_\beta$ are the elements $\alpha$ and $\beta$ of the data vector $\mathbf{S}$ and $C$ is the sample covariance matrix defined with elements

\[
C_{ij} = \langle (S_i - \langle S_i \rangle) (S_j - \langle S_j \rangle) \rangle,
\]

and the terms $\theta_i$ and $\theta_j$ are parameters $i$ and $j$ of the model. We multiply the inverse of the covariance matrix by the Kaufman-Hartlap factor (Kaufman, 1967; Hartlap et al., 2007) $(N_{\text{sim}} - 2 - N_S)/(N_{\text{sim}} - 1)$ where $N_S$ is the length of the data vector $\mathbf{S}$ and $N_{\text{sim}}$ is the number of simulations used to estimate the covariance matrix.

Usually the partial derivatives used to calculate the Fisher matrix are calculated using analytical expressions. However, for some summary statistics like the MST, the statistics have to be derived via simulations. The partial derivative as a function of parameter $\theta$ of the summary statistics (with data vector $\mathbf{S}$) is estimated using

\[
\frac{\partial \mathbf{S}}{\partial \theta} \approx \frac{\mathbf{S}(\theta + d\theta) - \mathbf{S}(\theta - d\theta)}{2d\theta},
\]

where $d\theta$ is a small deviation from a fiducial $\theta$. Most of the parameters considered in this Chapter will use this estimator for their partial derivative estimates however for neutrino mass another estimator is needed. Since we want the partial derivative when $M_\nu = 0$ eV.

\[\text{https://www.4most.eu/cms}/\]
we can no longer use the estimator above since this would require simulations where \( M_\nu < 0 \text{ eV} \). Instead we use the following estimators which are estimated solely from values of \( M_\nu \geq 0 \),

\[
\begin{align*}
\left[ \frac{\partial S}{\partial M_\nu} \right]_1 & \approx \frac{S(M_\nu) - S(M_\nu = 0)}{M_\nu}, \\
\left[ \frac{\partial S}{\partial M_\nu} \right]_2 & \approx \frac{-S(2M_\nu) + 4S(M_\nu) - 3S(M_\nu = 0)}{2M_\nu}, \\
\left[ \frac{\partial S}{\partial M_\nu} \right]_3 & \approx \frac{S(4M_\nu) - 12S(2M_\nu) + 32S(M_\nu) - 21S(M_\nu = 0)}{12M_\nu}.
\end{align*}
\] (3.4)

The estimators above are designed to make use of simulations calculated for \( M_\nu = 0.1, 0.2 \) or 0.4 eV. The first Equation (the non-symmetric two-point derivative; Equation 3.4) can be used for all the values of \( M_\nu \), the second Equation (the non-symmetric three-point derivative; Equation 3.5) can be used only when \( M_\nu = 0.1 \) or 0.2 eV and the third Equation (the non-symmetric four-point derivative; Equation 3.4) can be used only when \( M_\nu = 0.1 \) eV. The derivation of these estimators can be found in Appendix A.3.

The posterior distribution is then assumed to follow a multivariate Gaussian (Heavens, 2009, i.e. with Gaussian errors for each parameter) defined as

\[
L(\theta) = \frac{1}{(2\pi)^{M/2} \sqrt{\det F}} \exp \left( -\frac{1}{2} (\theta - \theta_{\text{Fid}})^\top \cdot F \cdot (\theta - \theta_{\text{Fid}}) \right),
\] (3.7)

where \( \theta \) and \( \theta_{\text{Fid}} \) are parameters from a \( \nu \Lambda \)CDM model of length \( M \), where the former are free parameters and latter is set to Fiducial parameter values.

### 3.3.2 Summary Statistics in Real and Redshift Space

In redshift space, redshift space distortions (RSD; Kaiser, 1987) caused by the peculiar velocities of galaxies alter the observed redshifts of galaxies. This shift in coordinates is defined as

\[
x_{\text{RSD}} = x + \frac{1 + z}{H(z)} (v \cdot e),
\] (3.8)

where \( x \) is the real space coordinate, \( x_{\text{RSD}} \) is the redshift space coordinate, \( v \) the peculiar velocity, \( z \) is the redshift and \( H(z) \) is the Hubble expansion rate at redshift \( z \). Finally, \( e \) is the basis vector defining the axis in the line-of-sight. In the analysis presented in this Chapter the line-of-sight is taken to be the \( z \)-axis, therefore the basis vector is defined as \( e = (0, 0, 1) \).
3.3.2.1 Power Spectrum Multipoles

The 1D power spectrum is measured in real space using Fast Fourier Transforms. In redshift space we instead compute the power spectrum monopole ($\ell = 0$) and quadrupole ($\ell = 2$) defined as

$$P_\ell(k) = (2\ell + 1) \int_0^1 P(k, \mu) L_\ell(\mu) d\mu,$$

where $L_\ell$ is the Legendre polynomial and $\mu$ is the cosine of the angle between the line-of-sight and tangential components of the Fourier mode vector $\mathbf{k} = (k_x, k_y, k_z)$,

$$\mu = \frac{|k_z|}{\sqrt{k_x^2 + k_y^2 + k_z^2}}.$$  

3.3.2.2 Minimum Spanning Tree Statistics

The MST is constructed in 3D comoving coordinates $\mathbf{x}$ in real space and $\mathbf{x}_{\text{RSD}}$ in redshift space. Before constructing the MST we first merge haloes which are separated by a minimum distance $4\pi \simeq 13 \, h^{-1}\text{Mpc}$ (this is equivalent to the maximum Fourier mode $k_{\text{max}} = 0.5 \, h\text{Mpc}^{-1}$ considered in the power spectrum measurements). This creates a catalogue of nodes, which is a collection of grouped and ungrouped haloes. The MST is then constructed from the node catalogue. Throughout this Chapter redshift space MST statistics are distinguished from real space measurements with a subscript $z$.

The probability distribution functions (PDF) of the MST statistics degree $d$, edge length $l$, branch length $b$ and branch shape $s$ (Naidoo et al., 2020) are measured. The edges are the lines of the MST graph, the degrees are the number of edges attached to each node and the branches are edges connected in continuous chains by nodes of degree $d = 2$. For the branches we measure their length (i.e. the sum of the lengths of member edges) and the shape (i.e. the straight line distance between branch ends divided by the branch length). Note that throughout this Chapter we will refer to the PDF of $\sqrt{1 - s}$ as $s$, meaning straighter branches are ones where $s \simeq 0$ while larger values indicate more curved branches. Furthermore, to remove non-Gaussian features in the tails of the PDF we only include regions of the data vector where the mean of the cumulative distribution function (CDF) for the Fiducial simulations lies in the range $0.05 < \text{CDF} < 0.95$ (with the exception of $d$ where we include the PDF in the range $1 \leq d \leq 4$). The publicly available PYTHON package MiSTree (see Appendix A.2; Naidoo, 2019) is used to construct and measure the statistic of the MST.
COSMOLOGY AND NEUTRINOS WITH THE MST

Table 3.1: We summarise the Quijote simulations used in this study and highlight the deviations from the Fiducial cosmological parameters. We use the standard simulations for each set of simulations with the exception of the massive neutrino simulations ($M_\nu^+, M_\nu^{++}$ and $M_\nu^{+++}$) where we use the paired fixed simulations.

<table>
<thead>
<tr>
<th>Name</th>
<th>Deviation from Fiducial</th>
<th>Realisations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fiducial</td>
<td>n/a</td>
<td>15000</td>
</tr>
<tr>
<td>$\Omega_m^+$</td>
<td>$\Delta \Omega_m = +0.01$</td>
<td>500</td>
</tr>
<tr>
<td>$\Omega_m^-$</td>
<td>$\Delta \Omega_m = -0.01$</td>
<td>500</td>
</tr>
<tr>
<td>$\Omega_b^+$</td>
<td>$\Delta \Omega_b = +0.002$</td>
<td>500</td>
</tr>
<tr>
<td>$\Omega_b^-$</td>
<td>$\Delta \Omega_b = -0.002$</td>
<td>500</td>
</tr>
<tr>
<td>$h^+$</td>
<td>$\Delta h = +0.02$</td>
<td>500</td>
</tr>
<tr>
<td>$h^-$</td>
<td>$\Delta h = -0.02$</td>
<td>500</td>
</tr>
<tr>
<td>$n_s^+$</td>
<td>$\Delta n_s = +0.02$</td>
<td>500</td>
</tr>
<tr>
<td>$n_s^-$</td>
<td>$\Delta n_s = -0.02$</td>
<td>500</td>
</tr>
<tr>
<td>$\sigma_8^+$</td>
<td>$\Delta \sigma_8 = +0.015$</td>
<td>500</td>
</tr>
<tr>
<td>$\sigma_8^-$</td>
<td>$\Delta \sigma_8 = -0.015$</td>
<td>500</td>
</tr>
<tr>
<td>$M_\nu^+$</td>
<td>$\Delta M_\nu = +0.1$</td>
<td>500</td>
</tr>
<tr>
<td>$M_\nu^{++}$</td>
<td>$\Delta M_\nu = +0.2$</td>
<td>500</td>
</tr>
<tr>
<td>$M_\nu^{+++}$</td>
<td>$\Delta M_\nu = +0.4$</td>
<td>500</td>
</tr>
</tbody>
</table>

3.3.3 Quijote Simulations

The Quijote simulations (Villaescusa-Navarro et al., 2019) are a large set of N-Body simulations designed to quantify the information content of cosmological observables and to train ML algorithms. The simulations are constructed in boxes of length $L_{\text{box}} = 1 \, h^{-1}\text{Gpc}$, using $512^3$ dark matter particles and $512^3$ neutrino particles (for simulations with massive neutrinos). A detailed Table of the parameters used for the Quijote simulations can be found in Table 1 of Villaescusa-Navarro et al. (2019). The simulations are based on a fiducial $\Lambda$CDM cosmology (based on Planck Collaboration et al. 2018) with fiducial parameters: matter density $\Omega_m = 0.3175$, baryon density $\Omega_b = 0.049$, Hubble constant $h = 0.6711$, $n_s = 0.9624$, amplitude of the linear power spectrum at $8 \, h^{-1}\text{Mpc}$ $\sigma_8 = 0.834$, the sum of neutrino masses $M_\nu = 0 \, \text{eV}$ and the dark energy equation of state $w = -1$. This analysis is performed on haloes with masses larger than $10^{13.5} \, h^{-1}\text{M}_\odot$. To determine the dependence on each parameter we use 500 simulations that deviate a single parameter from the fiducial values. To construct the covariance matrix we use 15,000 simulations constructed with the fiducial cosmology (the Fiducial simulations). This is summarised in Table 3.1 where we use the standard simulations with the exception of the massive neutrinos simulations ($M_\nu^+, M_\nu^{++}$ and $M_\nu^{+++}$) where we use the paired fixed simulations as these are produced with Zel'dovich initial conditions (instead of second-order Lagrangian
Figure 3.1: The correlation matrix for the MST in redshift space. The components of the MST statistics in redshift space: the degree $d_z$, edge length $l_z$, branch length $b_z$ and branch shape $s_z$; are labelled and divided by dotted black lines. The matrix is shown to have significant non-diagonal features. The most striking is the correlation between the edge length $l_z$ and branch length $b_z$ which show several fading lines of positive correlation. These originate from correlations between branches formed from two edges, three edges and so on, becoming fainter for branches formed from more edges. For the degree $d_z$ and $l_z$ we see virtually no correlation while there are correlations between $d_z$ and the two branch statistic: length $b_z$ and shape $s_z$. These correlations appear to be strongly tied to $d_z = 2$ which can be explained by the fact that branches are defined by having intermediate nodes with degree $d_z = 2$. Lastly we see a weak correlation between $b_z$ and $s_z$ and negligible correlations between other MST statistics (i.e between $l_z$ and $d_z$ and between $l_z$ and $s_z$).

perturbation theory) which better models the neutrino and dark matter initial conditions (see Villaescusa-Navarro et al., 2019, for further details). The MST partial derivatives measured on other parameters for both the normal and paired fixed simulations were found to be consistent, demonstrating this change in simulation type is only to improve the accuracy of the neutrino modelling and does not alter the results in any other way.

3.4 RESULT

In this Section the following results are discussed: (1) the covariance matrix for the MST statistics and the internal correlations and correlations with the power spectrum, (2) the partial derivatives for the power spectrum and MST statistics and (3) parameter constraints for a $\nu\Lambda$CDM model obtained from individual and combined measurements of the MST
COSMOLOGY AND NEUTRINOS WITH THE MST

Figure 3.2: The correlation matrix for the power spectrum multipoles and MST in redshift space. The power spectrum multipoles and components of the MST statistic are divided by dotted black lines. A black line is used to indicate the separation between the power spectrum multipoles and MST statistics. The correlation matrix for the power spectrum multipoles is shown to be sparse and mostly diagonal. Positive correlations between multipoles are only found for corresponding Fourier modes. The correlations with the MST is shown to be negligible with only a small inverse correlation between edges and the power spectrum monopole. This is consistent with what we expect since smaller edge lengths should correspond to larger Fourier modes.

and power spectrum.

3.4.1 Covariance Matrix

The covariance matrix is constructed from Equation 3.2 using data vectors measured from 15000 Fiducial simulations. In Figure 3.1 the correlation matrix for the MST is shown and in Figure 3.2 the correlation matrix for the combined data vector of the power spectrum multipoles and MST statistics is shown; in both cases they are constructed from measurements conducted in redshift space. Unlike the correlation matrix for the power spectrum, the correlation matrix for the MST contains several non-diagonal features. One of the most striking features is the correlation between the edge lengths $l_z$ and branch length $b_z$ which show ‘waves’ of positive correlations between short edges and short branches followed by negative correlations and then positive correlation for longer edges and branches. These positive correlations stem from the correlations between branches formed from 2 edges, 3 edges and so on. For branches formed from more edges these correlations become weaker.
CHAPTER 3

Figure 3.3: Derivatives for the power spectrum monopole are shown for the 6 $\Lambda$CDM parameters ($h$, $n_s$, $\Omega_b$, $\Omega_m$, $\sigma_8$ and $M_\nu$). Error bars are obtained using Equation 3.11 and Equation 3.13 for $M_\nu$. The sensitivity to each parameter can be assessed by the significance of deviations away from $\partial_0 P_0(k) = 0$.

As branches formed from more than 3 edges are more rare. Other correlations in the MST statistics appear to stem from the definition of branches which are defined by intermediate nodes with degree $d = 2$. As a result we see strong correlations between the degree and branches. The correlations between branch length and shape is weak but appears to be indicating that longer branches are more curved while short ones tend to be straighter. The correlations between the power spectrum multipoles and MST are weak. This is clearest for the monopole and edge lengths which show an inverse correlation between edges and Fourier modes; which is completely consistent with the relation between Fourier modes and real space. Furthermore the large scale modes of the monopole (the first half of the data vector) show positive correlations with longer edges. This indicates that most of the large scale clustering information is stored in the large edges of the MST.

3.4.2 Fisher Matrix and Partial Derivatives

The Fisher matrix is calculated from Equation 3.1 from the data vector $S$. In this Section the data vectors are either the power spectrum $P(k)$, power spectrum monopole and quadrupole $P_{\ell=0,2}(k)$ or combinations of the MST statistics: degree $d$, edge length $l$, branch length $b$ and branch shape $s$. To calculate this we require measurements of the derivatives $\partial_0 S$ which are estimated using simulations summarised in Table 3.1. For each set of simulations (consisting of 500 individual simulations and 15000 simulations for the Fiducial set) the mean and standard deviation of the summary statistics are obtained. The
Figure 3.4: Derivatives for the power spectrum quadrupole are shown for the 6 $\nu$CDM parameters ($h$, $n_s$, $\Omega_b$, $\Omega_m$, $\sigma_8$ and $M_\nu$). Error bars are obtained using Equation 3.11 and Equation 3.13 for $M_\nu$. The sensitivity to each parameter can be assessed by the significance of deviations away from $\partial_0 P_2(k) = 0$.

Figure 3.5: Derivatives for the MST degree $d_s$ are shown for the 6 $\nu$CDM parameters ($h$, $n_s$, $\Omega_b$, $\Omega_m$, $\sigma_8$ and $M_\nu$) in redshift space. Error bars are obtained using Equation 3.11 and Equation 3.13 for $M_\nu$. The sensitivity to each parameter can be assessed by the significance of deviations away from $\partial_0 d_s = 0$ (dotted black line).
Figure 3.6: Derivatives for the MST edge length $l_z$ are shown for the 6 $\nu$LCDM parameters ($h$, $n_s$, $\Omega_b$, $\Omega_m$, $\sigma_8$, and $M_\nu$) in redshift space. Error bars are obtained using Equation 3.11 and Equation 3.13 for $M_\nu$. The sensitivity to each parameter can be assessed by the significance of deviations away from $\partial l_z = 0$ (dotted black line).

Figure 3.7: Derivatives for the MST branch length $b_z$ are shown for the 6 $\nu$LCDM parameters ($h$, $n_s$, $\Omega_b$, $\Omega_m$, $\sigma_8$, and $M_\nu$) in redshift space. Error bars are obtained using Equation 3.11 and Equation 3.13 for $M_\nu$. The sensitivity to each parameter can be assessed by the significance of deviations away from $\partial b_z = 0$ (dotted black line).
derivatives for the parameters are then obtained using Equation 3.3 with the exception of neutrino mass where three estimators are used (Equations 3.4, 3.5 and 3.6). The uncertainty for the symmetric estimator (Equation 3.3) is given by,

$$\text{Var}(\partial_\theta \mathbf{S}) \simeq \frac{\Delta \mathbf{S}(\theta + d\theta)^2 + \Delta \mathbf{S}(\theta - d\theta)^2}{4d\theta^2}$$ \quad (3.11)$$

where $\Delta \mathbf{S}^2 = \text{Var}(\mathbf{S}) / N_{\text{Real}}$ and $N_{\text{Real}}$ is the number of simulation realisations. For the non-symmetric derivative estimator (Equations 3.4, 3.5 and 3.6) used for $M_\nu$ the uncertainties are defined by the following Equations respectively:

$$\text{Var}\left(\left[\partial_{M_\nu} \mathbf{S}\right]_1\right) \simeq \frac{1}{M^2_\nu}\left[\Delta \mathbf{S}(M_\nu)^2 + \Delta \mathbf{S}(M_\nu = 0)^2\right], \quad (3.12)$$

$$\text{Var}\left(\left[\partial_{M_\nu} \mathbf{S}\right]_2\right) \simeq \frac{1}{4M^2_\nu}\left[\Delta \mathbf{S}(2M_\nu)^2 + 16\Delta \mathbf{S}(M_\nu)^2 + 9\Delta \mathbf{S}(M_\nu = 0)^2\right], \quad (3.13)$$

$$\text{Var}\left(\left[\partial_{M_\nu} \mathbf{S}\right]_3\right) \simeq \frac{1}{144M^2_\nu}\left[\Delta \mathbf{S}(4M_\nu)^2 + 144\Delta \mathbf{S}(2M_\nu)^2 + 1024\Delta \mathbf{S}(M_\nu)^2 + 441\Delta \mathbf{S}(M_\nu = 0)^2\right]. \quad (3.14)$$

In Appendix A.3, derivations of the estimators for the partial derivatives (Equations 3.3, 3.4, 3.5, 3.6) are provided. The most accurate estimators are those for which the errors are given to higher orders of $d\theta$. For most of the parameters the derivative $\partial_\theta \mathbf{S}$ is

---

Figure 3.8: Derivatives for the MST branch shape $s_z$ are shown for the 6 $\nu$CDM parameters ($h$, $n_s$, $\Omega_b$, $\Omega_m$, $\sigma_8$ and $M_\nu$) in redshift space. Error bars are obtained using Equation 3.11 and Equation 3.13 for $M_\nu$. The sensitivity to each parameter can be assessed by the significance of deviations away from $\partial_\theta s_z = 0$ (dotted black line).
Table 3.2: Constraints on $M_{\nu}$ using different estimators for $\partial_{M_{\nu}} S$ at $z = 0$: estimator 1 (Equation 3.4) is shown using $M_{\nu} = 0.1, 0.2$ and $0.4$ eV, estimator 2 (Equation 3.5) is shown using $M_{\nu} = 0.1$ and $0.2$ eV and estimator 3 (Equation 3.6) is shown using $M_{\nu} = 0.1$ eV. Typically estimator 3 has been used in previous studies but in this paper we use estimator 2 with $M_{\nu} = 0.1$ eV since the accuracy of this estimator is consistent with Equation 3.4 used for the other parameters. With the exception of results derived from Equation 3.4, the least accurate estimator using $M_{\nu} = 0.2$ and $0.4$ eV, all the estimators in real and redshift space show tighter constraints for the MST than the power spectrum. Furthermore, when they are combined the constraints appear to be dominated by the MST.

