

Islam, Salma (2019) *Exploring the parameters of peculiar velocity fields*. PhD thesis.

https://theses.gla.ac.uk/41045/

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses <u>https://theses.gla.ac.uk/</u> research-enlighten@glasgow.ac.uk

UNIVERSITY OF GLASGOW

DOCTORAL THESIS

Exploring the parameters of peculiar velocity fields

Author: Salma Islam

Supervisor: Dr. Martin A. HENDRY

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

in the

Astronomy & Astrophysics Group School of Physics & Astronomy

February 2019







Declaration of Authorship

I, Salma ISLAM, declare that this thesis titled, 'Exploring the parameters of peculiar velocity fields' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date: 25/2/2019

In the name of God, the most Gracious, the most Merciful

أَوَلَمْ يَرُ ٱلَّذِينَ كَفَرُوَا أَنَّ ٱلسَّمَوَنِ وَٱلْأَرْضَ كَانَا رَتْقَا فَفَنَقْنَهُمَا وَجَعَلْنَا مِنَ ٱلْمَآءِ كُلَّ شَيْءٍ حَيٍّ أَفَلَا يُؤْمِنُونَ (٣

Have those who disbelieved not considered that the heavens and the earth were a joined entity, and We separated them and made from water every living thing? Then will they not believe? - Surat Al-Anbiyaa (Chapter 21, Verse 30)

ٱللَّهُ ٱلَّذِى رَفَعَ ٱلسَّمَوَٰتِ بِغَيْرِ عَمَدِ تَرَوْنَهَا ثُمَّ ٱسْتَوَىٰ عَلَى ٱلْعَرْشِ وَسَخَرَ ٱلشَّمْسَ وَٱلْقَمَرِ كُلُّ يَجْرِى لِأَجَلِ مُّسَمَّى َّيُدَبِّرُ ٱلْأَمَرَ يُفَصِّلُ ٱلْآدَنِتِ لَعَلَكُم بِلِقَاءِ رَبِّكُمْ تُوَقِنُونَ ()

It is Allah who erected the heavens without pillars that you [can] see; then He established Himself above the Throne and made subject the sun and the moon, each running [its course] for a specified term. He arranges [each] matter; He details the signs that you may, of the meeting with your Lord, be certain. - Surat Ar-Ra'd (Chapter 13, Verse 2)

وٱلسَّمَاءَ بَنَيْنَهَا بِأَيْدٍ وَإِنَّالَمُوسِعُونَ (*)

And the heaven We constructed with strength, and indeed, We are [its] expander. - Surat Adh-Dhariyaat (Chapter 51, Verse 47)

المَّا أَشْهَد تُهُمْ خَلْقَ ٱلسَّمَوَنِ وَٱلْأَرْضِ وَلَا خَلْقَ أَنفُسِهِمْ وَمَا كُنتُ مُتَخِذَ ٱلْمُضِلِّينَ عَضُدًا (٥)

I did not make them witness to the creation of the heavens and the earth or to the creation of themselves, and I would not have taken the misguiders as assistants. - Surat Al-Kahf (Chapter 18, Verse 51)

UNIVERSITY OF GLASGOW

Abstract

Astronomy & Astrophysics Research Group School of Physics & Astronomy

Doctor of Philosophy

Exploring the parameters of peculiar velocity fields

by Salma Islam

The main focus of this work is to make use of a novel tool in the cosmologist's toolbox when it comes to constraining the parameters of the peculiar velocity fields of the nearby Universe called ROBUST, whose unique properties and lack of reliance on secondary distance indicators sets it apart from other available constraining techniques, rendering it potentially very useful for future upcoming surveys such as the LSST and the SKA.

While ROBUST proves itself more than adequate in constraining parameters in a mock controlled environment with the IRAS PSCz survey, it begins to struggle when applied to the real-world 2MRS survey, primarily due to an inherent fault in the survey that causes it to not function properly with the program. These problems persist even when we begin to make use of one of the ancillary tools developed in conjunction with ROBUST, namely relative entropy, despite it once again continuing to function adequately across multiple mock realisations.

It is the conclusion of this work that while ROBUST is not successful in recovering values for the cosmological parameters we seek to constrain, this does not necessarily negate its viability for use with upcoming surveys, as it has proven itself successful in determining exclusion intervals on the value of the linear redshift distortion parameter β for real world surveys that are in very good agreement with the generally small values computed by contemporary velocity-velocity constraining tehniques such as VELMOD and χ^2 -minimisation, while also confidently ruling out the results of older density-density constraining techniques such as POTENT that favour values closer to unity.

Keywords: cosmology, galaxies, data analysis, dark matter, redshift surveys, peculiar velocities, luminosity functions, ROBUST method, χ^2 minimisation, linear redshift distortion parameter, relative entropy

Acknowledgements

This thesis is the culmination of nearly 4 years worth of work and research which would not have been possible without the following resources from our fellow astronomers and cosmologists.

This research has made use of the NASA/IPAC Extragalactic Database (NED) which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

We acknowledge the work of astronomers supporting the LEDA database (http:// leda.univ-lyon1.fr/), the Asiago Supernova Catalogue (http://web.pd.astro.it/ supern/snean.txt), The STScI Digitized Sky Survey (http://archive.stsci.edu/ cgi-bin/dss_form), and the CfA List of Supernovae (http://cfa-www.harvard.edu/ cfa/ps/lists/Supernovae.html).

This publication also makes use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation.

This research has also made use of the VizieR catalogue access tool, CDS, Strasbourg, France. The original description of the VizieR service was published in A&AS 143, 23 (source [141] and the references therein).

In addition, this research has made use of the most recent data runs of the Sloan Digital Sky Survey. Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS web site is www.sdss.org.

SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, the Chilean Participation Group, the French Participation Group, Harvard-Smithsonian Center for Astrophysics, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) / University of Tokyo, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatário Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

Most importantly, my undying thanks goes out to my supervisor Martin A. Hendry, whose infinite patience with me either in my naive overeagerness or in my darker and more difficult periods wrangling with this project has been invaluable. Without your understanding and your insights none of this would have been possible and I am more grateful to you than I can possibly articulate. Thank you so very much.

Also no acknowledgements section would be complete without giving my thanks to all of my friends here in Glasgow, particularly Paul, Naomi, Chris and Adam for always helping me to keep seeing the funny side of life when things became difficult, and supplying me with endless laughter and support when I needed it the most. You guys are the best.

I also wish to extend special thanks and appreciation towards my family, be they here in the UK or abroad scattered across the globe. Your unwavering faith in me and in my abilities to do this and contribute to the scientific community in some fashion has been nothing short of astonishing, and I am truly humbled and grateful to be surrounded by people who love me so much and have my back no matter what. I love every single one of you and hope to see you all again soon.

Last and not least, I give thanks to Allah (swt) for blessing me with the soundness of mind and spirit that I might be able to peer into Your Universe that You have created and attempt to figure out some aspect of how it works. I'm not entirely sure I've been very successful, but my efforts over the past few years have made me all the more aware of Your ingenuity as the Perfect Designer and has fuelled my desire to look further into Your heavens and see what else I can learn and explore and in so doing, hopefully become closer to You. Alhamdulillah.

Contents

Declaration of Authorship	ii
Abstract	iv
Acknowledgements	v
Contents	vii
List of Figures	xi
List of Tables	xxvii
Abbreviations	xxxi
Physical Constants	xxxv
Symbols	xxxvii

Preface 1 1 A Brief Overview of Cosmology 3 1.1 Cosmology in the Modern Day 3 1.2The Concept of Universal Expansion and the Cosmological Principle . . . 4 1.3'de Sitter Space' and the FLRW Metric 9 A Question of Spacetime Curvature and the Fate of the Universe . 12 1.3.11.4 141.5The Mystery of Dark Matter 181.6191.7The Breakthrough: Dark Energy and the Double Dark Theory 2329 2 A Cosmologist's Toolbox 2.1292.2Hubble's Law 312.2.1332.3Apparent, Absolute Magnitudes with the Distance Modulus Law 342.4

2.6 The Tully-Fisher Relation for Spiral Galaxies 2.7 The Fundamental Plane and Faber-Jackson Relation for Elliptical Gala 2.8 Different Coordinate Systems 2.8.1 Celestial and Equatorial Coordinate Systems 2.8.2 The Galactic Plane 2.8.3 The Supergalactic Plane 2.8.4 The Goury et. al Supernova Data Set 2.9.1 The Toury et. al Supernova Data Set 2.9.2 The Union2.1 Compilation 2.9.3 The SAI Supernova Catalogue 2.10.1 RAS PSCz Galaxy Catalogue 2.10.2 2MASS 2.10.3 Sloan Digital Sky Survey 3 Probing Peculiar Velocity Fields of the Nearby Universe 3.1 The Importance of Probing Peculiar Velocity Fields 3.2 Defining Gravitational Instability 3.3 Exploring Cosmological Parameters 3.3.1 Bulk Flow 3.3.2 The Mass Power Spectrum 3.3.3 The Linear Redshift Distortion Parameter β 3.4.1 Basis Function Methods 3.4.2 Iterative Method - POTENT 3.4.3 Iterative Method - VELMOD 3.4.4 <t< th=""><th></th><th>2.5</th><th>NIa as Standard Candle Field Probes</th><th>37</th></t<>		2.5	NIa as Standard Candle Field Probes	37
2.7 The Fundamental Plane and Faber-Jackson Relation for Elliptical Gala 2.8 Different Coordinate Systems 2.8.1 Celestial and Equatorial Coordinate Systems 2.8.2 The Galactic Plane 2.8.3 The Supergalactic Plane 2.9 An overview of supernovae catalogues 2.9.1 The Union2.1 Compilation 2.9.2 The Union2.1 Compilation 2.9.3 The SAI Supernova Catalogue 2.10.1 IRAS PSCz Galaxy Catalogue 2.10.2 2MASS 2.10.3 Sloan Digital Sky Survey 3 Probing Peculiar Velocity Fields of the Nearby Universe 3.1 The Importance of Probing Peculiar Velocity Fields 3.2 Defining Gravitational Instability 3.3 Exploring Cosmological Parameters 3.3.1 Bulk Flow 3.3.2 The Mass Power Spectrum 3.3.3 The Linear Redshift Distortion Parameter β 3.4 Constraining β 3.4.2 Iterative Method - VELMOD 3.4.3 Iterative Method - VELMOD 3.4.4 Best of Both Worlds - UMV 3.4.5 Applying χ^2 -minimisation		2.6	The Tully-Fisher Relation for Spiral Galaxies	39
2.8 Different Coordinate Systems 2.8.1 Celestial and Equatorial Coordinate Systems 2.8.2 The Galactic Plane 2.8.3 The Supergalactic Plane 2.9 An overview of supernovae catalogues 2.9.1 The The Tonry et. al Supernova Data Set 2.9.2 The Union2.1 Compilation 2.9.3 The SAI Supernova Catalogue 2.10 An overview of redshift and peculiar velocity surveys 2.10.1 IRAS PSCz Galaxy Catalogue 2.10.2 2MASS 2.10.3 Sloan Digital Sky Survey 3 Probing Peculiar Velocity Fields of the Nearby Universe 3.1 The Importance of Probing Peculiar Velocity Fields 3.2 Defining Gravitational Instability 3.3 Supering Cosmological Parameters 3.3.1 Bulk Flow 3.3.2 The Mass Power Spectrum 3.3.3 The Linear Redshift Distortion Parameter β 3.4.1 Basis Function Methods 3.4.2 Iterative Method - POTENT 3.4.3 Iterative Method - VELMOD 3.4.4 Best of Both Worlds - UNV 3.4.5 Applying χ^2 -minimisation		2.7	The Fundamental Plane and Faber-Jackson Relation for Elliptical Galaxies	43
2.8.1Celestial and Equatorial Coordinate Systems2.8.2The Galactic Plane2.8.3The Supergalactic Plane2.9An overview of supernova catalogues2.9.1The Tonry et. al Supernova Data Set2.9.2The Union2.1 Compilation2.9.3The SAI Supernova Catalogue2.10.4noverview of redshift and peculiar velocity surveys2.10.5The SAI Supernova Catalogue2.10.6IRAS PSC2 Galaxy Catalogue2.10.7Sloan Digital Sky Survey3Probing Peculiar Velocity Fields of the Nearby Universe3.1The Importance of Probing Peculiar Velocity Fields3.2Defining Gravitational Instability3.3Exploring Cosmological Parameters3.3.1Bulk Flow3.3.2The Mass Power Spectrum3.3.3The Linear Redshift Distortion Parameter β 3.4.1Basis Function Methods3.4.2Iterative Method - POTENT3.4.3Iterative Method - VELMOD3.4.4Best of Both Worlds - UMV3.4.5Applying χ^2 -minimisation3.4.6Only the Beginning - A Brief Overview of More Recent Method3.4.7Attempts to constrain β to date4Introducing the ROBUST Method4.1Underlying Theory of ROBUST4.2I betarning β with ROBUST4.3Constraining β with ROBUST4.4Strength of ROBUST4.5Applying χ^2 to the IRAS PSCz SurveyAside - choosing the appropriate software5.1Using Real-World Sne Set		2.8	Different Coordinate Systems	45
2.8.2 The Galactic Plane 2.8.3 The Supergalactic Plane 2.9 An overview of supernovae catalogues 2.9.1 The Union2.1 Compilation 2.9.2 The Union2.1 Compilation 2.9.3 The SAI Supernova Catalogue 2.10 An overview of redshift and peculiar velocity surveys 2.10.1 IRAS PSCz Galaxy Catalogue 2.10.2 2MASS 2.10.3 Sloan Digital Sky Survey 3 Probing Peculiar Velocity Fields of the Nearby Universe 3.1 The Importance of Probing Peculiar Velocity Fields 3.2 Defining Gravitational Instability 3.3 Bulk Flow 3.3.1 Bulk Flow 3.3.2 The Mass Power Spectrum 3.3.3 The Linear Redshift Distortion Parameter β 3.4.1 Basis Function Methods 3.4.2 Iterative Method - POTENT 3.4.3 Iterative Method - VELMOD 3.4.4 Best of Both Worlds - UMV 3.4.5 Applying χ^2 -minimisation 3.4.6 Only the Beginning - A Brief Overview of More Recent Method 3.4.7 Attempts to constrain β to date <th></th> <th></th> <th>2.8.1 Celestial and Equatorial Coordinate Systems</th> <th>45</th>			2.8.1 Celestial and Equatorial Coordinate Systems	45
2.8.3 The Supergalactic Plane 2.9 An overview of supernova catalogues 2.9.1 The Tonry et. al Supernova Data Set 2.9.2 The Union2.1 Compilation 2.9.3 The SAI Supernova Catalogue 2.10 An overview of redshift and peculiar velocity surveys 2.10.1 IRAS PSCz Galaxy Catalogue 2.10.2 2MASS 2.10.3 Sloan Digital Sky Survey 3 Probing Peculiar Velocity Fields of the Nearby Universe 3.1 The Importance of Probing Peculiar Velocity Fields 3.2 Defining Gravitational Instability 3.3 Exploring Cosmological Parameters 3.3.1 Bulk Flow 3.3.2 The Mass Power Spectrum 3.3.3 The Linear Redshift Distortion Parameter β 3.4.1 Basis Function Methods 3.4.2 Iterative Method - POTENT 3.4.3 Iterative Method - VELMOD 3.4.4 Best of Both Worlds - UNV 3.4.5 Applying χ^2 -minimisation 3.4.6 Only the Beginning - A Brief Overview of More Recent Method 3.4.7 Attempts to constrain β to date 4 I			2.8.2 The Galactic Plane	47
2.9An overview of supernovae catalogues2.9.1The Tonry et. al Supernova Data Set2.9.2The Union2.1 Compilation2.9.3The SAI Supernova Catalogue2.10An overview of redshift and peculiar velocity surveys2.10.1IRAS PSCz Galaxy Catalogue2.10.22MASS2.10.3Sloan Digital Sky Survey3 Probing Peculiar Velocity Fields of the Nearby Universe 3.1The Importance of Probing Peculiar Velocity Fields3.2Defining Gravitational Instability3.3Exploring Cosmological Parameters3.3.1Bulk Flow3.3.2The Mass Power Spectrum3.3.3The Linear Redshift Distortion Parameter β 3.4.1Basis Function Methods3.4.2Iterative Method - POTENT3.4.3Iterative Method - VELMOD3.4.4Best of Both Worlds - UMV3.4.5Applying χ^2 -minimisation3.4.6Only the Beginning - A Brief Overview of More Recent Method3.4.7Attempts to constrain β to date4Introducing the ROBUST Method4.1Underlying Theory of ROBUST4.2Testing for Completeness, T_c 4.2.1Defining T_c with Two Limits4.3Iterativing L Error on β 4.4Strength of ROBUST5Applying χ^2 to the IRAS PSC2 SurveyAside - choosing the appropriate software5.1Using Real-World Sne Sets5.2Constraining β with the RS Sne Set5.2Constraining β with the RS Sne Set </th <th></th> <th></th> <th>2.8.3 The Supergalactic Plane</th> <th>48</th>			2.8.3 The Supergalactic Plane	48
2.9.1The Tonry et. al Supernova Data Set2.9.2The Union2.1 Compilation2.9.3The SAI Supernova Catalogue2.10An overview of redshift and peculiar velocity surveys2.10.1IRAS PSCz Galaxy Catalogue2.10.22MASS2.10.3Sloan Digital Sky Survey3 Probing Peculiar Velocity Fields of the Nearby Universe 3.1The Importance of Probing Peculiar Velocity Fields3.2Defining Gravitational Instability3.3Exploring Cosmological Parameters3.3.1Bulk Flow3.3.2The Mass Power Spectrum3.3.3The Linear Redshift Distortion Parameter β 3.4.1Basis Function Methods3.4.2Iterative Method - POTENT3.4.3Iterative Method - VELMOD3.4.4Best of Both Worlds - UMV3.4.5Applying χ^2 -minimisation3.4.6Only the Beginning - A Brief Overview of More Recent Method3.4.7Attempts to constrain β to date4Introducing the ROBUST Method4.1Underlying Theory of ROBUST4.2Testing for Completeness, T_c 4.3.1Defining T_c with Two Limits4.3Lostraining β with ROBUST4.4Strength of ROBUST5Applying χ^2 to the IRAS PSC2 SurveyAside - choosing the appropriate software5.1Using Real-World Sne Set5.2Constraining β with the RS Sne Set5.2Constraining β with the RS Sne Set5.2Constraining β with the RS Sne Set <th></th> <th>2.9</th> <th>An overview of supernovae catalogues</th> <th>51</th>		2.9	An overview of supernovae catalogues	51
2.9.2The Union2.1 Compilation2.9.3The SAI Supernova Catalogue2.10An overview of redshift and peculiar velocity surveys2.10.1IRAS PSCz Galaxy Catalogue2.10.22MASS2.10.3Sloan Digital Sky Survey3Probing Peculiar Velocity Fields of the Nearby Universe3.1The Importance of Probing Peculiar Velocity Fields3.2Defining Gravitational Instability3.3Exploring Cosmological Parameters3.3.1Bulk Flow3.3.2The Mass Power Spectrum3.3.3The Linear Redshift Distortion Parameter β 3.4.1Basis Function Methods3.4.2Iterative Method - POTENT3.4.3Iterative Method - VELMOD3.4.4Best of Both Worlds - UMV3.4.5Applying χ^2 -minimisation3.4.6Only the Beginning - A Brief Overview of More Recent Method3.4.7Attempts to constrain β to date4Introducing the ROBUST Method4.1Underlying Theory of ROBUST4.2Lefning T_c with Two Limits4.3Constraining β with ROBUST4.4Strength of ROBUST5Applying χ^2 to the IRAS PSCz SurveyAside - choosing the appropriate software5.1Using Real-World Sne SetsAn Alternative to Tonry - the Radburn-Smith Sne subset5.2Constraining β with the RS Sne Set5.2.1The Curiosity of Large χ^2 Values5.2.2Analysing Linear Regression			2.9.1 The Tonry et. al Supernova Data Set	51
2.9.3 The SAI Supernova Catalogue2.10 An overview of redshift and peculiar velocity surveys2.10.1 IRAS PSCz Galaxy Catalogue2.10.2 2MASS2.10.3 Sloan Digital Sky Survey3 Probing Peculiar Velocity Fields of the Nearby Universe3.1 The Importance of Probing Peculiar Velocity Fields3.2 Defining Gravitational Instability3.3 Exploring Cosmological Parameters3.3.1 Bulk Flow3.3.2 The Mass Power Spectrum3.3.3 The Linear Redshift Distortion Parameter β 3.4 Constraining β 3.4.1 Basis Function Methods3.4.2 Iterative Method - POTENT3.4.3 Iterative Method - VELMOD3.4.4 Best of Both Worlds - UMV3.4.5 Applying χ^2 -minimisation3.4.6 Only the Beginning - A Brief Overview of More Recent Method3.4.7 Attempts to constrain β to date44.1 Underlying Theory of ROBUST4.2.1 Defining T_c with Two Limits4.3 Constraining β with ROBUST4.4 Strength of ROBUST5.1 Using Real-World Sne SetsAn Alternative to Tonry - the Radburn-Smith Sne subset5.1 Using Real-World Sne SetsAn Alternative to Tonry - the Radburn-Smith Sne subset5.2 Constraining β with the RS Sne Set5.2.1 The Curiosity of Large χ^2 Values5.2.2 Analysing Linear Regression			2.9.2 The Union2.1 Compilation	52
2.10 An overview of redshift and peculiar velocity surveys2.10.1 IRAS PSCz Galaxy Catalogue2.10.2 2MASS2.10.3 Sloan Digital Sky Survey3 Probing Peculiar Velocity Fields of the Nearby Universe3.1 The Importance of Probing Peculiar Velocity Fields3.2 Defining Gravitational Instability3.3 Exploring Cosmological Parameters3.3.1 Bulk Flow3.3.2 The Mass Power Spectrum3.3.3 The Linear Redshift Distortion Parameter β 3.4 Constraining β 3.4.1 Basis Function Methods3.4.2 Iterative Method - POTENT3.4.3 Iterative Method - POTENT3.4.4 Best of Both Worlds - UMV3.4.5 Applying χ^2 -minimisation3.4.6 Only the Beginning - A Brief Overview of More Recent Method3.4.7 Attempts to constrain β to date4 Introducing the ROBUST Method4.1 Underlying Theory of ROBUST4.2 Testing for Completeness, T_c 4.2.1 Defining f_c with Two Limits4.3 Constraining β with ROBUST4.4 Strength of ROBUST4.4 Strength of ROBUST4.4 Strength of ROBUST5 Applying χ^2 to the IRAS PSCz SurveyAside - choosing the appropriate software5.1 Using Real-World Sne SetsAn Alternative to Tonry - the Radburn-Smith Sne subset5.2 Constraining β with the RS Sne Set5.2.1 The Curiosity of Large χ^2 Values5.2.2 Analysing Linear Regression			2.9.3 The SAI Supernova Catalogue	55
2.10.1IRAS PSCz Galaxy Catalogue2.10.22MASS2.10.3Sloan Digital Sky Survey3 Probing Peculiar Velocity Fields of the Nearby Universe 3.1The Importance of Probing Peculiar Velocity Fields3.2Defining Gravitational Instability3.3Exploring Cosmological Parameters3.3.1Bulk Flow3.3.2The Mass Power Spectrum3.3.3The Linear Redshift Distortion Parameter β 3.4Constraining β 3.4.1Basis Function Methods3.4.2Iterative Method - POTENT3.4.3Iterative Method - VELMOD3.4.4Best of Both Worlds - UMV3.4.5Applying χ^2 -minimisation3.4.6Only the Beginning - A Brief Overview of More Recent Method3.4.7Attempts to constrain β to date4Introducing the ROBUST Method4.1Underlying Theory of ROBUST4.2Testing for Completeness, T_c 4.2.1Defining T_c with Two Limits4.3Constraining β with ROBUST4.4Strength of ROBUST4.4Strength of ROBUST5Applying χ^2 to the IRAS PSC2 SurveyAside - choosing the appropriate software5.1Using Real-World Sne SetsAn Alternative to Tonry - the Radburn-Smith Sne subset5.2Constraining β with the RS Sne Set5.2.1The Curiosity of Large χ^2 Values5.2.2Analysing Linear Regression		2.10	An overview of redshift and peculiar velocity surveys	55
2.10.22MASS 2.10.32.10.3Sloan Digital Sky Survey3 Probing Peculiar Velocity Fields of the Nearby Universe 3.1The Importance of Probing Peculiar Velocity Fields 3.2Defining Gravitational Instability 3.3Exploring Cosmological Parameters 3.3.1Bulk Flow 3.3.2The Mass Power Spectrum 3.3.3The Linear Redshift Distortion Parameter β 3.4Constraining β 3.4.1Basis Function Methods 3.4.2Iterative Method - POTENT 3.4.3Iterative Method - VOTENT 3.4.43.4.4Best of Both Worlds - UMV 3.4.5Applying χ^2 -minimisation 3.4.6Only the Beginning - A Brief Overview of More Recent Method 3.4.74.1Underlying Theory of ROBUST 4.2.1Defining T_c with Two Limits 4.3Constraining β with ROBUST 4.2.14.3Interducing the ROBUST Method 4.44.4Strength of ROBUST 4.3.1Determining the Error on β 4.45Applying χ^2 to the IRAS PSC2 Survey Aside - choosing the appropriate software 5.1Using Real-World Sne Sets 5.25.2.2Analysing Linear Regression			2.10.1 IRAS PSCz Galaxy Catalogue	57
2.10.3Sloan Digital Sky Survey3Probing Peculiar Velocity Fields of the Nearby Universe3.1The Importance of Probing Peculiar Velocity Fields3.2Defining Gravitational Instability3.3Exploring Cosmological Parameters3.3.1Bulk Flow3.3.2The Mass Power Spectrum3.3.3The Linear Redshift Distortion Parameter β 3.4Constraining β 3.4.1Basis Function Methods3.4.2Iterative Method - POTENT3.4.3Iterative Method - VELMOD3.4.4Best of Both Worlds - UMV3.4.5Applying χ^2 -minimisation3.4.6Only the Beginning - A Brief Overview of More Recent Method3.4.7Attempts to constrain β to date4Introducing the ROBUST Method4.1Underlying Theory of ROBUST4.2Testing for Completeness, T_c 4.3.1Determining the Error on β 4.4Strength of ROBUST4.4Strength of ROBUST4.4Strength of ROBUST5Applying χ^2 to the IRAS PSCz SurveyAside - choosing the appropriate software5.1Using Real-World Sne Sets5.2Constraining β with RS Sne Set5.2.1The Curiosity of Large χ^2 Values5.2.2Analysing Linear Regression			2.10.2 2MASS	57
3 Probing Peculiar Velocity Fields of the Nearby Universe 3.1 The Importance of Probing Peculiar Velocity Fields 3.2 Defining Gravitational Instability 3.3 Exploring Cosmological Parameters 3.3.1 Bulk Flow 3.3.2 The Mass Power Spectrum 3.3.3 The Linear Redshift Distortion Parameter β 3.4.1 Basis Function Methods 3.4.2 Iterative Method - POTENT 3.4.3 Iterative Method - VELMOD 3.4.4 Best of Both Worlds - UMV 3.4.5 Applying χ^2 -minimisation 3.4.6 Only the Beginning - A Brief Overview of More Recent Method 3.4.7 Attempts to constrain β to date 4 Introducing the ROBUST Method 4.1 Underlying Theory of ROBUST 4.2 Testing for Completeness, T_c 4.3.1 Determining the Error on β 4.3.1 Determining the Error on β 4.4 Strength of ROBUST 4.4 Strength of ROBUST 5 Applying χ^2 to the IRAS PSCz Survey Aside - choosing the appropriate software 5.1 Using Real-World Sne Sets An Alternative to Tonry - the Radburn-Smith Sne subset 5.2 Constraining β with the RS Sne Set 5.2.1 The Curiosity of Large χ^2 Values 5.2.2 Analysing Linear Regression			2.10.3 Sloan Digital Sky Survey	60
3.1 The Importance of Probing Peculiar Velocity Fields 3.2 Defining Gravitational Instability 3.3 Exploring Cosmological Parameters 3.3 Exploring Cosmological Parameters 3.3 Exploring Cosmological Parameters 3.3 The Mass Power Spectrum 3.3.1 Bulk Flow 3.3.2 The Mass Power Spectrum 3.3.3 The Linear Redshift Distortion Parameter β 3.4 Constraining β 3.4.1 Basis Function Methods 3.4.2 Iterative Method - POTENT 3.4.3 Iterative Method - VELMOD 3.4.4 Best of Both Worlds - UMV 3.4.5 Applying χ^2 -minimisation 3.4.6 Only the Beginning - A Brief Overview of More Recent Method 3.4.7 Attempts to constrain β to date 4 Introducing the ROBUST Method 4.1 Underlying Theory of ROBUST 4.2 Testing for Completeness, T_c 4.2.1 Defining T_c with Two Limits 4.3 Constraining β with ROBUST 4.3.1 Determining the Error on β 4.4 Strength of ROBUST	3	Pro	ing Peculiar Velocity Fields of the Nearby Universe	33
3.2 Defining Gravitational Instability 3.3 Exploring Cosmological Parameters 3.3 Exploring Cosmological Parameters 3.3.1 Bulk Flow 3.3.2 The Mass Power Spectrum 3.3.3 The Linear Redshift Distortion Parameter β 3.4 Constraining β 3.4.1 Basis Function Methods 3.4.2 Iterative Method - POTENT 3.4.3 Iterative Method - VELMOD 3.4.4 Best of Both Worlds - UMV 3.4.5 Applying χ^2 -minimisation 3.4.6 Only the Beginning - A Brief Overview of More Recent Method 3.4.7 Attempts to constrain β to date 4 Introducing the ROBUST Method 4.1 Underlying Theory of ROBUST 4.2 Testing for Completeness, T_c 4.2.1 Defining T_c with Two Limits 4.3 Constraining β with ROBUST 4.3.1 Determining the Error on β 4.4 Strength of ROBUST 4.4 Strength of ROBUST 5 Applying χ^2 to the IRAS PSCz Survey Aside - choosing the appropriate software 5.1 Us		3.1	The Importance of Probing Peculiar Velocity Fields	63
3.3 Exploring Cosmological Parameters 3.3.1 Bulk Flow 3.3.2 The Mass Power Spectrum 3.3.3 The Linear Redshift Distortion Parameter β 3.4 Constraining β 3.4.1 Basis Function Methods 3.4.2 Iterative Method - POTENT 3.4.3 Iterative Method - VELMOD 3.4.4 Best of Both Worlds - UMV 3.4.5 Applying χ^2 -minimisation 3.4.6 Only the Beginning - A Brief Overview of More Recent Method 3.4.7 Attempts to constrain β to date 4 Introducing the ROBUST Method 4.1 Underlying Theory of ROBUST 4.2 Testing for Completeness, T_c 4.2.1 Defining T_c with Two Limits 4.3 Constraining β with ROBUST 4.3.1 Determining the Error on β 4.4 Strength of ROBUST 5 Applying χ^2 to the IRAS PSCz Survey Aside - choosing the appropriate software 5.1 Using Real-World Sne Sets An Alternative to Tonry - the Radburn-Smith Sne subset 5.2 Constraining β with the RS Sne Set 5.2.1		3.2	Defining Gravitational Instability	64
3.3.1 Bulk Flow 3.3.2 The Mass Power Spectrum 3.3.3 The Linear Redshift Distortion Parameter β 3.4 Constraining β 3.4.1 Basis Function Methods 3.4.2 Iterative Method - POTENT 3.4.3 Iterative Method - VELMOD 3.4.4 Best of Both Worlds - UMV 3.4.5 Applying χ^2 -minimisation 3.4.6 Only the Beginning - A Brief Overview of More Recent Method 3.4.7 Attempts to constrain β to date 4 Introducing the ROBUST Method 4.1 Underlying Theory of ROBUST 4.2 Testing for Completeness, T_c 4.2.1 Defining T_c with Two Limits 4.3 Constraining β with ROBUST 4.3.1 Determining the Error on β 4.4 Strength of ROBUST 4.4 Strength of ROBUST 5 Applying χ^2 to the IRAS PSCz Survey Aside - choosing the appropriate software 5.1 Using Real-World Sne Sets An Alternative to Tonry - the Radburn-Smith Sne subset 5.2 Constraining β with the RS Sne Set 5.2.1 The		3.3	Exploring Cosmological Parameters	67
3.3.2 The Mass Power Spectrum 3.3.3 The Linear Redshift Distortion Parameter β 3.4 Constraining β 3.4.1 Basis Function Methods 3.4.1 Basis Function Methods 3.4.2 Iterative Method - POTENT 3.4.3 Iterative Method - VELMOD 3.4.4 Best of Both Worlds - UMV 3.4.5 Applying χ^2 -minimisation 3.4.6 Only the Beginning - A Brief Overview of More Recent Method 3.4.7 Attempts to constrain β to date 4.1 Underlying Theory of ROBUST 4.2 Testing for Completeness, T_c 4.2.1 Defining T_c with Two Limits 4.3.1 Determining the Error on β 4.3.1 Determining the Error on β 4.4 Strength of ROBUST 4.4 Strength of ROBUST 5 Applying χ^2 to the IRAS PSCz Survey Aside - choosing the appropriate software 5.1 Using Real-World Sne Sets An Alternative to Tonry - the Radburn-Smith Sne subset 5.2 Constraining β with the RS Sne Set 5.2.1 The Curiosity of Large χ^2 Values			3.3.1 Bulk Flow	67
3.33 The Linear Redshift Distortion Parameter β 3.4 Constraining β 3.4.1 Basis Function Methods 3.4.1 Basis Function Methods 3.4.2 Iterative Method - POTENT 3.4.3 Iterative Method - VELMOD 3.4.4 Best of Both Worlds - UMV 3.4.5 Applying χ^2 -minimisation 3.4.6 Only the Beginning - A Brief Overview of More Recent Method 3.4.7 Attempts to constrain β to date 4 Introducing the ROBUST Method 4.1 Underlying Theory of ROBUST 4.2 Testing for Completeness, T_c 4.2.1 Defining T_c with Two Limits 4.3 Constraining β with ROBUST 4.3.1 Determining the Error on β 4.4 Strength of ROBUST 4.4 Strength of ROBUST 5 Applying χ^2 to the IRAS PSCz Survey Aside - choosing the appropriate software 5.1 Using Real-World Sne Sets An Alternative to Tonry - the Radburn-Smith Sne subset 5.2 Constraining β with the RS Sne Set 5.2.1 The Curiosity of Large χ^2 Values 5			3.3.2 The Mass Power Spectrum	68
3.4 Constraining β 3.4.1 Basis Function Methods 3.4.2 Iterative Method - POTENT 3.4.3 Iterative Method - VELMOD 3.4.4 Best of Both Worlds - UMV 3.4.5 Applying χ^2 -minimisation 3.4.6 Only the Beginning - A Brief Overview of More Recent Method 3.4.7 Attempts to constrain β to date 4 Introducing the ROBUST Method 4.1 Underlying Theory of ROBUST 4.2 Testing for Completeness, T_c 4.2.1 Defining T_c with Two Limits 4.3.1 Determining the Error on β 4.3.1 Determining the Error on β 4.4 Strength of ROBUST 5 Applying χ^2 to the IRAS PSCz Survey Aside - choosing the appropriate software 5.1 Using Real-World Sne Sets An Alternative to Tonry - the Radburn-Smith Sne subset 5.2 Constraining β with the RS Sne Set 5.2.1 The Curiosity of Large χ^2 Values 5.2.2 Analysing Linear Regression			3.3.3 The Linear Redshift Distortion Parameter β	69
3.4.1Basis Function Methods3.4.2Iterative Method - POTENT3.4.3Iterative Method - VELMOD3.4.4Best of Both Worlds - UMV3.4.5Applying χ^2 -minimisation3.4.6Only the Beginning - A Brief Overview of More Recent Method3.4.7Attempts to constrain β to date4Introducing the ROBUST Method4.1Underlying Theory of ROBUST4.2Testing for Completeness, T_c 4.2.1Defining T_c with Two Limits4.3Constraining β with ROBUST4.4Strength of ROBUST5Applying χ^2 to the IRAS PSCz SurveyAside - choosing the appropriate software5.1Using Real-World Sne SetsAn Alternative to Tonry - the Radburn-Smith Sne subset5.2Constraining β with the RS Sne Set5.2.1The Curiosity of Large χ^2 Values5.2.2Analysing Linear Regression		3.4	Constraining β	71
3.4.2Iterative Method - POTENT3.4.3Iterative Method - VELMOD3.4.4Best of Both Worlds - UMV3.4.5Applying χ^2 -minimisation3.4.6Only the Beginning - A Brief Overview of More Recent Method3.4.7Attempts to constrain β to date4Introducing the ROBUST Method4.1Underlying Theory of ROBUST4.2Testing for Completeness, T_c 4.2.1Defining T_c with Two Limits4.3Constraining β with ROBUST4.3.1Determining the Error on β 4.4Strength of ROBUST5Applying χ^2 to the IRAS PSCz SurveyAside - choosing the appropriate software5.1Using Real-World Sne SetsAn Alternative to Tonry - the Radburn-Smith Sne subset5.2Constraining β with the RS Sne Set5.2.1The Curiosity of Large χ^2 Values5.2.2Analysing Linear Regression			B.4.1 Basis Function Methods	72
3.4.3Iterative Method - VELMOD3.4.4Best of Both Worlds - UMV3.4.5Applying χ^2 -minimisation3.4.6Only the Beginning - A Brief Overview of More Recent Method3.4.7Attempts to constrain β to date4Introducing the ROBUST Method4.1Underlying Theory of ROBUST4.2Testing for Completeness, T_c 4.2.1Defining T_c with Two Limits4.3Constraining β with ROBUST4.4Strength of ROBUST5Applying χ^2 to the IRAS PSCz SurveyAside - choosing the appropriate software5.1Using Real-World Sne SetsAn Alternative to Tonry - the Radburn-Smith Sne subset5.2Constraining β with the RS Sne Set5.2.1The Curiosity of Large χ^2 Values5.2.2Analysing Linear Regression			3.4.2 Iterative Method - POTENT	73
3.4.4 Best of Both Worlds - UMV 3.4.5 Applying χ^2 -minimisation 3.4.6 Only the Beginning - A Brief Overview of More Recent Method 3.4.7 Attempts to constrain β to date 4 Introducing the ROBUST Method 4.1 Underlying Theory of ROBUST 4.2 Testing for Completeness, T_c 4.2.1 Defining T_c with Two Limits 4.3 Constraining β with ROBUST 4.3.1 Determining the Error on β 4.4 Strength of ROBUST 4.4 Strength of ROBUST 5 Applying χ^2 to the IRAS PSCz Survey Aside - choosing the appropriate software 5.1 Using Real-World Sne Sets An Alternative to Tonry - the Radburn-Smith Sne subset 5.2 Constraining β with the RS Sne Set 5.2.1 The Curiosity of Large χ^2 Values 5.2.2 Analysing Linear Regression			3.4.3 Iterative Method - VELMOD	75
3.4.5 Applying χ^2 -minimisation			B.4.4 Best of Both Worlds - UMV	77
3.4.6 Only the Beginning - A Brief Overview of More Recent Method 3.4.7 Attempts to constrain β to date 4 Introducing the ROBUST Method 4.1 Underlying Theory of ROBUST 4.2 Testing for Completeness, T_c 4.2.1 Defining T_c with Two Limits 4.3 Constraining β with ROBUST 4.3.1 Determining the Error on β 4.4 Strength of ROBUST 5 Applying χ^2 to the IRAS PSCz Survey Aside - choosing the appropriate software 5.1 Using Real-World Sne Sets An Alternative to Tonry - the Radburn-Smith Sne subset 5.2 Constraining β with the RS Sne Set 5.2.1 The Curiosity of Large χ^2 Values 5.2.2 Analysing Linear Regression			B.4.5 Applying χ^2 -minimisation	79
3.4.7 Attempts to constrain β to date 4 Introducing the ROBUST Method 4.1 Underlying Theory of ROBUST 4.2 Testing for Completeness, T_c 4.2.1 Defining T_c with Two Limits 4.3 Constraining β with ROBUST 4.3.1 Determining the Error on β 4.4 Strength of ROBUST 5 Applying χ^2 to the IRAS PSCz Survey Aside - choosing the appropriate software 5.1 Using Real-World Sne Sets An Alternative to Tonry - the Radburn-Smith Sne subset 5.2 Constraining β with the RS Sne Set 5.2.1 The Curiosity of Large χ^2 Values 5.2.2 Analysing Linear Regression			3.4.6 Only the Beginning - A Brief Overview of More Recent Methods . 8	80
4 Introducing the ROBUST Method 4.1 Underlying Theory of ROBUST 4.2 Testing for Completeness, T_c 4.2.1 Defining T_c with Two Limits 4.3 Constraining β with ROBUST 4.3.1 Determining the Error on β 4.4 Strength of ROBUST 4.4 Strength of ROBUST 5 Applying χ^2 to the IRAS PSCz Survey Aside - choosing the appropriate software 5.1 Using Real-World Sne Sets An Alternative to Tonry - the Radburn-Smith Sne subset 5.2 Constraining β with the RS Sne Set 5.2.1 The Curiosity of Large χ^2 Values 5.2.2 Analysing Linear Regression			3.4.7 Attempts to constrain β to date $\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots$	82
4.1 Underlying Theory of ROBUST 4.2 Testing for Completeness, T_c 4.2.1 Defining T_c with Two Limits 4.3 Constraining β with ROBUST 4.3 Constraining the Error on β 4.4 Strength of ROBUST 4.4 Strength of ROBUST 5 Applying χ^2 to the IRAS PSCz Survey Aside - choosing the appropriate software 5.1 Using Real-World Sne Sets An Alternative to Tonry - the Radburn-Smith Sne subset 5.2 Constraining β with the RS Sne Set 5.2.1 The Curiosity of Large χ^2 Values 5.2.2 Analysing Linear Regression	4	Intr	ducing the ROBUST Method 8	35
4.2 Testing for Completeness, T_c 4.2.1 Defining T_c with Two Limits 4.2.1 Defining T_c with ROBUST 4.3 Constraining β with ROBUST 4.3 Constraining β with ROBUST 4.3.1 Determining the Error on β 4.4 Strength of ROBUST 4.4 Strength of ROBUST 5 Applying χ^2 to the IRAS PSCz Survey Aside - choosing the appropriate software 5.1 Using Real-World Sne Sets An Alternative to Tonry - the Radburn-Smith Sne subset 5.2 Constraining β with the RS Sne Set 5.2.1 The Curiosity of Large χ^2 Values 5.2.2 Analysing Linear Regression		4.1	Jnderlying Theory of ROBUST	86
4.2.1 Defining T_c with Two Limits 4.3 Constraining β with ROBUST 4.3.1 Determining the Error on β 4.4 Strength of ROBUST 5 Applying χ^2 to the IRAS PSCz Survey Aside - choosing the appropriate software 5.1 Using Real-World Sne Sets An Alternative to Tonry - the Radburn-Smith Sne subset 5.2 Constraining β with the RS Sne Set 5.2.1 The Curiosity of Large χ^2 Values 5.2.2 Analysing Linear Regression		4.2	Testing for Completeness, T_c	88
4.3 Constraining β with ROBUST 4.3.1 Determining the Error on β 4.4 Strength of ROBUST 5 Applying χ^2 to the IRAS PSCz Survey Aside - choosing the appropriate software 5.1 Using Real-World Sne Sets An Alternative to Tonry - the Radburn-Smith Sne subset 5.2 Constraining β with the RS Sne Set 5.2.1 The Curiosity of Large χ^2 Values 5.2.2 Analysing Linear Regression			4.2.1 Defining T_c with Two Limits $\ldots \ldots \ldots$	93
4.3.1 Determining the Error on β 4.4 Strength of ROBUST 5 Applying χ^2 to the IRAS PSCz Survey Aside - choosing the appropriate software 5.1 Using Real-World Sne Sets An Alternative to Tonry - the Radburn-Smith Sne subset 5.2 Constraining β with the RS Sne Set 5.2.1 The Curiosity of Large χ^2 Values 5.2.2 Analysing Linear Regression		4.3	Constraining β with ROBUST	95
4.4 Strength of ROBUST 5 Applying χ^2 to the IRAS PSCz Survey Aside - choosing the appropriate software			4.3.1 Determining the Error on β	99
5 Applying χ^2 to the IRAS PSCz Survey Aside - choosing the appropriate software		4.4	Strength of ROBUST	01
Aside - choosing the appropriate software	5	Apr	ying χ^2 to the IRAS PSCz Survey 10)5
5.1 Using Real-World Sne Sets An Alternative to Tonry - the Radburn-Smith Sne subset 5.2 Constraining β with the RS Sne Set 5.2.1 The Curiosity of Large χ^2 Values 5.2.2 Analysing Linear Regression		Asid	- choosing the appropriate software	05
An Alternative to Tonry - the Radburn-Smith Sne subset5.2Constraining β with the RS Sne Set5.2.1The Curiosity of Large χ^2 Values5.2.2Analysing Linear Regression		5.1	Jsing Real-World Sne Sets	07
5.2 Constraining β with the RS Sne Set			An Alternative to Tonry - the Radburn-Smith Sne subset	07
5.2.1 The Curiosity of Large χ^2 Values		5.2	Constraining β with the RS Sne Set $\ldots \ldots \ldots$	08
5.2.2 Analysing Linear Regression			5.2.1 The Curiosity of Large χ^2 Values	08
			5.2.2 Analysing Linear Regression	11