<table>
<thead>
<tr>
<th>$M_{\nu}$</th>
<th>Est. 1 (Eq. 3.4)</th>
<th>Est. 2 (Eq. 3.5)</th>
<th>Est. 3 (Eq. 3.6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(k)$</td>
<td>0.1 eV 0.2 eV 0.4 eV</td>
<td>0.4 eV 0.2 eV 0.1 eV</td>
<td>0.1 eV 0.2 eV 0.39</td>
</tr>
<tr>
<td>MST</td>
<td>0.22 0.41 0.66</td>
<td>0.13 0.24 0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>$P(k) +$ MST</td>
<td>0.18 0.3 0.45</td>
<td>0.11 0.2 0.09</td>
<td>0.29</td>
</tr>
<tr>
<td>$P_{\ell=0,2}(k)$</td>
<td>0.45 0.49 0.71</td>
<td>0.34 0.35 0.12</td>
<td>0.12</td>
</tr>
<tr>
<td>MST$_x$</td>
<td>0.28 0.52 0.83</td>
<td>0.16 0.31 0.14</td>
<td>0.21</td>
</tr>
<tr>
<td>$P_{\ell=0,2}(k) +$ MST$_x$</td>
<td>0.23 0.33 0.5</td>
<td>0.14 0.21 0.11</td>
<td>0.11</td>
</tr>
</tbody>
</table>

determined by the symmetric derivative estimator Equation 3.3 which has errors of order $\mathcal{O}(d\theta^2)$, while the most accurate estimator for $\partial_{M_{\nu}} S$ is given by Equation 3.6 which has errors of order $\mathcal{O}(d\theta^3)$. However, in order for the estimators to be consistent for all parameters the correct estimator to use for $M_{\nu}$ is Equation 3.5 which has errors of order $\mathcal{O}(d\theta^2)$. Previous studies (Villaescusa-Navarro et al., 2019; Uhlemann et al., 2019; Hahn et al., 2019) have used Equation 3.6 (while Massara et al. 2020 uses Equation 3.5) consequently to facilitate comparisons to these studies, the results obtained using this estimator is provided in Table 3.3. However, throughout this Chapter we will refer to results obtained using Equation 3.5 with $M_{\nu} = 0.1$ eV.

The partial derivatives for the data vectors in redshift space (as a function of the $\nu$CDM parameters: $h$, $n_s$, $\Omega_b$, $\Omega_m$, $\sigma_8$ and $M_{\nu}$) are shown: for the power spectrum monopole $P_{\ell=0}(k)$ in Figure 3.3, the power spectrum quadrupole $P_{\ell=2}(k)$ in Figure 3.4, for the distribution of degree $d_z$ in Figure 3.5, for the distribution of edges $l_z$ in Figure 3.6, for the distribution of branches $b_z$ in Figure 3.7 and for the distribution of branch shapes $s_z$ in Figure 3.8. Depending on the estimators used the error bars are calculated using either Equation 3.11, 3.12, 3.13 or 3.14.

For large $k$ there are more Fourier modes to be measured in the simulations. Therefore the power spectrum will have smaller uncertainties at larger $k$. For all the derivatives we find that they tend towards 0 for small scales. This makes it difficult to determine the sensitivity to individual parameters from the derivatives alone. For both multipoles the constraints on $M_{\nu}$ are expected to be weak since for significant portions of the data vector they are indistinguishable from $\partial_{M_{\nu}} P_{\ell=0,2} \approx 0$.

In Figure 3.5 the derivatives show the degree $d_z$ has some small sensitivity to $h$, $n_s$
and $\Omega_\text{m}$ but is largely noise dominated. In Figures 3.6 and 3.7 the derivatives for the edge length $l_z$ and branch length $b_z$ show the statistics are sensitivity to all parameters and are not noise dominated. Lastly, in Figure 3.8 the derivatives for the branch shape $s_z$ show there is some sensitivity to $h$, $n_s$ and $\Omega_\text{m}$, while for the remaining parameters the derivatives are dominated by noise.

### 3.4.3 Sensitivity to Neutrinos and $\Lambda$CDM

In this Section the forecast constraints for parameters from a $\nu$CDM universe are obtained and discussed using Fisher matrices.

#### 3.4.3.1 Sensitivity to the Neutrino Partial Derivative Estimator

The partial derivative of the summary statistics as a function of neutrino mass can be estimated using Equations 3.4, 3.5 and 3.6. In Appendix A.3 these estimators are generally derived for the derivative $f'(x)$ of function $f(x)$ for simulations evaluated at $x = 0, \delta x$, $2\delta x$ and $4\delta x$. Estimator 1 (Equation 3.4) is shown to have errors of order $O(dx)$. The derivative can then be estimated by either taking $dx = \delta x$, $2\delta x$ or $4\delta x$ but since the error on the estimator is of order $O(dx)$ we immediately know that the most accurate will be the one using $\delta x$. Estimator 2 (Equation 3.5) is shown to have errors of order $O(dx^2)$ and therefore is more accurate than estimator 1 since $dx < 1$. The derivative can then be estimated using either $dx = \delta x$ or $2\delta x$ and we can immediately see that again using $\delta x$ provides the most accurate estimate (with $2dx$ having errors 4 times larger). Estimator 3 (Equation 3.6) is the most accurate estimator since it has errors of order $O(dx^3)$ and can only be used for $dx = \delta x$. However since the estimator used for the other parameters (Equation 3.3) has errors of order $O(dx^2)$ we will use estimator 2 with $\delta x$. For the neutrino mass simulations this is equivalent to using 3.5 with $M_\nu = 0.1$ eV.

In Table 3.2 the constraints on neutrino mass $\Delta M_\nu$ are compared using the different estimators for the power spectrum and MST in real and redshift space. The intrinsic accuracy of the estimators appear to show that less accurate estimators are associated with poorer constraints. The most accurate estimators, i.e. estimator 2 and 3 with $M_\nu = 0.1$ eV, are shown to be fairly consistent; with estimator 3 generally providing tighter constraints. In most cases the MST constraints on $M_\nu$ are stronger than ones obtained from the power spectrum, with the exception of those obtained using estimator 1 (Equation 3.4), i.e. the least accurate estimator for $M_\nu = 0.2$ or $0.4$ eV. Since the more accurate estimators obtained using Equation 3.5 and 3.6 find the MST to be the most constraining we can be fairly confident of this main result.
Separate and combined constraints for parameters from the $\nu\Lambda$CDM model determined from measurements of the power spectrum (multipoles in redshift space) and MST at redshift $z = 0, 0.5$ and 1. The constraints are obtained using Equation 3.3 for all of the parameter except $M_\nu$ which are obtained using Equation 3.5 and Equation 3.6. The former is used for the analysis presented in this Chapter and for comparison with Massara et al. 2020 while the latter is provided to allow comparisons with results obtained by Villaescusa-Navarro et al. 2019, Hahn et al. 2019 and Uhlemann et al. 2019. For the standard $\Lambda$CDM parameters we obtain competitive constraints from the MST at all redshifts, with the exception of $\Omega_m$ and $\sigma_8$ where the power spectrum dominates, however for $M_\nu$ the MST dominates. When measurements from the different redshifts are combined we find the MST is competitive for all parameters, including $\Omega_m$ and $\sigma_8$, but still dominates constraints on $M_\nu$.

<table>
<thead>
<tr>
<th>$z$</th>
<th>$\Delta h$</th>
<th>$\Delta n_s$</th>
<th>$\Delta \Omega_b$</th>
<th>$\Delta \Omega_m$</th>
<th>$\Delta \sigma_8$</th>
<th>$\Delta M_\nu$ [eV]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(k)$</td>
<td>0</td>
<td>0.19</td>
<td>0.2</td>
<td>0.015</td>
<td>0.024</td>
<td>0.036</td>
</tr>
<tr>
<td>MST</td>
<td>0</td>
<td>0.1</td>
<td>0.08</td>
<td>0.012</td>
<td>0.039</td>
<td>0.063</td>
</tr>
<tr>
<td>$P(k) +$ MST</td>
<td>0</td>
<td>0.07</td>
<td>0.05</td>
<td>0.007</td>
<td>0.007</td>
<td>0.012</td>
</tr>
<tr>
<td>$P_{\ell=0.2}(k)$</td>
<td>0</td>
<td>0.11</td>
<td>0.07</td>
<td>0.012</td>
<td>0.014</td>
<td>0.013</td>
</tr>
<tr>
<td>MST$_z$</td>
<td>0</td>
<td>0.09</td>
<td>0.07</td>
<td>0.01</td>
<td>0.041</td>
<td>0.062</td>
</tr>
<tr>
<td>$P_{\ell=0.2}(k) +$ MST$_z$</td>
<td>0</td>
<td>0.06</td>
<td>0.04</td>
<td>0.007</td>
<td>0.009</td>
<td>0.012</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$z$</th>
<th>$\Delta h$</th>
<th>$\Delta n_s$</th>
<th>$\Delta \Omega_b$</th>
<th>$\Delta \Omega_m$</th>
<th>$\Delta \sigma_8$</th>
<th>$\Delta M_\nu$ [eV]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(k)$</td>
<td>0.5</td>
<td>0.17</td>
<td>0.17</td>
<td>0.014</td>
<td>0.02</td>
<td>0.032</td>
</tr>
<tr>
<td>MST</td>
<td>0.5</td>
<td>0.1</td>
<td>0.07</td>
<td>0.01</td>
<td>0.03</td>
<td>0.047</td>
</tr>
<tr>
<td>$P(k) +$ MST</td>
<td>0.5</td>
<td>0.06</td>
<td>0.05</td>
<td>0.007</td>
<td>0.01</td>
<td>0.012</td>
</tr>
<tr>
<td>$P_{\ell=0.2}(k)$</td>
<td>0.5</td>
<td>0.12</td>
<td>0.07</td>
<td>0.012</td>
<td>0.013</td>
<td>0.01</td>
</tr>
<tr>
<td>MST$_z$</td>
<td>0.5</td>
<td>0.09</td>
<td>0.08</td>
<td>0.009</td>
<td>0.033</td>
<td>0.056</td>
</tr>
<tr>
<td>$P_{\ell=0.2}(k) +$ MST$_z$</td>
<td>0.5</td>
<td>0.06</td>
<td>0.05</td>
<td>0.006</td>
<td>0.011</td>
<td>0.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$z$</th>
<th>$\Delta h$</th>
<th>$\Delta n_s$</th>
<th>$\Delta \Omega_b$</th>
<th>$\Delta \Omega_m$</th>
<th>$\Delta \sigma_8$</th>
<th>$\Delta M_\nu$ [eV]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(k)$</td>
<td>1</td>
<td>0.18</td>
<td>0.19</td>
<td>0.015</td>
<td>0.026</td>
<td>0.03</td>
</tr>
<tr>
<td>MST</td>
<td>1</td>
<td>0.1</td>
<td>0.08</td>
<td>0.009</td>
<td>0.039</td>
<td>0.048</td>
</tr>
<tr>
<td>$P(k) +$ MST</td>
<td>1</td>
<td>0.07</td>
<td>0.05</td>
<td>0.007</td>
<td>0.009</td>
<td>0.008</td>
</tr>
<tr>
<td>$P_{\ell=0.2}(k)$</td>
<td>1</td>
<td>0.14</td>
<td>0.11</td>
<td>0.013</td>
<td>0.017</td>
<td>0.014</td>
</tr>
<tr>
<td>MST$_z$</td>
<td>1</td>
<td>0.1</td>
<td>0.07</td>
<td>0.009</td>
<td>0.033</td>
<td>0.044</td>
</tr>
<tr>
<td>$P_{\ell=0.2}(k) +$ MST$_z$</td>
<td>1</td>
<td>0.06</td>
<td>0.05</td>
<td>0.006</td>
<td>0.011</td>
<td>0.009</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$z$</th>
<th>$\Delta h$</th>
<th>$\Delta n_s$</th>
<th>$\Delta \Omega_b$</th>
<th>$\Delta \Omega_m$</th>
<th>$\Delta \sigma_8$</th>
<th>$\Delta M_\nu$ [eV]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(k)$</td>
<td>0, 0.5, 1</td>
<td>0.093</td>
<td>0.085</td>
<td>0.009</td>
<td>0.013</td>
<td>0.016</td>
</tr>
<tr>
<td>MST</td>
<td>0, 0.5, 1</td>
<td>0.053</td>
<td>0.039</td>
<td>0.006</td>
<td>0.011</td>
<td>0.01</td>
</tr>
<tr>
<td>$P(k) +$ MST</td>
<td>0, 0.5, 1</td>
<td>0.041</td>
<td>0.028</td>
<td>0.004</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>$P_{\ell=0.2}(k)$</td>
<td>0, 0.5, 1</td>
<td>0.058</td>
<td>0.042</td>
<td>0.007</td>
<td>0.009</td>
<td>0.007</td>
</tr>
<tr>
<td>MST$_z$</td>
<td>0, 0.5, 1</td>
<td>0.05</td>
<td>0.038</td>
<td>0.005</td>
<td>0.012</td>
<td>0.01</td>
</tr>
<tr>
<td>$P_{\ell=0.2}(k) +$ MST$_z$</td>
<td>0, 0.5, 1</td>
<td>0.035</td>
<td>0.027</td>
<td>0.004</td>
<td>0.006</td>
<td>0.005</td>
</tr>
</tbody>
</table>
Figure 3.9: Fisher matrix constraints for $\nu\Lambda$CDM parameters from the MST edge and branch length distributions in redshift space at $z = 0$. The constraints from the edge length distribution MST$_z(l)$ is shown with blue dashed line, from the branch length distribution MST$_z(b)$ with purple dotted lines and from combining the edge length and branch length distributions MST$_z(l, b)$ with black lines and contours.
Figure 3.10: Fisher matrix constraints for $\nu\Lambda$CDM parameters from the MST edge and branch length distributions in combination with the distribution of degree and branch shape in redshift space at $z = 0$. The constraints from the edge and branch length distributions $\text{MST}_z(l, b)$ is shown with black dashed lines; from the edge, branch length and degree distributions $\text{MST}_z(l, b, d)$ in green lines and from the edge, branch length, degree and branch shape distribution $\text{MST}_z(l, b, d, s)$ in red lines and contours. The contours show that adding the degree does very little to improve the constraints from the edge and branch length distribution while the branch shape tightens constraints slightly on most of the parameters and in particular $\Omega_m$. 

--- $\text{MST}_z(l, b)$ 
- $\text{MST}_z(l, b, d)$ 
- $\text{MST}_z(l, b, d, s)$
Figure 3.11: Fisher matrix constraints on $\nu$CDM parameters from the power spectrum and MST separately and combined in redshift space at $z = 0$. The constraints from the power spectrum multipoles $P_{\ell=0,2}(k)$ are shown with blue dashed lines, from the four MST statistics $\text{MST}_z(l, b, d, s)$ with red dotted lines and the combination of the power spectrum multipoles and MST $P_{\ell=0,2}(k) + \text{MST}_z(l, b, d, s)$ with purple lines and contours. Constraints for $\sigma_8$ and $\Omega_m$ are dominated by the power spectrum multipoles while $M_\nu$ is dominated by the MST. For $h$, $n_s$ and $\Omega_b$, the constraints are competitive and improve roughly by a factor of $\sim 2$ when combined in comparison to the power spectrum multipoles.
Figure 3.12: Fisher matrix constraints on $\nu$ΛCDM parameters from the power spectrum and MST separately and combined from redshift $z = 0, 0.5$ and 1 in redshift space. The constraints from the combined power spectrum multipoles $P_{\ell=0, 2}(k) [z = 0, 0.5, 1]$ is shown with blue dashed lines, from the combined MST MST$_{z}(l, b, d, s) [z = 0, 0.5, 1]$ is shown with dotted red lines and from the combined power spectrum multipoles and MST $P_{\ell=0, 2}(k) + MST_{z}(l, b, d, s) [z = 0, 0.5, 1]$ is shown with purple lines and contours. The constraints on all parameters are more competitive but similar trends as Figure 3.11 can be seen; the power spectrum multipoles still provide slightly better constraints on $\Omega_m$ and $\sigma_8$ (although the MST is now competitive) while the MST still dominates constraints on $M_\nu$. In combination the constraints on all parameters are greatly improved.
3.4.3.2 Constraints from the Minimum Spanning Tree

The constraints from different combinations of the MST statistics at $z = 0$ are shown in Figures 3.9 and 3.10. In Figure 3.9 the individual and combined constraints from the edge and branch length distribution is shown. The two statistics are shown to have competitive constraints between each other but in combination they significantly reduce their individual parameter degeneracies. This is best illustrate by looking at the posterior distributions for $n_s$ and $h$, for $\Omega_b$ and $n_s$ and for $\sigma_8$ and $\Omega_b$. Breaking the degeneracies in these parameters appear to be translating to far better combined constraints for the other parameters. In Figure 3.10 the edge length and branch length distributions are first combined with the degree and then the branch shape. This shows that the degree is adding very little information to the constraints while combining the branch shape appears to be improving the constraints moderately (in particular for $\Omega_m$). These results are consistent with the derivative estimates shown in Figures 3.5, 3.6, 3.7 and 3.8 where the derivatives for $l_z$ and $b_z$ are the clearest and least noise dominated. With the exception of a few parameters for $s_z$, the derivatives for $d_z$ and $s_z$ are found to be heavily noise dominated.

3.4.3.3 Combined Constraints from the Minimum Spanning Tree and Power Spectrum

The separate and combined constraints from the MST and power spectrum are shown in redshift space at $z = 0$ in Figure 3.11. For the constraints obtained from different redshifts please refer to Table 3.3. The individual constraints from the MST and power spectrum multipoles shown in Figure 3.11 are found to be competitive for $h$, $n_s$ and $\Omega_b$. For $\Omega_m$ and $\sigma_8$ the constraints are dominated by the power spectrum but for $M_\nu$ the constraints are dominated by the MST. When combined some degeneracies are broken and we find an improvement of roughly a factor of $\sim 2$ on the parameters $h$, $n_s$, $\Omega_b$ and $M_\nu$ in comparison to the power spectrum. This highlights the importance of including the MST in future galaxy redshift surveys to test and constrain $\Lambda$CDM parameters and to determine the sum of neutrino mass $M_\nu$ – a key scientific goal of many future surveys.

In Figure 3.12 the separate and combined constraints from the MST and power spectrum measured across three different redshifts ($z = 0, 0.5, 1$) are shown. This expands the total volume used to make this measurement from $1 \, h^{-3}\text{Gpc}^3$ for a single redshift to $3 \, h^{-3}\text{Gpc}^3$. Combining measurements of the MST from different redshifts reduces constraints on all parameters. For $\Omega_m$ and $\sigma_8$ the constraints from the MST alone are now competitive with constraints from the power spectrum. The same cannot be said for the power spectrum and $M_\nu$ where the MST still dominates.
3.5 DISCUSSION

In this Chapter, we have applied the Fisher matrix formalism using the Quijote simulations to calculate the information content of the MST statistics in a $\nu$CDM model. The Quijote simulations are a large suite of $N$-body simulations designed to test the information content of summary statistics and to test AI/ML algorithms. In this Chapter we use a subset of the full suite of simulations (summarised in Table 3.1) and measure the power spectrum (multipoles in redshift space) and MST statistics in real and redshift space on haloes with mass $\geq 10^{13.5} h^{-1} M_\odot$.

From the MST we measure the distribution of degree $d$, edge length $l$, branch length $b$ and branch shape $s$. In Naidoo et al. (2020) it was shown that $l$ was the most constraining MST statistics. However, this analysis was based on COLA simulations of length $L_{\text{box}} = 250 h^{-1} \text{Mpc}$ and limited to the most massive 5000 haloes. By using the Quijote simulations we are able to expand this analysis to a wider set of cosmological parameters ($h$, $n_s$, $\Omega_b$, $\Omega_m$, $\sigma_8$ and $M_\nu$ compared to only $A_s$, $\Omega_m$ and $M_\nu$), over a larger volume ($L_{\text{box}} = 1 h^{-1} \text{Gpc}$) and using larger halo catalogues (of the order of $10^5$). In Figures 3.9 and 3.10 the constraints from the MST are shown to be dominated by the distribution of edge and branch lengths. The degree is shown to add very little information while the branch shape provide moderate improvements on a few parameters, in particular $\Omega_m$.

In Figure 3.11 and Table 3.3 the MST and power spectrum constraints are compared and combined. The power spectrum (and multipoles in redshift space) provides much stronger constraints on $\Omega_m$ and $\sigma_8$. On the other hand the MST dominates the constraints on $M_\nu$: at $z = 0$ the power spectrum multipoles gives $\Delta M_\nu = 0.34 \text{eV}$ while the MST gives $\Delta M_\nu = 0.16 \text{eV}$, roughly a factor of $\sim 2$ improvement. For the other parameters we find that combining the two breaks several degeneracies which improves constraints on $h$, $n_s$ and $\Omega_b$ by roughly a factor of $\sim 2$ in comparison to the individual power spectrum constraints. When combining measurements from different redshifts ($z = 0$, 0.5 and 1) the constraints from the MST are competitive for all parameters. In particular the power spectrum no longer dominates constraints on $\Omega_m$ and $\sigma_8$ while the MST still dominates constraints on $M_\nu$. For the power spectrum multipoles the constraints on $M_\nu$ are $\Delta M_\nu = 0.155 \text{eV}$ while for the MST this is $\Delta M_\nu = 0.079 \text{eV}$.

In this Chapter we show that adding the MST to measurements of the power spectrum greatly improves constraints for parameters of the standard $\Lambda$CDM model and dominate the constraints on the sum of neutrino masses $M_\nu$. The MST is significantly more sensitive to the effects of neutrinos on the distribution of haloes. This sensitivity appears to come from the MST sensitivity to the small scale clustering in the cosmic web where neutrino free-streaming effects are important. The MST could prove pivotal in determining the sum of neutrino masses, a key and fundamental objective of future galaxy surveys. This demonstrates the importance of measuring more than just the power spectrum (or two-
point statistics) and provides a powerful argument for making measurements of the MST on current and future galaxy surveys such as DESI. In future studies we will investigate the relationship between the MST statistics and biasing. This would follow the analysis of Hahn et al. (2019) who use nuisance parameters (in their Fisher matrix analysis of the Quijote simulations) to incorporate a linear bias for their measurements of the bispectrum.
“The writing gets done away from the keyboard and away from the studio in my head, in solitude. And then I come in and hopefully have something, then I wrestle with sounds and picture all day long. But the ideas usually come from a more obscure place, like a conversation with a director, a still somebody shows you, or whatever.”

– Hans Zimmer
Chapter 4

Applying the Minimum Spanning Tree to Test $\Lambda$CDM with BOSS

4.1 Abstract

We compare the minimum spanning tree (MST) statistics from the real Baryon Oscillation Spectroscopic Survey (BOSS) sample to the $\Lambda$CDM PATCHY mocks. The MST statistics: the distribution of degrees $d$, edge lengths $l$, branch lengths $b$ and branch shapes $s$, are measured. The individual and combined MST statistics from the BOSS real galaxies are shown to be consistent with PATCHY mocks. We find one exception: the distribution of edge lengths $l$ for the LOWZ NGC region which is discrepant at the $\sim 2.5\sigma$ level. The source of this discrepancy will require further study but is suspected to arise from a redshift dependent offset in the density. In cosmology galaxy weights are assigned and used to propagate corrections to systematic errors to the calculation of summary statistics. However, how we should incorporate these weights into the MST has remained an open problem since the MST can only incorporate binary weights. In this Chapter, we devise a method to apply weights to the MST. We first describe an iterative stochastic integer (SI) weighting scheme which is shown to reproduce the non-integer weighted two-point correlation function for BOSS. Since the nodes for the MST either exist or do not, the MST can in practice only incorporate binary weights. We then construct a stochastic binary (SB) weighting scheme which is used on the MST. The SB weight is simply defined to be $w_{SB} = 1$ if the SI weight is $w_{SI} \geq 1$. This weighting scheme results in a bias since the sum of non-integer weights $W$ and the sum of SB weights $W_{SB}$ are related by $W \geq W_{SB}$. When comparing the real galaxy sample and the PATCHY mock samples, we find this bias to be different, which is accounted by sub-sampling the larger sample to ensure the sum of the weights are roughly equal. Since the MST statistics are probability distribution functions they will have Poisson noise near the tails of the distributions. We show that once the tails are removed, the MST data vectors can be fully approximated with Gaussian errors. Future parameter studies can therefore assume a Gaussian likelihood for the MST.
4.2 INTRODUCTION

The Baryon Oscillation Spectroscopic Survey (BOSS) advances the photometric Sloan Digital Sky Survey (SDSS) by carrying out a spectroscopic survey of over a million galaxies up to a redshift of \( z \lesssim 0.8 \). The main cosmological objectives for the survey were to measure the baryonic acoustic oscillation (BAO; Cole et al., 2005; Eisenstein et al., 2005) feature, to measure redshift space distortions (RSD; Kaiser, 1987) and to measure the two-point correlation function (2PCF) of BOSS galaxies (Alam et al., 2017). While these measurements remain consistent with \( \Lambda \)CDM other statistics may reveal discrepancies. As a result, the survey has also been used to measure higher order statistics such as the bispectrum (Gil-Marín et al., 2017) and three-point correlation function (Slepian et al., 2017), both of which have provided further tests of \( \Lambda \)CDM and provided substantial improvements on the constraints of cosmological parameters.

Recently Naidoo et al. (2020) (Chapter 2) and Naidoo et al. (in prep; Chapter 3) showed that the minimum spanning tree (MST) could be used to improve constraints on cosmological parameters and was particularly sensitive to the sum of neutrino masses. The MST is the minimum weighted graph that connects a set of points (in this case galaxies) in a single structure that has no loops. The graph was first introduced to astronomy by Barrow et al. (1985) and has been used for many years as a filament finder. It is this sensitivity to the cosmic web that allows the MST to incorporate the cosmic web into cosmological analysis. On the practical side, the calculation is extremely time efficient (the algorithm scales by \( O(n \log n) \) rather than \( O(n^N) \) for a naive brute force implementation of an \( N \)-point correlation function) meaning in practice the MST requires far less computing power than calculating even the second-order correlation function (i.e. 2PCF). The main drawback of the method is that MST statistics cannot be calculated analytically and instead require large suites of \( N \)-body simulations. Furthermore, unlike correlation functions, small scales are challenging to mask and the statistics are sensitive to the density of the point distribution.

In this Chapter we will explore whether the MST measured on real BOSS galaxies are consistent with the MST measured on mock galaxies (specifically the PATCHY mocks; Kitaura et al., 2016) and determine: (1) how to include galaxy weights into the construction of the MST, (2) how to mitigate non-linear and fiber collision effects on small scales and (3) whether the noise of the MST statistics are Gaussian and whether future parameter studies can use a Gaussian likelihood.

4.3 METHOD

Below we detail the methods used to calculate the 2PCF multipoles (monopole and quadrupole) and the MST statistics measured in this Chapter. The 2PCF is calculated to test the weighting scheme devised in this study for the MST.
4.3.1 Two-Point Multipoles

The 2PCF is estimated using the Landy & Szalay (1993) estimator (an unbiased estimator with fast convergence properties; Szapudi & Szalay 1998),

$$\xi(r, \mu) = \frac{N_D^2 DD(r, \mu)}{N_D^2 RR(r, \mu)} - \frac{2 N_R DR(r, \mu)}{N_D RR(r, \mu)} + 1,$$

where $N_D$ is the sum of the data point weights and $N_R$ is the sum of the random point weights. Random point catalogues are constructed to follow the survey’s footprint and redshift selection function. The random points are used to represent a homogeneous distribution. Distances between pairs of points are binned according to their separation $r$ and the cosine of the angle between the line vector joining the points $r$ and the line-of-sight vector $x$.

$$\mu = \frac{x \cdot r}{|x| |r|}.$$ (4.2)

A galaxy $i$ is given a weight $w_i$ which is used to propagate corrections to systematic errors from the calculation of summary statistics. Pairs of points (point $i$ and $j$) with separation $r$ and cosine of angle $\mu$ are counted, with weights equal to the product of the individual point weights (i.e. $w_i w_j$ by default assumed to be 1). Identical to the notation provided in Chapter 1, $DD(r, \mu)$ represents counts for data-to-data pairs, $DR(r, \mu)$ represents counts for data-to-random pairs and $RR(r, \mu)$ represents counts for random-to-random pairs. The monopole and quadrupole of the 2PCF can then be calculated using

$$\xi_\ell(r) = (2\ell + 1) \int_0^1 \xi(r, \mu) L_\ell(\mu) d\mu,$$ (4.3)

where $L_\ell$ is the Legendre polynomial. These are computed using TwoFast\(^1\), a brute force 2PCF estimator written in C++ and parallelised using OpenMPI.

4.3.2 Minimum Spanning Tree

The MST is constructed in 3D comoving distance using the public PYTHON package MiSTree (see Appendix A.2; Naidoo, 2019). Following Naidoo et al. (2020) we measure the probability distribution $P$ of the MST degree $d$, edge length $l$, branch length $b$ and branch shape $s$. Note throughout this Chapter we will plot and measure the distribution of $\sqrt{1-s}$ and will refer to this as $s$, meaning straighter branches are ones where $s \simeq 0$ while larger values indicate more curved branches. In Naidoo et al. (2020) we limited the effects of small scales by removing edges in the $k$-nearest neighbour graph below a certain length $l_{\text{min}}$. However, this resulted in strange features in the graph at small scales, so in this Chapter

\(^1\)https://github.com/knaidoo/twofast/
we will instead employ a similar method to Alpaslan et al. (2014) which was also applied in Chapter 3 (Naidoo et al. in prep). In Alpaslan et al. (2014) the MST is constructed on groups which in this Chapter are used to ‘mask’ small scales.

We define a group to be a collection of galaxies where each galaxy is separated with at least one other galaxy by a distance $r < r_{\text{min}}$ where $r_{\text{min}} = 10 \, h^{-1}\text{Mpc}$. This distance was chosen to limit the non-linear effects which become important on small scales and were not modelled in the construction of PATCHY (which were designed for BOSS’s core scientific mission to measure BAO on large scales). The position of a group is taken to be the mean of the positions of the member galaxies. A catalogue of nodes is then constructed which includes the positions of groups and the positions of ungrouped galaxies. The MST is then constructed from this catalogue of nodes.

### 4.3.3 Weights for Point Processes

It is common for cosmological data sets to assign galaxies weights to propagate corrections to systematic errors (such as fiber collisions, imaging depth, etc) into the measurement of the 2PCF and other summary statistics.

Galaxy weights can be and are typically non-integer values, however for the MST non-binary weights are difficult to incorporate. In order to construct a MST graph a point either exists, and is part of the MST, or it does not. Weights greater than one further complicate the matter, since this essentially means a galaxy represents more than one galaxy. However placing more than one point at a single place will introduce edges of length zero in the MST which is physically meaningless so galaxies with weights greater than one have to be assigned a weight of one.

In the following Section we introduce two weighting schemes which were devised for this study: stochastic integer weights which allows one to iteratively apply a non-integer weight to a point process and stochastic binary weights which are applied to the MST.

**Stochastic Integer Weights**

Stochastic integer (SI) weights $w_{\text{SI}}$ are assigned to a galaxy with non-integer weight $w$,

$$w_{\text{SI}} = \begin{cases} \lfloor w \rfloor, & \text{for } u < \lceil w \rceil - w, \\ \lceil w \rceil, & \text{otherwise}. \end{cases}$$

where $u$ is a random number between $[0, 1]$, $\lfloor x \rfloor = \text{floor}(x)$ is the floor function (rounds down to the nearest integer) and $\lceil x \rceil = \text{ceil}(x)$ is the ceiling function (rounds up to the nearest integer). A galaxy with a weight $w = 1.2$ will be assigned a SI weight $w_{\text{SI}} = 1$ at 80% of the time and $w_{\text{SI}} = 2$ at 20% of the time. The mean weight of the galaxy over many iterations will be 1.2.
Stochastic Binary Weights

While SI weights offer a means to introduce weights to the MST, as was previously discussed, the MST cannot handle weights $w > 1$. Since small scales are removed by merging points that are separated by a distance smaller or equal to $r_{\text{min}}$, points with $w > 1$ can be thought of as a collection of points separated by $r = 0$ and therefore, they are automatically merged into a single point. The result of applying SI weights to the MST is a stochastic binary (SB) weight,

$$w_{\text{SB}} = \begin{cases} 0, & \text{for } w_{\text{SI}} = 0, \\ 1, & \text{for } w_{\text{SI}} \geq 1. \end{cases}$$

(4.5)

The sum of the non-integer weights $W$ is not less than the sum of the SB weights $W_{\text{SB}}$.

$$W \geq W_{\text{SB}}.$$  

(4.6)

This introduces a bias. However, this bias is only relevant when making comparisons between two distributions and can be thought of as a reduction in the initial sample density. When comparing two data sets, as we will in this study, applying the SB weights to the real data and mocks leads to an offset in their relative total weights. If we were using non-integer weights this offset would not be present. To adjust this we apply a density correction where a random fraction $f$ of the larger sample is drawn to match the weights between samples. For example, take the sum of the SB weights for sample A to be $W_{\text{SB}}^A$ and for sample B to be $W_{\text{SB}}^B$. If $W_{\text{SB}}^B > W_{\text{SB}}^A$ we define,

$$f = \frac{W_{\text{SB}}^A}{W_{\text{SB}}^B},$$

(4.7)

where $f < 1$. To match the weights we sample galaxies from sample B by first assigning each galaxy a random number $u$ between $[0, 1]$ and then only keeping galaxies for which $u < f$. The sum of the SB weights for the sampled galaxies from sample B will now be $\approx W_{\text{SB}}^A$.

4.4 BOSS GALAXIES AND MOCKS

In this study we use the real BOSS galaxy sample and the BOSS PATCHY mocks\footnote{https://data.sdss.org/sas/dr12/boss/lss/dr12_multidark_patchy_mocks/} which are constructed using augmented Lagrangian perturbation theory (a combination of second-order Lagrangian perturbation theory on large scales and the spherical collapse model on small scales) and a stochastic biasing scheme which is calibrated with the large BigMultiDark N-body simulation (Kitaura et al., 2014). The BOSS data is divided into four regions
defined by two regions on the sky and two redshift ranges. The two regions on the sky are defined by their relation on the galactic plane: above the galactic plane we have the North Galactic Cap (NGC) and below we have the South Galactic Cap (SGC). The data is further divided into the two redshift regions: $0.2 < z \leq 0.43$ for LOWZ and $0.43 < z < 0.75$ for CMASS. Typically LOWZ and CMASS are divided at $z = 0.43$ including galaxies for LOWZ with redshifts $z < 0.2$, however the lower redshift limit $z_{\text{min}} = 0.2$ and upper redshift limit $z_{\text{max}} = 0.75$ are enforced by the limits of the PATCHY mocks which span a redshift range of $0.2 < z < 0.75$. The sky masks applied to define the four regions are defined on a HEALPix map and were initially constructed in the analysis by Loureiro et al. (2019). For this analysis the maps were degraded to an $N_{\text{side}} = 256$. Each galaxy from the real sample is assigned a non-integer weight,

$$w = w_{\text{see}}w_{\text{sys}}(w_{\text{col}} + w_{\text{nz}} - 1),$$  \hspace{1cm} (4.8)

while the mock PATCHY galaxies are assigned integer weights,

$$w = w_{\text{col}} + w_{\text{nz}} - 1. \hspace{1cm} (4.9)$$

Here $w_{\text{col}}$ is the fiber collision weight, $w_{\text{nz}}$ is the redshift failure weight, $w_{\text{see}}$ is the seeing weight and $w_{\text{sys}}$ is the systematic weight. The mocks retain integer weights since the effects of the non-integer weights $w_{\text{see}}$ and $w_{\text{sys}}$ are not forward modelled. In both cases we ignore the often-used Feldman-Kaiser-Peacock (FKP) weights (Feldman et al., 1994), used for power spectrum estimation, as this ensures that the mock catalogues have integer weights and therefore do not require an iterative stochastic weighting scheme.

### 4.5 RESULTS

In the following Section we test the SI weights by applying it to calculations of the 2PCF multipoles, we describe the process by which we apply the SB weights to construct the MST and we compare the MST applied to the real BOSS galaxies to the MST applied to the BOSS $\Lambda$CDM PATCHY mocks.

#### 4.5.1 Two-Point Multipoles with Stochastic Integer Weights

The 2PCF multipoles are calculated for the real BOSS galaxies and the PATCHY mocks, displayed in Figure 4.6. For the real sample they are first calculated using the non-integer weights (shown with purple lines) and then iterated over 10 times using the SI weights and averaged (shown with dashed magenta line). The 2PCF monopole $\xi_0(r)$ and quadrupole $\xi_2(r)$ are shown to be equivalent for the non-integer weights and for the SI weighting scheme. Since the PATCHY mocks have integer weights the SI weighting scheme is exactly
equal to the non-integer weights. This is compared to the multipoles calculated on 100 of the PATCHY mocks. As has previously been shown in Kitaura et al. (2016) and Ross et al. (2017) the mocks are in agreement with real data. The only exception appears in the monopole for CMASS NGC at \( r \gtrsim 125 \, h^{-1} \text{Mpc} \) where there appears to be an excess in the real data. This is similar to the excess reported in Ross et al. (2017) which is claimed to be statistically insignificant. Since the MST measurements probe smaller scales this inconsistency is unlikely to affect the results discussed in this paper.

### 4.5.2 Stochastic Binary Weights and Density Corrections

To ensure comparison between any MST statistics we need to ensure that each sample has roughly equivalent densities, since an offset in the density will result in shifts in the lengths of edges in the MST. This is further complicated in this study by groupings performed to remove small scales. The stages required to arrive to the final catalogues of nodes used to compare the real BOSS galaxies and PATCHY mocks are described below:

1. Assign SB weights: Galaxies from the real and mock catalogues are assigned SB weights. For the mocks the SB weights are deterministic since the weights are inte-
Figure 4.2: The stochastic binary (SB) weighted histogram for BOSS real and mock galaxies vs redshift. The PATCHY mocks are shown with green envelopes (1σ confidence limits shown in dark green and 2σ confidence limits in light green) and the mean for the real galaxies is shown in red. On the left are the NGC regions and on the right the SGC regions. The redshift regions LOWZ and CMASS are separated by the vertical dashed grey line. In the top panels are the SB weighted histograms $N_{\text{Galaxies}}$ and in the lower panels the relative difference from the PATCHY mocks $\Delta N_{\text{Galaxies}}$. The two distributions are shown to be offset; for LOWZ $N_{\text{Galaxies}}$ is larger for real galaxies while for CMASS $N_{\text{Galaxies}}$ is smaller for real galaxies.

Figure 4.3: The density-corrected stochastic binary (SB) weighted histogram for BOSS real and mock galaxies vs redshift. The PATCHY mocks are shown with blue envelopes (1σ confidence limits shown in dark blue and 2σ confidence limits in light blue) and the mean for the real galaxies is shown in purple. On the left are the NGC regions and on the right the SGC regions. The redshift regions LOWZ and CMASS are separated by the vertical dashed grey line. In the top panels are the density-corrected SB weighted histograms $N_{\text{Galaxies}}$ and in the lower panels the relative difference from the PATCHY mocks $\Delta N_{\text{Galaxies}}$. 
Chapter 4

Figure 4.4: The total stochastic binary weights ($N_{\text{Galaxies}}$; left of sub-plots) and total nodes ($N_{\text{Nodes}}$; right of sub-plots) for BOSS real and mock galaxies in the final catalogues. The PATCHY mocks are shown in blue and the real samples are shown in purple (the mean with the dashed line and $1\sigma$ confidence region is shown with a pale purple envelope). This is shown for the four BOSS survey regions for LOWZ on the right, CMASS on the left, NGC on the top and SGC on the bottom. The number of nodes for all regions are shown to be consistent, however the number of galaxies in LOWZ NGC in the real data appears significantly lower, suggesting the scalar density correction applied previously (shown in Figure 4.3) is insufficient.

1. Compare total weights: The total SB weights are compared to determine the offset $f_{\text{Galaxies}}$ from Equation 4.7. In Figure 4.2 the total SB weights (equivalent to $N_{\text{Galaxies}}$) is shown with respect to redshift. For the mocks this is shown with green envelopes ($1\sigma$ confidence limits shown in dark green and $2\sigma$ confidence limits in light green) and the mean for the real galaxies is shown in red. For LOWZ the real data is shown to have a larger total SB weight while for CMASS the real data is shown to have a smaller total SB weight.

2. Density correction on galaxies: A fraction $f_{\text{Galaxies}}$ of galaxies are randomly drawn from the larger sample. This ensures the total SB weights are roughly equal. This is shown in Figure 4.3 where the mocks are shown with blue envelopes ($1\sigma$ confidence limits shown in dark blue and $2\sigma$ confidence limits in light blue) and the mean for the real galaxies is shown in purple.

3. Grouping: Groups of galaxies are identified (to remove and mask small scales) in the density-corrected galaxy sample. Catalogues of nodes are constructed, which include groups and ungrouped galaxies.

4. Compare node counts: The real data and mocks are fundamentally different at small scales. In the real data fiber collisions prevent galaxies from being measured while in the mocks the modelling for small scale clustering does not consider astrophysical effects (such as baryonic feedback). For this reason we expect to find a different number of nodes despite previously applying the density correction. For all the survey regions we find a smaller number of nodes in the mock catalogues, this is most
likely due to the fact that galaxies in the real data cannot be observed due to fiber collisions and therefore fewer galaxies are merged into groups in the real data. Similar to before, we now determine the offset $f_{\text{Nodes}}$ from Equation 4.7.

6. Density correction on nodes: A fraction $f_{\text{Nodes}}$ of nodes are randomly drawn from the larger sample. This ensures the final catalogue of nodes are roughly at the same density. See Figure 4.4 to see the final count of galaxies and the final count of nodes once these corrections have been applied. There is good agreement between the mocks and real data except for the number of galaxies in LOWZ NGC where the real data contains significantly fewer galaxies.

### 4.5.3 Minimum Spanning Tree: Real BOSS Galaxies vs ΛCDM PATCHY Mocks

The MST is constructed on the nodes for all 2048 PATCHY mocks and the nodes for the real BOSS galaxies. For the real BOSS galaxies 100 SB weight realisations were used and the MST was constructed on each. The mean of $P$ is then taken for the four MST statistics. In Figure 4.5 we show the distribution of $d$, $l$, $b$ and $s$ measured from the real data in purple and measured from the PATCHY mocks in blue ($1\sigma$ confidence limits shown in dark blue and $2\sigma$ confidence limits in light blue). The four statistics appear to be consistent with the ΛCDM PATCHY mocks, with the exception of the distribution of $l$ for the LOWZ NGC where the peak occurs for longer edges. These distributions allow us to set cuts on $P$ to remove the tails of the distribution. For $l$, $b$ and $s$ we simply only consider $0.05 < \text{CDF} < 0.95$ (where CDF is the cumulative distribution function of $P$) while for $d$ we consider only $d \leq 4$.