	5.3	A More Recent Alternative - the Unified Supernova Catalogue \ldots Constraining β with the USC Sne Set \ldots \ldots	. 116 . 117
6	Pro	bing the Peculiar Velocity Field of the PSCz with ROBUST	121
	6.1	Implementing ROBUST with the PSCz Survey	. 121
	6.2	Applying Mock Method 1a) - 'Faint Limit Only Gaussian' Mock Magnitude	s123
	6.3	Applying Mock Method 1b) - 'Faint Limit Only Schechter' Mock Magnitude	s132
		6.3.1 Defining the Schechter Luminosity Function	. 132
	6.4	Applying Mock Method 2a) - 'Faint+Bright Limit Gaussian' Mock Mag-	
		nitudes	. 139
		6.4.1 A question of ΔZ	. 140
	6.5	Applying Mock Method 2b) - 'Faint+Bright Limit Schechter' Mock Mag-	
		nitudes	. 149
7	App	olying ROBUST to the IRAS PSCz Survey	159
	7.1	Making use of MC Simulations and Bootstrap Resampling	. 159
	7.2	Preparing the PSCz for use with ROBUST	. 166
	7.3	The problem of magnitude-galaxy matching	. 168
8	Pro	bing the Peculiar Velocity Field of 2MASS with ROBUST	171
	8.1	An aside: Thinking in cubes	. 171
	8.2	Carrick et. al and the 2M++ velocity field	. 174
	8.3	Preparing 2MASS for use with ROBUST in conjunction with $2M++$. 176
	8.4	An aside: Distance convergence loops	. 177
	8.5	Identifying the magnitude limits of the 2MRS survey	. 180
	8.6	Applying ROBUST	. 187
		8.6.1 Redshift Restricting the 2MRS	. 189
		8.6.2 Restricting cz_{corr} instead of cz_{obs}	. 192
		8.6.3 Experimenting with μ instead of u_{β}	. 196
		8.0.4 Communing the χ^2 results of Carrick	. 200
9	AR	OBUST Alternative - Relative Entropy	207
	9.1	A N _ Chair is a label in D latin E the	. 208
	9.2	A New Statistic, χ , and Applying Relative Entropy	. 210 019
	9.5	Determining the Error on β	. 213
10	Pro	bing the Peculiar Velocity Field of the PSCz with Relative Entropy	215
	10.1	Implementing Relative Entropy with the PSCz Survey	. 215
	10.2	Applying Mock Method 1a) - 'Faint Limit Only Gaussian' Mock Magnitude	s218
		10.2.1 The curiosity of small $S[(\zeta, \chi)]$ values	. 218
	10.2	10.2.2 A question of grid resolution size	. 222
	10.3	Applying Mock Method 1b) - 'Faint Limit Only Schechter' Mock Magnitude	s229
	10.4	Applying Mock Method 2a) - 'Faint+Bright Limit Gaussian' Mock Mag-	99 <i>6</i>
	10 5	Applying Mock Method 2b) (Faint + Bright Limit Schechter' Mock Mer	. 200
	10.9	nitudes	240
			. 2-10

11 Applying Relative Entropy to 2MASS

 $\mathbf{245}$

	11.1	Redshift Restricting the 2MRS	5
	11.2	Restricting cz_{corr} instead of cz_{obs}	9
	11.3	Experimenting with μ instead of u_{β}	1
12	Exp	erimenting with Distance Dependent Luminosity Functions 25	5
	12.1	Applying a Distance Dependent Luminosity Function	3
		12.1.1 IRAS PSCz Survey	3
		12.1.2 Redshift Restricted 2MRS	3
		12.1.3 2MRS: Using μ_{β} instead of u_{β}	3
	12.2	Applying a Universal Luminosity Function	7
		12.2.1 Redshift Restricted 2MRS	7
		12.2.2 2MRS: Using μ_{β} instead of u_{β}	1
13	Con	clusions and Future Avenues 279	9
	13.1	More Recent All-Sky Redshift-Distance Surveys to Consider for Future Use28	3
		13.1.1 Cosmic-flows3	3
		13.1.2 GLADE	7
		13.1.1 Cosmic-nows3	о 7

Bibliography

289

List of Figures

1.1	Velocity-Distance Relation among Extra-Galactic Nebulae. Radial veloc- ities, corrected for solar motion, are plotted against distances estimated from involved stars and mean luminosities of nebulae in a cluster. The black discs and full line represent the solution for solar motion using the nebulae individually; the circles and broken line represent the solution combining the nebulae into groups; the cross represents the mean veloc- ity corresponding to the mean distance of 22 nebulae whose distances	
	could not be estimated individually [84].	4
1.2	Images of the Hubble Deep Field and Hubble Deep Field South. Though both images represent deep observations of small regions of space in the northern and southern hemispheres respectively, the similarities between	
	the images lend credence to the concept of a homogeneous Universe	7
1.3	Images of the Hubble Ultra Deep Field and Hubble eXtreme Deep Field. Much like the photos of the HDF and HDF-S, the similarities between the images continue to lend credence to the concept of a homogeneous	
	Universe, especially holding true even as we probe very deep into space	
	and far back in time with the XDF.	8
1.4	Depiction of the three possible Universal curvatures in relation to different densities of the Universe, denoted as Ω_0 here. The density of the Universe also determines its geometry. If the density of the Universe exceeds the critical density (see Section 1.4 for an introduction to that variable), then the geometry of space is closed and positively curved like the surface of a sphere. This implies that initially parallel photon paths converge slowly, eventually cross, and return back to their starting point (if the Universe lasts long enough). If the density of the Universe is less than the critical density, then the geometry of space is open, negatively curved like the surface of a saddle. If the density of the Universe is flat like a sheet of paper. Thus, there is a direct link between the geometry of the universe and its fate. Image reproduced from the works of NASA and the WMAP team [138].	13
1.5	Rotational velocities for M31, as a function of distance from the centre of	10
	the galaxy as provided from the works of Rubin and Ford [165]	19
1.6	A slice from the CfA galaxy redshift survey completed in 1982, depict- ing the affectionately named 'Matchstick Man' in a non-random, non- isotropic observed distributon of galaxies, apparent filaments and voids	
	[86].	21
1.7	Pie chart depicting the distribution of baryonic matter, dark matter and dark energy in the Universe, as provided by Spergel et. al [182] and the	
	team behind WMAP at NASA [206]	23

1.8	The evolution in detail of the CMBR over the past decade	26
2.1	A typical B-band light curve for Type Ia supernovae, based on the observations of 22 SNe Ia [18]. The timescale used here is in units of days with the y-axis denoting changes in observed magnitude over time. Typically approximately 10 days after the initial explosion (the beginning of the curve) the peak B-band luminosity of the supernova is observed, denoted at $t_B=0$. The constraining of errors and the breadth of this luminosity peak is integral for accurate distance estimations to such supernovae. This peak is maintained for several days before the supernova starts to dim considerably as the explosive energy is radiated into space, denoted by the steep plunge in the curve after $t_B \approx 15$, eventually flattening out into a continual linear decline in brightness that has been observed to continue out to a t_B of 300 to 400 days [18]	30
2.2	The TF relations constructed from the H-band and I-band data of Bothun et. al [16] and Pildis, Schombert and Eder respectively [150]. The usual TF relation is apparent for massive galaxies in panel (a), however it breaks down at the low mass end. Panel (b) exhibits no 'H I TF relation' for massive galaxies, though there does appear to be a slight one for lower mass objects. The Baryonic TF relation which follows by summing stellar and gas mass nicely recovers a continuous relation over the entire observed mass range as seen in panel (c), however utilising lower mass-to-light ratios for the galaxy stars in panel (d) causes a noticeable discontinuity in slope, implying that the higher mass-to-light ratios adopted in panel (c) are more appropriate. Figure has been reproduced from the works of McCaugh [120]	42
2.3	Diagram of the celestial sphere with respect to the ecliptic. The ecliptic is at an angle of 23.5° to the celestial equator (otherwise known as the Earth's visible horizon), indicative of the tilt of the Earth and the conse- quent path the Sun draws along the celestial sphere throughout the year as a result. Right ascension and declination angles (otherwise known as equatorial coordinates) are measured with respect to the celestial equa- tor with the vernal (spring) equinox as the origin point, while celestial latitude and longitude angles are measured with respect to the ecliptic,	42
2.4	again with the vernal equinox as the origin point [135] Diagram of the galactic sphere with respect to the celestial sphere, specifically the celestial equator. The centre of the galaxy at RA=17h 45.6m and DEC=-28.94° acts as the origin point of the plane. With the galactic North Pole located at RA=12h 51.4m and DEC=+27.13° [12], this causes the galactic plane as whole to be inclined at an angle of approximately 63° to the celestial equator [198]	46 47

2.5	The distribution of galaxies projected on the sky in the IRAS and ORS samples, with the supergalactic plane being seen prominently as a thick clustering line along the centre of the great circle. This is an Aitoff projection in supergalactic coordinates, with SGL=90°; SGB=0° (close to the Virgo cluster) in the centre of the map. Objects within 2000kms ⁻¹ are shown as circled crosses; objects between 2000 and 4000kms ⁻¹ are	
	indicated as crosses, and dots mark the positions of more distant objects. Only catalogued galaxies from both the IRAS and ORS samples are used, leading to very prominent zones of avoidance in both figures. Figure re- produced from the work of Lahay et. al [112].	49
2.6	All-sky Aitoff projection map showing the positions of ~5700 galaxies with measured redshifts less than 3000 kms ⁻¹ . The figure is plotted in equatorial coordinates, with the center of the figure at 6h and 0°. The dotted lines denote the location of the galactic plane and lines of $b=\pm 20^{\circ}$. The Virgo cluster is the densest collection of objects on the left side of the figure. The supergalactic plane runs through Virgo, and can be traced most of the way around the sky. Figure reproduced from the works of	10
	Salzer and Havnes [167].	50
2.7	Distribution of nearby groups of galaxies over the celestial sphere in Supergalactic coordinates. Group of galaxies shown here are within 10-16	
	Supergalactic plane (horizontal line) [59]	50
2.8	Probability contours for the most statistically viable values of Ω_{Λ} vs. Ω_{M}	50
	as obtained using the Type Ia supernovae of the Tonry et. al dataset (black lines), presented at the 1, 2 and 3σ confidence levels respectively. The adopted a prior assumption of $\Omega_M h=0.20\pm0.03$ from the 2dF galaxy redshift survey and the works of Percival et. al [148] is also presented in grey out to the same number of confidence contours. The assumption of a flat Universe is depicted via the dashed straight line that restricts the summation of Ω_M and Ω_{Lambda} from exceeding 1. The intersection of both these sets of probability contours and our flat Universe assumption dashed line lends itself to values of $\Omega_M \sim 0.3$ and $\Omega_\Lambda \sim 0.7$, in very good agreement with the values obtained by WMAP. Figure reproduced from the works of Tonry et. al [189]	59
2.9	Evolution of the Union compilations from Union1 (2008) to Union2.1 (2012) and the various datasets integrated into the compilations over time.	52
2.10	Figures obtained from the Supernova Cosmology Project [185] Probability contours for the most statistically viable values of Ω_{Λ} vs. Ω_{M} as obtained using the Union2.1 compilation (blue), presented at the 1, 2 and 3σ confidence levels respectively. Previous attempts to constrain parameters by means of the CMBR and BAOs are also presented to the same number of confidence contours in orange and green respectively. The assumption of a flat Universe is depicted via the dashed straight line that restricts the summation of Ω_{M} and Ω_{Lambda} from exceeding 1. The intersection of all three of probability contours with the flat Universe assumption dashed line continues to lend itself to values of $\Omega_{M} \sim 0.3$ and $\Omega_{\Lambda} \sim 0.7$, in very good agreement with the values obtained by WMAP.	53
	Figure obtained from the Supernova Cosmology Project [185]	54

2	.11	An Aitoff projection of the distribution of galaxies in the IRAS PSCz catalogue as a function of distance. The gaps in the map are due to the Milky Way obscuring that region of the sky and is termed the 'Zone of Avoidance'. The IRAS PSCz is considered complete and uniform in flux density to within a few parameters to high galaxies latitudes and to within	
2	.12	10% at low galactic latitudes otherwise [170]	58
2	.13	Current sky coverage of the SDSS DR13 when combined with the DR12 and all previous data runs as of 2015 [175].	61
2	.14	Current sky coverage of the SDSS as of 2018. Sky coverage of all pre- vious data runs including DR13 are included here, with DR14 additions	0.1
		represented in blue [176].	61
4	.1	Plot of the variable Z plotted vs. absolute magnitude M for a theoretical galaxy survey. The procedure for calculating r_i and n_i in Equations 4.8 and 4.9 is also illustrated. Figure reproduced from the work of Bauzy [159]	89
4	.2	ζ -Z diagram for two values of the limiting apparent magnitude m^* . For m^* greater than the completeness limit m_{lim} , the number of galaxies fainter than $M_{lim}(Z)$ is underestimated due to incompleteness, inferring a systematic lack of points (or galaxies) with a value of $\hat{\zeta}_i$ close to unity. This effect is particularly visible at high Z (i.e. distant galaxies). Figure	
4	.3	reproduced from the works of Rauzy [159]	91
		in both sets. Figure reproduced from the work of Rauzy [159]	92
4	.4	Schematic illustrating the construction of regions S_1 and S_2 when a bright limit is introduced. These regions are uniquely defined for a slice of fixed width, δZ , in corrected distance modulus, and for 'trial' bright and faint apparent magnitude limits m_*^b and m_*^f , respectively. Also shown are the true bright and faint apparent magnitude limits m_{lim}^b and m_{lim}^f , within which the rectangular regions S_1 and S_2 contain a joint distribution of M and Z that are separable (i.e. independent). Figure reproduced from the works of Johnston et. al [98].	93
4	.5	Performance of the T_c statistic applied to the SDSS-DR1 early-type ellip- tical galaxies. In the lefthand panel T_c is computed by fixing the bright magnitude limit to equal the published value of 14.55mag. The behaviour of T_c is consistent magnitude completeness up to and including the pub- lished faint limit of 17.45mag with the statistic dropping rapidly thereafter - indicating the onset of incompleteness. In the righthand panel T_c is com- puted using only a faint limit. T_c deviates significantly from its expected value for a complete set at magnitudes which are much brighter than the published faint magnitude limit, though T_c exhibits an even stronger freefall one the published faint limit is exceeded. Figure reproduced from the works of Jonhston et. al [98].	95
			00

4.64.74.8	Observed coefficient of correlation $\rho(\zeta_{\beta}, u_{\beta})$ for the IRAS 1.2Jy survey sample as a function of the parameter β , returning a value of $\beta^*=0.1$ - 0.15 as the value for which ζ and the underlying peculiar velocity field u_{β} are completeley separable. Confidence levels of rejection on β at the 1, 2 and 3σ levels are denoted in dotted lines, and the means by which these confidence levels are constructed are discussed in Section 4.3.1. Figure reproduced from the works of Rauzy & Hendry [160]	. 98 . 99 . 100
5.1	Flowchart illustrating typical $\chi^2\text{-minimisation}$ methodology using real	
5.2	world data. Results of utilising real-world Sne sets to constrain β for the PSCz. Upper panel depicts the χ^2 plot indicating the optimal value of β , with χ^2_{min} at $\beta = 0.553 \pm 0.04$ providing the optimal value of β at the 1σ error level. The lower panel depicts the comparison between observed and interpolated velocities. Areas of individual plot points are proportional to the associated reconstruction errors of each individual Sne used, with the red line indicating the linear regression (or goodness-of-fit) between	. 106
5.3	both sets of velocities	. 109
5.4	Results of utilising the USC Sne set to constrain β for the PSCz. Upper panel depicts the χ^2 plot indicating the optimal value of β , with χ^2_{min} at $\beta = 0.444 \pm 0.04$ providing the optimal value of β at the 1σ error level. The lower panel depicts the comparison between observed and interpolated velocities. Areas of individual plot points are proportional to the associated reconstruction errors of each individual Sne used, with the red line indicating the linear regression (or goodness-of-fit) between both sets of velocities.	. 112
6.1	Flowchart illustrating typical ROBUST methodology for use with gener- ated mock magnitudes assigned to galaxies utilising either just a faint, or faint+bright magnitude limits during generation. Note that m_f and m_b stands for faint and bright apparent magnitude limit respectively [Meth- ods 1a), 1b), 2a) and 2b)]	. 124

6.2	Flowchart illustrating typical mock magnitude generation and assignment	25
6.3	Initial results of Method 1a). Upper panel depicts the ρ plots for all	120
	50 mocks generated, lower panel depicts the zero-intercept of each mock	
	plot as a function of trial value of β plotted as a filterarm with best-fit Gaussian (red line) modelled to the data. Optimal value of β returned	
	for Method 1a) is $\beta = 0.55 \pm 0.06$, as determined from the Gaussian	
	model. Luminosity function used during data generation is Gaussian of	
	the form $N[-21,1]$	27
6.4	Results of Method 1a) redone with 200 mocks instead of 50. Upper panel	
	depicts the ρ plots for all 200 mocks generated, lower panel depicts the zero intercept of each mock plot as a function of trial value of β plotted	
	as a histogram with best-fit Gaussian (red line) modelled to the data.	
	Optimal value of β returned for Method 1a) is $\beta = 0.55 \pm 0.06$, as	
	determined from the Gaussian model. Luminosity function used during	
~ ~	data generation is Gaussian of the form $N[-21,1]$	128
6.5	Method 1a) results with varied mock set sizes. Left-hand panels illustrate	
	of optimal value of β with best-fit Gaussian (red line). Set sizes used	
	are noted on each plot. Luminosity function generated continues to be	
	Gaussian of form N[-21,1]	29
6.6	Method 1a) results with varied mock generated Gaussian LF widths. Left-	
	hand panels illustrate the ρ plots for all 50 mocks, right-hand panels depict the histograms of optimal value of β with best-fit Gaussian (red line). LF	
	widths generated are noted on each plot. Entirety of the PSCz survey is	
	utilised for this analysis.	31
6.7	A sample Schechter distribution with $M^*=-21$ and slope $\alpha=1.09$, with	
68	the CDF generated from such a distribution normalised on $[0,1]$ Initial regults of Method 1b). Upper papel depicts the conjust for all	33
0.8	50 mocks generated, lower panel depicts the zero-intercept of each mock	
	plot as a function of trial value of β plotted as a histogram with best-fit	
	Gaussian (red line) modelled to the data. Optimal value of β returned	
	for Method 1b) is $\beta = 0.55 \pm 0.01$, as determined from the Gaussian model. Lyminosity function used during data generation is Schechter of	
	the form estimated $S[-21,1.09]$	135
6.9	Method 1b) results with varied mock set sizes. Left-hand panels illustrate	
	the ρ plots for all 50 mocks, right-hand panels depict the histograms	
	of optimal value of β with best-fit Gaussian (red line). Set sizes used	
	Schechter of form estimated S[-21.1.09].	136
6.10	Method 1b) results with varied mock Schechter power law slopes α . Left-	
	hand panels illustrate the ρ plots for all 50 mocks, right-hand panels	
	depict the histograms of optimal value of β with best-fit Gaussian (red	
	$\operatorname{Inte}_{\mathcal{I}}$, α values generated are noted on each plot. Entirety of the PSCZ survey is utilised for this analysis.	37
6.11	Sample scatter plots of the ROBUST variable ζ plotting against distance	
	moduli μ for the case where the correct magnitude limits have been de-	
	fined (top panel) and where the incorrect limits have been defined (bottom	
	paner)	-39

6.126.13	Schematic illustrating the construction of regions S_1 and S_2 when a bright limit is introduced. These regions are uniquely defined for a slice of fixed width, δZ , in corrected distance modulus, and for 'trial' bright and faint apparent magnitude limits m_*^b and m_*^f , respectively. Also shown are the true bright and faint apparent magnitude limits m_{lim}^b and m_{lim}^f , within which the rectangular regions S_1 and S_2 contain a joint distribution of M and Z that are separable (i.e. independent). Figure reproduced from the works of Johnston et. al [98]	. 141
	Gaussian (red line) modelled to the data. Optimal value of β returned for Method 2a) is $\beta = 0.56 \pm 0.07$, as determined from the Gaussian model. Luminosity function used during data generation is Gaussian of the form N[-21,1], with m_f and m_b given as 14.3 and 8 respectively. Bright limit was applied arbitrarily during S_1 , S_2 set generation.	. 142
6.14	Scatter plot of ζ vs. μ for a faint limit of 14.3 and an arbitrary bright	1.10
6 15	limit of 8	. 143
0.10	an assumed bright limit of 8 and $\Delta Z=5$, binned into 10 intervals on [0,1]. The thick red line denotes our expected number of galaxies per histogram bin for the distribution to be uniform	1/13
6.16	Method 2a) results with varied bright limits applied. Left-hand panels	. 140
	illustrate the ρ plots for all 50 mocks, right-hand panels depict the his- tograms of optimal value of β with best-fit Gaussian (red line). LF is of the form N[-21,1], with m_f and m_b noted on each plot. Entirety of the PSCz is utilised for this analysis with $\Delta Z=5$.	. 145
6.17	Monitoring of ζ behaviour with varied Gaussian bright limit. Left-hand panels illustrate the (ζ, μ) scatter plot for all galaxies used, right-hand panels depict the histograms of ζ value distribution over [0,1]. The thick red line denotes our expected number galaxies per histogram bin for a	
6.18	uniform distribution	. 146
	is applied in the presence of both a faint and bright limit (represented by blue lines), creating potential 'imbalance' and possibly affecting the separability of M and Z within those regions for a randomly selected galaxy on the M - Z plane, highlighted in green	147
6.19	Method 2a) results with varied ΔZ applied. Left-hand panels illustrate the ρ plots for all 50 mocks, right-hand panels depict the histograms of optimal value of β with best-fit Gaussian (red line). LF is of the form N[-21 1] with m_{c} m, and ΔZ used noted on each plot. Entirety of the	. 1.11
	PSCz is utilised for this analysis	. 148

6.20	Initial results of Method 2b). Upper panel depicts the ρ plots for all 50 mocks generated, lower panel depicts the zero-intercept of each mock plot as a function of trial value of β plotted as a histogram with best-fit Gaussian (red line) modelled to the data. Optimal value of β returned for Method 2a) is $\beta = 0.53 \pm 0.01$, as determined from the Gaussian model. Luminosity function used during data generation is Schechter of the form S[-21,1.09], with m_f and m_b given as 14.3 and 7 respectively. Bright limit was applied retroactively during S_1 , S_2 set generation.	. 150
6.21	Scatter plot of ζ vs. μ for a faint limit of 14.3 and an assumed bright limit of 7 for mock generated Schechter magnitudes.	. 151
6.22	Histogram of ζ scatter points generated from a fixed faint limit of 14.3, an assumed bright limit of 7 and $\Delta Z=1$, binned into 10 intervals on [0,1]. The thick red line denotes our expected number of galaxies per histogram bin for the distribution to be uniform	151
6.23	Method 2b) results with varied bright limits applied. Left-hand panels illustrate the ρ plots for all 50 mocks, right-hand panels depict the histograms of optimal value of β with best-fit Gaussian (red line). LF is of the form S[-21,1.09], with m_f and m_b noted on each plot. Entirety of the PSCz is utilised for this analysis with $\Delta Z=5$	153
6.24	Monitoring of ζ behaviour with varied Schechter bright limit. Left-hand panels illustrate the (ζ, μ) scatter plot for all galaxies used, right-hand panels depict the histograms of ζ value distribution over [0,1]. The thick red line denotes our expected number of galaxies per histogram bin for a	. 100
6.25	uniform distribution	. 154
7.1	Flowchart illustrating typical ROBUST methodology using real world data and magnitude-position galaxy matching. Note that m_B indicates B-band apparent magnitude.	. 160
7.2	Computed CDFs for $\rho(\zeta, \beta)$ utilising one mock trial with the symmetrical 68%, 95% and 99.7% confidence intervals marked off with blue, red and green dotted lines respectively. Top panel CDF is generated from applying MC simulations, bottom panel CDF is generated from applying bootstrap resampling. LF function utilised is Gaussian of the form N[-21,1] with no bright limit applied.	. 162
7.3	Computed CDFs for $\rho(\zeta, \beta)$ utilising one mock trial with the symmetrical 68%, 95% and 99.7% confidence intervals marked off with blue, red and green dotted lines respectively. Top panel CDF is generated from applying MC simulations, bottom panel CDF is generated from applying bootstrap resampling. LF function utilised is Schechter of the form S[-21,1.09] with no bright limit applied.	. 163