Considering only the cut regions of $P$ we construct the covariance matrix $C$ for the individual survey regions using three quarters of the PATCHY mocks (1536). The correlation matrix shown in Figure 4.6 is very similar in features to that measured from the Quijote simulations in Naidoo et al. (in prep) (Chapter 3). Unlike the power spectrum the covariance for the MST is highly non-diagonal; we see significant correlations between the four statistics. The presence of non-diagonal elements substantiates the need for large sets of mock catalogues for the accurate calculation of covariance matrices.

We evaluate the compatibility of the BOSS MST measurements to those obtained from PATCHY mocks. To do this we use the calculated covariance matrices constructed from a subset of the PATCHY mocks (3/4 of the total set, i.e. 1532) and calculate the mean $P_{\text{Mean}}$ of the MST statistics for this subset. We then calculate the $\chi^2$,

$$\chi^2 = (P - P_{\text{Mean}}) \cdot C^{-1} \cdot (P - P_{\text{Mean}})$$

(4.10)

where $P$ is the PDF of the separate or combined MST statistics and $C$ is the covariance matrix (obtained from a subset of the PATCHY mocks, i.e. 1536). If the variance on the
Figure 4.5: The MST statistics for BOSS galaxies in purple and PATCHY mocks in blue (1σ confidence limits shown in dark blue and 2σ confidence limits in light blue). From left to right we show the distribution of d, l, b and s. From top to bottom we show the distribution from the four BOSS survey regions: LOWZ NGC, LOWZ SGC, CMASS NGC and CMASS SGC. In the sub-plot panels the relative deviation from the mean of the PATCHY distribution is shown. The grey shaded regions show where we cut the tails of P for the four statistics and different survey regions (i.e. for l, b and s where CDF < 0.05 and CDF > 0.95 and for d where d > 4). We find good agreement between the distributions for all four statistics.
Figure 4.6: Correlation matrix for PATCHY mocks from the MST statistics for the four survey BOSS regions (LOWZ in the top row, CMASS in the bottom row, NGC in the left column and SGC in the right column). The individual MST statistics are divided by dotted black lines which are indicated by the labels on the axes.
Figure 4.7: The $\chi^2$ for the MST statistics for real and PATCHY galaxies. From top to bottom we show the $\chi^2$ distribution for the four survey regions LOWZ NGC, LOWZ SGC, CMASS NGC and CMASS SGC. From left to right we show the $\chi^2$ obtained for $d$, $l$, $b$, $s$ and all (combined data vector for the four MST statistics). The $\chi^2$ for the real galaxies are indicated with vertical purple lines, the theoretical $\chi^2$ distribution is indicated with dashed black lines and the measured $\chi^2$ from 512 PATCHY mocks are shown with blue histograms for the individual MST statistics and with grey histograms for the combined MST statistics. In most cases we find the $\chi^2$ to show good agreement between real data and mocks, however the $\chi^2$ for $l$ in the LOWZ NGC regions is discrepant at the $\sim 2.5\sigma$ level.
data vectors are roughly Gaussian the measured $\chi^2$ obtained from the mocks will follow the theoretical $\chi^2$ PDF,

$$\text{PDF}_{\chi^2}(x) = \frac{x^{k/2-1} \exp\left(-\frac{x}{2}\right)}{2^{k/2} \Gamma(k/2)},$$

(4.11)

where $k$ is the degree of freedom (equal to the length of the data vector) and $\Gamma$ is the gamma function.

In Figure 4.7 we compare the $\chi^2$ obtained for the four MST statistics (individually and combined) for the real BOSS galaxies (purple line), the remaining unused PATCHY mocks (blue and grey histograms) and the theoretical $\chi^2$ distribution (dashed black line). The PATCHY $\chi^2$ follow the theoretical $\chi^2$ closely. This indicates that a Gaussian is sufficient to describe the noise on the MST statistics and suggest that the Poisson noise in the tails of $P$ have been correctly mitigated by our choice of cuts. Future work could explore this further by applying additionally normality tests such as the Kolmogorov-Smirnov test (Kolmogorov, 1933; Smirnov, 1948).

The $\chi^2$ obtained from the real data are found to be compatible with those obtained from the PATCHY mocks and the theoretical $\chi^2$ distribution for $d$, $b$ and $s$. The one exception is the $\chi^2_l$ for LOWZ NGC which is significant at the $\sim 2.5\sigma$ level. A deeper analysis is needed to clarify whether this is a systematic issue or evidence of new physics. When the four MST statistics are combined LOWZ NGC is completely compatible with PATCHY which either means the good fits in $d$, $b$ and $s$ are masking the poor fit in $l$ or more information is reducing the significance. A possible systematic which may be causing this discrepancy is a redshift dependent offset in the density. A scalar density correction was used in this study which may not be sufficient for this particular data set.

### 4.6 DISCUSSION

In this Chapter the MST statistics measured on BOSS real galaxies are compared to the $\Lambda$CDM PATCHY mocks. We find that three of the four surveys regions are completely consistent, with one exception: the distribution of edges $l$ in LOWZ NGC which is found to be inconsistent with the $\Lambda$CDM mocks at the $\sim 2.5\sigma$ level. Determining the source of this discrepancy will require further study.

We describe how to incorporate non-integer weights to point processes using an iterative stochastic integer weighting scheme. We show this scheme can be used to reproduce the weighted 2PCF multipoles for the BOSS survey. For the MST this scheme is modified to be binary, so all $w \geq 1$ are assigned a stochastic binary weight $w_{SB} = 1$. This introduces a density bias, since the sum of non-integer weights will now be equal or larger than the sum of binary integer weights. This bias is important when comparing two samples as the SB weights can introduce different levels of bias to the two samples. To account for this we introduce a density correction which randomly draws from the larger sample to ensure that
the two samples have roughly equal total weights. This is further complicated by the fact that galaxies separated by a distance less than $r_{\text{min}} = 10 \, h^{-1}\text{Mpc}$ are merged into single points. This step is required to avoid small scale issues, such as the effects of fiber collisions on the real data and inaccuracies in the small scale clustering of PATCHY. Once this is carried out the new catalogue of nodes (containing grouped galaxies and ungrouped galaxies) is not guaranteed to be of equal density. In this study, the real data was found to contain more nodes than in PATCHY. This is likely due to galaxies not being observed at small scales in the real data due to fiber collisions. To correct for this we reapply the density correction which leaves us with a final catalogue of nodes that incorporates both non-integer weights iteratively and matched densities. The density correction will only work if the difference between the real and mock redshift selection functions can be defined completely by a scalar, which appears to be the case for three of the four survey regions. If this is not the case a more complex scheme would need to be applied. In the case of LOWZ NGC, the discrepancy in $l$ could be evidence that a scalar correction is not sufficient. In the lower left panel of Figure 4.3 the density-corrected sample of real galaxies appears to be tilting to lower values at higher redshift, suggesting that the scalar density-correction is insufficient and a redshift dependent density correction is required. This is even clearer in Figure 4.4 where the number of galaxies for this region appear significantly lower to the number of galaxies in the mocks.

Finally we construct the covariance matrix from 1532 PATCHY mocks and calculate the $\chi^2$ of the MST statistics with the remaining 512 PATCHY mocks. In Figure 4.7 the $\chi^2$ distributions for the MST statistics show the noise in the statistics can be described by a Gaussian. Future studies could look to determine whether further normality test agree with this conclusion. Assuming the result holds, future studies will be able to use a Gaussian likelihood to constrain cosmological parameters from the MST.

The PATCHY mocks used in this study are produced with some of the BOSS survey systematics, including its redshift selection function and some observing systematics including fiber collisions and redshift failures. The MST provides a further test of this forward modelling and could be used to test the accuracy of mock catalogues produced for future galaxy redshift surveys.

This work will be applicable to future spectroscopic surveys such as the Dark Energy Spectroscopic Instrument (DESI) that will map the positions of $\sim 35$ million galaxies and quasars (an order of magnitude larger than BOSS). Of particular relevance for the MST will be DESI’s bright galaxy survey that will map the positions of $\sim 10$ million galaxies and will sample the distribution of galaxies at low redshift $z \lesssim 0.5$ at higher densities. This will provide a much larger and denser data set (therefore increasing the sensitivity to the cosmic web) to apply the MST to test $\Lambda$CDM and to constrain cosmological parameters. Future work could explore other statistics and methods including machine learning, higher order moments and Minkowski functionals (Minkowski, 1903) to determine how their statistics
differ from the MST and whether the parameter degeneracies for these alternative statistics can be exploited or added for future parameter estimation.
“**Batman/Bruce Wayne:** People are dying, Alfred. What would you have me do?

**Alfred:** Endure, Master Wayne. Take it. They'll hate you for it, but that’s the point of Batman, he can be the outcast. He can make the choice that no one else can make, the right choice.

**Batman/Bruce Wayne:** Well today I found out what Batman can’t do. He can’t endure this. Today you get to say “I told you so.”

**Alfred:** Today, I don't want to... but I did bloody tell you.”

– The Dark Knight (2008)
5.1 ABSTRACT

We re-analyse the cosmic microwave background (CMB) Cold Spot (CS) anomaly with particular focus on understanding the bias a mask (contaminated by Galactic and point sources) may introduce. We measure the coldest spot, found by applying the spherical Mexican hat wavelet transform on 100,000 cut-sky (masked) and full-sky CMB simulated maps. The CS itself is barely affected by the mask; we estimate a 94% probability that the CS is the full-sky temperature minima. However, $\sim$48% (masked fraction of the mask) of full-sky minima are obscured by the mask. Since the observed minima are slightly hotter than the full-sky ensemble of minima, a cut-sky analysis would have found the CS to be significant at $\sim$2.2$\sigma$ with a wavelet angular scale of $R = 5^\circ$. Nonetheless, comparisons to full-sky minima show the CS significance to be only $\sim$1.9$\sigma$ and $<2\sigma$ for all R. The CS on the last scattering surface may be hotter due to the integrated Sachs-Wolfe effect in the line-of-sight. However, our simulations show that this is on average only $\sim$10% (about $-10 \mu K$ but consistent with zero) of the CS temperature profile. This is consistent with $\Lambda$ and Cold Dark Matter reconstructions of this effect based on observed line-of-sight voids.

5.2 INTRODUCTION

The cosmic microwave background (CMB) Cold Spot (CS) anomaly was discovered by Vielva et al. (2004) using the spherical Mexican hat wavelet (SMHW) (Cayón et al., 2001) on WMAP data. The anomaly has persisted (Cruz et al., 2007; Bennett et al., 2011) and was later verified by Planck (Planck Collaboration et al., 2016b).

Inoue & Silk (2006, 2007) claimed the integrated Sachs-Wolfe (ISW) (Sachs & Wolfe, 1967) and Rees-Sciama (RS) (Rees & Sciama, 1968) effects of a large void at redshift $z \sim 1$ could explain the entire feature (Nadathur et al. (2014) show the RS is subdominant in all cases). However, pencil beam surveys (Bremer et al., 2010; Granett et al., 2010) have effectively ruled out the possibility of such a large void at high redshift (i.e. $0.5 < z < 1$). Studies of the galaxy distribution in the relevant region using photo-z initially appeared to indicate that a single spherical/elliptical void exists along the line-of-sight (LOS) at lower
VOIDS AND THE COLD SPOT: THE ISW OR MORE?

redshift (see Szapudi et al., 2015; Kovács & García-Bellido, 2016). Several studies have shown this is insufficient to explain the CS (see Nadathur et al., 2014; Zibin, 2014; Marcos-Caballero et al., 2016). Naidoo et al. (2016) found that a model using multiple voids could only explain a fraction of the feature. This was recently confirmed by Mackenzie et al. (2017) who observed three voids along the LOS and came to the same conclusion. Hints of a stronger than expected ISW signal have been found in some stacked void studies (Granett et al., 2008; Cai et al., 2014; Kovács et al., 2017; Kovács, 2018), leading to speculation that the causal relation between the CS and the LOS voids may be much greater than that predicted by the ISW. However, Ilić et al. (2013), Hotchkiss et al. (2015) and Nadathur & Crittenden (2016) have found no such excess and obtain results consistent with ΛCDM.

The use of a mask in the SMHW analysis of the CS, to minimise contribution from the Galaxy and point sources, is common practice (see Vielva et al., 2004; Zhang & Huterer, 2010; Nadathur et al., 2014; Planck Collaboration et al., 2016b). Because the SMHW transform integrates across the sky, contributions from masked areas will leak to neighbouring regions. Thus a more aggressive mask than the original is applied to the filtered map (see Zhang & Huterer, 2010; Rassat et al., 2014). While the application of a mask is sometimes unavoidable, Rassat et al. (2014) show that many CMB anomalies, including the CS, are no longer significant when carried out without the use of a mask on full sky LGMCA CMB maps (Bobin et al., 2014). Furthermore, the CS’s inability to be detected by other filters (see Zhang & Huterer, 2010; Marcos-Caballero et al., 2017) has placed doubt on its significance. However, this is often argued to be due to the SMHW sensitivity to what makes the CS anomalous, i.e. its high transition from cold to hot.

In this paper, we investigate the effects of masking on the detection and resulting significance of the CS and the expected contribution of the ISW to the CS profile.

5.3 METHOD

In the following analysis we use the Planck SMICA CMB map and the Planck Common Field mask1.

5.3.1 Spherical Mexican Hat Wavelet

The SMHW is defined according to an angular scale $R$ as:

$$\Psi(\theta; R) = A_{\text{wav}}(R) \left( 1 + \left(\frac{y}{2}\right)^2 \right)^2 \left( 2 - \left(\frac{y}{R}\right)^2 \right) \exp\left(-\frac{y^2}{2R^2}\right).$$  \hspace{1cm} (5.1)

1Available from http://pla.esac.esa.int/pla/#home.
where \( y \equiv 2 \tan(\theta/2) \) and \( \theta \) is the angular separation between two points, \( \hat{n} \) and \( \hat{n}' \), on a sphere. \( A_{\text{wav}}(R) \) is a normalisation constant defined as:

\[
A_{\text{wav}}(R) = \left[ 2\pi R^2 \left( 1 + \frac{R^2}{2} + \frac{R^4}{4} \right) \right]^{-1/2}.
\] (5.2)

The filtered temperature, i.e. the SMHW value of a point at \( \hat{n} \) as the transform is applied to an area with an angular radius of \( \theta \), is given by:

\[
\Delta T_{\text{wav}}(\theta; \hat{n}, R) = \int_0^\theta \Delta T(\hat{n}'); \Psi(\theta'; R) d\Omega',
\] (5.3)

where \( \hat{n}' \) are pixels located within an angular distance \(< \theta \) from point \( \hat{n} \). Such pixels are found by using the HEALPix function \texttt{query_disc}. The SMHW of a single pixel, \( \Delta T_{\Psi}(\hat{n}) \), is then calculated by integrating Equation 5.3 across the whole sky or up to an angular radius of \( \theta \simeq 4R \) (since \( \Psi(\theta \gtrsim 4R; R) \simeq 0 \)):

\[
\Delta T_{\Psi}(\hat{n}) = \Delta T_{\text{wav}}(\pi; \hat{n}, R) \simeq \Delta T_{\text{wav}}(4R; \hat{n}, R).
\] (5.4)

In order to remove contamination from Galactic foregrounds and point sources a mask is applied. In order to do this we must first calculate an occupancy fraction (Zhang & Huterer, 2010), which determines the contribution of masked regions to the wavelet trans-
form. This is given approximately by:

$$
\mathcal{N}(\hat{n}; R) \simeq \int_0^{4R} \mathcal{M}(\hat{n}'); \Psi^2(\theta' ; R) d\Omega',
$$

(5.5)

where $\mathcal{M}(\hat{n})$ and $\mathcal{N}(\hat{n})$ are the mask and occupancy fraction value, respectively, at a point $\hat{n}$. Similarly to Equation 5.4, we integrate only up to $\theta = 4R$ rather than $\theta = \pi$ for the exact solution since $\Psi (\theta \gtrsim 4R ; R) \simeq 0$.

The SMHW is applied to the full CMB map. Pixels with a mask and occupancy fraction of $\mathcal{M} < 0.9$ or $\mathcal{N} < 0.95$ respectively are then masked to remove areas of the map where contaminated sources may contribute significantly to the result. This means the effective mask applied to the map is considerably larger than the mask $\mathcal{M}$, with $\sim 48\%$ ($\sim 66\%$ for $\mathcal{M} > 0.9$) unmasked pixels (see Figure 5.1).

### 5.3.2 Simulating Cosmic Microwave Background and Integrated Sachs-Wolfe Maps

Using CLASS\(^2\) (Blas et al., 2011) we generate $C_\ell$ based on best fit Planck TT, TE, EE + lowP + lensing + ext cosmological parameters (see Planck Collaboration et al., 2016a). We deliberately turn off the late ISW effect (for $z < 10$), giving $C_\ell$ for the primordial CMB. $C_\ell$ for only the late ISW effect are calculated separately. We then generate primordial CMB maps, $\Delta T_P$, and ISW maps, $\Delta T_{ISW}$, using the HEALPix software (Górski et al., 2005a) at $N_{\text{side}} = 128$ and add them,

$$
\Delta T(\hat{n}) = \Delta T_P(\hat{n}) + \Delta T_{ISW}(\hat{n}),
$$

(5.6)

to give a full CMB map ($\Delta T$). The motivation for generating these maps separately is to allow us to investigate the ISW contribution to the coldest spots in CMB realisations. Since the major contribution to the $\Delta T_{ISW}$ occurs at $z < 1.4$ the correlation between $\Delta T_P$ and $\Delta T_{ISW}$ is expected to be small.

### 5.3.3 Searching for the Coldest Spots

To search for the coldest spots in our simulated maps we apply the SMHW transform to $\Delta T$ maps downgraded from $N_{\text{side}} = 128$ to 16. This is carried out with and without a mask. Using the location of the coldest pixel in the downgraded map ($N_{\text{side}} = 16$) we measure $\Delta T_{\text{wav}}(\theta; R)$ (where $R = 5^\circ$), $\Delta T(\theta)$ and $\Delta T_{ISW}(\theta)$ (i.e. the average $\Delta T_i$ of $i$ in concentric rings of the coldest spot) on the original $N_{\text{side}} = 128$ map. This was carried out on 100,000 simulations. We will refer to the coldest spots identified in unmasked and masked maps as full-sky and cut-sky minima respectively.

To understand the role of masking we additionally measure the angular separation $\alpha$

\(^2\)Software is available from [http://class-code.net/](http://class-code.net/).
between the full-sky and cut-sky minima. The two are only considered to be equivalent if \( \alpha = 0 \), since even a slight misalignment will introduce a bias. We apply the exact same procedure to the Planck SMICA map using the Planck Common Field mask.

A Frequentist, rather than a Bayesian, approach is applied as we are determining the CS consistency with \( \Lambda \)CDM rather than doing model comparisons where the alternative would be better suited.

5.4 RESULTS

5.4.1 Masked vs Unmasked Coldest Spot

The full-sky and cut-sky minima are compared in Figure 5.2. Using the Planck Common Field mask, we find that these are equivalent only \( \sim 48\% \) of the time, as one would expect given that this is the effective fraction of the map that is removed by the mask. Since cut-sky minima are not always equal to the full-sky minima the use of a mask biases \( \Delta T_\Psi \), causing it to be on average \( \sim +0.93 \mu K \) hotter using the Common Field mask. This is because cut-sky minima are on average \( \sim +1.78 \mu K \) hotter than the full-sky minima. Interestingly, colder cut-sky minima (i.e. \( \Delta T_\Psi < -18 \mu K \)) are more likely to be equivalent to the full-sky minima. This becomes particularly interesting for the CMB CS.

5.4.2 The Cold Spot in Planck Data

The CS has a \( \Delta T_\Psi \simeq -19.3 \mu K \) with a significance of \( \sim 2.2\sigma \) when masked. To make a comparison between the full-sky minima in simulations we must first understand whether the CS is indeed our CMB’s full-sky minima. Without any prior knowledge of the CS’s \( \Delta T_\Psi \) the probability that the cut-sky minima is equivalent to the full-sky minima (\( P(\text{full}) \)) is \( \simeq 0.48 \). However, the probability increases as \( \Delta T_\Psi \) decreases. The conditional probability that a cut-sky minima similar to the CS (i.e. \( -19.5 \mu K < \Delta T_\Psi < -19 \mu K \)) is equivalent to the full-sky minima (\( P(\text{full} | \Delta T_\Psi^{CS}) \)) is actually \( \simeq 0.94 \). This means we can be fairly certain that the CS is the CMB’s full-sky minima. In Figure 5.2 the CS’s \( \Delta T_\Psi \) is shown and lies well within the \( 2\sigma \) distribution of full-sky minima in simulations. The CS’s significance in comparison to full-sky minima is \( \sim 1.9\sigma \) (which corresponds to a P-value \( \sim 3\% \)). In Figure 5.3 the CS’s \( \Delta T(\theta) \) and \( \Delta T_{\text{wav}}(\theta) \) are compared to the 1 and \( 2\sigma \) contours of the cut-sky and full-sky minima in simulations (indicated by black lines and blue contours respectively). The comparison illustrates precisely how the observed profiles are biased. For \( \Delta T(\theta) \) the main difference occurs near the center (\( \theta < 5^\circ \)) where full-sky minima appear slightly colder. This appears to be more pronounced in \( \Delta T_{\text{wav}}(\theta) \), where the distribution is found to be consistently colder for all values of \( \theta \).
5.4.3 The Cold Spot’s Significance vs. Mask Size

Using the SILC CMB map (Rogers et al., 2016, specifically using the $N = 5$ map) and corresponding mask we test the effect of the size of the mask on the CS’s significance. The mask for the SILC CMB map is relatively small such that even the effective mask has $\sim 88\%$ unmasked pixels ($f_{\text{sky}}$). We gradually enlarge this mask by masking away a wider galactic strip and run the same procedure. In Figure 5.4 we plot the CS’s significance in comparison to cut-sky minima (shown in black) and compare the CS’s significance to the full-sky minima (shown in blue) as a function of $f_{\text{sky}}$. The CS significance in comparison to cases where the full-sky and cut-sky minima are equivalent always remains $< 2\sigma$. But in comparison to cut-sky minima the significance becomes larger as $f_{\text{sky}}$ decreases. Rather unsurprisingly, a larger mask will make it harder to find the full-sky minima and will also make it more likely that a hotter cut-sky minima is measured. The net effect is that a full-sky minima measured in a cut-sky analysis will have a boosted significance due to the size of the mask. This appears to be the case for the CS.
Figure 5.3: The 1 and 2σ contours (dark and light shades respectively) for the \( \Delta T(\theta) \) (top left) (the average \( \Delta T \) in concentric rings from the cut/full-sky minima’s center) and \( \Delta T_{\text{wav}}(\theta) \) (top right) profiles, are shown in blue for cut-sky minima in 100,000 simulations. The 1 and 2σ contours for cut-sky minima are marked as dashed and dotted black lines, respectively. The CS’s \( \Delta T(\theta) \) and \( \Delta T_{\text{wav}}(\theta) \) are shown (measured on Planck’s SMICA map) as the dark blue dashed line. The subtle shift in the full-sky \( \Delta T(\theta) \) profile around \( \theta < 5^\circ \) shown on the left panel appears to lead to colder final temperatures shown on the right panel. The difference between the mean of the full-sky and cut-sky \( \Delta T \) (left) and \( \Delta T_{\text{wav}} \) (right) profiles are indicated with a superscript full and cut, respectively, in the bottom panels (note the scale on the bottom panels).

Figure 5.4: The significance of the CS is measured in comparison to the distribution of cut-sky minima (shown in black) as a function of the mask size \( f_{\text{sky}} \) (unmasked fraction of the sky). The significance of the CS is shown in blue in comparison to the full-sky minima observed in a cut-sky. As \( f_{\text{sky}} \) decreases it is more likely that the full-sky minima is obscured by the mask and that the cut-sky minimum measured is hotter. Consequently these two effects increase the significance of the CS. The vertical red dash-dotted line indicates the \( f_{\text{sky}} \) of the Planck Common Field mask.
5.4.4 The Integrated Sachs-Wolfe for the Coldest Spots

The ISW contribution to the coldest spots in simulations was measured and is shown in Figure 5.5. Here we display the mean and 1σ contours for all the full-sky minima and the most extreme 3% (which approximately corresponds to the CS’s p-value). The profiles are poorly constrained and very similar, with the more extreme case tending to be slightly more negative. The result illustrates that it is very likely that the ISW plays a minor role in the CS profile: ~ 10% of the full profile. The reconstructed ISW profiles (Rassat et al., 2014; Nadathur et al., 2014; Finelli et al., 2016; Planck Collaboration et al., 2016c) appear to be consistent with the predicted ISW shown in Figure 5.5. The presence of prominent voids in the LOS (see Szapudi et al., 2015; Kovács & García-Bellido, 2016) are therefore precisely what we would expect from ΛCDM.

5.4.5 Dependence on Angular Scale

Up to this point we have used a preselected angular scale, \( R = 5° \), where the CS was measured to be most significant by Planck Collaboration et al. (2016b). However, our conclusions for the CS significance may not necessarily hold true for other angular scales. To test this, \( R \) is varied between 4° and 7°, roughly equaling the range of \( R \) over which Zhang & Huterer (2010) found the CS to be significant. The same procedure is carried out as before except with a smaller number of realisations (10,000).

In Table 5.1 we summarise these results. The probability, \( P(\text{full}) \), is roughly equal to the fraction of unmasked pixels of the effective mask. However, \( P(\text{full}|T^C_{\Psi}S) \) is found to be > 0.85 for the angular scales considered. When the cut-sky significance > 2σ the
Table 5.1: The probability that the full-sky and cut-sky minima (cold spot) are equivalent for any $\Delta T_{\Psi}$ and for the CS’s $\Delta T_{\Psi}^{CS}$ is indicated by $P(\text{full})$ and $P(\text{full}|\Delta T_{\Psi}^{CS})$, respectively, for each angular scale $R$. The significance of the CS is shown in comparison to the cut-sky and full-sky minima in CMB realisations. For each value of $R$ (except $R = 5$ where 100,000 realisation were previously made) 10,000 CMB realisations were simulated.

| $R$ [$^\circ$] | $P(\text{full})$ | $P(\text{full}|\Delta T_{\Psi}^{CS})$ | Cut-sky [$\sigma$] | Full-sky [$\sigma$] |
|---------------|----------------|-----------------------------------|-------------------|-------------------|
| 4             | 0.51           | 0.94                              | 1.95              | 1.65              |
| 4.5           | 0.50           | 0.96                              | 2.18              | 1.85              |
| 5             | **0.48**       | **0.94**                          | **2.19**          | **1.91**          |
| 5.5           | 0.46           | 0.96                              | 2.19              | 1.89              |
| 6             | 0.45           | 0.94                              | 2.08              | 1.76              |
| 6.5           | 0.44           | 0.91                              | 1.85              | 1.50              |
| 7             | 0.42           | 0.86                              | 1.53              | 1.13              |

The probability is even higher ($> 0.93$). This makes it appropriate to compare the CS to full-sky minima in simulations where it is $< 2\sigma$ for $4^\circ < R < 7^\circ$. Combined with previous studies (e.g. Vielva et al., 2004) means the CS is $< 2\sigma$ for all angular scales.

5.5 CONCLUSIONS

We measure the cut-sky and full-sky minima (cold spot) in 100,000 simulations using the Planck Common Field mask which has a similar $f_{\text{sky}}$ to the WMAP KQ75 and Planck U74 masks used in Zhang & Huterer (2010) and Nadathur et al. (2014) respectively. The probability of observing the full-sky minima is found to be $\sim 0.48$ (which roughly equals the unmasked fraction of the effective mask). At other positions the cut-sky minima is not equivalent to the full-sky minima and this biases the distribution of minima (see Figure 5.2). This appears to have a significant effect only at $\Delta T_{\Psi} > -18 \mu K$; at the CS’s $\Delta T_{\Psi} \simeq -19.3 \mu K$ there is a $\sim 0.94$ probability that we are observing the CMB’s full-sky minima.

We argue that the CS is detected as an anomaly, with a significance of $\sim 2.2\sigma$, because the full-sky minimum is not always measured when using a mask resulting in an ensemble of cold spots which are slightly hotter than the full-sky ensemble. Correcting for this bias, by comparing to full-sky minima, reduces the significance to $\sim 1.9\sigma$. We emphasize that the CS itself does not change due to the mask; rather, the ensemble to which it is compared is colder when the mask is removed. The difference in $\Delta T(\theta)$ and $\Delta T_{\text{w}a\text{v}}(\theta)$ of the cut-sky and full-sky minima is subtle (see Figure 5.3). But, a colder $\Delta T(\theta)$ for $\theta < 5^\circ$ results in colder $\Delta T_{\Psi}$. This result is true for all angular scales (see Table 5.1) and would presumably remain for any model that can reproduce the CMB temperature $C_{\ell}$. In this sense these results are model independent.

By varying the size of the mask, we find that the cut-sky minima is often not equal to the full-sky minima due to the latter’s frequent obstruction by the mask. The inclusion of
these hotter cut-sky minima appear to be driving the CS’s significance. The CS can only be considered an anomaly if it is not the full-sky minimum itself as this would require a more extreme feature within the mask. This is unlikely, since such features are not seen in maps with a smaller mask or in full sky reconstructed maps (Rassat et al., 2014). This places the significance of the CS in serious doubt, as such it is becoming increasingly likely that the CS is simply a Gaussian fluctuation which has appeared anomalous in isolation. The ‘look-elsewhere’ effect, a phenomenon that occurs when a statistical significant observation arises by chance due to the size of the parameter space, is particularly relevant in this case. Consider the numerous non-Gaussianity test performed on the CMB, we should expect (based on the ‘look-elsewhere’ effect) for statistically significant anomalies, like the CS, to arise purely by chance. The only remaining thing which could cast doubt on this result is the rather curious discovery of voids in the LOS (Szapudi et al., 2015; Mackenzie et al., 2017). The relation between the CS and voids is recast in terms of the relation between the CS and the ISW where we assume that a negative ISW is associated with an underdensity or voids at low redshift in the LOS and vice versa.

The contribution from the ISW (predicted by ΛCDM) to the coldest spots is poorly constrained and consistent with zero, but leans towards a negative contribution (see Figure 5.5). On average it amounts to ~ 10% of the full profile. Measurements of large voids in the LOS and ISW reconstructions are consistent with this result. Since reconstructed ISW profiles (see Nadathur et al., 2014; Finelli et al., 2016; Kovács & García-Bellido, 2016) appear to be below the mean shown in Figure 5.5, it is possible that the ISW is amplifying the significance of the CS. This would mean the primordial CS profile is even less significant than measured. Alternative models, which are not investigated here, may explain the slightly higher than expected causal relation between the observed and expected ISW of large voids seen in certain studies (Granett et al., 2008; Cai et al., 2014; Kovács et al., 2017; Kovács, 2018) but not all (Ilić et al., 2013; Hotchkiss et al., 2015; Nadathur & Crittenden, 2016). Whether this is the case could be studied in future and would have implications for the predicted ISW contribution to the CS.

In this paper we have demonstrated the importance of understanding the bias that masking, which is often unavoidable, can introduce on the analysis of CMB anomalies. It is important that this is accounted for in future studies.
“You know, as bad as those things are, at least they're predictable. It's the normal people that scare me.”

– The Last of Us (2013)
6.1 ABSTRACT

We construct full sky maps of the integrated Sachs-Wolfe (ISW) effect for the MICE Grand Challenge and Flagship lightcone simulations. These are computed using spherical Fourier Bessel transforms in the linear regime and is computed solely from the lightcone density distribution. This removes the need for computationally expensive simulation snapshots of the density, velocity and potential fields at regular intervals. The ISW maps are computed up to redshift $z = 5$. They are provided either in slices of $\Delta z = 1$ or for MICE in the redshift ranges $0 < z < 1.4$ and $1.4 < z < 5$ and for Flagship in the redshift ranges $0 < z < 2.2$ and $2.2 < z < 5$. The different ranges allow for the study and localisation of the ISW at different redshifts. Due to the finite size of the simulations the maximum Fourier mode measured is given by the fundamental frequency of the simulation $k_F$, while the minimum Fourier mode measured is set to $k_{\text{max}} = 0.1 \, h\text{Mpc}^{-1}$ to ensure the calculation remains in the linear regime. The ISW is only fully reproduced for spherical harmonic modes $\ell$ where the sampled Fourier modes lie between $k_F < k < k_{\text{max}}$. We measure the angular power spectra of the ISW maps and determine the ranges in $\ell$ for which the maps fully reproduce the expected ISW. The ISW maps are made public to facilitate future large scale structure and cosmic microwave background cross-correlation studies for current and future galaxy surveys.

6.2 INTRODUCTION

The integrated Sachs-Wolfe (Sachs & Wolfe, 1967) (ISW) effect is a secondary anisotropy imprinted onto the cosmic microwave background (CMB) by the evolution of gravitational potentials along the line-of-sight (LOS). The effect is most prominent at large angular scales but is sub-dominant to the other anisotropies of the CMB. For this reason we cannot reconstruct the ISW from the CMB alone and we instead must do so by tracing large scale structure (LSS) from galaxy redshift surveys.

There have been a number of successful ISW cross-correlation studies (for example see Giannantonio et al., 2008; Planck Collaboration et al., 2016c; Stölzner et al., 2018). Since
the ISW is dependent on the linear evolution of gravitational potentials it provides a direct
detection of dark energy and could be crucial in determining whether dark energy is more
than just a cosmological constant.

Studies of the stacked ISW imprint of large voids have consistently measured larger
than expected imprints (Granett et al., 2008; Cai et al., 2017; Kovács et al., 2019). Al-
though similar studies (Nadathur & Crittenden, 2016) find no excess they appear to be
measuring smaller and spectroscopically identified voids, while Granett et al. (2008), Cai
et al. (2017) and Kovács et al. (2019) identify voids from photometric surveys which prefer-
entially identify larger elongated voids along the LOS. Whether this result is a sign of new
physics remains unclear but a recent study has found that a modification to the growth
function (specifically assuming an inhomogeneous AvERA model with emerging curva-
ture; Kovács et al., 2020) can explain both the excess ISW imprint from large voids and
the Hubble constant tension (between measurements from Type Ia supernova (Riess et al.,
2019) and CMB (Planck Collaboration et al., 2018)). This suggest the excess ISW from
large voids could be part of a larger picture of tensions in cosmology which includes the
Hubble tension and $\sigma_8$ tension (Hildebrandt et al., 2017). For the role of the ISW in these
tensions to be further explored it is important that the ISW be simulated for a larger set of
simulations for a more varied set of cosmologies.

Simulations of the ISW and LSS generally take one of two approaches: either they are
produced using correlated full sky maps from theoretical angular power spectra (Man-
zotti & Dodelson, 2014) or they are produced from $N$-body simulations (Cai et al., 2010;
Watson et al., 2014; Carbone et al., 2016). The latter method is crucial for the accurate pro-
duction of galaxy mock catalogues but it is computationally expensive to use for the ISW
(because producing the ISW requires snapshots of the density, velocity and potential field).
Furthermore because the ISW is strongest on large angular scales they require large simu-
lations (i.e. $\sim$ Gpc in size), while for accurate galaxy mocks fine resolutions are required.
This requirement for both large and fine resolution makes studying correlated ISW and
LSS simulations difficult to produce. They have so far been limited to a few simulations
exploring a small parameter space (fiducial $\Lambda$CDM, massive neutrinos and single $\omega$CDM
models; Cai et al., 2010; Watson et al., 2014; Carbone et al., 2016; Adamek et al., 2020).

In this paper we provide an alternative approach in which we compute the ISW from
lightcone simulations using the spherical Fourier Bessel (SFB) transform (similar to the
equations derived by Shapiro et al., 2012, except with different boundary conditions). We
concentrate only on the linear regime (since this provides the most significant contribu-
tions to the ISW) and since this lowers the necessity for outputting computationally ex-
pensive snapshots. Using this method the ISW can then be computed from any current or
future lightcone simulation.

The Chapter is organised as follows. In Section 6.3 we introduce the theory behind
the ISW, provide an overview of the SFB transform and discuss the methodology used to
apply this to the MICE Grand Challenge (Croce et al., 2015) (based on WMAP cosmology Dunkley et al. (2009)) and Flagship simulations (based on Planck cosmology Planck Collaboration et al. (2018)). These are the main lightcone simulations designed to be used by many of the current and future galaxy redshift surveys (such as the Dark Energy Survey\(^1\), Dark Energy Spectroscopic Instrument\(^2\) and Euclid\(^3\)). In Section 6.4 we discuss the results and limitations of these maps and in Section 6.5 we provide concluding remarks.

6.3 THEORY

In Section 6.3.1 we outline the basic equations for the ISW and in Section 6.3.2 we outline the steps required to calculate the analytical auto- and cross-angular power spectrum for the density field and ISW. In Section 6.3.3 we then provide an overview of the SFB transform. Finally in Section 6.3.4 we outline how the ISW can be reconstructed from the spherical harmonics of the projected density field and the SFB coefficients derived from a set of density maps provided in redshift shells.

6.3.1 Integrated Sachs-Wolfe

The ISW (Sachs & Wolfe, 1967) is a secondary anisotropy imprinted onto the CMB by light passing through evolving potentials. It is defined as

\[
\left( \frac{\Delta T(\hat{n})}{T} \right)_{\text{ISW}} = \frac{2}{c^2} \int_{t_0}^{t_{LS}} \dot{\Phi}(\hat{n}, t) \, dt,
\]

where \(\hat{n}\) is the LOS direction, \(c\) is the speed of light, \(t\) is time, \(\Phi\) is the gravitational potential, \(\dot{\Phi}\) its time derivative, \(t_0\) is the observer’s time and \(t_{LS}\) is the last scattering time. This can be rewritten by changing the variable of integration to the comoving distance \(r\),

\[
\left( \frac{\Delta T(\hat{n})}{T} \right)_{\text{ISW}} = \frac{2}{c^3} \int_0^{R_{\text{max}}} \dot{\Phi}(\hat{n}, r) a(r) dr,
\]

where \(a\) is the scale factor \(a = 1/(1 + z)\) and \(R_{\text{max}}\) is the comoving distance to the last-scattering surface (LS). The comoving distance \(r\), during and after matter domination in \(\Lambda\)CDM, can be approximated to be,

\[
r(z) \simeq 3000 \, h^{-1}\text{Mpc} \int_0^z \frac{1}{\Omega_{m,0}(1 + z')^3 + \Omega_{\Lambda,0}} \, dz',
\]

where \(\Omega_{m,0}\) is the current matter density and \(\Omega_{\Lambda,0}\) the current dark energy density. In this paper we will consider only the linear contribution to the ISW; this has been shown to

\(^1\)http://www.darkenergysurvey.org
\(^2\)http://desi.lbl.gov/
\(^3\)http://www.euclid-ec.org/
be dominant in ΛCDM (Cai et al., 2010; Nadathur et al., 2014) over the non-linear Rees-Sciama effect (Rees & Sciama, 1968). The gravitational potential is related to the density contrast $\delta(x, t)$, at a point $x$ and time $t$, by the Poisson equation which can be written as

$$\nabla^2 \Phi(x, t) = \frac{3}{2} H_0^2 \Omega_{m,0} \frac{\delta(x, t)}{a}.$$ (6.4)

In the linear regime perturbations grow according to the linear growth factor $D(z)$ defined as

$$D(z) \propto H(z) \int_0^a \frac{1}{(a'H(a'))^3} da',$$ (6.5)

where $D(0) = 1$ and $H(z)$ is the Hubble expansion rate at redshift $z$ approximated during and after matter domination as

$$H(a) \approx H_0 \left( \Omega_{m,0} a^{-3} + \Omega_{\Lambda,0} \right)^{1/2},$$ (6.6)

where $H_0$ is the present expansion rate. Since we assume curvature is flat, the energy density for dark energy is given by $\Omega_{\Lambda} = 1 - \Omega_m$. In the linear regime the time evolution of linear density perturbations can be separated,

$$\delta(x, t) = D(t)\delta(x),$$ (6.7)

and therefore the Poisson equation can be rewritten as

$$\nabla^2 \Phi(x, t) = \frac{3}{2} H_0^2 \Omega_{m,0} \delta(x) \frac{D(t)}{a},$$ (6.8)

where the solution in Fourier space is given by

$$\Phi(k, t) = \frac{3}{2} H_0^2 \Omega_{m,0} \frac{\delta(k)}{|k|^2} \frac{D(t)}{a}.$$ (6.9)

The derivative is then given by

$$\Phi'(k, t) = \frac{3}{2} H_0^2 \Omega_{m,0} \frac{\delta(k)}{|k|^2} \frac{\partial}{\partial t} \left( \frac{D(t)}{a} \right),$$ (6.10)

where

$$\frac{\partial}{\partial t} \left( \frac{D(t)}{a} \right) = \frac{D(t)}{a} H(t) [f(t) - 1];$$ (6.11)
here \( f = d\ln D / d\ln a \), which can be approximated as \( f(z) = \Omega_m(z)^{0.6} \) (Peebles, 1980; Lahav et al., 1991) where

\[
\Omega_m(z) = \frac{\Omega_{m,0}(1+z)^3}{\Omega_{m,0}(1+z)^3 + \Omega_{\Lambda,0}}.
\] (6.12)

Finally, the time derivative of the potential can be re-expressed in the linear regime as

\[
\dot{\Phi}(k, t) = \Phi(k, t) H(t) [f(t) - 1].
\] (6.13)

### 6.3.2 Theoretical Angular Power Spectrum for the Density and Integrated Sachs-Wolfe

To compute the theoretical auto- and cross-angular power spectrum of the density field and ISW we express the 1D linear power spectrum in its dimensionless form,

\[
\Delta^2(k) = \frac{4\pi}{(2\pi)^3} k^3 P(k).
\] (6.14)