Extrapolated confidence intervals on β utilising values of $\rho(\zeta, \beta)$ generated from one mock trial with the symmetrical 68%, 95% and 99.7% confidence intervals marked off with blue, red and green dotted lines respectively. Top panel intervals are generated from applying MC simulations, bottom panel intervals are generated from applying bootstrap resampling. LF function utilised is Gaussian of the form N[-21 1] with no bright limit	
applied. Entirety of the PSCz was utilised for this analysis Extrapolated confidence intervals on β utilising values of $\rho(\zeta, \beta)$ generated from one mock trial with the symmetrical 68%, 95% and 99.7% confidence	. 164
intervals marked off with blue, red and green dotted lines respectively. Top panel intervals are generated from applying MC simulations, bottom panel intervals are generated from applying bootstrap resampling. LF function utilised is Schechter of the form S[-21,1.09] with no bright limit	
applied. Entirety of the PSCz was utilised for this analysis Aitoff projection of the Enzo Branchini PSCz data set - blue dots, against	. 165
the original survey retrieved via VizieR - green crosses	. 167
Embedding a blank mesh grid into an existing galaxy survey to begin generating a 3D velocity field map. Basic map grid points are indicated in green, while a random selection of galaxies within 200kms ⁻¹ of the survey are indicated in blue.	. 172
A selection of galaxies (marked in blue) that, due to their positions, reside within a certain mini-cubelet marked in solid green	174
Flowchart for a typical distance convergence loop	170
M-Z plots for the 2MRS galaxies in the K (top panel). H (middle panel)	. 110
and J infrared bands respectively (bottom panel)	. 181
T_c plots identifying the faint magnitude limits in the K, H and J in- frared bands for the 2MRS survey. Magnitude limits at the 3σ confidence level are marked off via the dashed orange lines for all bands, given as	
$K_{lim}=11.55, H_{lim}=11.86 \text{ and } J_{lim}=12.55 \text{ respectively.}$. 182
Monitoring of ζ behaviour at the K (top panels), H (middle panels) and J limits returned by T_c (bottom panels). Left-hand panels illustrate the (ζ, μ) scatter plot for all galaxies used, right-hand panels depict the histograms of ζ value distribution over [0,1]. The thick red line denotes our	
expected number of galaxies per histogram bin for a uniform distribution χ^2 track of ζ value distribution vs. expected number of values per histogram bin across varying values of apparent magnitude for the K (solid blue), H (solid red) and J (solid green) infrared bands of the 2MRS respectively. KS cut-off points after which the null hypothesis of uniform	. 183
mity is rejected at the 95% confidence level for all bands is represented by the associated coloured dashed vertical line. KS limits are given as	
$K_{lim} = 11.56$, $H_{lim} = 11.89$ and $J_{lim} = 12.55$.	. 184
Flowchart for the application of data slicing, distance convergence loops 1 DODUST at the 1 lass for the 20 UPC	100
and ROBUST methodology for the 2MRS survey	. 186
shown in blue, red and green respectively. \dots	. 187
Monitoring of (ζ, μ) distribution of the 2MRS galaxies in the K band for various trial values of β . The trial values of β used are noted on each panel	<mark>el</mark> .188
	Extrapolated confidence intervals on β utilising values of $\rho(\zeta, \beta)$ generated from one mock trial with the symmetrical 68%, 95% and 99.7% confidence intervals marked off with blue, red and green dotted lines respectively. Top panel intervals are generated from applying MC simulations, bottom panel intervals are generated from applying bootstrap resampling. LF function utilised is Gaussian of the form N[-21,1] with no bright limit applied. Entirety of the PSCz was utilised for this analysis Extrapolated confidence intervals on β utilising values of $\rho(\zeta, \beta)$ generated from one mock trial with the symmetrical 68%, 95% and 99.7% confidence intervals are generated from applying MC simulations, bottom panel intervals are generated from applying bootstrap resampling. LF function utilised is Schechter of the form S[-21,1.09] with no bright limit applied. Entirety of the PSCz was utilised for this analysis Aitoff projection of the Enzo Branchini PSCz data set - blue dots, against the original survey retrieved via VizieR - green crosses A selection of galaxies (marked in blue) that, due to their positions, reside surther a typical distance convergence loop

8.11	One-sided confidence intervals calculated on β via bootstrap resampling
	(top panel) and Monte Carlo simulations (bottom panel) for the K, H and
	J bands depicted in blue, red and green respectively. The 3σ confidence
	intervals for all three bands are denoted via the black dashed lines, indi-
	cating all values of $\beta > 0.64$ are to be excluded from consideration for
	the 2MRS survey with 99.7% confidence
8.12	Plots of ρ for the redshift restricted 2MRS survey across the K, H and J
	infrared bands, shown in blue, red and green respectively
8.13	Monitoring of (ζ, μ) distribution of the redshift restricted 2MRS galaxies
	in the K band for various trial values of β . The trial values of β used are
014	
8.14	One-sided confidence intervals calculated on β via bootstrap resampling (tap panel) and Mapta Carla simulations (bottom panel) for the K. H and
	Liberds depicted in blue, red and green respectively. The 2σ confidence
	intervals for all three bands are denoted via the black dashed lines indi-
	cating all values of $\beta > 0.6$ are to be excluded from consideration for the
	redshift restricted 2MRS survey with 99.7% confidence
8.15	Plots of ρ for the restricted 2MRS survey utilising corrected redshifts
	across the K, H and J infrared bands, shown in blue, red and green
	respectively
8.16	Monitoring of (ζ, μ) distribution of the restricted 2MRS galaxies utilising
	corrected redshifts in the K band for various trial values of β . The trial
	values of β used are noted on each panel
8.17	Plots of ρ for the redshift restricted 2MRS survey where β -dependent
	distance moduli μ_{β} were utilised in the calculation of $\rho(\zeta, \beta)$ instead of
	β -dependent peculiar velocities u_{β} across the K, H and J infrared bands,
	shown in blue, red and green respectively
8.18	Monitoring of (ζ, μ) distribution of the redshift restricted 2MRS galaxies
	where β -dependent distance moduli μ_{β} were utilised in the calculation of μ_{β} instead of β dependent moduli probability and it is the <i>K</i> hand for
	$\rho(\zeta,\beta)$ instead of β -dependent peculiar velocities u_{β} in the K band for
9 10	Various trial values of β . The trial values of β used are noted on each panel. 198 One sided confidence intervals calculated on β via boststrap recompling
0.19	(top papel) and Monte Carlo simulations (bottom papel) for the K H
	and I bands depicted in blue red and green respectively. The 3σ confi-
	dence intervals for all three bands are denoted via the black dashed lines
	indicating all values of $\beta \sim > 0.1$ are to be excluded from consideration
	for the redshift restricted 2MRS survey in the K and H band where μ_{β} is
	used to calculate $\rho(\zeta,\beta)$ instead of u_{β} with 99.7% confidence
8.20	Successful recovery of the value of $\beta = 0.445 \pm 0.04$ via χ^2 -minimisation
	using Method 1 (top panel), alongside a comparison of RS Sne velocities
	computed via interpolation in Method 1 $(x-axis)$ with those computed
	using the Hubble Law $(y$ -axis). Areas of individual plot points are pro-
	portional to the associated reconstruction errors of each individual Sne
	used, with the red line indicating the linear regression (or goodness-of-fit)
	between both sets of velocities

8.21	Successful recovery of the value of $\beta = 0.426 \pm 0.035$ via χ^2 -minimisation using Method 2 (top panel), alongside a comparison of RS Sne velocities computed via interpolation in Method 2 (x-axis) with those computed using the Hubble Law (y-axis). Observed 'squashing effect' is due to the existence of a singular outlier in the data. Areas of individual plot points are proportional to the associated reconstruction errors of each individual Sne used, with the red line indicating the linear regression (or goodness- of-fit) between both sets of velocities	. 203
9.1	Example illustrating a typical (ζ, χ) distribution for a complete data-set for an MGC mock catalogue. The left-hand distribution shows ζ and χ estimated at the apparent magnitude limit of the survey $m_{lim} = 20.0$ and appears to be a random uniform distribution. Correlations between ζ and χ are shown on the right-hand panel where ζ and χ have been estimated at $m_* = 20.5$ (beyond the limit of the survey). Figure reproduced from the work of Johnston [96].	. 211
9.2	Example of measuring the entropy of a typical (ζ, χ) distribution for a complete data-set taken at the $m_* = m_{lim} = 20.0$ of an MGC mock catalogue. We calculate the total entropy of this distribution by imposing a grid with a predetermined mesh size. In this example we have split the grid into 0.1×0.1 mesh. We then count the number of objects contained in each box, p_i , and calculate the relative entropy. We then determine the total entropy by summing each $p_1, p_2, \dots, p_i, \dots, p_n$. Figure reproduced from the work of Johnston [96]	212
9.3	Relative entropy S , vs. trial $\hat{\beta}$ (in this instance a statistic used as a measure of galaxy evolution) using MGC mocks for three different values of assumed 'true' β_{true} . For each trial $\hat{\beta}$ we have applied 50 mock realisations with successive values of $\beta_{true} = 1.0, 2.0$ and 3.0. In each case we can clearly see that the relative entropy approach minimises at the correct value. Figure reproduced from the work of Johnston [96].	. 213
10.1	Flowchart illustrating typical ROBUST/relative entropy methodology for use with generated mock magnitudes assigned to galaxies utilising either just a faint, or faint+bright magnitude limits during generation. Note that m_f and m_b stands for faint and bright apparent magnitude limit respectively [Methods 1a] (1b) (2a) and (2b)]	017
10.2	respectively [Methods 1a), 1b), 2a) and 2b)]	. 217

10.3 Three different instances of 5 randomly selected PSCz galaxies movements are being tracked in both the M -Z (left side panels) at ζ - χ planes (right side panels) respectively. An arbitrary mesh size was applied to the ζ - χ grid, as denoted by the grey lines. Each denotes the position of each galaxy for a value of β_{trial} being inc	whose nd the e of 10 circle reased
incrementally on the interval [0,1], with ten values of β_{trial} being u total	sed in 221
10.4 Results of varying grid resolution during use in relative entropy. Left panels illustrate the $S[(\zeta, \chi)]$ plots for all 50 mocks, right-hand depict the histograms of optimal value of β with best-fit Gaussia line). Grid resolutions applied are noted on each plot. Entirety PSCz survey is utilised during analysis.	-hand panels n (red of the
10.5 Monitoring of $\zeta - \chi$ behaviour with varied grid resolution. Panels the bivariate histograms of $\zeta - \chi$ value distribution over [0,1]. The red line denotes our expected number galaxies per histogram bin uniform distribution.	depict e thick a for a
10.6 Method 1a) results with varied mock set sizes. Left-hand panels trate the $S[(\zeta, \chi)]$ plots for all 50 mocks, right-hand panels depi histograms of optimal value of β with best-fit Gaussian (red line sizes used are noted on each plot. Luminosity function generated on use to be Gaussian of form N[-21,1].	illus- ct the). Set contin- 226
10.7 Monitoring of $\zeta - \chi$ behaviour with varied mock set size. Panels the bivariate histograms of $\zeta - \chi$ value distribution over [0,1]. The red line denotes our expected number galaxies per histogram bin	depict e thick for a
10.8 Method 1a) results with varied mock Gaussian LF widths. Left panels illustrate the $S[(\zeta, \chi)]$ plots for all 50 mocks, right-hand depict the histograms of optimal value of β with best-fit Gaussia line). LF widths generated are noted on each plot. Entirety of the	227 -hand panels n (red PSCz
 survey is utilised for this analysis. 10.9 Monitoring of ζ-χ behaviour with varied mock Gaussian LF widths els depict the bivariate histograms of ζ-χ value distribution over The thick red line denotes our expected number galaxies per hist 	228 . Pan- . [0,1]. . ogram
bin for a uniform distribution	
10.11Method 1b) results with varied mock set sizes. Left-hand panels trate the $S[(\zeta, \chi)]$ plots for all 50 mocks, right-hand panels depi histograms of optimal value of β with best-fit Gaussian (red line) available. Set sizes used are noted on each plot. Luminosity fu	illus- ct the when nction
generated continues to be Schechter of form S[-21,1.09]. 10.12Monitoring of ζ - χ behaviour with varied mock set size. Panels the bivariate histograms of ζ - χ value distribution over [0,1]. The red line denotes our expected number galaxies per histogram bin uniform distribution.	

10.13Method 1b) results with varied Schechter power law slope.	Left-hand
panels illustrate the $S[(\zeta, \chi)]$ plots for all 50 mocks, right-h	and panels
depict the histograms of optimal value of β with best-fit Ga	ussian (red
line) when available. Power law slopes used are noted on each	1 plot 234
10.14Monitoring of ζ - χ behaviour with varied Schechter power laws	slope. Pan-
els depict the bivariate histograms of $\zeta - \chi$ value distribution	over $[0,1]$.
The thick red line denotes our expected number galaxies per	histogram
bin for a uniform distribution.	
10.15Initial results of Method 2a). Upper panel depicts the $S[(\zeta, \chi)]$)] plots for
all 50 mocks generated, lower panel depicts the minimum of	each mock
plot as a function of trial value of β plotted as a histogram w	vith best-fit
Gaussian (red line) modelled to the data. Optimal value of β r	eturned for
Method 2a) is $\beta = 0.55 \pm 0.27$, as determined from the Gauss	sian model.
Luminosity function used during data generation is Gaussian	of the form
N[-21,1], with m_f and m_b given as 14.3 and 8 respectively.	Bright limit
was applied arbitrarily during S_1, S_2 set generation	
10.16 Method 2a) results with varied bright limits. Left-hand panel	ls illustrate
the $S[(\zeta, \chi)]$ plots for all 50 mocks, right-hand panels depi	ct the his-
tograms of optimal value of β with best-fit Gaussian (red line	e). LF is of
the form N[-21,1], with m_f and m_b noted on each plot. Enti	rety of the
PSCz is utilised during analysis with $\Delta Z=5, \ldots, \ldots$	
10.17 Monitoring of ζ - γ behaviour with varied bright limit. Panels	depict the
bivariate histograms of $\zeta - \chi$ value distribution over [0,1]. Th	e thick red
line denotes our expected number galaxies per histogram bin fo	r a uniform
distribution.	
10.18Initial results of Method 2b), depicting the $S[(\zeta, \gamma)]$ plots for a	ll 50 mocks
generated Optimal value of β returned for Method 2b) is	$\beta = 0.5$
as determined from analysing plot minima. Luminosity fur	oction used
during data generation is Schechter of the form estimated	S[-21.1.09].
with m_f and m_b given as 14.3 and 7 respectively. Bright limit y	was applied
arbitrarily during S_1 , S_2 set generation. Entirety of the PSC	z was used
for this analysis.	
10.19Method 2b) results with varied bright limits. Panels illustrate t	he $S[(\ell, \gamma)]$
plots for all 50 mocks. LF is of the form S[-21,1,09], with m_f and	d m_b noted
on each plot. Entirety of the PSCz is utilised during analysis	with $\Delta Z=5$. 241
10 20 Monitoring of ζ - γ behaviour with varied bright limit Panels	depict the
bivariate histograms of $\zeta_{-\gamma}$ value distribution over $[0, 1]$. The	e thick red
line denotes our expected number galaxies per histogram bin for	r a uniform
distribution	242
11.1 Flowchart for the application of data slicing, distance convergence	gence loops
and the ROBUST methodology in conjunction with relative	entropy for
the 2MRS survey.	
11.2 Plots of $S[(\zeta, \chi)]$ for the redshift restricted 2MRS survey across	ss the K, H
and J infrared bands, shown in blue, red and green respective	ly

11.3 One-sided confidence intervals calculated on relative entropy produced β via Monte Carlo simulations for the K, H and J bands depicted in blue, red and green respectively. The 3σ confidence intervals for all three bands are denoted via the black dashed lines, indicating all values of $\beta > 0$ are to be excluded from consideration for the redshift restricted 2MRS survey
in the K, H and J bands with 99.7% confidence
noted on each panel
respectively
values of β used are noted on each panel
in the K, H and J bands with 99.7% confidence
shown in blue, red and green respectively
11.10One-sided confidence intervals calculated on relative entropy produced β via Monte Carlo simulations for the K, H and J bands depicted in blue, red and green respectively. The 3σ confidence intervals for all three bands are denoted via the black dashed lines, indicating all values of $\beta > 0$ are to be excluded from consideration for the restricted 2MRS survey in the K, H and J bands with 99.7% confidence when μ_{β} is utilised instead of $u_{\beta}.254$
12.1 Results of applying mock distance-dependent apparent magnitudes to the positions of the PSCz galaxies on the value of β_{opt} as returned by RO-BUST (upper panel) and relative entropy (lower panel) over several trial values of β on the interval [0,1] over 50 mock realisations. Grid mesh size applied for calculating relative entropy set at a value of 10 over all trials. 259
12.2 Results of applying mock distance-dependent apparent magnitudes to the positions of the redshift restricted 2MRS galaxies on the value of β_{opt} as returned by ROBUST (upper panel) and relative entropy (lower panel) over several trial values of β on the interval [0,1] for the K (blue), H (red) and J (green) infrared bands respectively. Grid mesh size applied
for calculating relative entropy set at a value of 10

	12.4 Monitoring of (ζ, χ) distribution of the redshift restricted 2MRS galaxies in the K band for various trial values of β when a distance-dependent luminosity function is applied. The trial values of β used are noted on
. 262	each panel.
	12.5 One-sided confidence intervals calculated on β via Monte Carlo simula- tions for both ROBUST (upper panel) and relative entropy (lower panel) for the K H and I bands depicted in blue red and green respectively
26 4	The 3σ confidence intervals for all three bands are denoted via the black dashed lines, indicating all values of $\beta > \sim 0.3$ are to be excluded from consideration for the redshift restricted 2MBS survey with 99.7% confidence
	12.6 Results of applying mock distance-dependent apparent magnitudes to the positions of the redshift restricted 2MRS galaxies where μ_{β} is utilised instead of u_{β} on the value of β_{opt} as returned by ROBUST (upper panel) and relative entropy (lower panel) over several trial values of β on the interval [0,1] for the K (blue), H (red) and J (green) infrared bands re-
. 265	spectively. Grid mesh size applied for calculating relative entropy set at a value of 10.
966	12.7 Monitoring of (ζ, μ) distribution of the μ_{β} reliant redshift restricted 2MRS galaxies in the K band for various trial values of β when a distance- dependent luminosity function is applied. The trial values of β used are
. 200	12.8 Monitoring of (ζ, χ) distribution of the μ_{β} reliant redshift restricted 2MRS galaxies in the K band for various trial values of β when a distance- dependent luminosity function is applied. The trial values of β used are
. 267	noted on each panel
. 268	2MRS survey with 99.7% confidence
	12.10Results of applying mock universal apparent magnitudes to the positions of the redshift restricted 2MRS galaxies on the value of β_{opt} as returned by ROBUST (upper panel) and relative entropy (lower panel) over several trial values of β on the interval [0, 1] for the K (blue) H (and) and I (green)
. 269	infrared bands respectively. Grid mesh size applied for calculating relative entropy set at a value of 10 .
	12.11Monitoring of (ζ, μ) distribution of the redshift restricted 2MRS galaxies
. 270	in the K band for various trial values of β when a universal luminosity function is applied. The trial values of β used are noted on each panel.
	12.12 Monitoring of (ζ,χ) distribution of the redshift restricted 2MRS galaxies
. 271	in the K band for various trial values of β when a universal luminosity function is applied. The trial values of β used are noted on each panel.

12.13One-sided confidence intervals calculated on β via Monte Carlo simula- tions for both ROBUST (upper panel) and relative entropy (lower panel) for the K, H and J bands depicted in blue, red and green respectively. The 3σ confidence intervals for all three bands are denoted via the black dashed lines, indicating all values of $\beta > \sim 0.3$ are to be excluded from consideration for the redshift restricted 2MBS survey with 99 7% confidence 272
12.14Monitoring of (ζ, μ) distribution of the μ_{β} reliant redshift restricted 2MRS galaxies in the K band for various trial values of β when a universal luminosity function is applied. The trial values of β used are noted on
each panel 273
12.15Besults of applying mock universal apparent magnitudes to the positions
of the μ_{β} reliant redshift restricted 2MRS galaxies on the value of β_{opt} as returned by ROBUST (upper panel) and relative entropy (lower panel) over several trial values of β on the interval [0,1] for the K (blue), H (red) and J (green) infrared bands respectively. Grid mesh size applied
for calculating relative entropy set at a value of 10
12.16Monitoring of (ζ, χ) distribution of the μ_{β} reliant redshift restricted 2MRS
galaxies in the K band for various trial values of β when a universal luminosity function is applied. The trial values of β used are noted on
each panel. $\ldots \ldots 275$
12.17One-sided confidence intervals calculated on β via Monte Carlo simula- tions for both ROBUST (upper panel) and relative entropy (lower panel) for the K, H and J bands depicted in blue, red and green respectively. The 3σ confidence intervals for all three bands are denoted via the black dashed lines, indicating all values of $\beta > \sim 0.3$ are to be excluded from consid- eration for the μ_{β} reliant redshift restricted 2MRS survey with 99.7% confidence
13.1 The density of galaxies in the GLADE catalogue, presented as an Aitoff projection on the sky. Figure reproduced from the GLADE documentation works of Dálya et. al [57]

List of Tables

1.1	Summary of possible Universe curvatures and geometries for use in the FLRW metric [121].		13
1.2	The development of cosmological models to the present date [157]	•	20
2.1	The magnitude limits of the 2MASS survey for unconfused sources outside of the Galactic Plane ($ b > 10^{\circ}$), and outside of any confusion-limited areas of the sky outside of the Galactic Plane [24]		59
2.2	The optimal optical wavelengths at which SDSS's five filters work best. The sensitivity of each filter falls off slowly at shorter and longer wave- lengths. [173]. The optical magnitude limits listed here are at 95% com- pleteness for point sources [175].		61
3.1	Summary of recent Bulk Flow measurements. * Surface Brightness Fluc- tuations. * Mark III dataset. [†] Tully-Fisher Measurement. • Supernovae Type 1a. ° Fundamental Plane Measurement. [‡] Brightest Cluster Galaxy Measurement. [#] Constrained Realisations. ^d Maximum Likelihood Ap-		
3.2	proach [210] [196] [62] [80] [126]. Summary of β values calculated from various basis and iterative methods. † Inconsistent flow fields (probably due to problematic calibration of Mark	•	69
3.3	111) [210]. β values obtained from χ^2 minimising. SBF represents surface brightness fluctuation, and PS denotes power spectrum [177]. The result highlighted in bold signifies the value of β we will use as a benchmark for our future	•	83
3.4	$\begin{array}{l} \text{computations.} & \dots & \dots \\ \beta \text{ values from a selection of more recent methodologies briefly introduced} \\ \text{in Section 3.4.6.} & \dots & \dots & \dots \\ \end{array}$		83 83
6.1	Summary of estimated values of β detailing the effect of varying mock galaxy sample sizes on computed β using ROBUST during Mock Method 1a) implementation	1	196
6.2	Summary of estimated values of β detailing the effect of varying mock generated Gaussian LF width σ on computed β using ROBUST during	. 1	120
	for this analysis.	. 1	130
6.3	Summary of estimated values of β detailing the effect of varying mock galaxy sizes on computed β using ROBUST during Mock Method 1b)	4	104
6.4	Implementation. Summary of estimated values of β detailing the effect of varying Schechter power law slope α on computed β using ROBUST during Mock Method 1b) implementation. Entirety of PSCz survey is utilised for this analysis.	. 1 . 1	134 138

6.5	Summary of estimated values of β detailing the effect of varying applied apparent bright magnitude limit on computed β using ROBUST during Mock Method 2a) implementation. Entirety of PSCz is utilised for this analysis with $\Delta Z=5$.	. 144
6.6	Summary of estimated values of β detailing the effect of varying applied ΔZ on computed β using ROBUST during Mock Method 2a) implementation. Entirety of PSCz is utilised for this analysis with m_f and m_b fixed at 14.3 and 8 respectively.	. 149
6.7	Summary of estimated values of β detailing the effect of varying applied apparent bright magnitude limit on computed β using ROBUST during Mock Method 2b) implementation. Entirety of PSCz is utilised for this analysis with $\Delta Z=5$	159
6.8	Summary of estimated values of β detailing the effect of varying applied ΔZ on computed β using ROBUST during Mock Method 2b) implementation. Entirety of PSCz is utilised for this analysis with m_f and m_b fixed at 14.3 and 6 respectively.	. 155
7.1	Summary table of extrapolated confidence intervals obtained on β using both MC simulations and bootstrap resampling for both an applied Gaussian and Schechter LF.	. 166
8.1	Summary of best fit values of β^* using different weighting schemes, methods of analysis and peculiar velocity datasets. Results obtained using luminosity weighting are indicated by (LW), whereas those obtained using number weighting are indicated by (NW). Unless explicitly indicated, all datasets were used for the method mentioned with the exception of Inverse VELMOD which used all individual galaxies from SFI++ [25] $$.	. 176
10.1	Summary of estimated values of β detailing the effect of varying arbitrarily applied grid resolutions to the construction of the ζ - χ grid for use in computation of β with relative entropy. Entirety of PSCz survey is utilised for this analysis	222
10.2	Summary of estimated values of β detailing the effect of varying mock galaxy sample sizes on computed β using ROBUST during Mock Method 1a) implementation.	. 225
10.3	Summary of estimated values of β detailing the effect of varying mock generated Gaussian LF width σ on computed β using relative entropy during Mock Method 1a) implementation. Entirety of PSCz survey is utilised for this analysis.	. 227
10.4	Summary of estimated values of β detailing the effect of varying mock galaxy sample sizes on computed β using relative entropy during Mock Method 1b) implementation	921
10.5	Summary of estimated values of β detailing the effect of varying Schechter power law slope α on computed β using relative entropy during Mock Method 1b) implementation. Entirety of PSCz survey is utilised for this analysis	. ∠JI 921
10.6	Summary of estimated values of β detailing the effect of varying applied apparent bright magnitude limit on computed β using relative entropy during Mock Method 2a) implementation. Entirety of PSCz is utilised for this analysis with $\Delta Z=5$.	. 231

10.7	Summary of estimated values of β detailing the effect of varying applied	
	apparent bright magnitude limit on computed β using relative entropy	
	during Mock Method 2b) implementation. Entirety of PSCz is utilised	
	for this analysis with $\Delta Z=524$	1

Abbreviations

ΛCDM	Lambda Cold Dark Matter Model
2dFS	2 D egree F ield S urvey
2M++	2MASS-XSC + 6dFGS + SDSS DR7 catalogues
2MASS	2 Micron All-Sky Survey
2MASS XSC	${\bf 2}$ Micron All-Sky Survey eXtended Source Catalog
2MRS	${\bf 2} \ {\bf MASS} \ {\bf R} {\rm edshift} \ {\bf S} {\rm urvey}$
$2 \mathrm{MPZ}$	2MASS Photometric Redshift(z) Catalog
2SLAQ	2dF- S DSS L RG A nd Q SO surveys
6dFGS	6 Degree Freedom Galaxy Survey
6 dFGSv	6 Degree Freedom Galaxy Survey, peculiar velocity
A1	1^{st} Amendment Supernovae Set
AGN	Active Galactic Nuclei
ASC	\mathbf{A} siago \mathbf{S} upernova \mathbf{C} atalog
BAO	Baryonic Acoustic Oscillations
BCG	\mathbf{B} rightest \mathbf{C} luster \mathbf{G} alaxy Measurement
BTFR	Baryonic Tully Fisher Relation
CBAT	Central Bureau for Astronomical Telegrams
CDF	Cumulative Distribution Function
CFi	i^{th} iteration of Cosmic-Flows
CfA	Centre for Astrophysics
CHDM	$\mathbf{Cold}/\mathbf{Hot} \ \mathbf{Dark} \ \mathbf{M}$ atter Model
CMBR	$\mathbf{C} \mathbf{o} \mathbf{s} \mathbf{m} \mathbf{i} \mathbf{c} \mathbf{r} \mathbf{o} \mathbf{w} \mathbf{a} \mathbf{v} \mathbf{e} \mathbf{B} \mathbf{a} \mathbf{c} \mathbf{k} \mathbf{g} \mathbf{r} \mathbf{o} \mathbf{u} \mathbf{d} \mathbf{R} \mathbf{a} \mathbf{d} \mathbf{i} \mathbf{a} \mathbf{t} \mathbf{i} \mathbf{o} \mathbf{n}$
CR	Constrained Realisations
DEC	Dec lination angle
DRi	i^{th} D ata R un of the SDSS

EFAR	Ellipticals FAR Away Project				
ENEAR	Early-type Near by Galaxies Survey				
ESO	European Southern Observatory				
FFT	$\mathbf{F} \text{ast } \mathbf{F} \text{ourier } \mathbf{T} \text{ransform}$				
FJR	\mathbf{F} aber- \mathbf{J} ackson \mathbf{R} elation				
FLRW	${f F}$ riedmann Lemaître Robertson Walker				
FORTRAN	FORmula TRANslation				
FP	\mathbf{F} undamental \mathbf{P} lane				
GSR	Galactic Standard of Rest				
GLADE	Galaxy List for the Advanced Detector \mathbf{E} ra				
GWGC	Gravitational Wave Galaxy Catalog				
HDF	$\mathbf{H} ubble \ \mathbf{D} eep \ \mathbf{F} ield$				
HDF-S	$\mathbf{H} \text{ubble } \mathbf{D} \text{eep } \mathbf{F} \text{ield -} \mathbf{S} \text{outh}$				
HDM	Hot Dark Matter Model				
HII	\mathbf{H} ydrogen \mathbf{II} (ionised hydrogen)				
IPAC	Infrared Processing and Analysis Centre				
IRAS PSCz	Infra Red Astronomical Satellite Point Source Catalogue				
ITF	Inverse Tully Fisher Relation				
KLT	\mathbf{K} arhunen- \mathbf{L} oève \mathbf{T} ransform				
KS	\mathbf{K} olmogorov- \mathbf{S} mirnov				
LEDA	\mathbf{L} yon-Meudon \mathbf{E} xtragalactic $\mathbf{D}\mathbf{a}$ tabase				
\mathbf{LF}	Luminosity Function				
LG	Local Group				
LIGO	${\bf L} aser {\bf I} nterferometer {\bf G} ravitational-Wave {\bf O} bservatory$				
LP, LP10K	Las Campanas Observatory/Palomar 10,000km s^{-1} Cluster Survey				
LRG	$\mathbf{L} uminous \ \mathbf{R} ed \ \mathbf{G} alaxies$				
LSST	\mathbf{L} arge \mathbf{S} ynoptic \mathbf{S} urvey \mathbf{T} elescope				
LW	\mathbf{L} uminosity \mathbf{W} eighting				
M3	Mark III Catalogue				
MGC	\mathbf{M} illennium \mathbf{G} alaxy \mathbf{C} atalogue				
ML	Maximum Likelihood				
NASA	National Aeronautics and Space Administration				
NED	\mathbf{N} ASA/IPAC E xtragalactic \mathbf{D} atabase				

NW	Number Weighting				
ORS	\mathbf{O} ptical \mathbf{R} edshift \mathbf{S} urvey				
PDF	\mathbf{P} robability \mathbf{D} ensity \mathbf{F} unction				
PS	Power Spectrum				
QSO	\mathbf{Q} uasi \mathbf{S} tellar \mathbf{O} bjects				
RA	$\mathbf{R} ight \ \mathbf{A} scension \ angle$				
RS	\mathbf{R} adburn- \mathbf{S} mith				
RW	\mathbf{R} obertson- \mathbf{W} alker				
SAI	\mathbf{S} ternberg \mathbf{A} stronomical \mathbf{I} nstitute				
\mathbf{SBF}	Surface Brightness Fluctuation				
SCDM	\mathbf{S} tandard \mathbf{C} old \mathbf{D} ark \mathbf{M} atter Model				
$\mathbf{SCI}, \mathbf{SCII}$	${f S}$ piral Galaxy Type ${f C}$ ${f I}$, or Type ${f C}$ ${f I}$				
SDSS	Sloan Digital Sky Survey				
SDSS-DR12Q	Sloan Digital Sky Survey Data Run 12 Quasar Catalog				
SEcat	\mathbf{S} FI + \mathbf{E} NEAR \mathbf{Cat} alogues				
SFI	${f S}$ piral ${f F}$ ield ${f I}$ -Band				
SGP	\mathbf{S} uper \mathbf{G} alactic \mathbf{P} lane				
SKA	\mathbf{S} quare \mathbf{K} ilometre \mathbf{A} rray				
SMAC	Streaming Motions of Abell Clusters sample				
SNIa	Super Novae TypeIa				
SNR	Signal-to-Noise Ratio				
SSC	SAI Supernova Catalog				
SSRS2	Southern Sky Redshift Survey 2				
\mathbf{TFR}	\mathbf{T} ully \mathbf{F} isher \mathbf{R} elation				
TMSS	Two Micron Sky Survey				
UDF	Ultra Hubble D eep F ield				
UGC	Uppsala General Catalogue of Galaxies				
UMV	Unbiased Minimal Variance				
USC	Unified Supernova Catalogue				
VIPERS	${\bf VIMOS}\ {\bf P}{\rm ublic}\ {\bf E}{\rm xtragalactic}\ {\bf R}{\rm edshift}\ {\bf S}{\rm urvey}$				
WALLABY	Widefield ASKAP L-Band Legacy All-Sky Blind SurveY				
WDM	\mathbf{W} arm \mathbf{D} ark \mathbf{M} atter Model				
WIMP	Weakly Interacting Massive Particle				

WMAPWilkinson Microwave Anisotropy ProbeXDFeXtreme Hubble Deep Field

Physical Constants

Speed of Light	c	=	$2.997 \ 924 \ 58 \times 10^5 \ \mathrm{km s^{-1}} \ \mathrm{(exact)}$
Gravitational Constant	G	=	$6.67408\times 10^{-11} {\rm m}^{3} {\rm kg}^{-1} {\rm s}^{-2}$
Pi	π	=	3.14159265359
Solar Mass	M_{\odot}	=	$(1.98855\pm0.00025)\times10^{30}\rm kg$
Electron volt	$1\mathrm{eV}$	=	$1.6\times 10^{-19} \rm J$
Parsec	1 pc	=	$3.0857\times 10^{16}\mathrm{m}$
Gigayear	1Ga	=	$3.16\times 10^{16} \rm s$
Astronomical Unit	1AU	=	149597870700m (exact)
Jansky	1Jy	=	$10^{-26} \mathrm{Wm}^{-2} \mathrm{Hz}^{-1}$
Angstrom	$1 { m \AA}$	=	10^{-10} m
Symbols