The auto-angular power spectrum \( C_{\ell g}^{gg} \) of the density field is given by

\[
C_{\ell g}^{gg} = 4\pi b^2 \int \frac{\Delta^2(k)}{k} [W_g(k)]^2 dk,
\] (6.15)

where

\[
W_g(k) = \int \Theta(r) j_\ell(kr) D(z) dr,
\] (6.16)

\( j_\ell \) is the spherical Bessel function and \( \Theta(r) \) is derived from the redshift selection function \( n(r) \) by

\[
\Theta(r) = \frac{r^2 n(r)}{\int x^2 n(x) dx},
\] (6.17)

which is integrated over the comoving distance \( r \). The cross-angular power spectrum \( C_{\ell g}^{\text{ISW} \rightarrow g} \) between the density field and ISW is defined to be

\[
C_{\ell g}^{\text{ISW} \rightarrow g} = 4\pi b \int \frac{\Delta^2(k)}{k} W_g(k) W_T(k) dk,
\] (6.18)

where

\[
W_T(k) = -\frac{3\Omega_{m,0} H_0^2}{k^2 c^3} \int j_\ell(kr) H(z) D(z) [f(z) - 1] dr.
\] (6.19)
Finally the auto-angular power spectrum for the ISW is given by

\[ C_{\ell}^{\text{ISW}} = \left( \frac{C_{\ell}^{\text{ISW}} - g}{C_{\ell}^{g g}} \right)^2, \]  

(6.20)

and the cosmic variance limited uncertainty is given by

\[ \Delta C_{\ell}^2 = \frac{2C_{\ell}^2}{2\ell + 1}. \]  

(6.21)

### 6.3.3 Spherical Fourier Bessel Transform

The three dimensional field \( f \) can be represented by its SFB coefficients,

\[ f(r, \theta, \phi) = \sum_{n=1}^{\infty} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} f_{n\ell m} R_{n\ell}(r) Y_{\ell m}(\theta, \phi), \]  

(6.22)

where \( r, \theta, \phi \) are spherical polar coordinates and the SFB coefficients can be calculated from

\[ f_{n\ell m} = \int_0^{R_{\text{max}}} \int_0^\pi \int_0^{2\pi} f(r, \theta, \phi) R_{n\ell}(r) Y_{\ell m}^*(\theta, \phi) r^2 \sin(\theta) dr d\theta d\phi. \]  

(6.23)

Here we will assume that \( f \) is defined up to a radius \( R_{\text{max}} \) where we will use zero-value boundary conditions (i.e. \( j_\ell(k_{n\ell} R_{\text{max}}) = 0 \) where \( k_{n\ell} = 1/q_{n\ell} \) and \( q_{n\ell} \) is the location of the \( n^{\text{th}} \) root of the spherical Bessel function; Wang et al., 2009) where

\[ R_{n\ell}(r) = \frac{1}{\sqrt{N_{n\ell}}} j_\ell(k_{n\ell} r), \]  

(6.24)

\( Y_{\ell m} \) are spherical harmonics and \( N_{n\ell} \) is a normalisation constant defined as

\[ N_{n\ell} = \frac{R_{\text{max}}^3}{2} j_{\ell+1}^2(k_{n\ell} R_{\text{max}}). \]  

(6.25)

Following Leistedt et al. (2012) we calculate \( f_{n\ell m} \) in two stages first deriving the spherical harmonic coefficients for a redshift slice at radius \( r \),

\[ f_{\ell m}(r) = \int_0^\pi \int_0^{2\pi} f(r, \theta, \phi) Y_{\ell m}^*(\theta, \phi) \sin(\theta) d\theta d\phi, \]  

(6.26)

and then noting that

\[ f_{n\ell m} = \int_0^{R_{\text{max}}} f_{\ell m}(r) R_{n\ell}(r) r^2 dr. \]  

(6.27)
Figure 6.1: The comoving distances in the MICE lightcone is shown for a range of redshifts in comparison to the MICE simulation box. In this plot the observer is located at the origin and the dotted black lines represent the MICE simulation box which is repeated without rotations. The corresponding comoving distance for integer redshifts is indicated with coloured dashed lines; shown in pale green for lower redshifts and dark purple for higher redshifts. Redshift \( z = 1.4 \) is indicated with a full green line and shows the maximum redshift that can be reached in the MICE lightcones without repeating the simulation box along the line-of-sight.

This allows us to take advantage of the fast spherical harmonic decomposition of HEALPix (Górski et al., 2005b). Similarly the field can be reconstructed in two steps,

\[
    f_{\ell m}(r) = \sum_{n=1}^{\infty} f_{n \ell m} R_{n \ell}(r),
    \tag{6.28}
\]

\[
    f(r, \theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} f_{\ell m}(r) Y_{\ell m}(\theta, \phi).
    \tag{6.29}
\]

6.3.4 Reconstructing the Integrated Sachs-Wolfe

The ISW can be recovered in a number of ways. In simulations, as has been carried out by Cai et al. (2010), Watson et al. (2014) and Carbone et al. (2016), the potential is determined directly using Fast Fourier Transforms of the density field. The ISW is then determined by tracing the trajectory of light through the derivative of the potential; this allows the full computation of Equation 6.2 including both the predominant linear contribution from the ISW and the subdominant contribution from the non-linear Rees-Sciama effect.

A second, more widely used, approach is to reconstruct the ISW from the projected
FULL SKY ISW MAPS FOR MICE AND FLAGSHIP

$Z_{\text{max}} = 1.4$ vs $Z_{\text{max}} = 6$

$Z_{\text{max}} = 2$ vs. $Z_{\text{max}} = 6$

$-19.95 \, \mu K \quad 19.95 \, \mu K$

Figure 6.2: Displaying the boundary effects from calculations of the SFB transform with different $R_{\text{max}}$. The ISW is calculated for MICE for the redshift range $0 < z < 1.4$. On the left the SFB coefficients M14 are used with an $R_{\text{max}}$ corresponding to $z = 1.4$ and on the right the SFB coefficients M20 are used with an $R_{\text{max}}$ corresponding to $z = 2$. For both maps we subtract the ISW map created using the SFB coefficients M60 with an $R_{\text{max}}$ corresponding to $z = 6$. This final map is taken to be the ground truth and the deviations shown in the plots show the level of disagreement. Some large modes will not be measured when $R_{\text{max}}$ is small, this results in larger discrepancies as can be seen in the map on the left.

density field,

$$A_{\ell m}^{\text{ISW}} = \frac{C_{\ell}^{\text{ISW}-g}}{C_{\ell}^{gg}} A_{\ell m}^{g},$$

(6.30)

where $A_{\ell m}^{\text{ISW}}$ and $A_{\ell m}^{g}$ are the spherical harmonics of the ISW and projected density field respectively. This has commonly been used for generating ISW maps from data (see for example Manzotti & Dodelson, 2014) but this method does not take into consideration any radial information in its computation and will presumably be less accurate.

A third approach, used here, is to compute the ISW using the SFB coefficients (Shapiro et al., 2012), which is a natural basis for data provided in spherical polar coordinates. We begin by first computing the SFB coefficients of the density field. In this paper, the density contrast of the simulations (determined from the dark matter particles) are provided on full sky HEALPix maps binned in (unequal) slices in redshift. The radial sizes of these bins vary so we integrate by taking the density contrast for each bin to be a step function (described below in Equation 6.32). Since we are only interested in the ISW, which is predominantly at large scales, the maps are downgraded to $N_{\text{side}} = 256$. Each map is then converted into its spherical harmonics coefficients $d_{\ell m}$ and divided by the linear growth function at the central redshift $z_{\text{mid}}$.

$$\delta_{\ell m} = \frac{d_{\ell m}(z_{\text{mid}})}{D(z_{\text{mid}})}.$$  

(6.31)
The cosmological parameters (current matter density $\Omega_{m,0}$, current baryon density $\Omega_{b,0}$, spectral tilt $n_s$, the amplitude of the linear power spectrum at $8\ h^{-1}\text{Mpc}\sigma_8$, the ‘little’ Hubble constant $h$ where $H_0 = 100h\ \text{km\ s}^{-1}\ \text{Mpc}^{-1}$) and box size $L_{\text{box}}$ used for the MICE and Flagship simulations. In both simulations curvature is zero and therefore the current dark energy density is $\Omega_\Lambda = 1 - \Omega_m$.

<table>
<thead>
<tr>
<th>Simulation</th>
<th>$\Omega_{m,0}$</th>
<th>$\Omega_{b,0}$</th>
<th>$n_s$</th>
<th>$\sigma_8$</th>
<th>$h$</th>
<th>$L_{\text{box}}\ [h^{-1}\text{Mpc}]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MICE</td>
<td>0.25</td>
<td>0.044</td>
<td>0.95</td>
<td>0.8</td>
<td>0.7</td>
<td>3072</td>
</tr>
<tr>
<td>Flagship</td>
<td>0.319</td>
<td>0.049</td>
<td>0.96</td>
<td>0.83</td>
<td>0.67</td>
<td>3780</td>
</tr>
</tbody>
</table>

Table 6.2: A summary of the calculated spherical Fourier Bessel coefficients. The table provides: the maximum redshift $z_{\text{max}}$, comoving radius $R_{\text{max}}$, and maximum spherical Bessel modes $\ell$ and $n$ used to calculate the SFB coefficients in this paper.

<table>
<thead>
<tr>
<th>Name</th>
<th>Simulation</th>
<th>$z_{\text{max}}$</th>
<th>$R_{\text{max}}\ [h^{-1}\text{Mpc}]$</th>
<th>$\ell_{\text{max}}$</th>
<th>$n_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>M14</td>
<td>MICE</td>
<td>1.4</td>
<td>3038</td>
<td>304</td>
<td>97</td>
</tr>
<tr>
<td>M20</td>
<td>MICE</td>
<td>2</td>
<td>3798</td>
<td>380</td>
<td>121</td>
</tr>
<tr>
<td>M60</td>
<td>MICE</td>
<td>6</td>
<td>6141</td>
<td>615</td>
<td>196</td>
</tr>
<tr>
<td>F60</td>
<td>Flagship</td>
<td>6</td>
<td>5657</td>
<td>566</td>
<td>181</td>
</tr>
</tbody>
</table>

The SFB coefficients are then computed using Equation 6.27:

$$
\delta_{\ell m} = \sum_{i=0}^{R_{\text{max},i}<R_{\text{max}}} \int_{R_{\text{min},i}}^{R_{\text{max},i}} \delta_{\ell m} R_{\text{nl}}(r) r^2 dr,
$$

(6.32)

where the integral is evaluated for each redshift slice from $R_{\text{min},i}$ and $R_{\text{max},i}$ (which are the minimum and maximum comoving radius for shell $i$) up to the first shell for which $R_{\text{max},i} > R_{\text{max}}$ (in this case it is integrated up to $R_{\text{max}}$).

The ISW spherical harmonics are calculated from

$$
A_{\ell m}^{\text{ISW}} = \frac{3\Omega_{m,0}H_0^2}{c^2} \sum_{n=1}^{R_{\text{max}}} \frac{\delta_{\ell m}}{k_{n\ell}^2} \int_0^{R_{\text{max}}} D(r) H(r) [1 - f(r)] R_{\text{nl}}(r) dr.
$$

(6.33)

The spherical harmonics $A_{\ell m}^{\text{ISW}}$ can then be transformed into a map. In our analysis we consider only $k$ modes that are smaller than $k_{\text{max}} = 0.1\ h\text{Mpc}^{-1}$. Following Shapiro et al. (2012), only $\ell \leq R_{\text{max}}k_{\text{max}}$ and $n \leq R_{\text{max}}k_{\text{max}}/\pi$ need to be considered.

GenISW is a publicly available Python and C++ package written for this analysis to calculate the theoretical angular power spectra (written in Python) and to calculate the SFB and ISW spherical harmonics coefficients (written in C++ and parallelised with OpenMPI).

---

4https://github.com/knaidoo/GenISW
Figure 6.3: The full theoretical $C_\ell^g$ (dotted line) and the $k$-limited $C_\ell$ (dashed-dotted line) for the density auto-angular power spectra $C_{\ell}^{gg}$ (top), density-ISW cross-angular power spectra $C_{\ell}^{ISW-g}$ (middle) and ISW auto-angular power spectra $C_{\ell}^{ISW}$ (bottom). On the left we show the measured and theoretical angular power spectra in the range $0 < z < 1.4$ (light blue) and $1.4 < z < 5$ (dark blue) for the MICE simulation and on the right we show the measured and theoretical angular power spectra in the range $0 < z < 2.2$ (light green) and $2.2 < z < 5$ (dark green) for the Flagship simulations. For each auto- and cross-$C_\ell$ we compare the full $C_\ell$, $k$-limited $C_\ell$ and the measured $C_\ell$ (full lines) in sub-panels that display the cosmic variance uncertainties ($1\sigma$ envelopes indicated by the black dashed lines and $2\sigma$ envelopes indicated by the black dotted lines). In all cases we can see that the measured $C_{\ell}^{gg}$ follow the full theoretical $C_\ell$ while the $C_{\ell}^{ISW-g}$ and $C_{\ell}^{ISW}$ follow the $k$-limited theoretical $C_\ell$ and its departure from the full theoretical $C_\ell$ can be used as diagnostic for the ranges in $\ell$ that these ISW maps accurately reproduce the correct angular power spectra (this is displayed in Table 6.3).
In this Section we present ISW maps produced using the SFB transform and measure the auto- and cross-angular power spectra to explore the systematics introduced at low $\ell$, where the ISW probes modes below the simulation’s fundamental frequency, and at high $\ell$, where non-linear and large $k$ modes become relevant. A summary of the SFB coefficients calculated are presented in Table 6.2.

### 6.4 RESULTS

#### 6.4.1 MICE and Flagship Lightcone Simulations

We use the MICE (Crocc et al., 2015) and Flagship lightcone simulations, provided as a set of full sky density contrast maps covering concentric shells of unequal spacings in comoving distance. This is often referred to as the ‘onion’ universe and this representation of large lightcone simulations provides a substantial compression of the data (Fosalba et al., 2008). For MICE the density contrast of the simulation are split across 400 maps while for Flagship they are provided over 333 maps. In both cases a significant proportion of the maps cover the low redshift region. The cosmological parameters used by the two simulations are summarised in Table 6.1; for MICE they follow the WMAP best fit parameters (Dunkley et al., 2009) while for Flagship they follow Planck best fit parameters (Planck Collaboration et al., 2018). In Figure 6.1 the simulation box is shown in comparison to the corresponding comoving distance to a range of redshifts in the lightcone, indicating when repeated simulation boxes are required.
6.4.2 Calculated Spherical Fourier Bessel Coefficients

The SFB coefficients calculated and used in this paper are summarised in Table 6.2. For MICE we calculate the SFB coefficients for three different $R_{\text{max}}$. The naming convention for the SFB coefficients uses an ‘M’ for MICE and an ‘F’ for Flagship, this is followed by the redshift corresponding to $R_{\text{max}}$ multiplied by 10. The top two (M14 and M20) in Table 6.2 are used to show the large angular effects of choosing the $R_{\text{max}}$ boundary. The SFB coefficients M60 and F60 are the final SFB coefficients used to construct the ISW maps. These coefficients are computed up to redshift $z = 6$ which is chosen as this limits the repetition of the simulation box to a maximum of 2 along the LOS. Ideally no repetition would be used however this would limit the analysis to below $z < 1.4$ for MICE and $z < 2.2$ for Flagship and would not allow for high redshift studies of the ISW.

6.4.3 Suppressed Power at Low-$\ell$

We test the effect of boundaries used to calculate the SFB coefficients by comparing the measured ISW calculated between $0 < z < 1.4$ from the M14 and M20 coefficients to those measured from the M60 coefficients; see Figure 6.2. The differences are limited to large angular features, demonstrating the importance of calculating the SFB coefficients with a larger $R_{\text{max}}$ to mitigate this effect as much as possible. For real data this effect will be most pronounced since $R_{\text{max}}$ is determined by the redshift range of the data. In this analysis we construct the ISW between a range of $0 < z < 5$ and to mitigate the boundary effect the SFB coefficients are computed with an $R_{\text{max}}$ corresponding to $z = 6$. This means the simulation boxes will be repeated at most twice along the LOS which is necessary to produce high redshift ISW maps.

Since the ISW is being constructed from a finite simulation we are only sensitive to Fourier modes present in the simulation, i.e. frequencies above the simulation’s fundamental frequency. For this reason our ISW maps have suppressed power at low $\ell$ where Fourier modes $k < k_F$ are important. We can calculate the theoretical effect by integrating Equations 6.16 and 6.19 only in the range $k_F < k < k_{\text{max}}$ (which we will refer to as the $k$-limited $C_\ell$). In Figures 6.3 and 6.4 the $k$-limited $C_\ell$ are indicated by the dotted-dashed coloured lines and the full $C_\ell$ are indicated by the dotted coloured lines.

6.4.4 Suppressed Power at High-$\ell$

We consider only the linear regime, i.e. Fourier modes below $k_{\text{max}} = 0.1\ h\text{Mpc}^{-2}$, above thus the equations derived are no longer valid and would start to include the Rees-Sciama effect. This effect, although sub-dominant becomes relevant at large $\ell$. For this reason we see that the angular power spectra measured from the maps (shown in Figure 6.3) dip below the expected ISW for all Fourier modes (dotted lines) and follow closely the dashed-dotted lines (which only include $k_F < k < k_{\text{max}}$). This difference can be used as a diagnostic
Table 6.3: A summary of the ISW maps made from the MICE and Flagship simulations displaying the names, simulation used, redshift range and effective $\ell$ range (i.e. where the theoretical $C_\ell$ measured in the range $k_F < k < k_{\text{max}}$ are within $1\sigma$ of the full $C_\ell$).

<table>
<thead>
<tr>
<th>Name</th>
<th>Simulation</th>
<th>$z_{\text{min}}$</th>
<th>$z_{\text{max}}$</th>
<th>Eff. $\ell_{\text{min}}$</th>
<th>Eff. $\ell_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MICE_00_14</td>
<td>MICE</td>
<td>0</td>
<td>1.4</td>
<td>3</td>
<td>132</td>
</tr>
<tr>
<td>MICE_14_50</td>
<td>MICE</td>
<td>1.4</td>
<td>5</td>
<td>8</td>
<td>315</td>
</tr>
<tr>
<td>MICE_00_10</td>
<td>MICE</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>115</td>
</tr>
<tr>
<td>MICE_10_20</td>
<td>MICE</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>248</td>
</tr>
<tr>
<td>MICE_20_30</td>
<td>MICE</td>
<td>2</td>
<td>3</td>
<td>8</td>
<td>382</td>
</tr>
<tr>
<td>MICE_30_40</td>
<td>MICE</td>
<td>3</td>
<td>4</td>
<td>9</td>
<td>471</td>
</tr>
<tr>
<td>MICE_40_50</td>
<td>MICE</td>
<td>4</td>
<td>5</td>
<td>10</td>
<td>532</td>
</tr>
<tr>
<td>Flagship_00_22</td>
<td>Flagship</td>
<td>0</td>
<td>2.2</td>
<td>2</td>
<td>134</td>
</tr>
<tr>
<td>Flagship_22_50</td>
<td>Flagship</td>
<td>2.2</td>
<td>5</td>
<td>7</td>
<td>379</td>
</tr>
<tr>
<td>Flagship_00_10</td>
<td>Flagship</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>106</td>
</tr>
<tr>
<td>Flagship_10_20</td>
<td>Flagship</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>237</td>
</tr>
<tr>
<td>Flagship_20_30</td>
<td>Flagship</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>359</td>
</tr>
<tr>
<td>Flagship_30_40</td>
<td>Flagship</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>438</td>
</tr>
<tr>
<td>Flagship_40_50</td>
<td>Flagship</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>493</td>
</tr>
</tbody>
</table>

for when non-linear effects become important. For the MICE ISW we see that the $C_{\text{ISW}}^\ell$ for the redshift range $0 < z < 1.4$ are linear for $\ell \lesssim 130$ and for the redshift range $1.4 < z < 5$ are linear for $\ell \lesssim 315$. Similarly, for the Flagship ISW the $C_{\text{ISW}}^\ell$ for the redshift range $0 < z < 2.2$ are linear for $\ell \lesssim 130$ and for the redshift range $2.2 < z < 5$ are linear for $\ell \lesssim 380$ (see Table 6.3 for precise ranges).

6.4.5 Amplified Power at Mid-$\ell$ from Periodic Boxes in the Lightcone

In Figure 6.4 we show the $C_{\text{ISW}}^\ell$ from both the MICE and Flagship simulations for contributions from the redshift range $0 < z < 5$. In both cases the simulation box has to be repeated along the LOS to span the required redshift range. This means features in the density field are measured more than once and are not independent. This causes the $C_{\text{ISW}}^\ell$ to have an extra power at mid-$\ell$ as can be seen in Figure 6.4. For this reason we would advise against using the maps created with repetitions of the simulation box in the LOS (i.e. maps for $0 < z < 5$) and to instead either use the ISW maps for MICE in the redshift ranges $0 < z < 1.4$ or $1.4 < z < 5$ and for Flagship in the redshift ranges $0 < z < 2.2$ and $2.2 < z < 5$. Alternatively one can use the maps created with $\Delta z = 1$. 
Figure 6.5: The ISW maps for MICE lightcone simulations for redshift ranges which span a comoving distance $< L_{\text{box}}$. We show the contribution from the redshift range $0 < z < 1.4$ at the top and from the redshift range $1.4 < z < 5$ at the bottom. The majority of the contribution to the ISW occurs at low redshift.
Figure 6.6: The ISW maps for Flagship lightcone simulations for redshift ranges which span a comoving distance $< L_{\text{box}}$. We show the contribution from the redshift range $0 < z < 2.2$ at the top and from the redshift range $2.2 < z < 5$ at the bottom. The majority of the contribution to the ISW occurs at low redshift.
6.4.6 Summary of Integrated Sachs-Wolfe Maps

Table 6.3 shows a summary of the ISW maps. For both simulations we produce maps measuring the ISW contribution up to redshift \( z = 5 \) and then break these up into redshift ranges where no repetition of the simulation box is needed in the LOS direction. For the MICE simulation this means limiting the redshift range to \( 0 < z < 1.4 \) and for the Flagship simulation limiting the redshift range to \( 0 < z < 2.2 \). The contribution beyond this is then measured separately (for MICE \( 1.4 < z < 5 \) and for Flagship \( 2.2 < z < 5 \)). These maps are shown in Figure 6.5 and 6.6, where the majority of the contribution occurs at low redshift. In addition we measure the ISW within redshift bins of \( \Delta z = 1 \). Table 6.3 shows the precise ranges in \( \ell \) for which the simulations produce unsuppressed \( C_\ell \) shown in the effective (Eff.) \( \ell_{\text{min}} \) and \( \ell_{\text{max}} \) columns. All maps are provided as HEALPix maps in FITS format\(^5\).

6.4.7 Missing Patches in Flagship

In producing these maps we noticed a very large cold imprint in the lower right region of the Flagship ISW map for the redshift range \( 0 < z < 2.2 \). This feature is particularly strong in the redshift range \( 1 < z < 2 \) where it is by far the most dominant feature but also indicated another lesser but also strong feature on the other side of the map. We were able to determine that the feature was caused by missing patches in a few of the Flagship density maps (maps 115 to 119) roughly at a redshift \( z \sim 1.65 \). These missing patches were registered as enormous and very empty voids and therefore produced enormous cold imprints in our maps. Rather curiously this topological defect cannot be seen from the \( C_\ell \), which perhaps shows the importance of looking at simulated data in many different ways. Once these defects are corrected the final ISW maps for Flagship will change but the analysis presented in this paper will remain the same as this feature does not effect the results or the pipeline discussed.

6.5 CONCLUSIONS

We generate full sky ISW maps for the MICE and Flagship lightcone simulations up to \( z = 5 \). The ISW maps are produced using SFB transforms following the equations derived by Shapiro et al. (2012) which we re-derive with zero-value boundary conditions (following the methods outlined in Sections 6.3).

The SFB coefficients are calculated in the linear regime (i.e. considering only Fourier modes \( k < k_{\text{max}} \) where \( k_{\text{max}} = 0.1 \)) as this allows us to make the linear growth assumption, i.e. that density perturbations grow according to the linear growth factor, and this in turn simplifies the computation of the time derivative of the gravitational potential.

\(^5\)Maps and ancillary data will be made public once the Chapter is submitted as a paper.
(Equation 6.13). The simulations are given as a series of density contrast maps provided in concentric shells. The SFB coefficients are constructed by first converting these maps into spherical harmonics and then computing the SFB coefficients by integrating across the shells up to the comoving distance corresponding to $z = 6$. The SFB coefficients are then used to calculate the ISW for an arbitrary set of redshift ranges between $0 < z < 6$ (using Equation 6.33). In this analysis we limit our computation of the ISW to $z < 5$ since the predominant contribution to the ISW occurs at low redshift and the galaxy redshift surveys to which the MICE and Flagship simulations will be used are predominantly at $z < 2$.

The final maps produced have several systematic issues which are explained in Section 6.4 and which we highlight here. Firstly due to the limited size of the simulation and the redshift range probed the simulation box has to be repeated along the LOS. This means density perturbations are sampled more than once and are not independent as we would expect in real data. This amplifies features in the ISW at mid-$\ell$ which is demonstrated in Figure 6.4. For this reason we split the ISW maps into redshift ranges that do not require repetition of the boxes along the LOS. For both sets of simulations we provide maps with redshift bins of $\Delta z = 1$ and in addition divide the ISW into the ranges $0 < z < 1.4$ and $1.4 < z < 5$ for MICE and $0 < z < 2.2$ and $2.2 < z < 5$ for Flagship. These sets of ISW maps produce the correct angular power spectra (see Figure 6.3) but suffer from other sources of systematics. There are two effects which both result in a suppression of power. The first is caused by the finite size of the simulation which means we only have Fourier modes larger than the fundamental frequency and the second is caused by enforcing the linear regime which means setting a maximum on the considered Fourier modes (i.e. $k_{\text{max}} = 0.1\ h\text{Mpc}^{-1}$). The combined effect of these two conditions means the angular power spectra is effectively limited to considering only Fourier modes between $k_F < k < k_{\text{max}}$. By computing the theoretical $C_\ell$ with all the Fourier modes and comparing to the $k$-limited $C_\ell$ (i.e. $k_F < k < k_{\text{max}}$) we are able to determine within which ranges in $\ell$ the ISW is correctly reproduced with the right amplitude. In Table 6.3 we summarise the ISW maps produced and the effective ranges in $\ell$ ($\text{Eff. } \ell_{\text{min}} < \ell < \text{Eff. } \ell_{\text{max}}$) where the modes are not suppressed.

These maps (which are public) will provide an important test of CMB and LSS cross-correlation studies from future galaxy redshift surveys such as the Dark Energy Survey, Dark Energy Spectroscopic Instrument, Euclid, the Nancy Grace Roman Space Telescope and the Rubin Observatory Legacy Survey of Space and Time which will map the distribution of galaxies to higher redshifts and across larger areas of the sky. Furthermore,
the maps will provide an important cross-check for ISW reconstruction techniques. The pipeline used to calculate all the data products in this paper can be found in the package GENISW[1]. This contains a PYTHON module pyGENISW which was used to calculate the theoretical $C_\ell$ and a C++ package which was used to calculate the SFB coefficients and the ISW spherical harmonics. This package can be used to construct the ISW from future lightcone simulations.

[1]https://github.com/knaidoo29/GenISW
“Every great magic trick consists of three parts or acts. The first part is called “The Pledge”. The magician shows you something ordinary: a deck of cards, a bird or a man. He shows you this object. Perhaps he asks you to inspect it to see if it is indeed real, unaltered, normal. But of course... it probably isn’t. The second act is called “The Turn”. The magician takes the ordinary something and makes it do something extraordinary. Now you’re looking for the secret... but you won’t find it, because of course you’re not really looking. You don’t really want to know. You want to be fooled. But you wouldn’t clap yet. Because making something disappear isn’t enough; you have to bring it back. That’s why every magic trick has a third act, the hardest part, the part we call “The Prestige”.”

7.1 THESIS SUMMARY

Galaxy redshift surveys will play a critical role in the advancement of cosmology in the next decade. These surveys will sample the distribution of galaxies to lower masses and to greater redshifts. Two-point statistics have proven to be immensely powerful in constraining cosmological parameters but is limited and cannot incorporate the complicated structure of the cosmic web for parameter analysis.

In the first part of this thesis I develop the minimum spanning tree (MST), a graph-based algorithm previously used for filament finding, to capture the cosmic web for cosmology.

In Chapter 2 the MST statistics are introduced and applied to a range of different simulations. In the first part the distribution of galaxies from an $N$-body simulation are compared to a Lévy-Flight distribution with roughly equal two-point statistics. From this the MST is shown to be sensitive to features in the cosmic web, such as voids and clusters, proving it is sensitive to features beyond two-point statistics. Using $N$-body simulations we then address what the MST is measuring and how the MST measured from dark matter particles changes when applied to haloes. Lastly, using haloes from COLA simulations the MST is shown to improve constraints on $\Omega_m$ and $A_s$ by $\sim 10\%$. We also show the MST provides additional information not present in either the power spectrum or bispectrum.

In Chapter 3 the Quijote simulations are used to calculate the covariance matrix for the MST which is shown to contain non-diagonal features. Fisher matrices are then calculated for the MST and power spectrum in real and redshift space for a $\nu$LCDM model. The MST is shown to be highly sensitive to the sum of neutrino masses $M_{\nu}$, providing a factor of $\sim 2$ improvement on the constraints on $M_{\nu}$ when compared to the power spectrum. This sensitivity to the sum of neutrino masses appears to stem from the MST’s sensitivity to small scale clustering where neutrino free-streaming effects become important. Furthermore, by combining the MST with measurements of the power spectrum the constraints on $h$, $n_s$ and $\Omega_b$ are improved by a factor of $\sim 2$ while the constraints on $\Omega_m$ and $\sigma_8$ are dominated by the power spectrum and the constraints on $M_{\nu}$ are dominated by the MST. When these measurements are combined across three different redshifts, the MST
CONCLUSIONS AND FUTURE WORK

constraints on all parameters are competitive with the power spectrum (including $\Omega_m$ and $\sigma_8$) and still dominate constraints on $M_\nu$.

Critical to the application of the MST in cosmology is determining how best to remove its sensitivity to small scales and how to incorporate galaxy weights used in cosmology to correct for survey systematics. In Chapters 3 and 4 the sensitivity to small scales is removed by grouping galaxies separated by a distance $r < r_{\text{min}}$. The MST is then constructed on a catalogue of nodes, containing groups and ungrouped galaxies. In Chapter 4 a method for incorporating galaxy weights to the MST is devised. This is achieved by averaging realisations where binary weights are assigned probabilistically based on the galaxy’s non-integer weight.

In Chapter 4 the MST is applied to real galaxies from the Baryon Oscillation Spectroscopic Survey (BOSS) and compared to the $\Lambda$CDM BOSS mocks. The MST constructed from real galaxies is found to be consistent with those constructed from $\Lambda$CDM mocks; the one exception is the distribution of edges for the LOWZ NGC region which is discrepant at the $\sim 2.5\sigma$ level. The source of this discrepancy remains unclear and will be explored in future work, but it is likely to be the result of a redshift dependent offset in the density. Lastly, we demonstrate that the noise in the MST statistics can be approximated with a Gaussian distribution, once the tails of the MST statistics have been removed. Future parameter studies can therefore rely on a simple Gaussian likelihood for the MST statistics.

The MST is constructed and its statistics measured using the public PYTHON package MiSTree (Naidoo, 2019, Appendix A.2) developed during and for the studies conducted in Chapters 2, 3 and 4. The package is designed to construct the MST in a range of different coordinate systems commonly found in astronomy. This will allow those using the package to concentrate on the physical interpretation of the constructed graph and will facilitate future MST studies inside and outside astronomy. An example of such a study is given in (Liu et al., 2020, a study on the role of magnetic fields in the early stages of star formation) where the MST and MiSTree are used to study the structure of molecular condensates measured by the Atacama Large Millimeter Array.

The second part of this thesis is focused on the integrated Sachs-Wolfe (ISW) effect: (1) on the role it plays in the cosmic microwave background (CMB) Cold Spot anomaly and (2) in deriving ISW maps for the MICE and Flagship lightcone simulations.

In Chapter 5 Gaussian maps of the ISW are produced. The maps are used to demonstrate that the Cold Spot feature is often correlated with a negative contribution from the ISW. As such, finding voids in the line-of-sight of the Cold Spot is to be expected in a $\Lambda$CDM model. Furthermore, we show there is a correlation between the size of the mask and the significance of the Cold Spot anomaly; if this is taken into consideration the Cold Spot anomaly is found to be significant to $< 2\sigma$.

In Chapter 6 ISW maps for the MICE and Flagship simulations are produced using spherical Fourier Bessel transforms. The maps are produced using HEALPix maps of the
density distribution provided in concentric shells. Since the ISW is dominated by linear effects, caused by the decay of gravitational potentials, we ignore the non-linear Rees-Sciama effect and compute the effect only in the linear regime. This means the ISW can be computed solely from lightcone simulations of the density distribution without needing regular snapshot outputs of the density, gravitational potential and velocity fields. The maps are made public to enable future cross-correlation studies of the CMB and large scale structure and will provide a test for ISW reconstruction techniques. The pipeline is also made public and will facilitate the creation of ISW maps from a broader range of simulations exploring a larger parameter space.

7.2 COSMOLOGY IN THE NEXT DECADE

Future galaxy surveys such as Dark Energy Spectroscopic Instrument\(^1\) (DESI), Euclid\(^2\), Nancy Grace Roman Space Telescope\(^3\) and Rubin Observatory Legacy Survey of Space and Time\(^4\) (LSST) will map the large scale distribution of galaxies to higher redshifts and at higher densities (see Figure 1.3 and 1.4 for more information on some of these upcoming surveys). These surveys will be pivotal in providing tighter constraints on cosmological parameters and will aim to do the following:

1. Reduce the uncertainties on the dark energy equation of state by an order of magnitude. In the process these surveys may reveal dark energy to be more than just a cosmological constant.

2. Determine the sum of neutrino masses \(M_\nu\). Currently measurements from Planck and baryonic acoustic oscillations from BOSS galaxies and the Lyman-\(\alpha\) forest have placed upper limits of \(M_\nu \lesssim 0.12\,\text{eV}\). This is tantalizingly close to the theoretical lower limit found by neutrino oscillation experiments of \(M_\nu \geq 0.06\,\text{eV}\). Determining a non-zero \(M_\nu\) is a discovery which we expect future surveys such as DESI to make.

In addition to mapping the 3D distribution of galaxies many of these surveys will also carry out weak lensing measurements and provide increasingly more accurate mass maps. Combining clustering measurements and weak lensing will enable redshift surveys to first be competitive and later surpass constraints from CMB measurements.

Lastly, the discoveries that will be made from ground based physics experiments may play a pivotal role in our understanding of cosmology. One result that would certainly inform cosmological studies is the discovery of the particle or particles responsible for

\(^{1}\)http://desi.lbl.gov/
\(^{2}\)http://www.euclid-ec.org/
\(^{3}\)https://roman.gsfc.nasa.gov
\(^{4}\)https://www.lsst.org/
dark matter. Such a discovery could come from particle accelerator experiments or dark matter detectors and would represent a major paradigm shift for fundamental physics.

7.3 FUTURE WORK

7.3.1 Parameter Constraints from the Minimum Spanning Tree

The MST work presented in this thesis has been designed with the intention of eventually being used to provide constraints on cosmological parameters. In this section I will briefly outline the steps required to fulfil this goal and the wider role the MST could play in understanding machine learning algorithms.

In Chapter 4 the MST measured on BOSS galaxies was shown to be consistent with $\Lambda$CDM PATCHY mocks. This means these mocks can be used to construct covariance matrices for parameter studies. The missing ingredient currently preventing this study is an emulator which can interpolate the MST statistics from simulations with different cosmological parameters. This is further complicated by the MST’s sensitivity to density and as such any emulator would need to be built using a halo occupation distribution prescription whose parameters would be free or fixed to match the BOSS density. In collaboration with Elena Massara, the Quijote simulations are being extended to higher resolutions; each simulation will consist of $1024^3$ dark matter particles and $1024^3$ neutrino particles in a box of length $L_{\text{box}} = 1 \ h^{-1}\text{Gpc}$. The new suite will contain 2000 simulations spanning a Latin hypercube in a $\nu\Lambda$CDM parameter space. These simulations will then be used to construct a MST emulator which will be used to constrain cosmological parameters from BOSS. The accuracy of the emulator can be tested against other large $N$-body suites such as Abacus (Garrison et al., 2018), Aemulus (DeRose et al., 2019) and MassiveNuS (Liu et al., 2018). Assuming the emulator is consistent, these other simulations could be used to expand the parameter space of the emulator to explore (for example) models with dynamical dark energy. The emulator and posterior estimation will likely follow the analysis performed by Rogers et al. (2019), i.e. using Gaussian Processes (GP) to emulate the MST statistics. A similar likelihood-free approach (Leclercq, 2018) could also be explored. This has the benefit of allowing us to determine where new simulations should be run to improve the accuracy of the posterior distribution.

In Chapter 4 the mocks and real data follow the BOSS redshift selection function $n(z)$ which is a complicated flux- and colour-limited galaxy sample. This means the MST is constructed from a non-uniform density distribution. Although this effect is mitigated by comparing like-for-like this is less than ideal as the MST graph constructed will be a product of both the $n(z)$ and underlying cosmic web distribution. In future work we will extract volume limited samples to remove any sensitivity to the $n(z)$. This could be carried out at several densities; at high density this could be used to explore small scales (by
sacrificing regions where the \( n(z) \) is low) and at low density this could be used to explore large scales (by sacrificing regions where the \( n(z) \) is high).

The MST could be applied to other spectroscopic surveys such as DESI, Euclid, Prime Focus Spectrograph\(^5\) (PFS), and the Roman Space Telescope, and even extended to photometric surveys such as Dark Energy Survey\(^6\) (DES), Euclid and LSST. However, the photometric redshift estimations would force the analysis to be performed in tomographic redshift slices. It is unclear whether the MST will still be sensitive to the cosmic web in this scenario or whether chance and random alignments will dominate the graph.

Furthermore, future galaxy surveys will provide increasingly accurate mass maps determined from weak lensing measurements. A study of the correlations between mass maps and the MST may help to reveal the MST’s accuracy in tracing the cosmic web. For mass map reconstructions the MST could be used as an additional source of information to improve accuracy.

The main motivation in using the MST ahead of other algorithms is its sensitivity to the cosmic web which is displayed by its previous use as a cosmic web classifier. This sensitivity is crucial and currently lacking from cosmological studies. However, there are other statistics that could be used and could be sensitive to information in different ways. In future we will explore its connection to other statistics, including higher order moments and Minkowski functionals (Minkowski, 1903), to determine whether there are different parameter degeneracies that can be exploited for parameter studies.

Lastly, the MST could play an interesting role in the growing effort to extract as much information as possible from large scale structure. Typically this effort has been motivated by applying artificial intelligence, machine learning and deep learning (AI/ML/DL) algorithms to simulations. Eventually, DL algorithms (Lecun et al., 2015) will be at a stage where parameters can be determined from the galaxy catalogue itself without requiring summary statistics. While these techniques are very sensitive to the non-linear cosmic web distribution and have enormous potential, it is difficult to interpret what is being learnt and therefore to develop physical insight. On the other hand the MST is very simple and it is more intuitive to make physical interpretations. As such the MST can be thought of as an intermediate statistics that is sensitive to the cosmic web but that remains physically interpretable. By studying the degeneracies between the MST and AI/ML/DL algorithms we could begin to understand what these algorithms are learning.

7.3.2 Exploring the Integrated Sachs-Wolfe for Different Cosmologies

Studies of the stacked imprint of the ISW from large voids have found the signal to be much larger than expected (Granett et al., 2008; Kovács et al., 2017). Coupled with the

\(^5\)https://pfs.ipmu.jp/index.html
\(^6\)http://www.darkenergysurvey.org
CONCLUSIONS AND FUTURE WORK

$H_0$ and $\sigma_8$ tension it is tempting to think these results are small hints for physics beyond $\Lambda$CDM but the significance of these tensions are still widely debated.

As of yet there are few $N$-body simulations which are produced with galaxy catalogues and corresponding ISW maps. Currently the simulations available are limited to either a fiducial $\Lambda$CDM cosmology (Cai et al., 2010; Watson et al., 2014) or, in the case of the DEMNUni simulations (Carbone et al., 2016), limited to a few cases with varying neutrino masses. A full exploration of the allowed parameter space has yet to be carried out. A large reason for this is the computational expense of this calculation. In earlier studies (Cai et al., 2010; Watson et al., 2014; Carbone et al., 2016) the ISW is calculated using regular snapshot outputs of the density, gravitational potential and velocity fields. While this allows the study of the non-linear ISW (i.e. the Rees-Sciama effect) we know this effect is sub-dominant and the computational cost does not appear to be justified if we want to greatly expand the parameter space.

In Chapter 6 we explain how the linear ISW can be calculated from solely the density distribution from lightcone simulations. This makes unnecessary the computationally expensive snapshot outputs of the density, gravitational potential and velocity fields. The ISW can then be calculated from existing simulations, greatly expanding the parameter space. This will help us to determine whether different cosmologies, especially those with dynamical dark energy, can explain the excess ISW imprint from large voids. These studies could be applied to the current redshift surveys: BOSS and DES, and the future galaxy surveys: DESI, Euclid and LSST.

The cosmic web and large scale structure will be a key frontier for advancements in cosmology in the next decade. In this thesis we demonstrate how to extract cosmic web information from the distribution of galaxies and develop products and pipelines to study the imprint of large scale structure on the CMB. These methods and the products produced will be particularly useful for the study of current and future galaxy redshift surveys.
“**Pumbaa**: Timon, ever wonder what those sparkly dots are up there?