α	Schechter power law exponent	
β	Linear redshift distortion parameter	
$\overline{\Delta M}$	Malmquist correction	
ΔZ	Change in corrected distance moduli	
δ	Mean mass density contrast	$\rm kgm^{-3}$
δ_g	Mass distribution of observed galaxies	
δ_{lmn}	Expansion coefficients	
δ_m	Mass fluctuations	
ϵ	Noise vector	
ζ	ROBUST statistic	
$\hat{\zeta}_i$	Unbiased estimator of ζ	
η	Disk mass-to-light ratio, or velocity width parameter	
θ	Longitude, Right Ascension angle	0
$\theta(x)$	Heaviside or unit step function	
λ	Wavelength	m
λ	Lagrangian multiplier	
μ	Distance Modulus	Mpc
μ_i,λ	Probability measures	
ν	χ^2 degrees of freedom	
ρ	Scalar mass density	$\rm kgm^{-3}$
$ ho(\zeta_eta,u_eta)$	Correlation coefficient	
$ ho_c$	Critical Density	$\rm kgm^{-3}$
$\rho(r,l,b)$	Galaxy spatial distribution function	
$\bar{ ho}(t)$	Scalar mean mass density	$\rm kgm^{-3}$
σ	Standard deviation	

xxxvii

σ_8	Averaged rms amplitude out to $8h^{-1}$ Mpc radius	
σ_B	Peak blue-band luminosity scatter	
Υ_o	Galaxy mass-to-light ratio	
Φ	Wiener filter or scalar potential	
$\Phi(M)$	Schechter luminosity function	
ϕ	Latitude, Declination angle	0
ϕ^*	Normalisation constant	
φ	Gravitational potential	
χ	Renormalised cumulative distribution of redshifts	
χ^2_{min}	Chisquared minimisation	
$\psi(m)$	Selection function in apparent magnitude	
Ω	Total Universe energy density	
Ω_{Λ}	Energy density of dark energy	
Ω_{ν}	Energy density of HDM (light neutrinos)	
Ω_b	Energy density of luminous baryonic matter	
Ω_m	Energy density of CDM (unless otherwise specified)	
$\Omega_m^{0.6},\Omega_m^{1.2}$	Amplitude of mass power spectrum	
ω	Angular coordinates	
(l,b)	Galactic longitude and latitude angles	
A	Normalisation factor	
$A_g(l,b)$	Galactic extinction correction term	
A(x)	Galaxy count up to limiting magnitude x	
a	Scale factor	
a_0	Present day value of scale factor $R(t)$	
a_{ij}	Weighting factor	
BF_{2M++}	Bulk flow motion arising from the $2M++$	${\rm km s^{-1}}$
b	Linear biasing parameter	
b_{apex}	Galactic latitude of velocity apex	0
C_{eta}	Correlation coefficient	
cz	Recessional velocity	${\rm km s^{-1}}$
cz_{LG}	Local Group redshift	$\rm km s^{-1}$
cz_{hel}	Heliocentric redshift	$\rm km s^{-1}$
D_n	Isophotal diameter	Mpc

d	Distance	Mpc
d_{ij}	Distance from the j^{th} supernova to the i^{th} galaxy	Mpc
d_L	Luminosity distance	Mpc
E	Event probability space	
E_i	Expectation or mean	
F(M)	Cumulative distribution of a luminosity function in ${\cal M}$	
$F_X(\zeta_i)$	Normalised CDF of ζ	
f	Growth factor	
$g_{\mu u}$	Metric tensor	
H_i	Hypothesis on probability measure	
H(t)	Hubble time	s
H(z)	Cumulative distribution of observed redshifts	
h	Hubble parameter	
$h(\mu, l, b)$	Line of sight distribution of μ	
I(0)	Central surface brightness	
I(1:2, E)	Mean weight of evidence for a hypothesis given x	
I_e	Effective elliptical galaxy surface brightness	
I(R)	Spiral galaxy disk surface brightness distribution	
jı	Spherical Bessel functions	
k	Curvature parameter	
k_{cor}	k-correction term	
k_{ln}	Set of wavenumbers	
L	Luminosity	W
L^*	Characteristic galaxy luminosity	
LG_{2M++}	Local Group motion arising from the $2M++$	$\rm km s^{-1}$
l_{apex}	Galactic longitude of velocity apex	0
M	Magnitude	
M_*	Reduced mass	kg
M^*	Characteristic absolute magnitude	
M_{lim}	Absolute magnitude limit	
m	Mass	kg
m_B	Apparent B-Band magnitudes	
m_b	Bright apparent magnitude limit	

m_{cor}	Corrected apparent magnitude	
m_f	Faint apparent magnitude limit	
m_{lim}	Apparent magnitude limit	
\mathcal{N}	Multiple noise realisations	
$\mathbf{N}[\mu,\sigma]$	Normal distribution with mean μ , standard deviation σ	
N_{gal}	Number of galaxies	
n_i	Unit vector of magnitude 1	
P(k)	Mass fluctuation power spectrum	
\mathbf{q}	Lagrangian coordinate	
\mathcal{R}	Spacetime curvature	
R	Point spread or response function	
R_D	Galaxy scale disk length	Mpc
R_e	Effective radius	Mpc
R(t)	Expansion scale factor	
r	Radius	m
r_i, n_i	Galaxy counts	
S	Flux	${\rm Wm}^{-2}$
$S[(\mu_1,\mu_2)]$	Relative entropy of μ_1 and μ_2	
$S\left[(\zeta_{eta},\chi) ight]$	Relative entropy of ζ and χ	
S_i	$i^{th} M$ -Z region	
8	Cosmological field	
T	Kinetic energy	kJ
T_c, T_v	Completeness statistics	
t	Time	S
U	Total energy	kJ
u_i°	Observed radial v_{pec}	$\rm km s^{-1}$
V	Potential energy	kJ
\mathbf{V}_{EXT}	Residual bulk flow	$\rm km s^{-1}$
V_{apex}	Rotational apex velocity	$\rm km s^{-1}$
V_{con}	Converted velocity	$\rm km s^{-1}$
V_i	Variance	
v	Peculiar velocity field	$\rm km s^{-1}$
v_{LG}	Peculiar velocity in the Local Group frame	$\rm km s^{-1}$

v_{pec}	Peculiar velocity	$\rm km s^{-1}$
$ ilde{W}(kR)$	Fourier transform of top-hat window of radius R	
$w^L(r)$	Luminosity-based galaxy weight	
Y_{lm}	Spherical harmonics	
Ζ	Corrected distance modulus	
z	Redshift	
z_u	Upper redshift limit	$\rm km s^{-1}$

Dedicated to my late grandfather Medhat Islam, and my father Sherief Islam, whose unceasing encouragement to look up into the heavens and wonder and contemplate my place in it has made me who I am today. I wouldn't change it for anything.

Also for my recently departed grandfather Ahmed Gaber Barakat who never got to see this published as he had wished, but never stopped believing in me and waiting to see me so that he could hold this and ask me all about it. I missed you by a few days Geddo, but look forward to meeting you again in a better place so that I can do just that and share with you all you want. I miss you so much.

Preface

In recent years the statistical analysis of galaxy redshift surveys has led to the constraining of several cosmological parameters in the current Λ CDM model, allowing us to learn more about the Universe as a whole, be it learning more about the underlying dark matter distribution in the nearby Universe, analysing and constraining the peculiar velocity fields of galaxies as a function of luminosity or by means of standard candles such as Type Ia supernovae, identifying all kinds of selection effects that inhibit our ability to effectively constrain parameters, the list is endless. This thesis aims to detail the work and research that has been done so far with regards to some of the statistical tools available (particularly a brief look into the use of χ^2 minimisation techniques, and a more in depth look into the application of the ROBUST method) to constrain the linear redshift distortion parameter β and the attempts made to confirm the results published by fellow colleagues in the field, a task that in itself was initiated as part of an MSc summer project that was meant to lay the foundations from which this particular doctorate of research was borne [95].

In this thesis we will study one particular aspect of dark matter cosmology: the peculiar velocity field. In being better able to define this field, we will have a better handle on the distribution of dark matter that generates it. In order to do this, we need to identify and constrain something called the **linear redshift distortion parameter**, upon which the peculiar velocity field is linearly dependent. Although several methods exist with which this parameter can be constrained, we will focus on two avenues:

- Applying a velocity-velocity interpolation technique such as χ^2 -minimisation using current redshift surveys and standard candles such as Type Ia supernovae to determine the most probable value of the parameter that fits observed data,
- Applying the statistical tool ROBUST as developed by Rauzy and Hendry [160] to determine the most probable value of the distortion parameter using galaxy luminosity functions.

While velocity-velocity interpolation techniques such as χ^2 -minimisation and VELMOD are nothing new, having been frequently used in the works of other cosmologists and astrophysicists alongside other methods to constrain a vast array of cosmological parameters such as the dark matter/energy densities Ω_M and Ω_Λ , they possess a limitation in that additional information in the form of secondary distance indicators (such as Type Ia Sne) are required in order for these comparisons to work. ROBUST is unique in that not only is it a novel approach to constraining parameters using the luminosity functions of the underlying data, but it also removes these limitations and makes full use of the surveys at hand to constrain whatever parameter we desire effectively.

It is the purpose of this thesis to explore the possibilities of ROBUST for use with real world datasets, as it has not seen much use outside of mock simulated runs. In particular we aim to investigate the following:

- While still in the simulated run stage, we shall explore the efficacy of ROBUST in recovering the linear redshift distortion parameter, and observe the sensitivity of the method to varying mock survey sizes, redshift reconstruction and distance errors.
- Considering that ROBUST requires galaxy luminosity functions to operate, we will explore how efficiently ROBUST returns a value for the distortion parameter when various luminosity functions are applied. In particular we will explore the effects when both Gaussian and Schechter luminosity functions (both of which are typical of various surveys and galactic environments) are used.
- Finally take ROBUST out of the mock simulation stage, and proceed to apply the method to real world datasets and observe the results returned. Specifically we will work with the IRAS PSCz survey of Enzo Branchini augmented with B-band magnitudes and attempt to reconstruct his distortion parameter value of $\beta=0.55\pm0.06$, and then proceed to utilise the 2MRS survey of Hudson et. al in conjunction with a velocity field map developed by Carrick et. al [25] to attempt to reconstruct their obtained distortion parameter value of $\beta=0.43\pm0.021$.
- Should time permit, we will also aim to make use of one of the ancillary tools made available when utilising ROBUST and its various statistics, namely relative entropy and fold in its functionality into our ROBUST analyses to see how well the results that it returns (be it for the PSCz or the 2MRS) help to either reinforce or reject the results for β that ROBUST returns on its own.

Chapter 1

A Brief Overview of Cosmology

1.1 Cosmology in the Modern Day

The avenue of cosmology is one that has been studied for many a century in many different guises, be it religious, philosophical or scientific. The word cosmology itself comes from the Greek: 'cosmologia', which means order or orderly arrangement, and 'logos', which means word, reason or plan. As such, cosmology means the study of the Universe in its totality, and has come a long way since those first Greek philosophers looked up at the sky and wondered how the heavens worked, and there has never been a better time than the present to be at the cosmological vanguard.

The reason for this is simple: cosmology has advanced in leaps and bounds over the past century, starting with Edwin Hubble's discovery that the Universe was expanding in 1929 [84], followed by observations by Fritz Zwicky in 1937 that there appeared to be missing mass in the Universe [217], and Vera Rubin's publishing of her observed velocity rotation curves of the Andromeda Galaxy in 1970 that defied Newtonian mechanics. More specifically, Rubin's observations as plotted in Figure 1.5 appeared to imply that for the galaxy as a whole to be rotating at the speeds she observed over increasing distance from the galactic centre, the galaxy needed to be embedded within a halo of invisible matter [165]. This supposed invisible or missing mass was confirmed in the 1980s when **dark matter** became recognised as the matter that stabilised the observed clustering of galaxies.

After several attempts to define a cosmological model to properly describe the Universe with dark matter, another cosmological landmark would be reached in 1998 with the startling discovery of accelerated Universal expansion, and the supposed **dark energy**



FIGURE 1.1: Velocity-Distance Relation among Extra-Galactic Nebulae. Radial velocities, corrected for solar motion, are plotted against distances estimated from involved stars and mean luminosities of nebulae in a cluster. The black discs and full line represent the solution for solar motion using the nebulae individually; the circles and broken line represent the solution combining the nebulae into groups; the cross represents the mean velocity corresponding to the mean distance of 22 nebulae whose distances could not be estimated individually [84].

that fuels it, paving the way for the now current Double Dark Theory and newer cosmological models that continue to be refined and tested as we see just how much more there is out there to discover and explore.

1.2 The Concept of Universal Expansion and the Cosmological Principle

As we touched upon in the introduction, Edwin Hubble's study of the velocities of extra-galactic nebulae, specifically their recessional velocities with increasing distance, led him to plotting the graph seen in Figure 1.1, from which he noted the existence of a roughly linear relation between their velocities and distances, with that relation appearing to dominate the distribution of velocities [84]. Put another way, the further away an extra-galactic nebula appeared to be from us as the observer, the faster it appeared to be moving away from us, with this bulk motion dominating over whatever so-called 'peculiar velocities' that nebula might additionally be exhibiting in its local neighbourhood due to gravitational influences from other nearby objects.

It is worth noting that while Hubble's 1929 publication was the one that gained the most traction in the scientific community at the time, with his observed linear relation ultimately being named after him, exactly the same results were noted and observed by both the American philosopher Vesto Slipher in 1917 [181] and by the Belgian priest Georges Lemaître in 1927 [118]. Lemaître in particular failed to gain any notice in the scientific community due to his results being published in a rather obscure Belgian journal – not receiving an English translation until 1931 [119].

As such, this relation would ultimately be defined as Hubble's Law (refer to Section 2.2 for a brief treatment of the equation itself and its integral use in our work), but would also serve as the means for a very important inference in cosmology, namely that our Universe is expanding, and has been doing so since the beginning of time.

When thinking about Universal expansion it would be incorrect to think of it in terms of a typical explosion, which has a physical origin in both time and space that we identify; because although the Universal expansion does indeed have an origin in time t (over 13Ga ago with the occurrence of the Big Bang), it has no physical origin in **space**. A more correct analogy would be to consider a hypothetical spherical balloon capable of expanding indefinitely: with all of our Universe as we know it existing on the surface of this balloon. This way, the Universal expansion has an origin in time t(when the balloon first started to expand), but no physical origin in space as a spherical surface mathematically has none. Consequently as this hypothetical 'balloon' continues to expand over time, everything will continue to move further away from each other, yet still allow for peculiar motions over relatively smaller cosmological distances (say over distances of several Mpc) to continue to occur without impediment. A good example of this would be that even though the Universe is indeed expanding, the Milky Way and Andromeda galaxies are due to collide with each other as Andromeda is moving towards us as opposed to further away, and this collision will happen ... though mercifully for us not for another 4 billion years at least.

This balloon analogy also provides us with two very interesting points to consider. If our balloon was to continue to expand to infinity there will come a point after which over a large enough scale distance, wherever you happen to be on this balloon – be it the top, or bottom or somewhere on its equator – the Universe immediately around you will start to look the same or, put another way, the Universe around you will be **homogenous** over a large enough scale distance. We can see proof of this in some of the images taken from the many iterations of the Hubble Deep Field (using the spacebased telescope of the same name) over varying patches of the sky, such as in Figures 1.2 and 1.3. Specifically, Figure 1.2 depicts images taken from the Hubble Deep Field, HDF (panel (a)) and the Hubble Deep Field South, HDF-S (panel (b)). The HDF is an image of a small region in the constellation Ursa Major, constructed from a series of observations by the Hubble Space Telescope in late 1995. It covers an area of about 2.6 arcminutes on a side, equivalent to about one 24-millionth of the whole sky, which is equivalent in angular size to a tennis ball at a distance of 100 metres [31]. Three years after the HDF observations were taken, a region in the southern celestial hemisphere was imaged in a similar way and was named the HDF-S. The key thing to take away from both these images is that whether you are observing a patch of sky in either the northern or southern hemispheres the distribution of galaxies observed is generally the same.

This is all the more apparent in Figure 1.3 with the results depicted from the Hubble Ultra Deep Field, UDF (panel(a)) and Hubble eXtreme Deep Field, XDF (panel(b)). The photo of the UDF is an image of a small area of space in the constellation Fornax, created using Hubble Space Telescope data from 2003 and 2004. By collecting faint light over many hours of observation, it revealed thousands of galaxies, both nearby and very distant, making it the deepest image of the Universe ever taken at that time. The XDF was assembled in 2012 by combining 10 years of NASA Hubble Space Telescope photographs taken of a patch of sky at the center of the original Hubble Ultra Deep Field, and is a small fraction of the angular diameter of the full moon [137]. Even though we are looking deeper into space with the XDF, observing galaxies as they were at least 13 billion years ago, its similarity to the UDF is unmistakable, lending further credence to the concept that the Universe we inhabit is indeed homogeneous.

The second point to consider with our balloon is that irrespective of where you are on the balloon, the Universe will continue to look the same and exhibit the same behaviours and large scale structures whether you choose to observe in front of you or behind you i.e., the Universe is also **isotropic** over the same large scale distances. Universal homogeneity and isotropy as a result lend credence to one of the key cornerstones of observational cosmology, namely the **cosmological principle**. The cosmological principle states that our place in the Universe is in no way special, or any more important than if we were to inhabit another location in the Universe entirely [121]. It is worth noting however that the cosmological principle is not exact, and as briefly touched upon in this section, the existence of peculiar motions and large-scale structures would cause homogeneity to break over a small enough distance, say, like within the Local Group of galaxies within which both the Milky Way and Andromeda amongst others are members. It is indeed these peculiar motions and distortions of these galactic motions over smaller cosmological distances that will make up the bulk of the analysis of this particular work.

Despite this, the concept of an expanding Universe does lead to an interesting puzzle. In order for us to be able to model and constrain this expansion, we will require an evolving



(B) The southern successor to the Hubble Deep Field, called Hubble Deep Field South.

FIGURE 1.2: Images of the Hubble Deep Field and Hubble Deep Field South. Though both images represent deep observations of small regions of space in the northern and southern hemispheres respectively, the similarities between the images lend credence to the concept of a homogeneous Universe.



(A) The Hubble Ultra Deep Field, crafted together from multiple Hubble Telescope images taken over the years of 2003/2004 of a small region of space in the constellation Fornax. This image is estimated to contain around 10000 galaxies [137].



(B) The Hubble eXtreme Deep Field, focusing on a small central region of the Hubble Ultra Deep Field and is the culmination of multiple observations of that small region of space taken over 10 years [137].

FIGURE 1.3: Images of the Hubble Ultra Deep Field and Hubble eXtreme Deep Field. Much like the photos of the HDF and HDF-S, the similarities between the images continue to lend credence to the concept of a homogeneous Universe, especially holding true even as we probe very deep into space and far back in time with the XDF. coordinate system with which to track real-space positions over increasing cosmological time.

1.3 'de Sitter Space' and the FLRW Metric

More specifically, we require a large-scale notion of space and time that allows us to relate observations we make here and now to physical conditions at some location that is distant in space and time. We thus require the interval or proper time between events (or measurements), which we can express in the form of a metric given by:

$$-ds^{2} = c^{2}d\tau^{2} = c^{2}dt'^{2} - dx'^{2} - dy'^{2} - dz'^{2} = g_{\mu\nu}dx^{\mu}dx^{\nu}, \qquad (1.1)$$

where dashed coordinates are local to the object, undashed refer to the global coordinates we will use, and $g_{\mu\nu}$ is the metric tensor, which is found in principle by solving Einstein's gravitational field equations [145]. Thankfully a simpler solution that closely matches what is observed in reality is available to us if we choose to consider the most symmetric form of this metric.

In order to do that let us first consider our own reality for a moment. Einstein postulated in 1905 that, as a consequence of his study of the electrodynamics of moving bodies, we actually inhabit a four dimensional Universe as opposed to just three, comprising of the three physical dimensions with which we are familiar - positional cartesian coordinates for example being defined by (x, y, z), and the additional fourth dimension being that of time t, with the fusing of these dimensions into one entity hereafter being referred to as **spacetime** [55]. As a result of this, in order to be able to properly define the spacetime coordinate of any object in our Universe, we can express it as a 4D coordinate of the form (x, y, z, t), and this idea can be easily extended to higher dimensional spacetime constructs if needed.

Now let us consider for a moment a 'hyper' 4D surface that exists in Euclidean 5D-space. As we just established a spacetime coordinate in 5D-space would be expressed as five positional coordinates, with the sixth one being time. To that effect in a 5D space the distance to an object, or in this instance the spacetime curvature \mathcal{R} , would be given by:

$$x^{2} + y^{2} + z^{2} + w^{2} - v^{2} = \mathcal{R}^{2}, \qquad (1.2)$$

where x, y, z, w and v refer to the positional coordinates of our surface. Consequently the metric of this spacetime curvature \mathcal{R} can be defined by:

$$ds^{2} = dx^{2} + dy^{2} + dz^{2} + dw^{2} - dv^{2}$$
(1.3)

9

Here, we have made the last positional coordinate v imaginary (hence the negative sign in Equation 1.2) since we are dealing with a theoretical 4D surface. This also serves to allow us to re-obtain the 4D Einstein spacetime signature with which we are familiar with in our own Universe (i.e. expressing every spacetime position in our Universe as a set of four coordinates) [145]. The square 4×4 diagonal matrix of this metric would collapse down to Einstein's special relativity case of the metric as given by:

(refer to Einstein's seminal work on general relativity [56] for an extensive look at the subject and all relevant derivations) which will be maximally symmetric, and as a consequence of that, would qualify for being a homogeneous interpretation of our own Universe since in this instance all our spacetime points are manifestly equivalent [145]. This special case of symmetrical spacetime was derived by Willem de Sitter in 1917/18, and was thus named **de Sitter Space** (refer to de Sitter's papers [48] and [49] for his in-depth derivations and treatment of the subject).

If we were to take the five dimensional coordinates expressed in de Sitter space and instead choose to express them in polar coordinates we would get:

$$v = \mathcal{R} \sinh \alpha$$

$$w = \mathcal{R} \cosh \alpha \cos \beta$$

$$z = \mathcal{R} \cosh \alpha \sin \beta \cos \gamma$$
 (1.4)

$$y = \mathcal{R} \cosh \alpha \sin \beta \sin \gamma \cos \delta$$

$$x = \mathcal{R} \cosh \alpha \sin \beta \sin \gamma \sin \delta,$$

which by the theoretical definition of polar coordinates will be an orthogonal coordinate system (i.e. the dot product of the partial derivatives of these coordinates will be equal to zero as they are all perpendicular to each other), which means that we can express the squared length of any vector expressed in this system as the sum of the individual derivatives squared, just as in Pythagorean calculations. As a result we can re-express our de Sitter metric as:

$$ds^{2} = -\mathcal{R}^{2}d\alpha^{2} + \mathcal{R}^{2}\cosh^{2}\alpha \left(d\beta^{2} + \sin^{2}(\beta)\left[d\gamma^{2} + \sin^{2}\gamma d\delta^{2}\right]\right), \qquad (1.5)$$

which, if we substitute β , γ and δ with polar coordinate notation with which we are more familiar: r, θ and ϕ respectively, and recall the definition of ds^2 from Equation 1.1 we get:

$$c^{2}d\tau^{2} = c^{2}dt^{2} - \mathcal{R}\cosh^{2}\left(\frac{ct}{\mathcal{R}}\right)\left(dr^{2} + \sin^{2}(r)\left[d\theta^{2} + \sin^{2}\theta d\phi^{2}\right]\right),$$
(1.6)

where

$$\alpha = \frac{ct}{\mathcal{R}}$$

This equation now gives us a different means of interpreting our de Sitter metric. Specifically we can rewrite it like this:

$$(\text{interval})^2 = (\text{time interval})^2 - (\text{scale factor})^2 (\text{comoving interval})^2, \qquad (1.7)$$

where the universal scale factor R(t) can now be defined as:

$$R(t) = \mathcal{R} \cosh\left(\frac{ct}{\mathcal{R}}\right) \tag{1.8}$$

It is the appearance of this scale factor R(t) in de Sitter's work in 1917/18 that led him to the conclusion that our Universe must be expanding just like Slipher and Lemaître, and it is indeed this 'de Sitter effect' that Hubble was attempting to observe in his own work in 1929 [84] [145].

Equation 1.7 in particular is actually the descriptive definition of what is known as the **Friedmann-Lemaître-Robertson-Walker Metric** (or FLRW Metric), which was developed in 1935 by Howard P. Robertson in collaboration with Alexander Friedmann, Georges Lemaître and Arthur Geoffrey Walker. It is an exact solution of Einstein's field equations of general relativity for the case of a homogeneous, isotropic, expanding or contracting Universe that is path connected, but not necessarily simply connected. The mathematical expression of the FLRW metric can be expressed as:

$$c^{2}d\tau^{2} = c^{2}dt^{2} - R^{2}(t)\left[dr^{2} + S_{k}^{2}(r)d\psi^{2}\right],$$
(1.9)

where $d\psi$ is the angle that separates two points on the sky, such that

$$d\psi^2 = d\theta^2 + \sin^2\theta d\phi^2$$

when expressed in spherical polar coordinates [145]. The function $S_k(r)$ is an interesting part of the FLRW metric due to its definition being reliant on a variable k, taken to describe the curvature of the Universe we are modelling. More specifically $S_k(r)$ allows for positive or negative curvature of the comoving part of the FLRW metric and can take on three different values:

$$S_k(r) \equiv \begin{cases} \sin r & (k = +1) \\ r & (k = 0) \\ \sinh r & (k = -1), \end{cases}$$
(1.10)

corresponding to a 'closed', 'flat' and 'open' Universe respectively [145]. But what do these different Universes mean or represent?

1.3.1 A Question of Spacetime Curvature and the Fate of the Universe

As raised in the previous section, the FLRW metric which describes a homogeneous and isotropic expanding Universe allows for the modelling of Universes with specific curvatures or geometries. The curvatures in question are as follows (with a graphical depiction of those curvatures shown in Figure 1.4) and are summarised in Table 1.1:

- 1. A value of k > 0, which results in a 'closed' spherical Universal geometry. In such a geometry the angles of a drawn triangle would add up to more than 180°, and the circumference of a circle would be less than $2\pi r$. In a closed Universe the currently observed rate of expansion would eventually be superseded by the gravitational forces of all the matter in that Universe, causing it to collapse back in on itself, possibly leading to another Big Bang.
- 2. A value of k = 0, which results in a 'flat' Universal geometry. This is the easiest one to understand and can be imagined easily on a piece of paper, where a drawn triangle will have internal angles that sum up to $\equiv 180^{\circ}$ and the circumference of a circle will equal to $2\pi r$ just like conventional mathematics dictates. In a flat model the Universe is expected to continue to expand ... but only up to a point. Specifically once the forces driving the Universal expansion become perfectly balanced with the gravitational pull of all existing matter, the expansion will halt with the Universe remaining at a fixed size until the end of time.
- 3. A value of k < 0, which results in an **'open'** hyperbolic or saddle-based geometry. In such a 'concave' geometry the angles of a drawn triangle sum to less than 180° and the circumference of a circle is larger than $2\pi r$. An open Universe model is inherently the most depressing, as it predicts that the Universe will continue to expand indefinitely with gravitational forces not being able to curtail the expansion at all. Over time all galaxies and large scale structure will continue to smooth out and move further and further away from each other, leading to the eventual 'heat-death' of the Universe. All the stars will run their course and exhaust their supplies

Curvature	Geometry	Angles of triangle	Circumference of circle	Type of Universe
k > 0	Spherical	$> 180^{\circ}$	$c < 2\pi r$	Closed
k = 0	Flat	$= 180^{\circ}$	$c = 2\pi r$	Flat
k < 0	Hyperbolic	$< 180^{\circ}$	$c > 2\pi r$	Open

TABLE 1.1: Summary of possible Universe curvatures and geometries for use in the FLRW metric [121].



FIGURE 1.4: Depiction of the three possible Universal curvatures in relation to different densities of the Universe, denoted as Ω_0 here. The density of the Universe also determines its geometry. If the density of the Universe exceeds the critical density (see Section 1.4 for an introduction to that variable), then the geometry of space is closed and positively curved like the surface of a sphere. This implies that initially parallel photon paths converge slowly, eventually cross, and return back to their starting point (if the Universe lasts long enough). If the density of the Universe is less than the critical density, then the geometry of space is open, negatively curved like the surface of a saddle. If the density of the Universe exactly equals the critical density, then the geometry of the Universe is flat like a sheet of paper. Thus, there is a direct link between the geometry of the universe and its fate. Image reproduced from the works of NASA and the WMAP team [138].

of hydrogen fuel, ejecting all their energy and heat into space, and there will no longer be any matter-dense regions remaining rich enough to kick-start further stellar evolution and large-scale structural development of any kind [121].

1.4 The Friedmann Equation and Critical Density

Having now defined a metric with which we can model and describe an expanding Universe of our choice, we now have the means to define the set of comoving coordinates we discussed at the end of Section 1.2 so that we can track the movements of objects across time as the Universe continues to expand. More specifically, with the introduction of the Universal scale factor R(t), we can now express all position vectors at a given time t as just the scaled versions of their values at a reference time t_0 such that:

$$\mathbf{x}(t) = R(t)\mathbf{x}(t_0) \tag{1.11}$$

If we were to differentiate this with respect to t to obtain the velocity vector we would get:

$$\dot{\mathbf{x}}(t) = \dot{R}(t)\mathbf{x}(t_0) = \left[\frac{\dot{R}(t)}{R(t)}\right]\mathbf{x}(t), \qquad (1.12)$$

where the characteristic time of the expansion, otherwise known as the **Hubble time** H(t) is given as:

$$H(t) = \frac{\dot{R}(t)}{R(t)},$$

and the parameter H_0 - commonly known as the Hubble constant - is the value of H(t) given at the current epoch [145]. We will discuss the attempts over the past century to define the true value of H_0 in Section 2.2.

Having now defined our comoving coordinate system we are now in a position where we can begin to describe our expanding Universe mathematically. While a rigorous treatment of this would require us to use the principles of general relativity and solve Einstein's gravitational field equations, we will proceed with a simpler, approximate, and heuristic treatment by considering the laws of energy conservation for any object in the Universe instead. More specifically, the law of energy conservation states that the total energy of an object U is given as the sum of its kinetic T and potential energies V, the latter corresponding in this case to its gravitational potential energy such that:

$$U = T + V \tag{1.13}$$

The kinetic energy of an object of mass m is easily given by:

$$T = \frac{1}{2}m\dot{\mathbf{x}}^2,\tag{1.14}$$

and its gravitational potential energy (assuming that this object is spherical and of density ρ , existing a small distance **x** away from a theoretical 'particle' of mass m) is given by:

$$V = -\frac{GMm}{\mathbf{x}} = -\frac{4\pi G\rho \mathbf{x}^2 m}{3},\tag{1.15}$$

where

$$M = \frac{4}{3}\pi\rho\mathbf{x}^3,$$

and G is Newton's gravitational constant given as $G = 6.672 \times 10^{-11} \text{m}^3 \text{kg}^{-1} \text{sec}^{-2}$. Substituting these into Equation 1.13 gives:

$$U = \frac{1}{2}m\dot{\mathbf{x}}^2 - \frac{4\pi}{3}G\rho\mathbf{x}^2m,$$
 (1.16)

and making use of our definition of comoving coordinates in Equation 1.12 we can rewrite this as:

$$U = \frac{1}{2}m\left[\dot{R}\mathbf{x}^2\right] - \frac{4\pi}{3}G\rho R^2 \mathbf{x}^2 m, \qquad (1.17)$$

Multiplying each side through by $2/mR^2\mathbf{x}^2$ and rearranging the terms will then give:

$$\left(\frac{\dot{R}^2}{R}\right) = \frac{8\pi G}{3}\rho - \frac{kc^2}{R^2} \tag{1.18}$$

which, substituting our expression for the Hubble time yields:

$$H^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{R^2}$$
(1.19)

where

$$kc^2 = -\frac{2U}{m\mathbf{x}^2}$$

,

and k continues to be the same Universal curvature parameter we defined earlier. Considering that we established that k will either be positive, negative or 0, we can rewrite this equation one last time to get:

$$H^{2} = \frac{8\pi G}{3}\rho - \frac{\text{const}}{R^{2}}$$
(1.20)

Equation 1.20 is known as the **Friedmann Equation**, which was developed by the Russian physicist and mathematician Alexander Friedmann in 1922 (albeit published in German [68], the 1999 English translation can be referred to here: [69]) and is one of the cornerstone equations of modern cosmology. It provides us with a useful means of describing an expanding Universe of specific density and curvature, and solving for the various parameters will help give us a gauge on the age of the Universe, the rate

of its expansion and its ultimate fate. While admittedly Friedmann derived Equation 1.20 using the more rigorous treatment we mentioned earlier: utilising the principles of general relativity and Einstein's gravitational field equations, the same result can be achieved through the principles of energy conservation, as we have just demonstrated [121].

Due to the existence of the variable k in the Friedmann equation we are now presented with something interesting to consider. Supposing that the Universe we are inhabiting is flat with k = 0, then the constant term on the right hand side vanishes. Rearranging Equation 1.20 to accommodate for this we get the following:

$$\rho_c = \frac{3H^2}{8\pi G} \tag{1.21}$$

where ρ_c is defined as the **critical density** necessary to yield a flat Universe. It is often common to redefine ρ_c to be dimensionless by defining it as the ratio of density to critical density via:

$$\Omega \equiv \frac{\rho}{\rho_c} = \frac{8\pi G\rho}{3H^2},\tag{1.22}$$

where Ω is taken to be the **total energy density** of the Universe, and by solving the Friedmann Equation for the present day value of the scale factor, R_0 , yields a value of $\Omega = 1$ for a flat Universe [145].

If we were also to define a dimensionless form of the Hubble parameter h, such that

$$h \equiv \frac{H_0}{100 \mathrm{km s}^{-1} \mathrm{Mpc}^{-1}}$$

where h can also be used to parametrise our ignorance of the true value of H_0 [81], we can solve Equation 1.21 for the critical density ρ_c such that:

$$\rho_c = 1.878 \times 10^{-26} h^2 \text{kgm}^{-3}$$

Measuring the density of our Universe to see how well it compares with ρ_c is no easy task, but the methods used can generally be boiled down to one of two approaches:

1. The accounting approach in which one attempts to estimate the mass of a given (large) volume of the Universe by measuring the masses of objects within the volume. Masses may be estimated directly (e.g. by the measurement of kinematic properties such as galaxy motions within clusters) or indirectly by assuming a relation between the luminosities and masses of individual galaxies within the volume. This indirect method suffers from our lack of knowledge of the fraction of dark matter present in and around galaxies (refer to Section 1.5 for an in depth

introduction to this curious cosmological entity). However, the technique can still be used, with an appropriate assumption about the luminous to dark matter ratio, to estimate the total mass in the volume [37].

2. The geometrical approach which makes use of the idea of converging/diverging parallel lines, in a smilar fashion to our depiction of different Universal geometries in Figure 1.4. For example, if the Universe is closed and the parallel lines converge, the observed density of distant galaxies should be less than that expected by extrapolating the local density of galaxies backwards in time. On the other hand, in an open Universe, the diverging parallel lines would cause the observed density of distant galaxies to be greater than expected [37].