**Timon**: Pumbaa, I don’t wonder, I know.

**Pumbaa**: Oh. What are they?

**Timon**: They’re fireflies. Fireflies that, uh... got stuck up in that big bluish-black thing.

**Pumbaa**: Oh, gee. I always thought they were balls of gas burning billions of miles away.

**Timon**: Pumbaa, with you, everything’s gas. ”

– The Lion King (1994)
**Appendix**

A./one.lf

**A.1 Gaussian Process Interpolation**

We will be modelling data vectors following a method similar to that of Rogers et al. (2019) and Bird et al. (2019) in which they emulated the 1D flux power spectrum of the Lyman-α forest using Gaussian Processes (GP). In this section we provide a brief introduction to GPs and outline their usage in Chapter 2. A comprehensive overview of GPs and their applications can be found in Rasmussen & Williams (2006), while an overview of their implementations for vectors can be found in Alvarez et al. (2011).

**A.1.1 Introduction**

GPs are a non-parametric kernel-based regression and interpolation method. In GPs we model the desired function \( f(x) \) as a stochastic process with a prior probability over all parametric functions. For a finite input data set \( X = \{x_1, ..., x_n\} \), this can be modelled as a multivariate Gaussian,

\[
\text{GP} = \mathcal{N} (m(X), K(X, X')) ,
\]  

(A.1.1)

with mean \( m(X) \) and covariance \( K(X, X') \). Given training data \( Y_1 \) at \( X_1 \), we model the posterior of the function \( f(x) \) at new positions \( X_2 \) as a multivariate Gaussian,

\[
P (Y_2|X_1, Y_1, X_2) = \mathcal{N} (\mu_{2|1}, S_{2|1}) ,
\]  

(A.1.2)

with mean \( \mu_{2|1} \) and covariance \( S_{2|1} \). Assuming that both \( Y_1 \) and \( Y_2 \) are drawn from the same multivariate Gaussian, as our prior on the function indicates (see Equation A.1.1), we can write the relation

\[
\begin{bmatrix}
  Y_1 \\
  Y_2
\end{bmatrix}
\sim \mathcal{N}
\left(
\begin{bmatrix}
  \mu_1 \\
  \mu_2
\end{bmatrix}
, 
\begin{bmatrix}
  K_{11} + l \sigma_n^2 & K_{12} \\
  K_{21} & K_{22}
\end{bmatrix}
\right),
\]  

(A.1.3)

where \( l \) is the identity matrix and \( \sigma_n \) is the standard deviation of the training data \( Y_1 \) (which is either known or fitted later). Thus assuming the mean function is zero we arrive
APPENDIX A.1

at the predicted mean and covariance,

\[
\begin{align*}
\mu_{2|1} &= \left( (K_{11} + I_n \sigma_n^2)^{-1} K_{12} \right)^\top Y_1, \\
S_{2|1} &= K_{22} - \left( (K_{11} + I_n \sigma_n^2)^{-1} K_{12} \right)^\top K_{12},
\end{align*}
\]  

(A.1.4) \hspace{1cm} (A.1.5)

where the dependence on \( K_{21} \) has been removed due to the symmetry \( K_{12} = K_{21}^\top \). Note that in practice we determine the GPs mean and standard deviation at a single new position and thus the standard deviation is simply a scalar – this means that \( K_{12} \) and \( K_{21} \) reduce to vectors and \( K_{22} \) to a scalar.

A.1.2 Kernel

GPs use kernels to weight the interdependency of points in parameter space. In our model we use a Gaussian kernel,

\[
\kappa(\theta_i, \theta_j) = \sigma_{GP}^2 \exp \left(-\frac{r^2}{2}\right).
\]  

(A.1.6)

Here,

\[
r = \frac{|\theta_{i,1} - \theta_{j,1}|^2}{2l_{GP,1}^2} + \frac{|\theta_{i,2} - \theta_{j,2}|^2}{2l_{GP,2}^2} + \frac{|\theta_{i,3} - \theta_{j,3}|^2}{2l_{GP,3}^2};
\]  

(A.1.7)

\( \sigma_{GP}, l_{GP,1}, l_{GP,2}, \) and \( l_{GP,3} \) are GP hyperparameters to be fitted with independent scale terms for each axis in the parameter space; and \( \theta = [10^9 A_s, \Omega_m, m_\nu] \). The covariance matrix \( K \) is then defined to have elements

\[
(K)_{ij} = \kappa(\theta_i, \theta_j) + \sigma_n^2 \delta_k(\theta_i, \theta_j),
\]  

(A.1.8)

with an additional noise term \( \sigma_n \).

A.1.3 Hyperparameter Optimization

The hyperparameters \( \phi = [\sigma_{GP}, l_{GP,1}, l_{GP,2}, l_{GP,3}] \) are optimized by maximising the likelihood function

\[
\mathcal{L}(D | \theta, \phi) = \sum_i^n \mathcal{L}(d_i | \theta, \phi),
\]  

(A.1.9)

where \( D \) are the ensemble of training data vectors, \( d_i \) is an element of a specific data vector and

\[
\mathcal{L}(d_i | \theta, \phi) = -\frac{1}{2} d_i^\top K^{-1} d_i - \frac{1}{2} \log |K| - \frac{n}{2} \log 2\pi.
\]  

(A.1.10)
Figure A.1.1: The positions in parameter space of simulations (grid, validation, fiducial and mocks) used in section 2.7. Note that for the Grid simulations each cross marks the point of five simulations.

A.1.4 Implementation and Validation

The GP hyperparameters are trained on the measurements of $P(k)$, the maximally compressed $B(k_1, k_2, k_3)$ and the MST statistics $d$, $l$, $b$, and $s$ (see section 2.7.2 for further details on these measurements) from the Grid simulations separately. In Figure A.1.1 we show the placement of the grid, fiducial, mock and validation (used only in this section) simulations in parameter space. To test that our GP interpolation is emulating the statistics accurately we calculate the residuals between the grid simulations (using the mean of five realisations made at each point in parameter space),

$$
\sigma_{\text{Residual}} = \frac{d - \mu_{\text{GP}}}{\sqrt{\sigma^2_{\text{Fiducial}} + \sigma^2_{\text{GP}}}}
$$

(A.1.11)

where $\mu_{\text{GP}}$ and $\sigma_{\text{GP}}$ are the GP mean and standard deviation evaluated at the same points in parameter space as $d$. We plot histograms of the residuals for the grid data vectors in Figure A.1.2 shown in orange. Notice that since the grid simulations are the mean of five simulations the distribution follows a Gaussian with mean 0 and standard deviation $1/\sqrt{5}$ (illustrated by the black dotted line). Furthermore to test that our GPs interpolation produces a good fit to simulations not present in the training data, we generate 25 new simulations (called the validation simulations) with randomly drawn cosmological parameters (shown in Figure A.1.1). We then again compare the residuals to that of our GPs interpola-
Figure A.1.2: The residuals between the statistics of $P(k)$ (top left), maximally compressed $B(k_1, k_2, k_3)$ (top right), MST degree (middle left), edge length (middle right), branch length (bottom left) and branch shape (bottom right) for the grid (shown by the orange histograms) and validation (shown by the blue histograms) simulations calculated from Equation A.1.11. Since the grid data vectors are the mean of five realizations the residuals are expected to follow a normal distribution of $\mathcal{N}(0, 1/\sqrt{5})$ (shown by the dotted black line), whilst the validation data vector are expected to follow a normal distribution of $\mathcal{N}(0, 1)$. We see that for most of the statistics the agreement is fairly good, with the exception of $B(k_1, k_2, k_3)$ which shows more spread than is expected.

We also find a good agreement (with the exception of $B(k_1, k_2, k_3)$) with a Gaussian with mean 0 and standard deviation 1 illustrated by the black full lines.
A.2 MISTREE: A PYTHON PACKAGE FOR CONSTRUCTING AND ANALYSING MINIMUM SPANNING TREES

A.2.1 Summary

The minimum spanning tree (MST), a graph constructed from a distribution of points, draws lines between pairs of points so that all points are linked in a single skeletal structure that contains no loops and has minimal total edge length. The MST has been used in a broad range of scientific fields such as particle physics (to distinguish classes of events in collider collisions, see Rainbolt & Schmitt (2017)), in astronomy (to detect mass segregation in star clusters, see Allison et al. (2009)) and cosmology (to search for filaments in the cosmic web, see Alpaslan et al. (2014)). Its success in these fields has been driven by its sensitivity to the spatial distribution of points and the patterns within. MiSTree, a public Python package, allows a user to construct the MST in a variety of coordinates systems, including Celestial coordinates used in astronomy. The package enables the MST to be constructed quickly by initially using a $k$-nearest neighbour graph ($k$NN, rather than a matrix of pairwise distances) which is then fed to Kruskal’s algorithm (Kruskal, 1956) to construct the MST. MiSTree enables a user to measure the statistics of the MST and provides classes for binning the MST statistics (into histograms) and plotting the distributions. Applying the MST will enable the inclusion of high-order statistics information from the cosmic web which can provide additional information to improve cosmological parameter constraints (Naidoo et al., 2020). This information has not been fully exploited due to the computational cost of calculating $N$-point statistics. MiSTree was designed to be used in cosmology but could be used in any field which requires extracting non-Gaussian information from point distributions.

A.2.2 Motivation

Studies of point distributions often measure their 2-point statistics (i.e. the distribution of distances between pairs of points) which are then compared to theoretical models. This is a powerful technique and has been used very successfully in the field of cosmology to study the early Universe and the large scale distribution of galaxies. Unfortunately this statistic can only fully describe a distribution that is Gaussian, if it is non-Gaussian then the 2-point is no longer sufficient. The conventional method to incorporate non-Gaussian information is to look at the distribution’s $N$-point statistic (if $N = 3$ we look at the distribution of triangles, if $N = 4$ we look at the distribution of quadrilaterals and so on). This method is well motivated as in principle all the information that can describe a distribution of points is contained within its $N$-point statistics (see Szapudi & Szalay, 1998). However, calculating $N$-point statistics even for $N > 3$ becomes quickly intractable for large data sets.
APPENDIX A.2

The MST offers an alternative approach; the MST graph draws lines between pairs of points so that all points are linked in a single skeletal structure that contains no loops and has minimal total edge length. Unlike \( N \)-point statistics, that typically scale by \( \mathcal{O}(n^N) \) for \( n \) points, the MST (computed using the Kruskal algorithm (Kruskal, 1956) which sequentially adds edges, from shortest to longest, with the condition that the added edge does not form a loop) can be constructed much faster (at best \( \mathcal{O}(n \log n) \)). While the MST does not contain all the information present in \( N \)-point statistics, it enables some of this information to be captured and allows the identification of skeletal patterns, as such it has found a broad range of applications in physics: such as finding filaments in the distribution of galaxies (Alpaslan et al., 2014), classifying particle physics collisions (Rainbolt & Schmitt, 2017) and mass segregation in star clusters (Allison et al., 2009). The MST has also been used in a number of other scientific fields such as computer science, sociology and epidemiology.

While algorithms to construct the minimum spanning tree are well known (e.g. Prim (1957) and Kruskal (1956)) implementations of these often require the input of a matrix of pairwise distances. For a large data set the creation of this matrix (with \( n^2 \) elements) can be a significant strain on memory while also making the construction of the MST slower (\( \mathcal{O}(n^2 \log n) \)).

A.2.3 MiSTree

MiSTree is a public Python package for the construction and analysis of the MST. The package initially creates a \( k \)-nearest neighbour graph (\( k \)NN, a graph that links each point to the nearest \( k \) neighbours, using scikit-learn’s kneighbors_graph function) which improves speed by limiting the number of considered edges from \( n^2 \) to \( kn \) (where \( k \ll n \)) and then runs the Kruskal algorithm (Kruskal, 1956) (using scipy’s minimum_spanning_tree function). The stages of the MST construction are shown in Figure A.2.1.

The MST can be constructed from data provided in 2/3 dimensions and in tomographic (on a unit sphere) or spherical polar coordinates. The weights of the edges are assumed to be the distances between points; i.e. the Euclidean distance for 2/3 dimension and spherical polar coordinates, and angular distances for tomographic coordinates. Furthermore, the package can very quickly measure the standard statistics:

- degree \( (d) \) – the number of edges attached to each node.
- edge length \( (l) \) – the length of edges in the MST.

While also being able to measure the statistics of branches, which are defined as chains of edges connected with degree \( = 2 \):

- branch length \( (b) \) – the sum of the lengths of member edges.
Figure A.2.1: An example of how MiSTree constructs the MST from a distribution of points (shown on the left). MiSTree first begins by constructing a $k$NN graph which links all points to their nearest $k$ neighbours (shown in the centre) and then runs the Kruskal algorithm to construct the MST (shown on the right).
Figure A.2.2: Histograms of the distribution of the MST statistics degree ($d$), edge length ($l$), branch length ($b$) and branch shape ($s$) for a Levy Flight and Adjusted Levy Flight distribution in comparison to a set of random distribution (details of which are provided in Naidoo et al. 2019) in 3 dimensions.

• branch shape ($s$) – the straight line distance between the tips of branches divided by the branch length.

The statistics calculated by MiSTree are extensively explored in Naidoo et al. (2020) and found to significantly improve constraints on cosmological parameters when tested on simulations.

A.2.4 Basic Usage

To construct the MST using MiSTree from a distribution of points in 2 dimensions you would use the following commands:

```python
import mistree as mist

# initialise MiSTree minimum spanning tree class
mst = mist.GetMST(x=x, y=y)
mst.construct_mst()
```

Once the MST is constructed it can either be used to look for features in the distribution or to measure statistics of the graph which in turn tell us about how points have been distributed. MiSTree can measure four statistics by default, which can be calculated directly after initialising the GetMST class (an example of the distribution of these statistics is shown in Figure A.2.2):
d, l, b, s = mst.get_stats()

The source code can be found on github\footnote{https://github.com/knaidoo29/mistree} while documentation and more complicated tutorials are provided \footnote{https://knaidoo29.github.io/mistreedoc/}.

A.2.5 Dependencies

Dependencies for MiSTree include the Python modules {\textsc{numpy}}\footnote{Oliphant, 2006}, \textsc{matplotlib}\footnote{Hunter, 2007}, \textsc{scipy}\footnote{Jones et al., 2001}, \textsc{scikit-learn}\footnote{Pedregosa et al., 2011} and \textsc{f2py}\footnote{Peterson, 2009} (the latter of which is used to compile Fortran subroutines).
A.3 NUMERICAL DERIVATIVE ESTIMATORS

The derivative of a function $f(x)$ can be estimated in several ways. In Section A.3.1 the symmetric two-point estimator is derived and in Section A.3.2 the non-symmetric two-, three- and four-point estimators are derived.

A.3.1 Symmetric Derivative Estimator

To derive the symmetric two-point derivative it is useful to first express the Taylor expansion of a function $f(x)$ at a point $x + dx$,

$$f(x + dx) = \sum_{n=0}^{\infty} \frac{dx^n}{n!} \frac{\partial^n f(x)}{\partial x^n}.$$  \hspace{1cm} (A.3.1)

If $dx$ is small the lowest order terms are dominant and are shown below up to $dx^4$,

$$f(x + dx) = f(x) + dx f'(x) + \frac{dx^2}{2} f''(x) + \frac{dx^3}{6} f'''(x) + \frac{dx^4}{24} f''''(x) + O(dx^5),$$  \hspace{1cm} (A.3.2)

where $' = \partial / \partial x$. Similarly the Taylor expansion at a point $x - dx$ is defined as

$$f(x - dx) = f(x) - dx f'(x) + \frac{dx^2}{2} f''(x) - \frac{dx^3}{6} f'''(x) + \frac{dx^4}{24} f''''(x) + O(dx^5).$$  \hspace{1cm} (A.3.3)

Combining Equation A.3.2 and A.3.3 gives

$$f(x + dx) - f(x - dx) = 2dx f'(x) + O(dx^3)$$  \hspace{1cm} (A.3.4)

and therefore the symmetric two-point derivative can be defined by rearranging the above,

$$f'(x) = \frac{f(x + dx) - f(x - dx)}{2dx} + O(dx^2).$$  \hspace{1cm} (A.3.5)

A.3.2 Non-Symmetric Derivative Estimators

There are cases where the symmetric derivative cannot be used, for example if the derivative of a function $f(x)$ needs to be calculated near a boundary. In the derivations that follow, we take the case where it is not possible to calculate $f(x)$ when $x < 0$. 
Two-Point Derivative Estimator

The most straightforward estimator simply requires rearranging Equation A.3.2 to give

\[ f'(x) = \frac{f(x + dx) - f(x)}{dx} + O(dx), \]  

(A.3.6)

although this estimator is simple and similar in form to the symmetric derivative estimator it has larger errors of order \( O(dx) \).

Three-Point Derivative Estimator

A more complex estimator makes use of three points rather than two. To derive this we first Taylor expand the function \( f(x) \) at a new point \( x + 2dx \) which is chosen to match the location of simulations used in Chapter 3,

\[ f(x + 2dx) = f(x) + 2df(x) + \frac{dx^2}{2} f''(x) + \frac{16dx^3}{6} f'''(x) \]

\[ + \frac{64dx^4}{24} f''''(x) + O(dx^5). \]  

(A.3.7)

To remove terms of order \( dx^2 \) Equations A.3.2 and A.3.7 can be combined in the following way

\[ f(x + 2dx) - 4f(x + dx) = -3f(x) - 2df(x) + O(dx^3), \]  

(A.3.8)

which gives the estimator

\[ f'(x) = \frac{-f(x + 2dx) + 4f(x + dx) - 3f(x)}{2dx} + O(dx^2). \]  

(A.3.9)

Unlike Equation A.3.6 this estimator has errors of order \( O(dx^2) \) similar to the symmetric derivative A.3.5.

Four-Point Derivative Estimator

The estimator can be improved further by using four points. To derive this we first Taylor expand the function \( f(x) \) at a new point \( x + 4dx \) which is chosen to match the location of simulations used in Chapter 3,

\[ f(x + 4dx) = f(x) + 4df(x) + \frac{16dx^2}{2} f''(x) + \frac{64dx^3}{6} f'''(x) \]

\[ + \frac{256dx^4}{24} f''''(x) + O(dx^5). \]  

(A.3.10)
To combine Equations A.3.2, A.3.7 and A.3.10 the scalars $A$, $B$ and $C$ are defined such that

\[ Af(x + dx) + Bf(x + 2dx) + Cf(x + 4dx) = (A + B + C)f(x) \]
\[ + (A + 2B + 4C)dx f'(x) + (A + 4B + 16C)\frac{dx^2}{2} f''(x) \]
\[ + (A + 8B + 64C)\frac{dx^3}{6} f'''(x) + O(dx^4). \] (A.3.11)

To ensure the terms of order $dx^2$ and $dx^3$ are removed

\[ A + 4B + 16C = 0, \] (A.3.12)
\[ A + 8B + 64C = 0, \] (A.3.13)

and therefore $A = 32$, $B = -12$ and $C = 1$. Equation A.3.11 then reduces to

\[ 32f(x + dx) - 12f(x + 2dx) + f(x + 4dx) = 21f(x) + 12dx f'(x) \]
\[ + O(dx^4), \] (A.3.14)

which gives the final estimator

\[ f'(x) = \frac{1}{12dx} \left[ f(x + 4dx) - 12f(x + 2dx) + 32f(x + dx) \right. \]
\[ \left. - 21f(x) \right] + O(dx^3). \] (A.3.15)
Abraham C., 1962, Memo, IBM Corp, York Town Heights, NY
REFERENCES


Bhavsar S. P., Ling E. N., 1988, PASP, 100, 1314


Bird S., Rogers K. K., Peiris H. V., Verde L., Font-Ribera A., Pontzen A., 2019, Journal of Cosmology and Astro-Particle Physics, 2019, 050


Campana R., Massaro E., Bernieri E., 2018a, Ap&SS, 363, 144


REFERENCES


Einstein A., 1916, Annalen der Physik, 49, 769

Einstein A., 1917, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin, pp 142–152

Einstein A., 1936, Science, 84, 506


197
REFERENCES


REFERENCES

Howlett C., Manera M., Percival W. J., 2015, Astronomy and Computing, 12, 109
Hubble E., 1929, Proceedings of the National Academy of Science, 15, 168
Hunter J. D., 2007, Computing in Science & Engineering, 9, 90
REFERENCES

Jones E., Oliphant T., Peterson P., 2001


Kaufman G., 1967


Kruskal J. B., 1956, Proceedings of the American Mathematical Society, 7, 48


Leavitt H. S., 1908, Annals of Harvard College Observatory, 60, 87

Leclercq F., 2018, Phys. Rev. D, 98, 063511


Minkowski H., 1903, Mathematische Annalen, 57, 447


REFERENCES


Pedregosa F., et al., 2011, Journal of machine learning research, 12, 2825

Princeton, N.J


Planck Collaboration et al., 2016c, A&A, 594, A21


Prim R. C., 1957, Bell System Technical Journal, 36, 1389

Rainbolt J. L., Schmitt M., 2017, JINST, 12, P02009


Scoccimarro R., 2015, Phys. Rev. D, 92, 083532


Smirnov N., 1948, The annals of mathematical statistics, 19, 279


REFERENCES

Zahn C. T., 1971, IEEE Transactions on computers, 100, 68
Zel'dovich Y. B., 1967, ZhETF Pisma Redaktsiiu, 6, 883
Zhang R., Huterer D., 2010, Astroparticle Physics, 33, 69