To date, both of these techniques return values for the density of the Universe entirely consistent with, or extremely close to, the critical density ρ_c , lending credence to the theory that we are actually inhabiting a flat Universe [37]. Further compelling evidence for us indeed existing in a flat Universe would be obtained from the results of WMAP and Planck (space-based probes analysing the primordial blackbody radiation of the early Universe) from 2003 through to 2016, which we will explore in more detail in Section 1.7.

As it stands, our current understanding of the Universe (which we will delve into in more detail in upcoming sections of this chapter) appears to suggest that it consists of three key components: the luminous baryonic matter that we can directly observe, dark matter - a curious addition to our cosmological modelling that is used to explain the strange additional peculiar motions most objects in the Universe appear to exhibit, and an illusive vacuum energy component called dark energy that is used to model the accelerated Universal expansion that was detected using supernovae in 1998. As a result of this, if we are to continue to presume that we are inhabiting a flat Universe, then the sum of the energy densities of these components must equal 1. Put another way:

$$\Omega \equiv \Omega_b + \Omega_m + \Omega_\Lambda \equiv 1, \tag{1.23}$$

where Ω_b , Ω_m , and Ω_Λ represent the energy densities of baryonic matter, dark matter and dark energy respectively. Constraining the values of these various energy densities and by proxy the value of H is one of the key targets of modern cosmology ... as the knowledge of the fate of our Universe depends upon it, after all.

1.5 The Mystery of Dark Matter

The quest for the identity of dark matter is an interesting one that has spanned well over 70 years, where it was first noted by F. Zwicky in 1937 when, while studying nebulae clusters, he reported a discrepancy between the nebulae mass calculated from observed individual nebulae luminosities, and the average individual mass calculated from observed nebulae cluster sizes. This discrepancy was on the order of a factor of 500, which was significantly larger than the variational factor of 3 which had been noted so far for Kapteyn stellar systems [217]. This was echoed in similar studies by F. D. Kahn and L. Woltjer in Princeton University [101] whereupon when they attempted to estimate the reduced mass of the Milky Way and M31 'Andromeda' galaxies (i.e. to calculate their masses as if they were one 'fused' object as opposed to two entities separated by a considerable distance) using an estimate of their shared centre of gravity and applying Kepler's third law, they calculated a reduced value of $M_* \ge 1.8 \times 10^{12} M_{\odot}$, which was six times larger than observed values [101]. So where was all this missing matter? The idea of hidden or invisible matter began to take a vague shape. This idea began to gain more support when in 1970, scientist Vera Rubin published her observed velocity curves for the Andromeda Galaxy. Newtonian laws predicted that all bodies moving around a centre must move more slowly with increasing distance from the centre. Consequently it was expected that the velocities of the monitored HII regions of Andromeda would decrease with increasing distance from the central core of the galaxy [165]. This was not what Rubin saw however, as can be noted in Figure 1.5. What she saw in fact was completely contradictory: the bodies moving around the outskirts of the galaxy were moving at approximately the same speed as the bodies orbiting near its centre (approximately $270 \text{kms}^{-1} \pm 10 \text{kms}^{-1}$ [165]), therefore suggesting that some extra matter had to exist that would provide these bodies with the 'extra motion' that they were exhibiting [166].

By 1974, all of these peculiarities led to one apparently inescapable alternative, as noted by Einasto et. al [54] in their discussion of dynamical evidence for the existence of massive coronas in galaxies:

"A longstanding unresolved problem in galactic astronomy is the mass discrepancy observed in clusters of galaxies. The virial mass of the cluster per galaxy and the mass-luminosity ratio are considerably larger than the corresponding quantities for individual galaxies. This discrepancy cannot be a result of expansion or be because of the recent origin of clusters: these ideas contradict our present knowledge of the physical evolution and ages of



FIGURE 1.5: Rotational velocities for M31, as a function of distance from the centre of the galaxy as provided from the works of Rubin and Ford [165].

galaxies. Therefore it is necessary to adopt an alternative hypothesis: that the clusters of galaxies are stabilised by hidden matter." [54]

By the 1980s this hidden matter, or dark matter as it were, was accepted by scientists to exist, but many questions as to its exact identity, spatial distribution, energy density and how to model it still remained unanswered.

1.6 A Question of Cosmological Models

The first question that was posed as to the identity of dark matter was whether it was baryonic or non-baryonic in nature. Attempts to model a Universe where the dark matter was due to purely baryonic adiabatic fluctuations was ruled out relatively quickly in the 1980s when the resultant models failed to predict what was actually being observed [157]. More specifically, for any model in which only baryons and adiabatic fluctuations were considered, they did not generate enough significant thermal fluctuations or variations in the Cosmic Microwave Background Radiation (CMBR) necessary to generate the large scale galactic structures and galaxy formations that we observe in the current day [157].

Acronym	Cosmological Model	Flourished
HDM	Hot Dark Matter, $\Omega_m = 1$	1978 - 1984
SCDM	Standard Cold Dark Matter, $\Omega_m = 1$	1982 - 1992
CHDM	Cold+Hot Dark Matter, $\Omega_m = 0.7$, $\Omega_{\nu} = 0.2$ -0.3	1994 - 1998
ΛCDM	Cold Dark Matter, $\Omega_m \sim \frac{1}{3}$, $\Omega_\Lambda \sim \frac{2}{3}$	1996-today

TABLE 1.2: The development of cosmological models to the present date [157].

So now the next question was ... what kind of non-baryon could be a dark matter candidate?

The first attempt to answer this question came in 1982 in the form of the theorist Zel'dovich with his proposal of a hot dark matter theory. Hot dark matter (HDM) is postulated to be composed of high energy particles formed shortly after the Big Bang that are travelling at ultra-relativistic speeds or, very very close to the speed of light, c. To that effect Zel'dovich theorised that the Universe was dominated by light neutrinos making up most of the dark matter distribution [215]. However, the small scale fluctuations predicted by this model would have been damped out by the relativistic motion of these neutrinos, ultimately predicting a galaxy distribution that would be much more anisotropic and non-homogeneous than the one that was being observed, such as in Figure 1.6, leading to this model being eventually ruled out in the later 1980s [157]. With the ruling out of neutrinos with masses on the order of a few tens of eV, the focus moved to another potential candidate: Weakly Interacting Massive Particles (WIMPS). This covered several possibilities: the neutrino once again, this time on a massive scale (an order of 100eV), and the gravitino, the theoretical supersymmetric partner of the graviton. This was the beginnings of the cold dark matter model (CDM), that dealt with particles that moved more sluggishly in the early Universe. At the same time, theorists began to suggest an Ω_b of 0.2, where Ω_b is the critical energy density ratio for baryonic matter as discussed at the end of Section 1.4. In other words, the observable Universe was thought to consist of 20% regular matter and 80% cold dark matter. At the time, this model was consistent with the inferred proportions of luminous and dark matter and gas in clusters, and it also predicted the characteristic luminosities of the bright galaxies observed in the day.

In 1984, the proposed CDM model began to take firmer hold, with the dark matter candidates including axions (another hypothetical elementary particle) and stable photinos (the theoretical WIMP supersymmetric partner of the photon), which better fit the observed mass range of galaxies, and was also consistent with the large-scale clustering, superclustering and voids being observed in the distribution of galaxies in redshift surveys [15].



Copyright SAO 1998

FIGURE 1.6: A slice from the CfA galaxy redshift survey completed in 1982, depicting the affectionately named 'Matchstick Man' in a non-random, non-isotropic observed distributon of galaxies, apparent filaments and voids [86].

One such set of surveys was the CfA (Centre for Astrophysics) surveys, which were started in 1977, and its first run being completed in 1982. The first slice of the survey is depicted in Figure 1.6, where the galaxy distribution is clearly anything but random and isotropic, appearing to be distributed on 'bubble-like surfaces' [86] to create filaments and voids, and an affectionately nicknamed 'Matchstick Man'.

However, the CDM model was not without its inconsistencies, particularly over smaller cosmological scales or where actual galaxy formation was concerned. Some of these challenges included, but are not limited to:

- The 'Cuspy Halo Problem' as noted by Gentile et. al, where the density distributions of dark matter halos in CDM simulations are seen to be much more peaked than what is observed in galaxies when their rotation curves are observed [71].
- The 'Missing Satellites Problem' of Klypin et. al, where CDM simulations predicted much larger numbers of small dwarf galaxies orbiting larger galaxies like the Milky Way than are observed [103].
- The 'Disk of Satellites' problem as discussed by Pawlowski et. al, where they noted that dwarf galaxies around the Milky Way and Andromeda galaxies are observed

to be orbiting in thin, planar structures whereas CDM simulations predict that they should be distributed more randomly about their parent galaxies [144].

• The 'Galaxy Morphology' problem noted by Kormendy et. al. Specifically they noted that if galaxies grew in a hierarchical fashion over time, then massive galaxies would require many mergers, with CDM simulations predicting that that many major mergers would indelibly create a classical bulge in the newly formed central core of the merged galaxy. However they indicate that about 80% of observed galaxies are bulgeless, and giant pure disc galaxies are commonplace, in complete contradiction to CDM simulations [107].

While some solutions have been proposed to these problems, there remains uncertainty as to whether these challenges can indeed be solved without resorting to abandoning the CDM model in its entirety, so another possible alternative needed to be found.

In the 1990s, such an alternative cosmological model was indeed proposed: the warm dark matter model (WDM or CHDM - Cold/Hot Dark Matter), which was designed to bridge the gap between the relativistic inconsistencies of the HDM models and galaxy clustering and hierarching inconsistencies of the CDM models while still predicting a Universe similar to what we observe today [179]. It was hoped that the WDM model would resolve the issues of these competing models by warming and smoothing out the particles that constitute the CDM [136], but the publication of the Wilkinson Microwave Anisotropy Probe (WMAP) results in 2003 rather put a stop to that (see the work of Spergel et. al [182] for a discussion of the preliminary results and observations from the probe for that first year run. Additional media and maps can be found at the following link provided by the Goddard Space Flight Centre: https://www.nasa.gov/ centers/goddard/news/topstory/2003/0206mapresults.html). In particular, their calculations for the re-ionisation period of the Universe, the period of time in the early Universe when hydrogen became cool enough to interact with itself, put serious constraints on the WDM model; ultimately jeopardising it. Based on the results of WMAP, the WDM model would have to rapidly increase its ionisation factor in order to match the new constraints and correlate with the calculated re-ionisation period of the Universe, and this unfortunately was something that the WDM model by its very design could not do, leading to its ruling out just like the HDM model [207].

The CDM model, challenges and all, was once again ready to be adopted as the most statistically probable model for the Universe, with the concept of cosmic inflation being used to explain the small-scale anisotropies observed by WMAP in the CMBR (refer to Section 1.7 for a brief summary of the relevant theory) until 1998. Recall from Section 1.4 where we utilised the Friedmann Equation to derive the critical energy density parameter



FIGURE 1.7: Pie chart depicting the distribution of baryonic matter, dark matter and dark energy in the Universe, as provided by Spergel et. al [182] and the team behind WMAP at NASA [206].

 Ω , and how its value is reliant on the observed curvature of the Universe k and the energy densities of its constituent components. In particular as mentioned in Section 1.3.1 for a presumed flat Universe, a slow-down in its expansion is expected before eventually stopping entirely and maintaining an equilibrium state with the gravitational forces exerted by all objects in the Universe. When a few astronomers took it upon themselves to study supernovae for such evidence that the Universal expansion was indeed slowing down, they discovered something rather startling. The Universal expansion was not slowing down at all. It was in fact, accelerating.

1.7 The Breakthrough: Dark Energy and the Double Dark Theory

It was the breakthrough of the decade. Not only was the expansion accelerating, but it had been accelerating for some time: starting at least when the Universe was 10 billion years old [70]. Now a new objective was set: attempt to identify what is causing this phenomenon. Over time, a description of the 'culprit' came into view: whatever it was accounted for over two-thirds of the cosmic energy density, rendering previous estimates of the dark matter energy density $\Omega_m = 0.8$ impossible. More specifically by utilising supernovae to constrain both Ω_m and this 'culprit', denoted as Ω_{Λ} , A. V. Fillipenko was able to compute values for both these parameters as $\Omega_m = 0.3$ and $\Omega_{\Lambda} = 0.7$ [63].

The 'culprit' also exhibited other curious properties: it was apparently gravitationally repulsive, did not appear to cluster in galaxies, and was last seen stretching space-time apart [23]. Theorists named the phenomenon **dark energy**, with a critical energy density denoted by Ω_{Λ} , and the Λ CDM model was born, otherwise known as the Double Dark Theory that described a Universe that consisted of both dark energy and cold dark matter alongside baryonic matter. However, theorists felt that the direct evidence provided by the supernovae was not compelling enough to state as fact that dark energy existed as a reason for the cosmic acceleration. That would change however with the data provided by WMAP in 2008 [157] as it provided two key pieces of information that would change the case for the existence of dark energy from possibility to very compelling.

The first piece of information was a result of combining data from WMAP, the Sloan Digital Sky Survey (SDSS) and other sources to report evidence of a phenomenon known as the Integrated Sachs-Wolfe effect, where they found that the gravitational repulsion of dark energy slowed down the collapse of overdense regions of matter in the Universe [66]. The second piece of evidence came from the WMAP (and further updated in 2013 with Planck) images of the Cosmic Microwave Background.

The CMBR is a relic from some 400,000 years after the Big Bang and is black-body radiation from the primordial plasma of that era. As the Universe cooled below about 3000K the plasma became transparent to photons, allowing them to propagate freely through space. To that effect the CMBR actually represents the surface of last scattering, beyond which the Universe becomes opaque and we are unable to look any further back into the past.

The thermal fluctuations mapped in the CMBR, known as the CMB anisotropy, are on the order of 10^{-5} K, and reflect the slight variations in density and motion of the early Universe. It is effectively a blueprint for the large-scale structures of galaxies and clusters that we see today. A key element of the CMBR lies in the angular size of these aforementioned anisotropies, whose intrinsic sizes are well determined by plasma physics. The conversion from physical scale into angular scale on the sky for these anisotropies will depend heavily on the curvature of the Universe (as discussed in Section 1.3.1) and the distance to the surface of last scattering [82]. More specifically over cosmological distances as large as the redshift-distance to the surface of last scattering 400,000 years after the Big Bang (taken to be approximately at a redshift of $z \approx 1100$ [187]), the spacetime curvature dominates, with small angular sizes on the order of 0.5° of these CMB anisotropies being equated with a closed Universe model with k = 1. Similarly larger observed angular sizes on the scale of approximately 1° of these anisotropies would indicate a flat k = 0 Universe [83], while anything larger than 1.7° for these anisotropies (the angular size that indicates the particle horizon length i.e. the maximum distance over which particles could have traveled to the observer in the age of the Universe [187]) would indicate an open Universe model with k = -1 [83].

The most recently published WMAP results have since confirmed an angular anisotropy size of 1°, indicating that the Universe is flat, and has estimated an energy density for Ω_{Λ} at 0.742±0.030 [157], and consequently has lent credence to the theory of inflation, where in the first few moments after the Big Bang, a false vacuum environment was generated that accelerated expansion and stretched away any large-scale spatial curvature that would have made the Universe anything other than flat. Such an inflation period could not have been driven without some form of energy or particle (an inflaton, as it is termed) that does not cluster in galaxies, is gravitationally repulsive, and stretches out the curvature of space-time, which is exactly what dark energy is defined to be. This helped to solidify the status of the Λ CDM model as the accepted cosmological model of the Universe today.

Although cosmologists are confident that the statistics continue to favour the ΛCDM model in most data runs from WMAP and Planck etc. [157], there continues to be an irony that up until the current day, the identity of dark matter continues to remain elusive, consequently leaving the certainty of ΛCDM always in doubt. While successful measurements of what has been called the 'Casimir Effect' by Steven Lamoreaux in 1996 [114] (said effect being defined by the Dutch physicist of the same name in 1948 as a small attractive force between two extremely close parallel uncharged perfectly conducting plates arising due to quantum vaccuum fluctuations of the surrounding electromagnetic field [27][26]) has lent credence to the existence of, and our ability to measure, vacuum energy in a laboratory environment; it becomes a far more challenging matter to measure this energy on the macro scales over which dark energy itself operates. It is also worth noting that since ACDM is effectively an extension of previously established CDM models, it also continues to suffer from the same sort of challenges that we discussed in brief in Section 1.6 such as the 'Cuspy Halo' and 'Missing Satellites' problems etc.. Similarly as pointed out by Bullock and Boylan-Kolchin in their review of small-scale challenges to the ΛCDM paradigm [21], other anomalies such as the observed planar and orbital configurations of Local Group satellites (what they termed the 'Satellite Planes' problem), and the tight baryonic/dark matter scaling relations obeyed by the galaxy population such that reobtaining the slope and scatter of the baryonic Tully Fisher relation predicted for such populations becomes problematic (termed the 'Regularity in the face of Diversity' problem)[156]; have been less thoroughly explored in the context of ΛCDM theory, necessitating the use of future surveys to discover faint dwarf galaxies and precisely measure their masses and density structures in order to test possible solutions for these challenges [21][156].



(A) The Cosmic Microwave Background as imaged by WMAP in 2003 [206]. The colour variations depicted here indicate slight variations in the photon temperatures across the sky - known as the CMB anisotropy - consequently reflecting slight variations in the density and motion of the early Universe. These variations, which occur at the level of 10⁻⁵K, reveal the blueprint for the large-scale structure of galaxies and clusters that we see today.



(B) The Cosmic Microwave Background as imaged by Planck in 2013 [151]. Unlike WMAP, the photon temperature variations depicted in this map correspond to fluctuations on the order of millionths of a degree Kelvin, an increase in sensitivity of at least a factor of 10, leading to even more accurate analyses of current day large-scale structures of galaxies and clustering, and providing greater insight into the nature of the density fluctuations present soon after the birth of the Universe [60].

FIGURE 1.8: The evolution in detail of the CMBR over the past decade.

As always with cosmology however, theories will continue to be refined as the search for the identity of dark matter and dark energy continues alongside continued attempts to constrain their energy densities so that we can fine tune our understanding of our own Universe and what fate indeed awaits it ... just going to show how much more there is out there to discover and explore.
Chapter 2

A Cosmologist's Toolbox

In order for any sort of cosmological analysis to be viable, we need to be able to determine the velocities of, and distances to objects in our Universe. Without such rulers and abilities to track the motion of the smallest particles to the largest galactic structures, none of the discoveries and theoretical models of our Universe presented up to this point would have been possible. Here we will introduce some of the various distance and velocity measuring techniques that are at our disposal, focusing in particular on the tools that we will make use of in future sections of this work.

2.1 Light Propagation and Redshift

Redshift is the term used to describe the amount by which the wavelength of light from a receding object in space is lengthened [5]. In its simplest form, if we choose to consider an object in a **static** Universe such as a star that emits light as it moves away from us, that movement will cause the observed wavelength of that emitted light to be stretched (reduced in frequency) while it travels towards us as the observers, while a star moving towards us would instead have the wavelength of the emitted light squashed (increased in frequency) when it is observed. All objects that emit light have a characteristic set of absorption and emission lines in their observed spectra. Consequently for a receding object we can expect these observed spectral lines to shift towards the lower frequency 'red' end of the spectrum, experiencing a **redshift**, while an incoming object would cause those observed spectral lines to shift towards the lower frequency the spectrum, exhibiting a **blueshift**. Since Edwin Hubble deduced that most objects in the Universe appear to be receding away from us, it has become the convention to define the redshift z of an object as:

$$z = \frac{\lambda_{obs} - \lambda_{em}}{\lambda_{em}},\tag{2.1}$$

where λ_{obs} and λ_{em} represent the wavelength of light at the time of observation and emission respectively. However this formulation for z is overly simplistic since our Universe is not static and is in fact expanding. Consequently there are actually three main factors that go into the composition of an object's observed redshift, namely:

1. Relativistic Doppler redshift, where the peculiar recessional velocities inherent to the object at the moment that light was emitted will cause changes in the wavelengths of light observed and is most valid over small scale distances in the Universe where $v \ll c$. More specifically if an object appears to be receding at a velocity v then its Doppler redshift is consequently given by:

$$z = \frac{v}{c},\tag{2.2}$$

where c is the speed of light. It should be noted that Equation 2.2 only holds true for low redshifts of z < 1 because at $z \sim 1$ and larger, the equation reduces to objects possessing speeds of c or larger which, by the rules of relativity is impossible. To accommodate for this Equation 2.1 can be rewritten as:

$$1 + z = \frac{\lambda_{obs}}{\lambda_{em}} \tag{2.3}$$

and the special relativity result as developed by Einstein for calculating an object's relativistic Doppler redshift from its recessional velocity is given as:

$$1 + z = \sqrt{\frac{1 + v/c}{1 - v/c}},$$
(2.4)

for which the expression v/c is taken to be small [55][121].

2. Cosmological redshift, where the observed expansion of the Universe will also cause the wavelengths of light emitted from an object to be further stretched with this redshift becoming more dominant over its relativistic Doppler counterpart over larger scale distances. The cosmological redshift can be expressed in terms of the scale factor *a* such that:

$$1 + z = \frac{a_{now}}{a_{then}},\tag{2.5}$$

where a_{now} denotes the present-day value of the scale factor, often denoted as a_0 , and a_{then} denotes the scale factor of the Universe at the time the light was emitted [200].

3. Gravitational redshift, also known as the 'Einstein shift', which is the phenomenon wherein which the wavelengths of light emitted from an object are stretched as they pass through strong gravitational fields, be they the gravitational fields of dwarf stars (which are relatively weak), or the much stronger fields emitted from the surface of neutron stars or the event horizon of a black hole as they travel towards us [3]. With regards to the former, the gravitational redshift can be approximated as:

$$z \sim \frac{GM}{rc^2},\tag{2.6}$$

where r is the photon's starting distance from the object emitting the gravitational field of mass M, while for stronger gravitational fields emitted by neutron stars and black holes the gravitational redshift is given by:

$$1 + z = \frac{1}{\sqrt{1 - \frac{2GM}{rc^2}}},$$
(2.7)

with such redshifts usually being very large [38].

The compilation of massive redshift surveys of objects in the observable Universe and their unmistakable importance in observational cosmology will be detailed in Section 2.10.

2.2 Hubble's Law

Edwin Hubble's discovery that the Universe was expanding in 1929 through the study of the line spectra and redshift of galaxies was one of the most iconic landmarks of the century for cosmology. In noting that the majority of observable bodies were indeed moving away from us, Hubble deduced the relation that the recessional velocity of a galaxy, denoted cz, is proportional to the distance that the light had travelled from the galaxy in the rest frame of the observer [177] such that:

$$cz = H_0 d, \tag{2.8}$$

where H_0 is the eponymous Hubble constant, and the subscripted "0" refers to the current epoch, since the constant is thought to change over time [81] and is denoted by:

$$H_0 = 100 h \rm km s^{-1} M p c^{-1} \tag{2.9}$$

where the h is used to parametrise our ignorance of the true value of H_0 [81]. Our uncertainty on the true value of H_0 dates back to when it was first defined by Hubble in 1929 [84]. His initial value for the expansion rate of the Universe was set at $500 \text{kms}^{-1} \text{Mpc}^{-1}$, from which the expansion age of the Universe was inferred to be 2 Ga. This proved to be problematic when geologists used radioactive carbon dating of rocks to prove that the Earth was at least 3 Ga old [87]. Further problems arose with this value of H_0 astronomically, because the size of the Milky Way was fairly well established by this time, and Hubble's calibration implied that the Milky Way was far larger than any other nearby galaxy except possibly Andromeda, which also wasn't true. In 1956, Allan Sandage published a new value for H_0 at $180 \text{kms}^{-1} \text{Mpc}^{-1}$ [93], and so began the great debate over the true value of the constant. Throughout the coming years, the value would continue to be lowered, where in 1958 Sandage would further constrain H_0 to a lower value of $75 \text{kms}^{-1} \text{Mpc}^{-1}$ [168] before lowering it yet again in a 1974 collaboration with G. A. Tammann to $H_0 = 55 \text{kms}^{-1} \text{Mpc}^{-1}$ [169]. Come the 1990's, the problem of the value of H_0 would be resolved by the HST Key Project: the telescope that bears Hubble's name. They determined a best fit value of $H_0 = 71 \pm 7 \text{kms}^{-1} \text{Mpc}^{-1}$, which most recent attempts to calculate other values for H_0 seem to fall close to [87]. For a comprehensive list of published values of H_0 over the past century complete with associated references, confidence errors and details of the methodologies used, refer to the list compiled by John Huchra at the Harvard-Smithsonian Center for Astrophysics [88] to see how the value has continued to fluctuate over time.

One of the key things that can be taken away from the list compiled by Huchra is that there are two general methodologies that can be used to infer the value of H_0 : direct and indirect inference. More specifically the value of H_0 can be determined indirectly from e.g. measurements of the CMBR at high redshift or from large scale structure surveys at intermediate and lower redshifts [142]. Conversely it can be measured directly by measuring the velocities and distances to so-called "standard candles" in the nearby Universe (objects in space whose distances are known to very high accuracy and have very clearly defined luminosities) such as Type Ia supernovae (refer to Section 2.5 for a brief introduction into their use as standard candle field probes). Both these methodologies have their advantages and disadvantages. Direct measurements are independent of the cosmological model being assumed such as Λ CDM, but are very prone to systematic errors related to local bulk flows in the nearby Universe or whatever assumptions are being presumed for the standard candles being utilised in the analysis [142]. Similarly, indirect measurements from the CMBR or large scale structure surveys are very robust and precise, but are completely reliant on the underlying cosmological model being correct - thus resulting in discrepancies being observed in values of H_0 that are measured at either high redshift or intermediate to lower redshifts [142].

A typical example of this would be results recently released from Planck in 2013 analysing the CMBR that point to a lower H_0 value of 67.3 ± 1.2 kms⁻¹Mpc⁻¹ [153], while direct measurements by Riess et. al utilising Type Ia supernovae in the nearby Universe produced a larger value of $H_0 = 73.8 \pm 2.4$ kms⁻¹Mpc⁻¹ [164], with the two values being inconsistent at the 2.4σ confidence level [142]. This discrepancy could be due to the underlying Λ CDM model being incorrect on a fundamental level as previously mentioned, or due to inaccuracies or biases in the methodologies being used.

With respect to the latter it must be noted that as cosmologists continue to refine the methodologies at their disposal, computed values of H_0 began to improve. In particular George Efstathiou used improved distance calibrations for Type Ia supernovae to recompute H_0 to 72.5 ± 2.5 kms⁻¹ Mpc⁻¹ [53], while Clarkson et. al showed that applying relativisitic corrections to the distance to the surface of last scattering of the CMBR increases the best-fit Planck value of H_0 by 5% [32], slowly beginning to bring computed vales of H_0 closer in agreement with the value obtained by the HST Key Project.

For the purpose of this work (and for the sake of scaling simplicity in our future computational endeavours) we will use a Hubble value of

$$H_0 = 100 \rm{kms}^{-1} \rm{Mpc}^{-1}, \qquad (2.10)$$

performing the necessary corrections and rescaling of real-space distances of galaxy surveys as required should the need arise.

2.2.1 Modifying Hubble's Law

Value of H_0 aside, this version of Hubble's Law as presented in Equation 2.8 poses a fundamental problem. It is incapable of describing the peculiar motions exhibited by galaxies as they recede at velocity cz through a varying mass density field such as those observed in reality, making accurate cosmological analyses of galactic kinematics and dynamics problematic at best. Additionally it is difficult to decompose the observed velocity of a galaxy into its receding component due to its redshift, and its peculiar component. If one is to accurately model observed galactic behaviour and large scale structure in the Universe and potentially verify the validity of the Λ CDM model, a modified version of this Law that allows for the accurate inclusion of these peculiar motions is absolutely essential.

Thankfully the cosmological principle does provide us with a solution. Specifically the cosmological principle emphasises that the typical size of the peculiar velocity is independent of the real-space position of the galaxy in space [121], and therefore logically the peculiar velocity should be independent of the real-space position which, as established in Equation 2.8, is proportional to the Hubble distance. As such, the law can now be modified to describe both the receding and peculiar forms of this motion as follows:

$$cz = H_0 d + v_{pec},\tag{2.11}$$

with H_0 remaining as stated in Equation 2.10. Now armed with this modified form of Hubble's Law, studying galaxy kinematics becomes a relatively simpler affair, and now opens the door to the sort of integral kinematic analyses upon which this work will be completely reliant, namely constraining the parameters of observed peculiar velocity fields in current redshift and redshift-distance surveys within a Λ CDM modelled methodology. The parameter values obtained by us in this work will serve as a means of testing and fine tuning the methodologies we will develop here as we compare them against parameter values computed in other works and collaborations. It is our hope that the methodologies discussed in future sections are successful, opening them for future use by others in various cosmological survey analyses and peculiar field velocity probes, without the modified Hubble Law none of which would be possible.

2.3 Apparent, Absolute Magnitudes with the Distance Modulus Law

The distance modulus of an object, denoted μ , is the difference between its apparent magnitude, i.e. its brightness as observed on Earth, and its absolute magnitude: what its brightness would be if it was placed at a distance of 10 parsecs away from Earth (or 10pc, where 1pc is defined as the distance at which 1 astronomical unit, 1AU - the distance between the Earth and the Sun, subtends an angle of 1 arcsecond on the sky), an equivalent distance of 32.6 light-years. If the distance modulus of an object is known, then we can calculate its distance as follows:

$$d = 10^{0.2(\mu+5)} \tag{2.12}$$

Similarly, we can rearrange Equation 2.12 to get

$$\mu = 5\log d - 5, \tag{2.13}$$

but since we will often be working in distances of Mpc throughout this work, Equation 2.13 needs to be reworked to reflect this:

$$\mu = 5\log d_L + 25, \tag{2.14}$$

where d_L is the luminosity distance of the objects we are using. More generally the d_L of an object can be calculated from its observed intrinsic luminosity L and its measured flux S via the equation:

$$d_L = \sqrt{\frac{L}{4\pi S}},\tag{2.15}$$

As can be expected, the accuracy of distances computed in this manner will rely heavily on the accuracy of the luminosity information of the available objects such as supernovae, consequently having a knock-on effect on the values of cosmological parameters constrained using this information.

2.4 A Question of Selection Effects: Malmquist Bias

One thing as cosmologists that we need to consider when dealing with any kind of cosmological survey is the influence of selection effects on the data we observe and analyse. In particular, consider an area of the sky that is filled with galaxies of varying brightnesses such as what we might see in any of the images of the Hubble Deep Field (as can be seen in Section 1.2). Observational equipment will always have a limiting faint magnitude limit below which no galaxy will be observed, as it is too faint to be seen. This holds true for faint galaxies that are very far away as well as for faint galaxies that are much closer and should be able to be seen if our equipment was sensitive enough. Conversely, very bright galaxies that are both far away and nearby will always be observed as they fall well above our limiting magnitude threshold, which consequently means that there will always be a bias towards luminous galaxies being observed more often than dim ones. This bias, known as Malmquist Bias and developed by the Swedish astronomer of the same name in 1922 and expanded thoroughly upon in 1925 [127] [128], will have an effect on the calculations of the average absolute magnitude and average distance to a group of stars. More specifically because of the luminous galaxies that are at a further distance, it will appear as if the sample of galaxies we are observing is farther away than it actually is, and that each galaxy is intrinsically brighter than it actually is [127]. In order for us to be able to perform any useful analyses we need to be able to properly define the kinds of Malmquist bias that exist and correct accordingly for them.

While the traditional Malmquist correction as defined by Malmquist himself $\overline{\Delta M}$ is given as:

$$\overline{\Delta M} = \overline{M} - M_0, \qquad (2.16)$$

where $\overline{\Delta M}$ is the difference between an average value of absolute magnitude \overline{M} and the true intrinsic value of absolute magnitude M_0 [127], it does lend itself to the bias appearing in three different forms. Averaging over an entire sample of galaxies to obtain its luminosity function would introduce an *integral bias*. Similarly, in the absence of interstellar extinction an absolute magnitude can be uniquely derived from an apparent magnitude and distance where we can average over any of the two variables while keeping the other fixed. This would introduce a *differential bias* that would consequently have two types: *magnitude-dependent* and *distance-dependent* depending on which variable you choose to keep fixed [22]. In particular, the traditional Malmquist correction of Equation 2.16 is the explicit definition of such a distance-dependent bias.

In order to be able to properly perform any sort of correction for this bias, the following assumptions need to be made:

- 1. There exists no interstellar absorption,
- 2. The luminosity function is independent of distance,
- 3. For a given area on the celestial sphere, the spatial density of stars depends only on distance,
- 4. There is completeness to an apparent magnitude limit m_{lim} ,
- 5. Stars are of the same spectral type, with intrinsic mean absolute magnitude M_0 and dispersion σ ,
- 6. The luminosity function can be approximated as a Gaussian with mean M_0 [22].

With these (admittedly very ideal) assumptions in mind, we can integrate the luminosity function over all distances and magnitudes greater than m_{lim} to obtain $\overline{\Delta M}$ such that:

$$\overline{\Delta M} = -\sigma^2 \left[\frac{\mathrm{d}\ln A(m_{lim})}{\mathrm{d}m_{lim}} \right], \qquad (2.17)$$

where $A(m_{lim})$ is the total number of galaxies brighter than m_{lim} [127] [22]. If one were to make the further simplifying assumption that the spatial density of galaxies is homogeneous then this relation can be simplified even further to yield:

$$\overline{\Delta M} = -1.382\sigma^2, \tag{2.18}$$

where σ is the observed dispersion as before [127] [22]. This is known as a **homogeneous Malmquist correction** but it does not hold well over large scales where the effects of clustering would render the spatial density of galaxies inhomogeneous. This can be solved however if one were to assume a prior distribution and likelihood on the logarithmic true distances of the galaxies instead, utilising the observed distribution of 'raw' distance estimates to provide a good approximation to the prior distribution assumed, such as done in the works of Landy & Szalay [115]. In principle this improved prior would take into account the effects of clustering and selection that render the observed spatial distribution inhomogeneous, thus leading to the definition of an **imhomogeneous Malmquist correction** which can better correct for what is observed in reality [79].

In future sections we will make use of a methodology of our own to analyse surveys, namely the ROBUST method as developed by Rauzy & Hendry [160], which by its very construction is such that no Malmquist corrections of any kind should be required. We will delve more deeply into the underlying theory and derivation of this method in Chapter 4.

2.5 SNIa as Standard Candle Field Probes

Ever since the first exploding star was observed in 1885, the urge to classify these 'supernovae' according to their emission spectra began to grow [43]. In 1941, Rudolph Minkowski concluded (based on a sample of 14 objects) that there existed at least two different types of supernovae: provisionally called Type I and Type II [133]. Type I supernovae were broadly classified by their lack of hydrogen emission lines in their observed spectra and their homogeneity in brightness, whereas Type II supernovae exhibited strong hydrogen emission lines and a completely heterogeneous range of brightnesses [133]. In 1964/65 Fritz Zwicky would expand this two-type classification to include several more categories with unique properties: namely Type III, IV and V, though these later additions would all continue to share the same feature of hydrogen emission lines with Type II [218]. While the discussion of the properties of the various classification typings of supernovae is an interesting avenue in and of itself, for the purpose of this work we will only focus on the supernovae type most commonly used in probing the nearby Universe, namely Type I, and more specifically a certain subset of Type I designated as Type Ia.

While there exist at least two agreed upon subsets of Type I supernovae: Type Ia and Ib where the former exhibits hydrogen emission lines in their spectra, while the latter does not, they both share one thing in common: their origins lie in the explosion of a white dwarf star [43]. White dwarf stars are the stellar core remnants of stars that, by the end of their evolutionary cycle, did not possess enough mass to collapse into a neutron star or a black hole and instead collapsed into an extremely dense stellar remnant composed primarily of electron-degenerate matter (high-energy electron plasma). In particular their masses do not generally exceed 1.44 times the mass of the Sun $(1.44M_{\odot})$ otherwise known as the Chandrasekhar limit [28], though that mass is packed into a volume equivalent to that of the Earth.

For Type Ia supernovae specifically to occur requires a white dwarf star to be part of a binary system, though the companion star in the system need not be a white dwarf itself. Due to the high density of the white dwarf its powerful gravitational force will cause it to draw mass from its orbiting companion onto itself, eventually leading to that white dwarf's mass exceeding the Chandrasekhar limit. Once that limit of $1.44M_{\odot}$ is exceeded a nuclear chain reaction is triggered which causes the white dwarf to explode [85]. Because Type Ia supernovae all trigger the same kind of nuclear chain reaction within a white dwarf star, the resultant brilliance of the explosion is also inherently the same. Put another way, Type Ia supernovae are characterised by a very high intrinsic luminosity which as a result makes them excellent distance indicators.

As a consequence of these high intrinsic luminosities, Type Ia supernovae possess characteristic light curves (i.e. plots of their luminosity as a function of time after the explosion) such as the ones depicted in Figure 2.1, where due to the high luminosity, the scatter in the peak blue-band luminosity σ_B on such a plot is assumed to be relatively small: about 0.4 – 0.5 magnitudes [17]. However by the mid 1990s, there was enough variation in observed luminosity peaks in SNIa data to begin to introduce uncertainties and limit the effectiveness of the supernovae as good distance indicators.

Thankfully these uncertainties can be resolved by using various fitting techniques to constrain the models and parameters of these light curves, consequently minimising the breadth of the blue-band luminosity peak and increasing the precision of distance indication to within 7 - 10% [35]. Examples of fitting techniques that can be used to constrain these light curves include SiFTO as developed by Conley et. al which models supernovae light curves by manipulating spectral templates (refer to Conley et. al's work [35] for an in depth discussion on the features of SiFTO), as well as SALT and SALT2 as developed by Guy et. al in 2005 and 2007 respectively, which empirically model SNIa luminosity variations as a function of phase, wavelength, a shape parameter, and a color parameter [73] [74]. Regardless of the fitting technique used, it is necessary to try to



FIGURE 2.1: A typical B-band light curve for Type Ia supernovae, based on the observations of 22 SNe Ia [18]. The timescale used here is in units of days with the y-axis denoting changes in observed magnitude over time. Typically approximately 10 days after the initial explosion (the beginning of the curve) the peak B-band luminosity of the supernova is observed, denoted at $t_B=0$. The constraining of errors and the breadth of this luminosity peak is integral for accurate distance estimations to such supernovae. This peak is maintained for several days before the supernova starts to dim considerably as the explosive energy is radiated into space, denoted by the steep plunge in the curve after $t_B \approx 15$, eventually flattening out into a continual linear decline in brightness that has been observed to continue out to a t_B of 300 to 400 days [18].

reduce the margin of error in the distance measurements of these supernovae as much as possible, as this will have a knock-on effect on the value of any cosmological parameter we are attempting to calculate within this work.

2.6 The Tully-Fisher Relation for Spiral Galaxies

The Tully-Fisher relation, as developed by R. Brent Tully and J. Richard Fisher [195], serves as an empirical relation between the intrinsic luminosity of a galaxy (or its inherent mass) and its rotational or angular velocity, indicating a positive correlation between the two variables. While the luminosity of a galaxy can be determined well enough from photometric observations, determining its rotational velocity requires a little thought. Consider the example of a rotating spiral galaxy that is almost edge-on with an observer on Earth in a Universe that is not expanding. As such a galaxy continues to rotate we would observe half of the spiral disk as blueshifted as it spins towards us, while the other half would be redshifted as it spins away. Consequently by making use of the Doppler shift and redshift equations from Section 2.1, the angular velocity for both halves of the galaxy can be determined, from which the average rotational velocity for the galaxy as a whole can be computed. Having noted throughout this work that the Universe is indeed expanding, we find that this fundamental concept continues to hold true regardless. In an expanding Universe a rotating spiral that is continuing to move further away from an observer would exhibit one half of its disk appearing to receded from us more slowly (i.e. the side of the galaxy rotating towards us exhibiting a smaller redshift) while the other half would appear to be receding from us more quickly, i.e. exhibiting a higher redshift. Applying the Doppler shift equations while taking into account the additional speed from Universal expansion would still give the galaxy angular velocities we require to make use of the empirical relation derived by Tully and Fisher.

As a result the Tully-Fisher relation (hereafter denoted as TF) serves as a very useful tool for determining galaxy distances once their luminosities have been determined from their rotational velocities, particularly their luminosity distances d_L and distance moduli μ ; helping to provide a more fundamental understanding of galactic structure in the Universe as a whole [195]. Specifically, the disk surface brightness distribution of a typical spiral galaxy can be modelled via an exponential law as:

$$I(R) = I(0) \exp\left[\frac{-R}{R_D}\right],$$
(2.19)

where I(0) is the central surface brightness and R_D is the disk scale length. Consequently the luminosity of the entire disk is achieved via integrating over the whole volume:

$$L_D = \int_{Disk} I(R) dA = \int_0^{2\pi} \int_0^\infty I(R) R dR d\theta = 2\pi I(0) R_D^2$$
(2.20)

While an exponential law models out to $R = \infty$, the galaxy luminosity will tend to converge after a few scale lengths, say $R = \alpha R_D$, at which point the rotational velocity of the galaxy should be at its maximum, V_{max} . For a spiral galaxy to remain stable, its total rotational velocity as calculated via Doppler shift should be equal to its gravitational acceleration as defined by Newtonian mechanics, therefore for an object of mass m within a rotating galaxy moving at velocity v at distance r from the centre of the galaxy:

$$\frac{mv^2}{r} = \frac{GM_rm}{r^2},\tag{2.21}$$

which simplifies to

$$v^2 = \frac{GM_r}{r} \tag{2.22}$$

where M_r is the mass of the galaxy inside and up to the defined radius r and G is the gravitational constant. Consequently substituting our maximum rotational velocity out

to αR_D yields:

$$V_{max}^2 = \frac{GM_{\alpha R_D}}{\alpha R_D} \tag{2.23}$$

Squaring Equation 2.23 and substituting in our final result from Equation 2.20 yields the following:

$$V_{max}^{4} = \frac{G^2 M_{\alpha R_D}^2}{\alpha^2 R_D^2} = \frac{G^2 M_{\alpha R_D}^2}{\alpha^2} \cdot \frac{2\pi I(0)}{L_D}$$
(2.24)

We can define the parameter η as the disk mass-to-light ratio via:

$$\eta = \frac{M_D}{L_D} \approx \frac{M_{\alpha R_D}}{L_D} \tag{2.25}$$

and substitute η into Equation 2.24 to obtain:

$$V_{max}^4 = \frac{2\pi I(0)G^2\eta^2 L_D^2}{\alpha^2 L_D},$$
(2.26)

and if we assume that η and I(0) are the same for all galaxies that we get the equivalence relation:

$$L_D \propto V_{max}^4$$
 (2.27)

In most cases the exponent of 4 tends to hold true, however generally speaking the values of η and I(0) are not equal for spiral galaxies, causing the exponent to vary slightly [78].

More specifically, it must be noted that there are various forms of the TF relation available, all of which are dependent on which variables one uses to relate to the other. In their work, Tully and Fisher made use of optical luminosities to derive their relation [195], but subsequent work has shown the relation to be tighter and exhibiting an exponential slope more in line with our theoretical approximation of $\alpha = 4$ in Equation 2.27 when defined using microwave to infrared (K band) radiation (a good proxy for stellar mass), and even tighter when luminosity is replaced by the galaxy's total baryonic mass (the sum of its mass in stars and gas) [130]. This latter form is known as the **Baryonic Tully-Fisher Relation** (BTFR), and as noted by Torres-Flores et. al, the relation states that the baryonic mass of a spiral galaxy is typically proportional to its velocity to the power of 3.5-4, slightly lower than the standard TF relation [190]. Additionally, more recent work performed by Zaritsky et. al has suggested that the BTFR may actually be better modelled by a linear relationship with a gradient of 3.5 ± 0.2 as opposed to a power law [208], bringing into doubt the validity of the relation.

This can perhaps be more easily illustrated in Figure 2.2, depicting distance moduli of galaxies (determined from their luminosities) vs. their rotational velocities for samples of galaxies where only their stars are considered, only the mass of gas within those galaxies are considered, or the total baryonic mass is considered with both low and



FIGURE 2.2: The TF relations constructed from the H-band and I-band data of Bothun et. al [16] and Pildis, Schombert and Eder respectively [150]. The usual TF relation is apparent for massive galaxies in panel (a), however it breaks down at the low mass end. Panel (b) exhibits no 'H I TF relation' for massive galaxies, though there does appear to be a slight one for lower mass objects. The Baryonic TF relation which follows by summing stellar and gas mass nicely recovers a continuous relation over the entire observed mass range as seen in panel (c), however utilising lower mass-to-light ratios for the galaxy stars in panel (d) causes a noticeable discontinuity in slope, implying that the higher mass-to-light ratios adopted in panel (c) are more appropriate. Figure has been reproduced from the works of McGaugh [129].

high mass-to-light ratios galaxies, depicted as Υ_o . In each panel the presumed ideal TF exponent of α =4 is plotted. While the TF relation appears to be in good agreement with the observed data in panel (a), the relation breaks considerably for lower mass stars, and shows very poor agreement across the board in panel (b) where only the mass of gas within the surveyed galaxies are considered. This agreement improves considerably however in panels (c) and (d) when the total baryonic mass of the galaxies are considered, suggesting that the disk mass is the fundamental quantity of interest in the TF relation, although the slightly better agreement with the theoretical slope of the higher mass-to-light ratio galaxies in panel (c) would appear to suggest that their use with the BTFR is more appropriate than galaxies with lower mass-to-light ratios [129].

The takeaway from this is simple. While both the TF and BTFR remain fundamentally useful tools for distance determination, the keeping of both relations in the general form of $L \propto V^4$ remains up for debate to this day [129].

Curiosities with the value of the power law (or linear) exponent put aside for a moment, there remains another problem with the Tully-Fisher relation. While the relation is an excellent cosmological probe for galaxy distances, its key drawback lies in the fact that it does not apply well to elliptical galaxies due to their not exhibiting large systematic rotational velocities. Additionally their stars move rapidly on a variety of often very complex orbits determined by the elliptical galaxy's gravitational potential [78]. To that effect another relation is required tailored specifically to ellipticals such that their parameters can also be usefully constrained. This is where the Faber-Jackson relation comes into play.

2.7 The Fundamental Plane and Faber-Jackson Relation for Elliptical Galaxies

The Faber-Jackson relation, as developed by S. M. Faber and Robert E. Jackson [61], is an early empirical power law relation between the luminosity L of an elliptical galaxy and its central stellar velocity dispersion, σ . σ is defined as the spread of velocities of stars within the elliptical, specifically where these stars will have individual orbital velocities around the elliptical's centre of mass. By measuring the radial velocities of these stars the velocity dispersion can be estimated and used to derive the elliptical's total mass via the virial theorem. In the case of distant galaxies where individual stars may be difficult to resolve σ can be determined from Doppler (or redshift) broadening in the spectrum of the integrated starlight [7]. Much like the TFR, the Faber-Jackson relation (hereafter denoted FJR) also demonstrates a positive correlation between the two variables meaning that it can also be utilised as a tool for determining distances to external galaxies. Specifically, if we were to observe the spectrum along the line of sight through the centre of an elliptical galaxy, we will see a central velocity dispersion, or spread of orbital velocities, σ_0 . Applying the virial theorem we can show that:

$$\sigma_0^2 \approx \frac{GM_{virial}}{5R} \tag{2.28}$$

This result will mainly depend upon the ellipticity, or triaxiality, of the galaxy in question but will generally simplify to:

$$\sigma_0^2 \propto \frac{GM_{virial}}{R},\tag{2.29}$$

where R is the radius of the galaxy. The value of R that will be utilised here will depend upon the surface brightness profile that is modelled. If we were to model that profile using the de Vaucouleurs law, then galaxy luminosity would be derived via:

$$I(R) = I(R_e) \exp\left[-b\left(\frac{R}{R_e}\right)^{\frac{1}{n}} - 1\right],$$
(2.30)

with n = 4, $b \approx 2n - 0.327$ and R_e being the effective radius containing half the luminosity of the galaxy in question. As with the derivation of the TFR, if we are to assume that the luminosity of the elliptical galaxy will begin to converge after a few scale disk lengths, expressing it as a multiple of R_e , then we can model the relation of galaxy luminosity as:

$$L \propto I_e R_e^2, \tag{2.31}$$

where I_e is the surface brightness at $R = R_e$. If we proceed to square Equation 2.29 and substitute in Equation 2.31, as well as our disk mass-to-light ration parameter η from Equation 2.25 we obtain:

$$\sigma_0^4 \propto \frac{G^2 M^2}{R_e^2} \propto \frac{G^2 \eta^2 L^2 I_e}{L}$$
 (2.32)

Assuming once again that η and I_e are the same for all ellipticals we obtain the Faber-Jackson Relation:

$$\mathbf{L} \propto \sigma_{\mathbf{0}}^{\mathbf{4}} \tag{2.33}$$

wherein which the brighter the elliptical galaxy is, the more massive it is and the faster the stars in its central region are moving [78]. However, this relation manifests a considerable amount of scatter from the modelled exponent of 4 due to, once again, η and I_e not being equal for all ellipticals. There are two possible ways to tweak this relation to account for this:

- 1. Define the radius of an elliptical galaxy out to a fixed isophotal value, effectively defining a standard galaxy size which would reduce the effect of variation in surface brightness profiles between galaxies. This is called the $D_n \sigma$ relation, where D_n is the isophotal diameter chosen.
- 2. Account for the variance in I_e across ellipticals by including the effective radius R_e as a term in the FJR. [78]

The latter method serves as the foundation from which the Fundamental Plane relation for ellipticals is defined, with the FJR now being understood to be a projection of that plane:

$$L \propto \sigma_0^{2.65} R_e^{0.65}$$
 (2.34)

Much like the TFR and BTFR however the FJR, and by extension the Fundamental Plane relation, are not necessarily perfectly modelled by an exponential power law with a set exponent. More specifically Minkowski noted in 1962 that the correlation between elliptical luminosity and velocity dispersion is poor and that extending the observations to more objects, especially at low and medium absolute magnitudes is important [134]. It becomes increasingly apparent that the value of the exponent is indeed heavily reliant on the range of galaxy luminosities that is being fitted, as Davies et. al reported an exponent value of $\gamma=2$ for low-luminosity elliptical galaxies [46], while Schechter calculated a value of $\gamma=5$ for luminous elliptical galaxies [171].

2.8 Different Coordinate Systems

Various coordinate systems are used throughout this work to define the positions of the galaxies and supernovae being used, where there is often a need to convert between one system and the other. While the means of transforming from one system to another will not be described either here or in later chapters, we will introduce the key concepts behind the different systems we will use here, alongside a broad description of some of their key properties and characteristics.

2.8.1 Celestial and Equatorial Coordinate Systems

The concept of the celestial sphere has been fundamental to positional astronomy from pre-Babylonian times through to the Middle Ages and beyond to the Early-Modern era; and still remains useful today [135]. At the centre of the sphere is the Earth, and the surface of the sphere acts as the reference against which all celestial bodies



FIGURE 2.3: Diagram of the celestial sphere with respect to the ecliptic. The ecliptic is at an angle of 23.5° to the celestial equator (otherwise known as the Earth's visible horizon), indicative of the tilt of the Earth and the consequent path the Sun draws along the celestial sphere throughout the year as a result. Right ascension and declination angles (otherwise known as equatorial coordinates) are measured with respect to the celestial equator with the vernal (spring) equinox as the origin point, while celestial latitude and longitude angles are measured with respect to the ecliptic, again with the vernal equinox as the origin point [135].

observed from the centre are located. Since the heavens appear to rotate around the Earth once every twenty-four hours, the celestial sphere is considered to possess this particular motion. The axis of rotation terminates in the north and south celestial poles; equidistant between the poles, the sphere is encircled by the celestial equator. At a fixed angle of 23.5° to the equator is another great circle of the celestial sphere, the ecliptic. This represents the annual path of the sun. The two points where the ecliptic and equator cross therefore mark the position of the sun at the spring (vernal) and autumn equinoxes [135]. Both the celestial and ecliptic systems are depicted in Figure 2.3.

As a result, the position of any point on the surface of the sphere (and hence that of any celestial body which is referred to it) can be given with reference to the equator or the ecliptic. In the equatorial co-ordinate system, position is specified by right ascension and declination with reference to the celestial equator, while in the celestial coordinate system celestial longitude and latitude angles are given with respect to the ecliptic. In



FIGURE 2.4: Diagram of the galactic sphere with respect to the celestial sphere, specifically the celestial equator. The centre of the galaxy at RA=17h 45.6m and DEC=-28.94° acts as the origin point of the plane. With the galactic North Pole located at RA=12h 51.4m and DEC=+27.13° [12], this causes the galactic plane as whole to be inclined at an angle of approximately 63° to the celestial equator [198].

either system right ascension/celestial longitude is the angular distance along the equator or ecliptic from the vernal equinox (taken to be in the direction of the constellation Ares); and declination/celestial latitude is the distance north or south of the equator or ecliptic along a great circle passing through the point in question and the two celestial poles [135]. The angles of right ascension and declination can be expressed either in degrees or in hours, arcminutes (') and arcseconds ("). The relationship between these two measurements is as follows:

$$1hr = 15^{\circ} \qquad 1' = \frac{1}{60}^{\circ}$$
 (2.35)

$$1'' = \frac{1}{60}' = \frac{1}{3600}^{\circ} \tag{2.36}$$

2.8.2 The Galactic Plane

The galactic plane in essence is much like the celestial coordinate system in that it is measured in degrees of galactic longitude (the equivalent of right ascension) and latitude (the equivalent of declination), except that it is a sphere that is centred on the galactic centre of the Milky Way as opposed to being centred on Earth. The centre of the Milky Way is given at right ascension 17h 45.6m and declination of -28.94° in the constellation of Sagittarius, with the North Pole of the Galactic Plane at right ascension and declination angles of 12h 51.4m and $+27.13^{\circ}$ respectively [12]. Consequently when you combine these two measurements together the galactic plane as a whole appears to be inclined at an angle of approximately 63° to the celestial equator as depicted in Figure 2.4.

2.8.3 The Supergalactic Plane

The supergalactic plane is somewhat harder to define, as can be expected of structures with poorly defined disks. Although it was noted by William Herschel more than 200 years earlier, the major planar structure of the local Universe, i.e. the supergalactic plane (SGP) was recognised by Vaucouleurs [197] following an analysis of the radial velocities of nearby galaxies [112].

Although the SGP is relatively easy to see in Figure 2.5, it is not as easy to describe geometrically. Lahav et. al reported in their studies of the SGP using data from the Optical Redshift Survey (ORS) and the Infrared Astronomical Satellite (IRAS) that the structure of the SGP is not well described by a homogeneous ellipsoid, although it does appear to be a flattened structure [112]. The directions of the principal axes also vary with radius, consequently causing the structure of the SGP to change shape with radius as well, varying between a flattened pancake and a dumbbell, the latter at a radius of $\sim 50 h^{-1}$ Mpc. This consequently calls into question the 'connectivity' of the plane beyond a distance of $\sim 40h^{-1}$ Mpc. However if we choose to consider the plane as a whole only out to that limiting distance of $\sim 40h^{-1}$ Mpc, the centre of the SGP is given in galactic coordinates at $l=137.37^{\circ}$, $b=0^{\circ}$, with its North Pole given at $l=47.37^{\circ}$, $b=+6.32^{\circ}$ [39] which, when combined together, results in the SGP being aligned almost at a right angle to the galactic plane. When plotted with respect to the celestial equator as in Figure 2.6, the supergalactic plane runs through the Virgo cluster (the dense collection of objects on the left side of the sphere), and extends northward, passing close to the north celestial pole. It can be traced around to the southern galactic cap, although the density of galaxies in the anti-Virgo region is significantly reduced compared to the north [167]. The general planar structure of the SGP and the clustering of large galactic structures and superclusters along the equator of the SGP can be seen in Figure 2.7.



FIGURE 2.5: The distribution of galaxies projected on the sky in the IRAS and ORS samples, with the supergalactic plane being seen prominently as a thick clustering line along the centre of the great circle. This is an Aitoff projection in supergalactic coordinates, with SGL=90°; SGB=0° (close to the Virgo cluster) in the centre of the map. Objects within 2000kms⁻¹ are shown as circled crosses; objects between 2000 and 4000kms⁻¹ are indicated as crosses, and dots mark the positions of more distant objects. Only catalogued galaxies from both the IRAS and ORS samples are used, leading to very prominent zones of avoidance in both figures. Figure reproduced from the work of Lahav et. al [112].





FIGURE 2.6: All-sky Aitoff projection map showing the positions of ~5700 galaxies with measured redshifts less than 3000 kms⁻¹. The figure is plotted in equatorial coordinates, with the center of the figure at 6h and 0°. The dotted lines denote the location of the galactic plane and lines of $b=\pm 20^{\circ}$. The Virgo cluster is the densest collection of objects on the left side of the figure. The supergalactic plane runs through Virgo, and can be traced most of the way around the sky. Figure reproduced from the works of Salzer and Haynes [167].



FIGURE 2.7: Distribution of nearby groups of galaxies over the celestial sphere in Supergalactic coordinates. Group of galaxies shown here are within 10-16 Mpc. Note the marked concentration of galaxies and clusters toward the Supergalactic plane (horizon-tal line) [59].

2.9 An overview of supernovae catalogues

Having discussed the importance of supernovae, specifically Type Ia, as cosmological field probes due to their intrinsic luminosities and high accuracy distance estimates in Section 2.5, developing extensive catalogues of as many nearby and distant supernovae as possible for further use only makes sense. By analysing the velocities of these supernovae once their distances are determined, various cosmological models and parameters can be tested and constrained. As such we will take a moment to briefly discuss and introduce some of the supernovae catalogues and datasets that we will be making use of throughout this work, and discuss some pertinent cosmological results that have been computed from their use.

2.9.1 The Tonry et. al Supernova Data Set

The Tonry et. al Supernova data set is a compilation of 230 Type Ia Sne developed during 2003 at varying redshifts between 0.1 and 1.8 which has been utilised in confirming the results of Riess [163], Perlmutter [149] and others that supernova luminosity distances imply an accelerating Universe. Specifically, the discovery and addition of 8 Type Ia Sne between redshifts z = 0.3 - 1.2 to this set has served to extend the redshift range over which Type Ia Sne can be consistently observed to $z \approx 1$, where the signature of cosmological effects has the opposite sign of some plausible systematic effects. As a result, these measurements not only provide another quantitative confirmation of the importance of dark energy as a cosmological indication of an acelerating expanding Universe, but also constitute a powerful qualitative test for the cosmological origin of this acceleration [189].

More specifically Tonry et. al utilised these supernovae to obtain a value for the dark matter energy density as $\Omega_m = 0.28 \pm 0.05$, and consequently constrained the dark energy density Ω_{Λ} via the variable: $\Omega_{\Lambda} - 1.4\Omega_m = 0.35 \pm 0.14$. These values are in good agreement with the WMAP probes of the CMBR and their computed values of $\Omega_m \sim 0.3$ and $\Omega_{\Lambda} \sim 0.7$ for a flat Λ CDM Universe where $\Omega = 1$ [189]. Figure 2.8 depicts Tonry et. al's constraining of both dark matter and dark energy densities in a presumed flat Universe (denoted by the dashed straight line which restricts the summation of Ω_m and Ω_{Λ} from ever exceeding 1) using all supernovae in their set that exhibited redshifts larger than z > 0.01 and extinction values less than 0.5 magnitudes.

The Tonry et. al set has been constructed from several datasets over the past two decades (refer to the seminal work of Tonry [189] for more information on the assembly



FIGURE 2.8: Probability contours for the most statistically viable values of Ω_{Λ} vs. Ω_M as obtained using the Type Ia supernovae of the Tonry et. al dataset (black lines), presented at the 1, 2 and 3σ confidence levels respectively. The adopted a prior assumption of $\Omega_M h=0.20\pm0.03$ from the 2dF galaxy redshift survey and the works of Percival et. al [148] is also presented in grey out to the same number of confidence contours. The assumption of a flat Universe is depicted via the dashed straight line that restricts the summation of Ω_M and Ω_{Lambda} from exceeding 1. The intersection of both these sets of probability contours and our flat Universe assumption dashed line lends itself to values of $\Omega_M \sim 0.3$ and $\Omega_{\Lambda} \sim 0.7$, in very good agreement with the values obtained by WMAP. Figure reproduced from the works of Tonry et. al [189].

techniques used), and can be viewed and downloaded in its entirety at http://vizier.u-strasbg.fr/viz-bin/VizieR-3?-source=J/ApJ/594/1/table8

2.9.2 The Union2.1 Compilation

The Union2.1 Compilation, developed in 2012 by Suzuki et. al [186] is an update of both the Union1 and Union2 Compilations of supernovae as developed by Kowalski et. al in 2008 [108] and Amanullah et. al in 2010 respectively [2], which now brings together data for 833 supernovae drawn from 19 datasets, as is noted in Figure 2.9 which plots the distance moduli of all supernovae in the initial Union1 and final Union2.1 datasets as a function of redshift. The reason for the augmentation is due to the fact that Type Ia SNe are an excellent probe of dark energy, as they measure the magnitude-redshift relation with very good precision over a wide range of redshifts, from z = 0 up to $z \sim 1.5$ and possibly beyond [186].



(B) Plot of distance moduli of updated Union2.1 compilation supernovae as a function of increasing redshift [186].

FIGURE 2.9: Evolution of the Union compilations from Union1 (2008) to Union2.1 (2012) and the various datasets integrated into the compilations over time. Figures obtained from the Supernova Cosmology Project [185].



FIGURE 2.10: Probability contours for the most statistically viable values of Ω_{Λ} vs. Ω_M as obtained using the Union2.1 compilation (blue), presented at the 1, 2 and 3σ confidence levels respectively. Previous attempts to constrain parameters by means of the CMBR and BAOs are also presented to the same number of confidence contours in orange and green respectively. The assumption of a flat Universe is depicted via the dashed straight line that restricts the summation of Ω_M and Ω_{Lambda} from exceeding 1. The intersection of all three of probability contours with the flat Universe assumption dashed line continues to lend itself to values of $\Omega_M \sim 0.3$ and $\Omega_{\Lambda} \sim 0.7$, in very good agreement with the values obtained by WMAP. Figure obtained from the Supernova Cosmology Project [185].

More specifically, much like the Tonry et. al datasets before it (which have indeed been incorporated into the Union compilations), the Union2.1 supernovae have been used to successfully constrain Ω_{Λ} to a value of $0.724^{+0.071}_{-0.077}$ at the 68% confidence level (refer to Figure 2.10), which continues to be in excellent agreement with the results obtained from WMAP and from BAOs - baryon acoustic oscillations, which are regular, periodic fluctuations in the density of the visible baryonic matter of the Universe [186].

It should be noted in Figure 2.9 that at redshift values larger than ~ 0.2 in both the Union1 (panel (a)) and Union2.1 datasets (panel (b)), the linear predictions of an isotropic homogeneous Universe begin to break as clustering effects and the influence of large scale galactic structures begin to affect the distribution of supernovae at higher redshifts. This highlights the importance for our work going forward and our assumptions of homogeneity and continuous isotropy as discussed in Section 1.2 that we continue to work at low redshifts and restrict ourselves to analysing the nearby Universe alone. To that effect the Union2.1 compilation will be filtered accordingly. The full list of supernovae in the Union2.1 (and indeed Union1 and Union2) compilations can be found at http://supernova.lbl.gov/Union

2.9.3 The SAI Supernova Catalogue

The Sternberg Astronomical Institute catalogue is a Russian repository of multiple supernovae catalogues such as LEDA, NED and the SDSS amongst others that, up until late 2014, was being updated regularly as new supernovae were found [191]. As of October 2014 the catalogue consists of 6545 SNe at varying distances and also includes the relevant data of the parent galaxies hosting those supernovae. The catalogue can be viewed and downloaded in its entirety at http://www.sai.msu.su/sn/sncat/

2.10 An overview of redshift and peculiar velocity surveys

Having established the integral equation behind calculating the redshift z of an object in space from its observed photometry in Section 2.1 and consequently inferring their recessional velocities, generating redshift surveys from optical observations and by extension, generating peculiar velocity surveys from those redshifts are just as essential to our understanding of the Universe as our use of Type Ia supernovae catalogues are in probing cosmological parameters. They are especially important since at any scale larger than relatively nearby, it is extremely difficult to calculate the true real-space position of an object but measuring its redshift is relatively simpler and can be directly done. As a result large redshift surveys of galaxies such as the 'work-in-progress' SDSS (Sloan Digital Sky Survey) or the 2dFS (2 degree Field Survey) or many others are essential in mapping the 'luminous' or baryonic distribution of matter in the Universe and by proxy, probing the distribution of the hidden or dark matter that directly influence the recessional and peculiar velocities observed. Additionally they serve as useful tools to investigate the evolution of galaxies and large scale galactic structures over time as one probes larger and larger redshifts [6]. To be more specific, the vital importance of galaxy redshift and peculiar velocity surveys can be summarised in a few key reasons:

- Redshift surveys have unprecedented quantity and quality: taking the 2dF survey as an example, it contains over a quarter million galaxies with a predicted final survey size of ~800,000 objects once the final runs of the SDSS are completed. The unprecedented number of objects in such a survey as well as the homogeneous selection criteria enable the precise statistical analysis of their distribution.
- The Universe at $z \approx 1000$ is well specified: data obtained from both WMAP and Planck over the years have established a set of cosmological parameters that may be taken as the *initial condition* of the Universe from the point-of-view of the structure evolution towards z = 0. In a sense, the origin of the Universe at $z \approx 1000$ (approximate redshift of the surface of last scattering of the CMBR) and the evolution of the Universe after that epoch to the present are now equally important, but they constitute well separable questions that particle and observational cosmologists focus on, respectively.
- Gravitational growth of dark matter is well understood: In addition, extensive numerical simulations of structure formation in the Universe has significantly advanced our understanding of the gravitational evolution of the dark matter component in the standard gravitational instability picture (refer to Section 3.2 for an introduction to this concept and the key derivations behind it). In fact, we even have very accurate and useful analytic formulae to describe the evolution deep in its nonlinear regime. Thus we can now directly address the evolution of visible objects from the analysis of their redshift surveys separately from the nonlinear growth of the underlying dark matter gravitational potentials.
- Formation and evolution of galaxies: In the era of precision cosmology among others, the scientific goals of research using galaxy redshift surveys are gradually shifting from inferring a set of values of cosmological parameters using galaxies as their probes to understanding the origin and evolution of galaxy distribution given a set of parameters accurately determined by the other probes like CMB and supernovae. [113]

For more information about the importance of redshift and peculiar velocity surveys, as well as in depth discussions on various topics and cosmological methodologies related to their use, refer to the 1995 seminal paper of Strauss and Willick [184], which also served as the starting point for reading and research into this particular work. While a full review of that paper is outwith the scope of this work due to its length, the first sections can be summarised thus: the relevant background cosmological concepts such as the Hubble Law and the growth of large scale structure out of small perturbations is covered in brief, followed by the basics of Big Bang cosmology such as the definition of the Robertson-Walker metric, the Friedmann equation, discussing the gravitational instability paradigm and delving into power spectra and their use in determining the underlying dark matter distribution of the nearby Universe. A history of redshift surveys as of the date of the paper is then discussed as well as many of the practical issues needing to be addressed in their quantitative analysis. This is followed by a qualitative tour of the structures that we see within 8000kms⁻¹ as revealed to us from these surveys, as well as the various statistical measures that can be inferred from these surveys [184].

We will now briefly discuss the key redshift surveys that we will be using throughout this particular work to constrain various cosmological parameters in later sections.

2.10.1 IRAS PSCz Galaxy Catalogue

The Infrared Astronomical Satellite Point Source Catalog Redshift Survey (IRAS PSCz) is a redshift survey of IRAS galaxies out to 0.6 Jy (Jy - or Jansky - being a unit of spectral flux density). It contains 15,411 galaxies (14,677 with redshifts) that span over 84% of the sky as can be seen in Figure 2.11, and boasts a level of completeness and uniformity to within a few percent at high latitudes, and to within 10% at low latitudes. The maximal sky coverage of the PSCz allows for indepth mapping of the local galaxy density field and by proxy allows for high accuracy probing of the local gravity field as well. Additionally the high level of completeness and flux uniformity of the survey within well-defined areas and redshift ranges has allowed for in depth statistical studies of the IRAS galaxy population and its distribution to be performed [170]. The IRAS PSCz was initially published by Saunders et. al in 2000 and has since been augmented with data from the CfA (amongst other survey sources) to include 15795 galaxies, all normalised and corrected to account for certain cosmological parameters and assumed Universal models. The full catalogue can be accessed via the Strasbourg Astronomical Data Center at http://cdsweb.u-strasbg.fr/Cats.html

2.10.2 2MASS

The 2 Micron All Sky Survey (2MASS) as developed by Skrutskie et al. [180] is a near all-sky infrared survey that improves on the accuracy and coverage of its predecessor: the Two Micron Sky Survey (TMSS) which was constructed in 1969. Initialised in 1997 and completed in 2001, it achieves this by uniformly scanning the entire sky in three near-infrared bands (see Table 2.1) to detect and characterise point sources brighter



FIGURE 2.11: An Aitoff projection of the distribution of galaxies in the IRAS PSCz catalogue as a function of distance. The gaps in the map are due to the Milky Way obscuring that region of the sky and is termed the 'Zone of Avoidance'. The IRAS PSCz is considered complete and uniform in flux density to within a few percent at high galactic latitudes and to within 10% at low galactic latitudes otherwise [170].

than about 1 mJy in each band, with signal-to-noise ratio (SNR) greater than 10, using a pixel size of 2.0", thus achieving an 80,000-fold improvement in sensitivity relative to earlier surveys [24].

The overall sky coverage of this survey comes to $\gg 95\%$ for galactic latitude $|b| > 10^{\circ}$ and approx. 95% for $|b| < 10^{\circ}$, with no gaps larger than 200 square degrees, essentially generating a point source catalogue that contains accurate positions and fluxes for nearly 300 million stars and other unresolved objects, and an extended source catalogue containing positions and total magnitudes for more than 1,000,000 galaxies and other nebulae, as can be seen in Figure 2.12. The immediate benefits of the survey include:

- An unprecedented view of the Milky Way nearly free of the obscuring effects of interstellar dust, which will reveal the true distribution of luminous mass, and thus the largest structures, over the entire length of the Galaxy.
- The first all-sky photometric census of galaxies brighter than $K_s=13.5$ mag, including galaxies in the 60°-wide 'Zone of Avoidance', where dust within the Milky Way renders optical galaxy surveys incomplete. The catalogue of more than 1,000,000 galaxies will provide a rich statistical database, including photometric measurements in three wavelengths and a few structural parameters for large samples of galaxies in differing environments, measured at wavelengths which are sensitive to the stellar populations dominating the luminous mass.

		Magnitude Limits		
Band	Wavelength (μm)	Point Sources (SNR = 10)	Extended Sources	
J	1.25	15.8	15.0	
Η	1.65	15.1	14.3	
\mathbf{K}_{s}	2.17	14.3	13.5	

TABLE 2.1: The magnitude limits of the 2MASS survey for unconfused sources outside of the Galactic Plane ($|b| > 10^{\circ}$), and outside of any confusion-limited areas of the sky outside of the Galactic Plane [24].



FIGURE 2.12: Full-sky distribution of extended sources in the 2MASS survey. The faint blue band in the extended source map traces the Galactic plane as represented by the Point Source Catalog. Intensity is proportional to source density. The images are a color composite of source density in the J (blue), H (green), and K_s (red) bands [180].

• The statistical basis to search for rare but astrophysically important objects, which are either cool, and thus extremely red (e.g., extremely low-luminosity stars and brown dwarfs), or heavily obscured at optical wavelengths (e.g., dust-obscured Active Galactic Nuclei (or AGNs) and globular clusters located in the Galactic plane). [24]

The survey can be accessed in its entirety alongside in-depth documentation and all publications released to date at

http://www.ipac.caltech.edu/2mass/overview/about2mass.html [25].

2.10.3 Sloan Digital Sky Survey

The Sloan Digital Sky Survey, or SDSS, is on ongoing optical redshift survey and is one of the most ambitious astronomical surveys ever undertaken. The survey aims to map onequarter of the entire sky in detail, determining the positions and absolute brightnesses of hundreds of millions of celestial objects across all optical wavebands (namely u, g, r i and z). More specifically, the SDSS operates by using five specific colour filters as it observes the night sky, with each one designed to only let in and observe optical light of a certain wavelength or colour. The individual colour names and wavelengths of the different filters are given in Table 2.2 alongside the effective optical magnitude limits determined by the survey for those filters (at 95% completeness for point sources). The SDSS will also measure the distances to more than a million galaxies and quasars at high redshifts [174].

The SDSS is continually being updated with new data runs, with DR14 being the most current data set at the time of submission of this work, though we have focused on its predecessor, DR13, for this section. For the sake of completeness however, we will include a projection of the additions to the sky provided by the DR14 in Figure 2.14, as an indication of the continuing progression and development of the scope of the SDSS.

Benefits of the SDSS include but are not limited to:

- Creating a 3D picture of the Universe through a volume one hundred times larger than that explored to date,
- Providing unprecedented information of the distribution of matter at the edge of the visible Universe through its recorded distances to over 100,000 quasars (extremely luminous active galaxies) at high redshift,
- Being the first large-area survey of its kind to use electronic light detectors, so the images it produces will be substantially more sensitive and accurate than earlier surveys, which relied on photographic plates. As such the results of the SDSS (which is predicted to be in excess of 15 terabytes at the end of the survey) will also be made electronically available to both the scientific community and the general public, both as images and as precise catalogues of all objects discovered. [174]

By systematically and sensitively observing a large fraction of the sky, the SDSS will have a significant impact on astronomical studies as diverse as the large-scale structure of the Universe, the origin and evolution of galaxies, the relation between dark and luminous matter, the structure of our own Milky Way, and the properties and distribution of the

Filter	Wavelength (Å)	Magnitude limit	
Ultraviolet (u)	35430	22.0	
Green (g)	47700	22.2	
$\operatorname{Red}(\mathbf{r})$	62310	22.2	
Near Infrared (i)	76250	21.3	
Infrared (z)	91340	20.5	

TABLE 2.2: The optimal optical wavelengths at which SDSS's five filters work best. The sensitivity of each filter falls off slowly at shorter and longer wavelengths. [173]. The optical magnitude limits listed here are at 95% completeness for point sources [175].



FIGURE 2.13: Current sky coverage of the SDSS DR13 when combined with the DR12 and all previous data runs as of 2015 [175].



FIGURE 2.14: Current sky coverage of the SDSS as of 2018. Sky coverage of all previous data runs including DR13 are included here, with DR14 additions represented in blue [176].

dust from which stars like our own Sun were created. The SDSS will be a new reference point, a field guide to the Universe that will be used by scientists for decades to come [174].

Data Release 13 (DR13) is the first data release of the fourth phase of the Sloan Digital Sky Survey. It includes SDSS data taken through June 25, 2015, and encompasses more than one-third of the entire celestial sphere, as seen in Figure 2.13. The total unique area covered by this DR reaches 14,555 square degrees, and encompasses over 1.2 billion objects.

This DR can be accessed in its entirety at http://www.sdss.org/dr13/data_access/.

For more information pertaining to the contents and observations of this data run, specifically the location and format of the data now publicly available, as well as providing references to the important technical papers that describe the targeting, observing, and data reduction methods used, refer to the work of Albareti et. al and the references therein [1].

Chapter 3

Probing Peculiar Velocity Fields of the Nearby Universe

3.1 The Importance of Probing Peculiar Velocity Fields

As we established in Section 2.2, over large scale cosmological distances the relation between redshift space and real space is easily defined via the Hubble Law. On smaller scales however, gravitational instabilities give rise to galaxy peculiar velocities which need to be corrected for and constrained in order to perform any useful cosmological analyses. To that effect the mapping of the distribution of galaxies and their peculiar velocity fields constitutes a major research area in modern astronomy. More specifically, the measurements of peculiar motions provide a fundamental tool to probe the mass distribution in the local Universe [20], with redshift surveys providing one of the only truly useful sources for directly determining that distribution. Once identified, the underlying cosmological model parameters driving that distribution (such as those that will be discussed later in this chapter) can be constrained. This can be achieved for example via methods that quantify the amplitude of clustering in the matter distribution such as baryonic acoustic oscillations (BAOs) - measurements of the spatial distribution of galaxies to determine the rate of growth of cosmic structure within the overall expansion of the Universe [36]. Put another way, and as noted by Schmoldt et. al [172], redshift surveys (and their associated reconstructed peculiar velocity fields) provide the only possibility to determine the 3D density fields of luminous matter. These in turn are crucial for studies of mass concentrations, the mass power spectrum, dynamical analyses to probe the relationship between dark and luminous matter, and many other areas of observational cosmology [172]. Consequently the importance of probing the peculiar
velocity fields of the Universe around us via redshift and redshift-distance surveys cannot be overstated enough for our continued evolving understanding of the Universe and future discoveries to come.

In this chapter we will briefly discuss the gravitational instability paradigm that gives rise to the velocity distortions observed in these surveys, while also discussing a selection of the various cosmological parameters that can be constrained from such a paradigm and how we intend to make use of one parameter in particular, the **linear redshift distortion parameter** β , in this work going forward.

3.2 Defining Gravitational Instability

Gravitational instability is a universally accepted concept that provides a theoretical framework for the formation of structures in the Universe (for an introduction and more extensive treatment of the subject, consult the seminal work of Strauss & Willick [184] and the references therein) and is what is responsible for generating the distortions that give rise to peculiar velocity fields as we know them today. In the early Universe, minute irregularities in the distribution of matter (and consequently the mass density field) caused gravitational fluctuations, where areas with more matter would exert a more powerful gravitational force on neighbouring regions, drawing in more matter and consequently causing an even bigger irregularity in the mass density field. It is these irregularities in the mass density field, and the corresponding gravitational fluctuations over large scales that cause bodies to exhibit the peculiar or 'extra' component in their total bulk motion that is observed in galaxy redshift surveys and so forth. These irregularities in the density field can be modelled by means of the equation:

$$\delta(\mathbf{r},t) = \frac{\rho(\mathbf{r},t) - \bar{\rho}(t)}{\bar{\rho}(t)},\tag{3.1}$$

where δ is the fractional mass density difference or contrast at different points in the field at a given point in time, and $\bar{\rho}(t)$ is the scalar mean mass density. If we assume that these density perturbations are small enough that they can be treated according to the linear theory of gravitationally evolving cosmological density and perturbation fields (developed by Jim Peebles in 1980 in his seminal book describing the large scale structure of the Universe [147]) then we can proceed to make a few more assumptions about our Universe, namely:

1. The Universe is a perfect CDM pressureless fluid (meaning all perturbation or fluctuation terms arising from pressure are deemed negligible) consisting only of baryonic matter and dark matter with no dark energy component (i.e. $\Omega_{\Lambda} = 0$),

- 2. The velocity field flow is irrotational (meaning it exhibits no vorticity over large scales and thus the field can be derived from a scalar potential only [184]),
- 3. The gravitational perturbations exhibited will still converge at large scales to the perturbations predicted by the Zel'dovich approximation (see Section 3.4 for an introduction to its underlying theory), despite its root mean square field value being larger than 1 on small scales. [131]

From these we can begin to define gravitational instability by first writing down the equations of mass continuity, force and gravitation (Poisson's equation for fluids) respectively in an expanding Universe such that:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0, \qquad (3.2)$$

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} + \nabla\varphi = 0, \qquad (3.3)$$

$$\nabla^2 \varphi = 4\pi G \rho, \tag{3.4}$$

where ρ is the scalar mass density field, **v** is the velocity field and φ is the gravitational potential and as per our first assumption, all terms dependent on pressure have been deemed negligible and thus dropped [184]. If we then proceed to expand these three equations to first order, convert to comoving coordinates and subtract the zeroth order solutions (which in itself involves subtleties having to do with the gravitational potential of a uniform universe [184]), both the mass continuity and force equations simplify to:

$$\frac{\partial\delta}{\partial t} + \frac{1}{a}\nabla\cdot\mathbf{v} = 0, \qquad (3.5)$$

and

$$\frac{\partial \mathbf{v}}{\partial t} + \frac{\dot{a}}{a}\mathbf{v} + \frac{1}{a}\nabla\varphi = 0 \tag{3.6}$$

respectively, where δ is the dimensionless density contrast established in Equation 3.1 and *a* is the scale factor. If we then proceed to take the time derivative of the continuity equation (Equation 3.2) and substitute that into the divergence of the force equation (Equation 3.3) in conjunction with the Poisson equation (Equation 3.4) we obtain:

$$\frac{\partial^2 \delta}{\partial t^2} + \frac{2\dot{a}}{a} \frac{\partial \delta}{\partial t} = 4\pi G \rho_0 \delta. \tag{3.7}$$

where the second left hand 'drag' term denotes the expansion of the Universe [184]. As noted by Jim Peebles this approximation satisfies our requirement that our velocity field be irrotational, and infers that the angular momentum of observed bound structures can only be gained by non-linear tidal interactions between different overdensity perturbations in higher order perturbation theory [147]. Since Equation 3.7 is a second-order partial differential equation that is dependent on time t only, this allows us to separate the spatial and temporal dependencies such that:

$$\delta = A(\mathbf{x})D_1(t) + B(\mathbf{x})D_2(t), \qquad (3.8)$$

where D_1 and D_2 are growing and decaying modes respectively relative to the scale factor a [184][177]. In particular the decaying mode is representative of the rotational modes of the velocity field which as a result have short damping scales in t (once again keeping with our second assumption that the velocity field be irrotational on large scales). In the case of our perfect pressureless CDM Universe as per our first assumption where $\Omega_{\Lambda} = 0$, analytic expressions for $D_1(t)$ and $D_2(t)$ can be obtained where in the special case of a flat Universe we find that:

$$\frac{\partial^2 \delta}{\partial t^2} + \frac{4}{3t} \frac{\partial \delta}{\partial t} = \frac{2}{3t^2} \delta, \tag{3.9}$$

which has an analytic solution in terms of power laws of t such that:

$$\delta(\mathbf{x},t) = A(\mathbf{x})t^{\frac{2}{3}} + B(\mathbf{x})t^{-1}.$$
(3.10)

For more general cosmological models, the solution for δ will rely on the values chosen for Ω_m and Ω_{Λ} . In particular the growth is seen to be faster with increasing Ω_m while for $\Omega_m < 1$ the 'drag' term of Equation 3.7 (i.e. the Universal expansion) starts to dominate over the gravitational attraction of matter causing the predicted gravitational clustering to halt at an approximate redshift of $z \approx 1/\Omega_m - 1$ [184]. As we progress into late times in t, the growing mode of Equation 3.8 starts to dominate with the decaying vorticity mode vanishing such that we can rewrite Equation 3.5 to model the peculiar velocity field as:

$$\nabla \cdot \mathbf{v} = -a\delta \frac{\dot{D}_1}{D_1} = -aH_0 f\delta, \qquad (3.11)$$

Here we are effectively modelling the divergence of the velocity field (i.e. its rate of flow or change) as being proportional to the fractional mass density contrast, where a_0 is a linear constant taken to be the present-day value of the scale factor R(t) (as defined in Section 1.3), H_0 is the present-day value of the Hubble constant as these values are thought to change over time and epoch [81], and f is a growth factor that takes into consideration whatever cosmological model we are using to describe the rate of expansion of the Universe (in this case, we are using the Λ CDM model). More specifically f is given as:

$$f \equiv \frac{1}{H_0 D_1} \frac{\mathrm{d}D_1}{\mathrm{d}t} = \frac{\mathrm{d}\ln D_1}{\mathrm{d}\ln a},$$
 (3.12)

and was estimated by Lahav et. al [111] as a function of Ω_m and Ω_{Λ} (the critical energy density ratios of baryonic matter and dark energy respectively) to be:

$$f(\Omega_m, \Omega_\Lambda) = \Omega_m^{0.6} + \frac{\Omega_\Lambda}{70} (1 + \frac{1}{2}\Omega_m).$$
(3.13)

However, since we are working at low redshift in the nearby Universe, the dynamic influence of Ω_{Λ} becomes negligible [111], so the growth factor reduces to:

$$f = \Omega_m^{0.6}, \tag{3.14}$$

meaning we can now re-write Equation 3.11 to obtain:

$$\nabla \cdot \mathbf{v} = -a_0 H_o \Omega_m^{0.6} \delta(\mathbf{r}), \qquad (3.15)$$

and by using the results of the theory of electrostatics; specifically the 'Divergence Theorem' that states that the flux of a vector field through a closed surface S is equal to the integral of the divergence of that field over a volume V for which S is a boundary [65], we can now model a solution for the peculiar velocity field as follows:

$$\mathbf{v}(r) = \frac{H_0 \Omega_m^{0.6}}{4\pi} \int \frac{\delta(\mathbf{r}')(\mathbf{r}' - \mathbf{r})}{|\mathbf{r}' - \mathbf{r}|^3} d^3 r', \qquad (3.16)$$

where \mathbf{r} is the position of the galaxy, and $\mathbf{r'}$ is a position in space [184]. This cornerstone equation essentially defines an integral over a volume of space within which the peculiar velocity field $\mathbf{v}(r)$ at every point is strictly dependent upon the mass distribution $\delta(\mathbf{r})$ everywhere else within that volume. The presence of the cubed denominator indicates that the influence of more distant matter on the peculiar velocity field at a given point diminishes rapidly as you integrate over the whole volume. Equation 3.16 serves as the fundamental basis from which many different cosmological parameters can be constrained and estimated as the peculiar velocity field is probed, three key statistical measures of which we will discuss in the next section.

3.3 Exploring Cosmological Parameters

3.3.1 Bulk Flow

This measure, defined as the average streaming motion within a certain volume, is probably the easiest statistic to estimate from the observed radial component of peculiar velocities. In the Cosmic Microwave Background radiation (CMB) restframe, the bulk motions are expected to converge to zero with increasing volume. The rate of convergence depends on the fluctuations in the matter distribution on various scales, i.e., the power spectrum of the large scale matter fluctuations. Put another way, the rate of convergence of the bulk flow will be heavily reliant on the variable Ω_m , which in turn will be reliant on the sort of cosmological model and Universal curvature and geometry being assumed here as was discussed previously in Chapter 1. This dependence on cosmological models has motivated several attempts to measure the dipole component of the local peculiar velocity field and to determine the volume within which the streaming motion vanishes. Theoretically, the mean square bulk velocity within a sphere of radius R, is given by

$$V_b = \frac{\Omega_m^{1.2}}{2\pi^2} \int_0^\infty P(k) \tilde{W}^2(kR) dk, \qquad (3.17)$$

where $\Omega_m^{1,2}$ is a scalar value representing the amplitude of the mass power spectrum (refer to Section 3.3.2), P(k) is the mass fluctuation power spectrum and $\tilde{W}(kR)$ is the Fourier transform of a top-hat window of radius R [210].

As of yet, bulk flow measurements have produced conclusive and consistent results only on scales $\leq 60h^{-1}$ Mpc, but failed to do so on scales $\geq 100h^{-1}$ Mpc (consult the overview literature of Zaroubi [210] and the references therein for a more extensive treatment on the subject). A summary of recent bulk flow measurements within top-hat windows of varying radius R are given in Table 3.1.

3.3.2 The Mass Power Spectrum

The power spectrum of mass and density fluctuations is the most common statistic used to quantify the large-scale structure of the Universe [147]. This statistic is useful for several reasons. If the initial fluctuations were a Gaussian random field as commonly assumed (in other words, a random field where the underlying defining variables of that field - in this case the initial fluctuations - can be modelled as Gaussian probability distributions), then the initial power spectrum fully characterises the statistical properties of the field, and it reflects the origin of fluctuations in the early Universe that go on to fuel the large-scale cosmic growth and structure observed in the present day [104]. To that end Equation 3.17 suggests that one can estimate the bias free , $\Omega_m^{1.2}$ weighted, matter power spectrum directly from the measured peculiar velocities (refer to Section 3.3.3 for a brief definition of the Kaiser linear biasing model applied here).

Most of these power spectrum estimations are determined by applying likelihood analyses; which assume that both the underlying velocity field and the errors are drawn from independent random Gaussian fields. The observed peculiar velocities then constitute a multi-variate Gaussian data set, albeit with sparse and inhomogeneous sampling; and

Survey	$V_b \ (\mathrm{km s^{-1}})$	$R~(h^{-1}{ m Mpc})$	Comments
Tonry $et. al$ [188]	350	30	SBF^*
Dekel $et. al$ [51]	370	60	$M3^{\star} (TF^{\dagger} + D_n - \sigma)$
Giovanelli $et. al$ [72]	200	60	SFI (TF)
Courteau <i>et.</i> al [40]	70	60	Shellflow (TF)
da Costa $et. al$ [42]	220	60	ENEAR $(D_n - \sigma)$
Riess et. al [162]	≈ 0	100	SNIa•
Colless $et. al$ [33]	≈ 0	~ 110	EFAR (FP^{\diamond})
Hudson <i>et.</i> al [92]	600	140	SMAC (FP)
Lauer & Postman $et. al$ [116]	700	150	$LP (BCG^{\ddagger})$
Willick [203]	700	150	LP10K (TF)
Dale et. al [44]	≈ 0	~ 150	SCI/SCII (TF)
Turnbull et. al [196]	≈ 250	~ 60	SNIa
Feindt $et. al$ [62]	≤ 240	90-240	SNIa
Hoffman $et. al [80]$	250	50	Cosmicflows-2 (CR^{*})
Magoulas $et. al [126]$	≈ 400	N/A	$6 dFGSv (ML^{\triangleleft})$

TABLE 3.1: Summary of recent Bulk Flow measurements. * Surface Brightness Fluctuations. * Mark III dataset. [†] Tully-Fisher Measurement. • Supernovae Type 1a. [◊]
 Fundamental Plane Measurement. [‡] Brightest Cluster Galaxy Measurement. [#] Constrained Realisations. [⊲] Maximum Likelihood Approach [210] [196] [62] [80] [126].

this probability distribution function is then reinterpreted as a likelihood function of the measured radial velocities given a model power spectrum. Maximising the likelihood with respect to the model free parameters yields a best fit power spectrum [210].

Estimations of the matter power spectrum exist for the Mark III ([104], [214]), SFI ([67]) and ENEAR catalogues ([212]), as well as additional estimations constrained from analysing the temperature fluctuations or anisotropies of the CMBR using both WMAP across its many year runs (refer to [183], [105] and [106] and their associated papers for more in depth discussions on the additional parameters constrained and the methodologies used) and more recently with Planck in 2014 and 2016 (refer to [152] and [154] and their associated papers for more extensive results). All of these measurements have consistently produced power spectra with amplitudes larger than those measured by other data sets, with galaxy redshift surveys for example usually favouring the standard Λ CDM model ($\Omega_m = 0.3$, $\Omega_{\Lambda} = 0.7$ and h=0.65) [210].

3.3.3 The Linear Redshift Distortion Parameter β

Although we know that the Universe consists of both luminous baryonic matter and nonbaryonic dark matter, only the luminous matter can be observed directly. Therefore if we are to determine the distribution of dark matter, we must establish a relationship between it and the distribution of luminous matter that we observe. The simplest argument would be that the distribution of observed galaxies also contains the information about the dark matter distribution [184]:

$$\delta_g(\mathbf{r}) = \delta(\mathbf{r}),\tag{3.18}$$

but on the macro scale, this argument is no longer valid due to the effects of dark matter causing galaxies to gather in filaments and clusters and superclusters. To combat this effect, it was suggested by Kaiser et. al that galaxies will only form at the high-density peaks of the mass density field [102], such that galaxy clusters are said to be biased with respect to the mass distribution. This came to be known as the **linear bias model**, where the galaxy and dark matter distribution are related as follows:

$$\delta_{galaxies}(\mathbf{r}) = b\delta_{darkmatter}(\mathbf{r}) \tag{3.19}$$

where b is a linear biasing parameter. It must be noted that Kaiser observed that the value of b was heavily reliant on the assumed mass density coherence length (a characteristic length over which the mass density field is assumed to be coherent and continuous). As a result of this Kaiser determined that his assumption that this linear biasing model remains independent of scale only really holds true out to distances of $5h^{-1}$ Mpc $\leq \mathbf{r} \leq 7h^{-1}$ Mpc [102], with the linear biasing parameter becoming increasingly dependent on redshift over larger scales (for an overview of the various linear biasing models that have been developed to accommodate for this dependence, refer to the works of Basilakos & Plionis [8] and the references therein for a comprehensive look at the subject).

Let us assume for now that we are working at small enough distances for the linear biasing model to remain independent of scale. In that situation the following expression can be applied:

$$\beta = \frac{f(\Omega_m, \Omega_\Lambda)}{b},\tag{3.20}$$

and substituting from Equation 3.14 we get:

$$\beta = \frac{\Omega_m^{0.6}}{b},\tag{3.21}$$

where β is the **redshift distortion parameter**. Following on the theorems of electrostatics and the divergence theorem as before, we can rewrite our solution for the peculiar velocity field to indicate its linear dependence on β :

$$\mathbf{v}(r) = \frac{H_0\beta}{4\pi} \int \frac{\delta(\mathbf{r}')(\mathbf{r}' - \mathbf{r})}{|\mathbf{r}' - \mathbf{r}|^3} d^3r'$$
(3.22)

It is this parameter β that we have selected for analysis in this work; wishing to investigate other methods for estimating β in order to better understand the nature of the peculiar velocity field, and consequently learn more about the mass distribution of a particular region of space. Knowing that more accurately will also give us a clearer picture of the distribution of dark matter due to its linear dependence on the observed peculiar velocities of galaxies. While we will be primarily focusing on various applications of the ROBUST method of Rauzy and Hendry [160] to constrain β , we will briefly discuss some of the various methods and techniques that have been applied in the past to constrain β across various surveys, and provide a summary of β values calculated to date using these methods.

3.4 Constraining β

There are several methods for correcting for redshift distortions and recovering real space density and peculiar velocity fields of galaxies, and thus using these data to constrain cosmological parameters such as β . These can be roughly divided into two types: 1. Basis Function Methods and 2. Iterative Methods [172], both of which make use of either linear theory or the Zel'dovich approximation, the latter of which was developed by the Soviet physicist of the same name in 1970.

Zel'dovich adapted existing linear perturbation theory by choosing to define the actual position \mathbf{r} of an object as a function of its Lagrangian coordinate \mathbf{q} (a set of coordinates that are invariant with time) and time t, such that $\mathbf{r} = \mathbf{r}(t, \mathbf{q})$. Where only the growing perturbations are considered, his solution for the position \mathbf{r} becomes:

$$\mathbf{r} = a(t)\mathbf{q} + b(t)\mathbf{p}(\mathbf{q}),\tag{3.23}$$

where $a(t)\mathbf{q}$ describes the cosmological expansion and the second term describes the perturbations [216]. The functions a(t) and b(t) are known; b(t) is growing faster than a(t) due to gravitational instability, and the vector function $\mathbf{p}(\mathbf{q})$ depends upon the initial perturbation. The simple form of Equation 3.23 which is linear in t implies that all objects (at least initially) move with constant velocity, with an allowance for multiple objects to have their own velocity trajectories that will intersect at some point in the future such that regions of high (if not infinite) density can form [178]. This is particularly convenient for a simple reason. Considering what happens when these trajectories intersect, the resultant collisions would cause effects such as multistream configurations (non-constant velocity flows) to occur that can continue to be modelled using $\mathbf{r} = \mathbf{r}(t, \mathbf{q})$, and would not be unlike the peculiar velocity fields we observe today (which themselves have their origin in minute irregularities in the mass density field in

the early Universe as per the paradigm of gravitational instability (see Section 3.2)). Put another way, the Zel'dovich approximation gives us another way (more specifically a *kinematical approach* [146]) to model the full motion of an object over time, while also allowing for relatively well predicted modelling of evolving large scale cosmic structure - far better than what can be achieved with the gravitational instability paradigm or Eulerian kinematics alone (the process of describing the position of an object as a set of coordinates in which the properties of the object - in this case its velocity - are assigned to points in space at each given time). With respect to the latter, the Zel'dovich approximation has the advantage of breaking down much later than Eulerian linear theory, allowing for predictive modelling over longer periods of cosmological time [146].

Consequently with given $\mathbf{r}(t, \mathbf{q})$, it is possible to calculate the distribution of velocity and density anywhere in space, with $\mathbf{r}(t, \mathbf{q})$ containing the whole picture of the motion as well as well-modelled predictions for large scale clustering effects that will form over large scales of cosmological time [216] [178]. Consequently the Zel'dovich approximation provides an intuitive way to understand the emergence of large scale cosmic structure, and accurately predicts the rich structure of voids, clusters, sheets and filaments observed in the Universe [201].

Regardless of the theory or approximations used however, both the Basis Function and Iterative methods used to constrain cosmological parameters are ultimately very limited in their ability to reconstruct the high-density regions of the real space density and peculiar velocity fields [172].

3.4.1 Basis Function Methods

When transforming the measured redshift space density field into a combination of angular and radial basis functions, the distortion is concentrated in the radial part and its correction becomes an algebraic matrix inversion problem. Some versions of this method transform the angular part of the density field into basis functions while expressing the radial in differential equations which are then solved numerically (see the work of Nusser and Davis [140]), while others transform both the angular and radial parts into basis functions, using a combination of spherical harmonics and spherical Bessel functions (refer to the works of Fisher et. al [64] and Zaroubi et. al [213]). The underlying theory of the latter approach lies in expressing the mass overdensity as a Fourier-Bessel expansion:

$$\delta(r,\omega) = \sum_{lmn} Y_{lm}(\omega) j_l(k_{ln}r) \delta_{lmn}, \qquad (3.24)$$

where ω denotes the angular coordinates, Y_{lm} are spherical harmonics, j_l are spherical Bessel functions, k_{ln} are a set of wavenumbers that depend on the boundary conditions assumed, and δ_{lmn} are the expansion coefficients [172]. Once the expansion coefficients are known, the linear theory velocity field is easily calculated in terms of these as:

$$\mathbf{v}(r,\omega) = H_0 \beta \sum_{lmn} \delta_{lmn} \nabla \left[Y_{lm}(\omega) \frac{j_l(k_{ln}r)}{k_{ln}^2} \right]$$
(3.25)

The problem is to determine the expansion coefficients δ_{lmn} from redshift survey data, especially correcting for distortions of redshift space (s, ω) relative to real space (r, ω) arising from the velocity field [172]. In addition, such a Fourier-Bessel expansion as above will contain spurious extra power from shot noise that needs to be eliminated. This shot noise arises due to the continuous discrete summations over multiple Bessel functions and spherical harmonics, which causes an artificial contribution or 'white noise' to start appearing in the data; which generally tends to the reciprocal of the square root of the number of Bessel functions and spherical harmonics summed [94]. This spurious power can be suppressed by means of a Wiener Filter Φ , as defined by the American mathematician of the same name during the 1940s and published in 1949 [202]:

$$\Phi = \frac{\text{power in signal}}{\text{power in (signal + noise)}}$$
(3.26)

but this in itself is not without its problems. One of the main drawbacks of the Wiener filter is that it suppresses the amplitude of the estimated signal. The suppression factor is roughly equal to:

$$\frac{\text{Signal}^2}{\text{Signal}^2 + \text{Noise}^2},\tag{3.27}$$

therefore in the limit of very poor signal-to-noise ratio data, which in the context of the sorts of cosmological parameters we are studying corresponds to galaxy peculiar velocities, and in view of the typically quite large uncertainties on galaxy distance estimators, the estimated field approaches a value of zero [211].

3.4.2 Iterative Method - POTENT

As stated in Section 3.2, in the linear regime of gravitational instability, a simple relation between peculiar velocity, v, and mass density contrast, δ_m , can be easily obtained from mass conservation, with the differential form of that relation being expressed as:

$$\nabla \cdot \mathbf{v} = -a_0 H_o \Omega_m^{0.6} \delta(\mathbf{r}), \qquad (3.28)$$

Equation 3.28 can be used to perform what is called a density-density comparison, which is typically performed in the following way:

- 1. Reconstruct the 3D velocity field from observed galaxy radial velocities,
- 2. Differentiate $\mathbf{v}(\mathbf{r})$ and use Equation 3.28 to compute δ_g . More specifically δ_g is the galaxy density contrast, which can be related to the mass density contrast $\delta_m(\mathbf{r})$ by means of our previously discussed linear biasing model over small scales such that $\delta_g = b_g \delta_m$, where b_g is the galaxy biasing parameter [20]. To that effect we can rewrite Equation 3.28 such that:

$$\nabla \cdot \mathbf{v} = -\left[\frac{a_0 H_o \Omega_m^{0.6}}{b_g}\right] \delta_g(\mathbf{r}) \tag{3.29}$$

3. Compare the computed δ_q to observed galaxy density fields [20].

Step 1 is the least trivial to implement and requires some additional theoretical assumptions. The POTENT method as developed by Bertschinger & Dekel [10] is one such density-density comparison technique that makes the assumption that the velocity field v(r) is irrotational, i.e.

$$\nabla \times v = 0 \tag{3.30}$$

Any vorticity mode is expected to decay in the linear regime as the Universe expands, therefore based on Kelvin's circulation theorem the velocity flow should remain vorticity-free in the quasi-linear regime (i.e. at larger cosmological scales) provided that the flow remains laminar [50]. Irrotationality implies that the velocity field can be derived from a scalar potential:

$$\mathbf{v}(x) = -\nabla\Phi(x),\tag{3.31}$$

so the radial velocity field u(x) should contain enough information for a full reconstruction [10][50]. To that effect the radial component of the velocity field, defined by $u(r, \theta, \phi)$ (the only component we can directly compute using redshift surveys, by subtracting the peculiar motions of us as the observer from the observed recessional velocity cz while also correcting for local bulk flow motions), can be used to calculate the velocity potential Φ via:

$$\Phi(r) = -\int_0^r u(r',\theta,\phi)dr', \qquad (3.32)$$

where differentiating Equation 3.32 via Equation 3.31 will then yield the full 3D velocity field [50].

However, since a continuous radial velocity field u(r) cannot be easily observed due to all the additional motions that need to be corrected for, noisy data arises for a nonuniform and sparsely sampled set of galaxies. Therefore an integral component of the POTENT method involves turning available data into a continuous radial velocity field via smoothing methods [184]. Dekel, Bertschinger & Faber use a tensor window function smoothing that takes into account the radial nature of the observed velocity field [10]. More specifically, they take the observed velocity field and cut it into spherical shells, then calculate the multipole coefficients of the radial velocity distribution across each shell, or 'window'. This has the advantage of retaining all details of the velocity field flow, and to calculate the statistical distribution of those multipoles is a relatively simple affair [161]. However this produces additional sources of error in the resultant smoothed velocity field:

- Statistical noise in the velocity field due to the error in the individual peculiar velocities. One minimises this noise by weighting each galaxy by the inverse square errors of those velocities.
- Malmquist bias (as discussed in Section 2.4), both homogeneous and inhomogeneous, resulting from the peculiar velocity errors.
- Sampling gradient bias, due to the inhomogeneous sampling of the velocity field within a smoothing window. This is minimised using equal volume weighting, wherein which a weighted mean is used to properly account for the relative contributions of each galaxy to the velocity field. This is achieved by weighting the average absolute magnitude or luminosity function of the galaxies (from which their real-space distances and consequently their peculiar velocities will be determined) by $1/V_{max}$ where V_{max} is the maximum volume over which the galaxies could have been seen. Brighter galaxies will have a larger volume over which they can be observed or detected, and thus they will be given a smaller weight since these brighter objects will be more fully sampled [14][184].

As a result of these concerns and smoothing effects, the many applications of POTENT to various data sets have consistently led to large values of β being constrained that are consistent with unity [20]. This is particularly problematic as it basically implies that all observed peculiar motions are driven exclusively by dark matter such that $\Omega_m \sim 1$, which is inconsistent with the values of $\Omega_m \sim 0.3$ constrained by WMAP and Planck via their analyses of the CMBR for our assumed Universal model of Λ CDM where $\Omega = 1$.

3.4.3 Iterative Method - VELMOD

While Equation 3.28 serves as the differential form of the relation between peculiar velocity and mass density contrast, the principles of the theory of electrostatics can be

applied as before to derive the integral form of that relation as:

$$\mathbf{v}(r) = \frac{\Omega_m^{0.6}}{4\pi} \int \mathrm{d}^3 \mathbf{r}' \frac{\delta_m(\mathbf{r}')(\mathbf{r}' - \mathbf{r})}{|\mathbf{r}' - \mathbf{r}|^3},\tag{3.33}$$

which serves as the basis of another form of iterative analysis, namely velocity-velocity comparisons [20]. In this approach, the observed galaxy distribution is used to infer the mass density field from which peculiar velocities are obtained and compared to the observed ones [210]. One of the most well known forms of this sort of comparison is the VELMOD method as derived by Willick et. al [205]. While we won't delve into the specific derivations of VELMOD and the extensive discussions of its applications (refer to the works of Willick et. al [205] and the references therein for that), we will briefly cover the main points of how VELMOD operates here. VELMOD uses a so-called "Method II approach" as defined in the works of Strauss & Willick in 1995, wherein which we take the Tully-Fisher observables (apparent magnitude m and velocity width η) and the redshift of an object and quantify the probability of observing the former given the latter for a particular model of the velocity field and a TF relation [184]. This probability is then maximised with respect to the free parameters of the velocity model and the TF relation [20]. Specifically, for a galaxy with given angular coordinates (l, b), redshift cz, apparent magnitude m and velocity width parameter $\eta \equiv \log_{10}(W) - 2.5$, where W is twice the rotational velocity of the galaxy, the joint probability of said galaxy having a certain apparent magnitude at a given redshift and velocity width parameter is defined by:

$$P(m|\eta, cz) \simeq \frac{1}{\sqrt{2\pi\sigma_e}} \exp\left[\left[-\frac{1}{2\sigma_e^2} \left\{m - \left[M(\eta) + 5\log w + 3 \times \frac{5}{\ln 10}\Delta_v^2\right]\right\}^2\right]\right], \quad (3.34)$$

where w is the solution to the equation cz = w + u(w), i.e. it is the distance inferred from the redshift and peculiar velocity model; $\Delta_v \equiv \sigma_v / [w(1+u')]$, where $u' = (\partial u / \partial r)_{r=w}$, is the effective logarithmic velocity dispersion; and

$$\sigma_e \equiv \left[\sigma_{TF}^2 + \left(\frac{5}{\ln 10}\right)^2 \Delta_v^2\right]^{\frac{1}{2}}$$
(3.35)

is the effective TF scatter, including the contribution due to σ_v [205]. This probabilistic approach allows a statistical treatment for effects such as small scale velocity noise, inaccuracy of the velocity model and the existence of triple-values regions that spoil the uniqueness of the redshift-distance mapping [20]. Also unlike POTENT, VELMOD doesn't require smoothing of the TF data which, along with the allowance for triplevalued regions and small-scale velocity noise, allows one to probe the velocity field of high density regions, thus allowing one to exploit the denser sampling of surveys such as the PSCz galaxy catalogue. Another convenient feature of VELMOD is that it does not require an a priori calibration of the TF relation, which is a common issue of concern in peculiar velocity studies. Instead, a fit of the parameters of the TF relation is performed simultaneously with a fit of the parameters of the velocity field [20].

Velocity-velocity methods such as VELMOD have been applied to many current galaxy catalogues and consistently yielded systematically lower values of β in the range of 0.4-0.6 irrespective of the velocity tracers, modelled gravity field and comparison techniques used [210][20]. However, some of these analyses have shown evidence for a poor match between models and data, which would render the estimate of β meaningless [20].

Examples of this can be found in the works of Davis, Nusser & Willick where they compared the gravity field derived from the IRAS 1.2Jy survey using ITF (Inverse Tully Fisher relation) with the peculiar velocities obtained from the Mark III catalogue, and noted that the coherent dipole residuals they found were indicative of significant discrepancy between the modelled and observed velocity fields [47]. To that end, Strauss & Willick also considered the IRAS 1.2Jy velocity predictions and the Mark III dataset and compared them using VELMOD, only being able to obtain a good fit to the data when they introduced a physically motivated, external quadrupole contribution to the modelled velocity field [205] [204] [20].

3.4.4 Best of Both Worlds - UMV

The unbiased minimal variance (UMV) method, as developed by Zaroubi et. al [209], acts as a sort of intermediary solution that is the best of, or compromise between, both worlds. With the UMV method, both velocity-velocity and density-density comparisons can be carried out within the same methodological framework. Similar to a Wiener Filter, the UMV estimator is derived by requiring the linear minimal variance solution for a given cosmological field, and an assumed prior model specifying the covariance matrix of that underlying field. More specifically for an underlying cosmological field s with a set of observations d (where d can either be directly sampled from s or from any field linearly related to s) we are interested in measurements that can be modelled mathematically as a linear convolution or mapping of the underlying field such that:

$$\boldsymbol{d} = \boldsymbol{o} + \boldsymbol{\epsilon} = \mathbf{R}\boldsymbol{s} + \boldsymbol{\epsilon},\tag{3.36}$$

where $\boldsymbol{o} = \mathbf{R}\boldsymbol{s}$, and \mathbf{R} is an $M \times N$ matrix that represents the response or point spread function, and $\boldsymbol{\epsilon}$ is the noise vector associated with the data [209]. Leading on from this the unbiased minimal variance estimator \boldsymbol{s}^{UMV} is defined such that $\boldsymbol{s}^{UMV} = \mathbf{H}\boldsymbol{d}$, where **H** is the $N \times M$ matrix that minimises the variance of the residual r, where $r = s - s^{UMV}$, satisfying the constraint:

$$\left[\boldsymbol{s}^{UMV}\right]_{\mathcal{N}} = \left[\mathbf{H}\boldsymbol{d}\right]_{\mathcal{N}} = \boldsymbol{s},\tag{3.37}$$

where $[...]_{\mathcal{N}}$ is the ensemble average over multiple noise realisations [209]. In effect we are seeking **H** that minimises:

$$\left\langle \left[\boldsymbol{r}\boldsymbol{r}^{+} + \boldsymbol{\lambda}\mathbf{H}\boldsymbol{d} \right]_{\mathcal{N}} \right\rangle = \left\langle \left[\left(\boldsymbol{s} - \boldsymbol{s}^{UMV} \right) \left(\boldsymbol{s} - \boldsymbol{s}^{UMV} \right)^{+} + \boldsymbol{\lambda}\mathbf{H}\boldsymbol{d} \right]_{\mathcal{N}} \right\rangle,$$
 (3.38)

where the ⁺ exponent denotes the complex conjugate of the transpose of the underlying signal and λ is a Lagrange multiplier [209]. Minimising Equation 3.38 with respect to **H** and using the result in conjunction with Equation 3.37 to solve for **H** and λ one obtains the solution for the UMV estimator:

$$\boldsymbol{s}^{UMV} = \left\langle \boldsymbol{s}\boldsymbol{o}^{+} \right\rangle \left\langle \boldsymbol{o}\boldsymbol{o}^{+} \right\rangle^{-1} \boldsymbol{d}$$
(3.39)

This estimator allows one to reconstruct an unbiased cosmological field at any point in space from sparse, noisy and incomplete data and to map it into a dynamically related cosmic field (to go from peculiar velocities to overdensities, for example [211]). However, unlike the Wiener filter, the minimisation is carried out with the added constraint of an unbiased reconstructed mean field [210]. Specifically for a group of catalogue data points consisting of a set of observed radial peculiar velocities u_i° , measured at positions $\mathbf{r_i}$ with estimated errors ϵ_i , assumed to be uncorrelated, then the observed velocities are thus related to the 3D underlying velocity field $v(\mathbf{r})$, or to its radial component u_i via the equation:

$$u_i^{\circ} = v(\mathbf{r}_i) \cdot \mathbf{r}_i + \epsilon_i \equiv u_i + \epsilon_i, \qquad (3.40)$$

As per the definitions of gravitational instability we established earlier, the peculiar velocity field v(r) and the density fluctuation field $\delta(\mathbf{r})$ are related via $\delta = f(\Omega_m)^{-1} \nabla \cdot v$, where $f(\Omega_m) \approx \Omega_m^{0.6}$ and Ω_m is the matter mean-density parameter [209][211]. Under the assumption of a specific theoretical prior for the power spectrum P(k) of the underlying density field, we can write the UMV estimator of the 3D velocity field as:

$$\mathbf{v}^{UMV}(\mathbf{r}) = \langle v(\mathbf{r})u_i \rangle \langle u_i u_j \rangle^{-1} u_j^{\circ}$$
(3.41)

and the UMV estimator of the density field as:

$$\delta^{UMV}(\mathbf{r}) = \left\langle \delta(\mathbf{r}) u_i \right\rangle \left\langle u_i u_j \right\rangle^{-1} u_j^{\circ} \tag{3.42}$$

As was touched upon briefly in Section 3.4.1, one of the drawbacks of a Wiener Filter is that it suppresses the amplitude of the estimated signal (see Equations 3.26 and 3.27). By contrast, the UMV estimator as defined above has been specifically designed to not alter the values of the reconstructed field at the locations of the data points, thus avoiding spurious suppression effects. An unbiased estimate of the reconstructed field at any point in space is then obtained by interpolating between the data points, according to the correlation function assumed a priori [211].

The fields reconstructed from this estimator are compared with those predicted from the IRAS PSCz galaxy redshift survey to constrain the value of β . For example, the analysis of the SEcat catalogue for the first time leads to consistent β values from the comparison of the density and the velocity fields yielding $\beta = 0.57^{+0.11}_{-0.13}$ and $\beta = 0.51 \pm 0.06$, respectively [210].

3.4.5 Applying χ^2 -minimisation

To constrain β , recall Equation 3.22 which describes the linear dependence of β on the mean density function Ω_m , and consequently its dependence on the peculiar velocity field. An extension of this theory suggests that the difference between the observed peculiar velocity values and those calculated via interpolation from the surrounding galaxy distribution and their respective velocities is about a factor of β . Using a χ^2 minimisation to determine this value of β is consequently quite useful.

The basic concept of this technique is simple, and is not unlike a simplified variant of a velocity-velocity comparison method. Consider a galaxy catalogue of peculiar velocities and 3D positional coordinates. To constrain β , a group of secondary distance indicators such as Type 1a supernovae (whose distances are typically known to within 8% as was discussed in Section 2.5) are embedded amongst these galaxies, and their peculiar velocities are interpolated based on the velocity field generated from the surrounding galaxies. The computed supernovae velocities are then rescaled by a factor of β due to our aforementioned linear dependence until we reach a value for which the computed velocities closest match that which are observed in the actual catalogue. More specifically,

$$u_{pec}^{trial}(i|\beta) = \beta u_{pec}^{pred}(SN_i|\beta = 1), \qquad (3.43)$$

where our re-scaled β will take a value between 0 and 1 incrementally increasing according to the number of trials we perform. We take this range [0,1] in particular because as previously mentioned in Section 3.4.2, a value of $\beta \sim 1$ represents a scenario in which there is no contribution to the peculiar velocity field from luminous matter; a problematic scenario at best given what we observe in the Universe. Similarly a value of $\beta \sim 0$ indicates that galaxies do not exhibit any peculiar velocities at all, which we again know is not true considering the large scale cosmic bulk flows observed in the Universe. As a result a β range of [0,1] represents the two extreme limits which we know cannot be true, but in between which the 'true' value of β must reside. Therefore the χ^2 calculation would be as follows:

$$\chi^{2} = \sum_{i=1}^{nsn} \frac{\left[v_{pec}^{obs}(i) - u_{pec}^{trial}(i|\beta)\right]^{2}}{\sigma_{i}^{2}},$$
(3.44)

where σ refers to the radial redshift reconstruction error and Sne distance errors involved in reconstructing the peculiar velocity fields and nsn signifies the number of Type Ia Sne being used for the analysis. Consequently χ^2 will be minimised at the value of β for which the differences between the observed and predicted $\beta\text{-dependent}$ peculiar velocities are minimised, with a confidence interval as defined by $\sigma_{\chi^2} = \beta_{opt} \pm (\chi^2_{min} + 1)$. The degrees of freedom ν inherent to this χ^2 statistic, i.e. the number of values in the final calculation of χ^2 that are free to vary, will determine what the minimum value of χ^2 should be. In this example that would be equal to the number of Sne used during the analysis minus the number of parameters held fixed which, in this instance, would be three: the two sources of error used in the construction of σ and the observed peculiar velocities v_{pec}^{obs} ; in other words $\chi^2_{min} = \nu = nsn - 3$. As a general rule of thumb, should the value that χ^2 minimises to be on the order of the number of degrees of freedom ν , then we can accept the value of β constrained here as the correct value, whereas should the minimum value of χ^2 be on the order of double the number of degrees of freedom or larger than the value of β constrained is more likely to be rejected unless a more thorough investigation of the sources of error involved in the analysis and their associated estimations is performed [199].

Table 3.3 in Section 3.4.7 presents previous attempts to constrain β using χ^2 -minimisation.

3.4.6 Only the Beginning - A Brief Overview of More Recent Methods

The methods that we have discussed in detail over the past few sections to constrain β are by no means exhaustive, though frequently used up to the present day (look to the 2015 works of Carrick et. al [25] for example and their use of VELMOD to constrain β for the 2M++ density field). As we continue to improve the techniques currently at our disposal we have also progressed in leaps and bounds in developing new experimental techniques to constrain β , amongst other cosmological parameters, and while delving into all of the most recent developments would make for a fascinating study in and of itself, that is not the purpose of this work. As such we will merely provide a brief descriptive overview of some of the newer methods available here, with links to the appropriate references for further reading. Table 3.4 in Section 3.4.7 will summarise some of the recent results for β obtained from these newer methodologies but, as the title of this section suggests, this is only the beginning. As we continue to find new ways to probe the peculiar velocity field of our nearby Universe and get a better handle on the cosmological parameters that fuel it, we will never be at a loss for new things to understand and explore.

Recent developments in β -constraining techniques include, but are not limited to:

- 1. Analysing the clustering signal of LRGs (Luminous Red Galaxies) in redshift surveys in order to recover the values of Ω_m and β . More specifically we make use of the Alcock-Paczyński cosmological test, which is an evaluation of the ratio of observed angular size of a galaxy to its radial/redshift size. The main advantage of this test is that it does not depend on the evolution of the galaxies, but only on the geometry of the Universe [123]. By taking into consideration the influence of linear clustering evolution constraints, β can be constrained effectively [132] [29].
- 2. Reconstructing the density and peculiar velocity fields of a redshift survey such as 2MASS by means of expanding them using Fourier-Bessel functions, while also making use of a distortion matrix to deconvolve the parameter β from the density field in particular [58].
- 3. Analysing the 3D power spectrum of redshifts from SDSS galaxies. This is achieved by applying the Karhunen-Loève Transform (KLT), which represents the spectrum as an infinite linear combination of orthogonal functions, analogous to a Fourier series representation of a function on a bounded interval, which are then solved in order to recover the signal of β amongst other parameters from the spectrum [155]. The key difference between using KLT and more traditional methods like Fast Fourier Transform (FFT) lies in the fact that KLT is better able to reconstruct a weak signal from extremely noisy data, whereas FFT would fail (refer to the work of Maccone [125] for an in depth look into the math and theory behind KLT, how it is derived, and its importance in astronomy and cosmology in general).
- 4. Making use of the clustering dipole of redshift surveys to determine β. More specifically, one wishes to compare the peculiar velocities of galaxies with their gravitational accelerations (as induced by the density field) in order to constrain β. To that effect, one can use the motion of the Local Group (LG) for that purpose. Its peculiar velocity is known from the dipole component of the CMBR, whereas its acceleration can be estimated with the use of the so-called clustering dipole of surveys such as 2MASS [11].
- 5. Measuring the growth rate of large scale structures around cosmic voids to constrain β . More specifically, by measuring the cross-correlation function between

the centres of observed voids in redshift surveys and the complete galaxy catalogue itself, it is expected to exhibit a clear anisotropy that is characteristic of the linear redshift distortion parameter β . By measuring the projected cross-correlation and then de-projecting it we are able to estimate the un-distorted cross-correlation function. With a sufficiently well-measured cross-correlation function one should be able to measure the linear growth rate of structure β by applying a simple linear Gaussian streaming model to the redshift space distortions [77] [75].

3.4.7 Attempts to constrain β to date

Table 3.2 and Table 3.4 present a selection of current attempts to constrain β using the methods we have discussed both in detail and in summary in this chapter, with Table 3.3 specifically showing previous attempts to constrain β using χ^2 minimisation techniques. The Radburn-Smith value of $\beta = 0.55 \pm 0.06$ [158] is highlighted in bold as it will be used as our benchmark value for future computations performed in this work.

It should be noted that the majority of values reported in Table 3.3 have been calculated using objects at low redshift, where the effects of the breakdown in isotropy and homogeneity at lower redshift may affect the value of β being constrained, if at all.

Having now discussed in broad strokes the importance and usefulness of probing peculiar velocity fields of the nearby Universe in astronomical and cosmological terms, in addition to exploring some of the various methods and analyses available to us to attempt to constrain various cosmological parameters such as the linear redshift distortion parameter β , we will now discuss an alternative method of our own. The ROBUST method as developed by Rauzy and Hendry [160] provides us with another independent means of constraining parameters such as β without the need for secondary distance indicators such as SNIa or the Tully-Fisher relation, and only utilises the galaxy catalogue in question.

It is the purpose of this work to explore the workings of ROBUST and apply its methodology to various mock and real-world data catalogues, comparing our calculated values with those obtained above (specifically using χ^2 -minimisation to benchmark compare our results with previously established values for various surveys) as a means of testing the suitability of ROBUST for use with future survey endeavours such as the LSST (Large Synoptic Survey Telescope) and future data runs of the SDSS. To be clear, we are not just looking to re-enact ROBUST as it has been used in prior works such as Rauzy and Hendry [160] and the works of Johnston et. al [98] but to instead extend its use to entire galaxy populations, i.e. in principle utilising the same galaxies that have been used to reconstruct a β -dependent estimate of the density field from e.g. IRAS

Method	Compared data	β		
	$\delta - \delta$ Comparison			
POTENT UMV	Mark III vs. IRAS 1.2Jy SEcat vs. PSCz	$\begin{array}{c} 0.89 \pm 0.12 \\ 0.57 \substack{+0.11 \\ -0.13} \end{array}$		
v - v Comparison				
VELMOD VELMOD ITF ITF UMV	Mark III vs. IRAS 1.2Jy SFI vs. PSCz Mark III vs. IRAS 1.2Jy SFI vs. IRAS 1.2Jy SEcat vs. PSCz	$\begin{array}{c} 0.50 \pm 0.07 \\ 0.42 \pm 0.07 \\ ?^{\dagger} \\ 0.6 \pm 0.1 \\ 0.51 \pm 0.06 \end{array}$		

TABLE 3.2: Summary of β values calculated from various basis and iterative methods. [†] Inconsistent flow fields (probably due to problematic calibration of Mark III) [210].

Data	eta
$D_n - \sigma$, IRTF, ESO, UGC	0.50 ± 0.06
SNIa, IRAS 1.2Jy	0.40 ± 0.15
SNIa, ORS	0.30 ± 0.10
SBF, IRAS $1.2Jy$	$0.42^{+0.10}_{-0.06}$
SBF, ORS	0.26 ± 0.08
SMAC, IRAS 1.2Jy	0.39 ± 0.17
SNIa, IRAS PSCz	$\textbf{0.55} \pm \textbf{0.06}$
SNIa, SBF, TF, 2MASS	0.49 ± 0.04
PS,SFI	$0.49\substack{+0.08\\-0.05}$
SNIa, IRAS PSCz	0.50
	Data $D_n - \sigma$, IRTF, ESO, UGC SNIa, IRAS 1.2Jy SNIa, ORS SBF, IRAS 1.2Jy SBF, ORS SMAC, IRAS 1.2Jy SNIA, IRAS PSCz SNIA, SBF, TF, 2MASS PS,SFI SNIA, IRAS PSCz

TABLE 3.3: β values obtained from χ^2 minimising. SBF represents surface brightness fluctuation, and PS denotes power spectrum [177]. The result highlighted in bold signifies the value of β we will use as a benchmark for our future computations.

Work	Data	β
Ross et. al $[132]$	2dF-SDSS LRG, QSO (2SLAQ)	0.45 ± 0.05
Chuang & Wang [29]	SDSS LRG	0.44 ± 0.15
Erdoğdu et. al [58]	2MASS, Fourier-Bessel	0.54 ± 0.12
Pope et. $al [155]$	SDSS, KLT	0.45 ± 0.12
Bilicki et. al [11]	2MASS XSC, LG motion	0.43 ± 0.03
Hawken et. al [77]	VIPERS, voids	$0.423\substack{+0.104 \\ -0.108}$
Hamaus $et. al$ [75]	SDSS, voids	$0.457_{-0.054}^{+0.056}$

TABLE 3.4: β values from a selection of more recent methodologies briefly introduced in Section 3.4.6.

PSCz or the 2MRS, to estimate and constrain β . This would be instead of comparing the β -dependent reconstructed velocity field with a sparsely sampled set of peculiar velocity estimates from e.g. Type Ia Sne.

We also intend to evaluate how well ROBUST can constrain β in the situation where we have a much less precise galaxy distance indicator (such as a broader luminosity function for galaxies in a redshift survey), but conversely the survey in question has a much larger sample of galaxies to which that is applied.

Chapter 4

Introducing the ROBUST Method

While we have dedicated sections of the previous chapter to detailing some of the various methods available to us to constrain cosmological parameters such as β , we have established that they are not without their problems. Basis Functions in particular suffer from shot noise that needs to be suppressed, which in itself introduces additional errors as the amplitude of the signal one is attempting to recover from the data is affected [172]. Moreover the smoothing techniques required for POTENT to reliably observe a continuous radial velocity field for its measurements are also fraught with additional sources of error being introduced such as Malmquist biases that need to be accounted for, with various prior distributions of luminosity functions and assumed likelihoods of distances being required [184] [79]. The same is true of UMV which will not function adequately without assumed priors on the covariance of the variables of the underlying cosmological field one is attempting to recover, despite it being constructed to be the 'happy medium' solution between POTENT and VELMOD [209]. While the aforementioned VELMOD is indeed constructed to not require any smoothing of the data used, its computed values for parameters such as β have been inconsistent with POTENT, and have shown indications of poor matches between models and data that require the use of additional quadrupole contributions to fix the mismatch.

The main takeaway from the above is simple. In order for us to be able to use galaxy peculiar velocity data to constrain cosmological parameters effectively, it would be ideal to make use of methods that require as little to no smoothing of the observed data as possible such as to minimise reconstruction errors and/or shot noise, while also making use of as few a priori assumptions and likelihoods on the underlying data as possible. This will ideally prevent additional biases or errors being introduced that are a result

of possibly assuming the wrong initial model or conditions. In keeping such smoothing methods and prior likelihoods and assumptions to a minimum the potential application of such a method can also be broadened considerably. Thankfully we are now in a position where such methods can indeed be constructed and tested, one such method of which, the ROBUST method of Rauzy & Hendry, we will discuss in detail below.

4.1 Underlying Theory of ROBUST

To begin with, let us consider a redshift-distance survey of galaxies that we require to be **complete** up to a limiting magnitude given by m_{lim} or, put another way, the selection function in apparent magnitude $\psi(m)$ is well defined by a sharp cut-off such that:

$$\psi(m) = \theta(m_{lim} - m),$$

where θ is the Heaviside function (otherwise known as a unit step function which, in this instance, will take a value of 0 for all observed magnitudes fainter than m_{lim} , and a value of 1 otherwise). If we then make the assumption that **the distribution of absolute magnitudes M, i.e. its luminosity function** F(M) **is independent of the spatial position** $\mathbf{r} = (r, l, b)$ **of the galaxies** (which makes sense when one considers the cosmological principle), then we are in a position where we can write the probability density function (or PDF) of such a survey as the product of the probabilities of those two variables such that:

$$dP \propto dP_{\mathbf{r}} \times dP_M = \rho(r, l, b)r^2 \cos b \, dl \, db \, dr \times f(M)dM, \tag{4.1}$$

where $\rho(r, l, b)$ is the spatial distribution function of the galaxies in polar coordinates [160]. Further incorporating our aforementioned requirement of completeness and accounting for selection effects we can rewrite Equation 4.1 such that:

$$dP = \frac{1}{A}h(\mu, l, b)\cos b \, dl \, db \, d\mu f(M)dM\theta(m_{lim} - m), \qquad (4.2)$$

where μ is the distance modulus, $h(\mu, l, b)$ is the line of sight distribution of μ , and A is a normalisation constant such that the integral of the PDF is equal to 1 [160]. The important thing to note here is that now, due to observational selection effects in apparent magnitude, a correlation will be introduced between absolute magnitude M and μ . We will make use of this in later sections of this chapter.

The milestone of the ROBUST method lies in the defining of a random variable ζ such that:

$$\zeta = \frac{F(M)}{F(M_{lim})},\tag{4.3}$$

where F(M) is the cumulative distribution function (or CDF) of absolute magnitudes such that:

$$F(M) = \int_{\infty}^{M} f(x) \mathrm{d}x,$$

and $M_{lim} \equiv M_{lim}(\mu)$ is the absolute magnitude limit at which a galaxy at a distance modulus μ can be observed in the survey due to instrumental precision or redshift selection effects [160]. Differentiating ζ yields:

$$\mathrm{d}\zeta = \frac{f(M)}{F\left[M_{lim}(\mu)\right]} \mathrm{d}M_{lim}(\mu)$$

which, due to the definition of ζ as the ratio of magnitudes up to an assumed magnitude limit will have a value on the interval [0,1]. If we then choose to substitute the volume element $d\mu d\zeta$ into Equation 4.2 (multiplying $d\mu$ into our previous differentiation of ζ) the PDF will reduce to the following:

$$dP = \underbrace{\frac{1}{A}h(\mu)F[M_{lim}(\mu)]d\mu}_{dP_{\mu}} \times \theta(\zeta)\theta(1-\zeta)d\zeta, \qquad (4.4)$$

where

$$A = \int h(\mu) F[M_{lim}(\mu)] \,\mathrm{d}\mu,$$

where θ remains the Heaviside function as before, and dP_{μ} now describes the observed spatial distribution function of the galaxies in the survey [160]. Equation 4.4 now allows us to deduce two very important features of our random variable ζ , namely:

- 1. ζ will be uniformly distributed on the interval [0,1],
- 2. ζ and μ are statistically independent, i.e. the distribution of ζ is independent of the spatial distribution of the galaxies in the survey, just like our initial assumption on the distribution of absolute magnitude made earlier [160].

We can make use of the first property to measure the completeness of a redshift survey up to a given apparent magnitude, while the second property can be used to fit peculiar velocity field models and constrain the cosmological parameters on which they are dependent. We will now delve into how both of these can be achieved.

4.2 Testing for Completeness, T_c

Let us first consider a **magnitude-complete** redshift-distance galaxy survey whose magnitudes have been corrected to account for Galactic extinction effects and have also been k-corrected (in others words, adjustments have been made to the photometric magnitudes and colours of the galaxies to take into account the effect of the redshift on the galaxy's spectrum; correcting the magnitudes to be in the rest frame of each galaxy [4]). These corrected magnitudes m_{cor} would be written as:

$$m_{cor} = m - k_{cor}(z) - A_q(l, b),$$

where z is the observed redshift and $A_g(l, b)$ is the Galactic extinction correction [159]. We can then infer the corrected distance moduli Z of the galaxies in the survey to be:

$$Z = m - M = \mu(z) + k_{cor}(z) + A_q(l, b), \qquad (4.5)$$

which we can substitute into our definition of ζ and Equation 4.4 to obtain:

$$\zeta = \frac{F(M)}{F\left[M_{lim}(Z)\right]},\tag{4.6}$$

and

$$dP = \frac{1}{A}h(Z,l,b)F[M_{lim}(Z)] dl db dZ \times \theta(\zeta)\theta(1-\zeta)d\zeta, \qquad (4.7)$$

with

$$A = \int h(Z, l, b) F[M_{lim}(Z)] dl db dZ,$$

and our two properties from before continuing to hold true, with the caveat that now instead of μ , both ζ and Z(l, b) are statistically independent.

In the case where our survey is indeed complete up to a given apparent magnitude, it becomes possible for ζ to be estimated without any prior knowledge of the cumulative luminosity function F(M). In particular, for an M-Z plot such as that in Figure 4.1, we can assign each galaxy on the plot a pair of coordinates of the form (M_i, Z_i) which is associated with the region $S_i = S_1 \cup S_2$, where S_1 and S_2 are defined as:

$$S_1 = \{ (M, Z) \text{ such that } M \le M_i \text{ and } Z \le Z_i \}$$
$$S_2 = \{ (M, Z) \text{ such that } M_i < M \le M_{lim}^i \text{ and } Z \le Z_i \}$$

The variables M and Z are considered independent in each region S_i since the cutoff in apparent magnitude is dominated by the S_1 and S_2 region border constraints $M \leq M_{lim}^i(Z_i)$ and $Z \leq Z_i$ [159]. The expected number of points r_i belonging to the



FIGURE 4.1: Plot of the variable Z plotted vs. absolute magnitude M for a theoretical galaxy survey. The procedure for calculating r_i and n_i in Equations 4.8 and 4.9 is also illustrated. Figure reproduced from the work of Rauzy [159].

region S_1 is then given by:

$$\frac{r_i}{N_{gal}} = F(M_i) \times \frac{1}{A} \int_{-\infty}^{Z_i} h(Z, l, b) \, \mathrm{d}Z \, \mathrm{d}l \, \mathrm{d}b, \tag{4.8}$$

where N_{gal} is the number of galaxies in the survey. Similarly n_i , the expected number of galaxies in the region $S_i = S_1 \cup S_2$ is given by:

$$\frac{n_i}{N_{gal}} = F\left[M_{lim}(Z_i)\right] \times \frac{1}{A} \int_{-\infty}^{Z_i} h(Z,l,b) \, \mathrm{d}Z \, \mathrm{d}l \, \mathrm{d}b \tag{4.9}$$

 r_i and n_i are easily determined from Figure 4.1 by simply counting the number of galaxies in each region, and an unbiased estimator of the variable $\hat{\zeta}$ can then be calculated for each galaxy such that:

$$\hat{\zeta}_i = \frac{r_i}{n_i + 1},\tag{4.10}$$

where $\hat{\zeta}_i$ can essentially be considered the normalised rank of each galaxy on the interval [0,1], where the absolute magnitudes are sorted by increasing order within each region S_i [160]. Utilising the rank-based statistics works of Efron & Petrosian [52] where we

test for the independence of rank-based variables, we can infer that when M and Z are independent of each other (as they are here due to how the regions S_1 and S_2 are constructed), then the normalised rank $\hat{\zeta}_i$ will equal any of its possible values with equal probability. Put another way, when M and Z are independent, it won't matter which galaxy on the M-Z plot is selected, as the probability of picking it will be the same, and consequently the probability of that galaxy being assigned a particular value of ζ will also be the same [52]. Moreover since we have also established that ζ is uniform on [0,1], then the mean (or expectation E_i) and variance V_i of ζ is given as:

$$E_i = \frac{1}{2}$$
 $V_i = \frac{1}{12} \cdot \frac{n_i - 1}{n_i + 1}$

Utilising all of the above we can now construct a completeness statistic, denoted T_c , given by:

$$T_{c} = \frac{\sum_{i=1}^{N_{gal}} \left(\hat{\zeta}_{i} - \frac{1}{2}\right)}{\left(\sum_{i=1}^{N_{gal}} V_{i}\right)^{\frac{1}{2}}}$$
(4.11)

where, much like ζ before it and on which it relies, T_c can be computed without any prior knowledge or assumptions made about the model of the luminosity function F(M), or indeed made about the spatial distribution h(Z, l, b) of these galaxies either. This gives T_c a unique advantage in that it also allows for clustering effects and galaxy evolution over time to be considered, as well as various selection functions to be applied, without any bias being introduced to either ζ or T_c [159].

We can now apply T_c to a survey over incremental increases of apparent magnitude and monitor how T_c behaves. Considering how we have constructed it we expect it to behave thusly: as long as the magnitude m^* we are testing at is brighter than m_{lim} (i.e. below the assumed completeness limit of the survey), then the sample we are testing is also presumed to be complete, and T_c is expected to have a mean of 0 and a variance of unity. However, if m^* becomes fainter than m_{lim} , then the incompleteness introduces a deficit of galaxies in Figure 4.1 whose M are fainter than $M_{lim}(Z)$, which will result in a lack of galaxies with values of $\hat{\zeta}_i$ that are close to unity, as illustrated in Figure 4.2. In his work defining this completeness statistic, Rauzy notes that T_c is expected to be systematically negative for all magnitudes greater (or fainter) than m_{lim} , with a plot of T_c being characterised by a plateau of zero mean for all m^* above m_{lim} , followed by a significant falloff beyond m_{lim} such that

$$T_c \simeq 0$$
 for $m^* \le m_{lim}$ $T_c < 0$ for $m^* > m_{lim}$



Random variable ζ

FIGURE 4.2: ζ -Z diagram for two values of the limiting apparent magnitude m^* . For m^* greater than the completeness limit m_{lim} , the number of galaxies fainter than $M_{lim}(Z)$ is underestimated due to incompleteness, inferring a systematic lack of points (or galaxies) with a value of $\hat{\zeta}_i$ close to unity. This effect is particularly visible at high Z (i.e. distant galaxies). Figure reproduced from the works of Rauzy [159].

where $T_c < -1$, $T_c < -2$ and $T_c < -3$ are taken to be the confidence levels of rejection at the 1σ , 2σ and 3σ levels respectively, as depicted in Figure 4.3 when T_c was applied to the elliptical and spiral galaxies of the Second Southern Sky Redshift Survey (SSRS2) to determine the survey's completeness limit [159].

It should be noted that although T_c by its very construction requires no prior knowledge of the luminosity function or spatial distribution of a redshift survey, and allows for the inclusion of clustering effects and galaxy evolution etc. without introducing any bias into T_c , it only remains viable for so long as the following conditions are met:

- 1. The distances of the galaxies are required, which necessitates that a cosmological model of some sort (defining H_0 , Ω_m , Ω_Λ etc.) be specified, and that the contributions of peculiar velocities to the individual redshifts are deemed to be negligible. Additionally the k-corrections and Galactic extinction corrections are essential.
- 2. The shape of the luminosity function of the galaxies must be invariant with time,



FIGURE 4.3: Completeness test T_c applied to the 1373 E/SO galaxies (top) and the 2780 spiral galaxies (bottom) of the SSRS2 sample with redshifts between $2500 < cz < 15000 \text{kms}^{-1}$. The systematic falloff at $m_{lim}=15.35$ is present in both sets. Figure reproduced from the work of Rauzy [159].

3. The luminosity function of the galaxy population must remain independent of the spatial positions of the galaxies [159].

For a more in depth discussion of statistical completeness tests and their applications in astronomy in general, refer to the work of Russell Johnston that gives a broad overview of the statistical tools developed over the past century [97], in addition to his collaborative efforts with Luís Teodoro and Martin Hendry that delve more specifically into the use of T_c and a variant thereof, denoted T_v , and how both can be expanded to include the use of both a faint and a bright limit to create more powerful tools to determine the true completeness limits of a survey while also characterising its systematic errors [98] [99] [100]. As we will be making use of the introduction of a bright completeness limit into our surveys in future chapters, we will briefly cover its definition and construction here



FIGURE 4.4: Schematic illustrating the construction of regions S_1 and S_2 when a bright limit is introduced. These regions are uniquely defined for a slice of fixed width, δZ , in corrected distance modulus, and for 'trial' bright and faint apparent magnitude limits m_*^b and m_*^f , respectively. Also shown are the true bright and faint apparent magnitude limits m_{lim}^b and m_{lim}^f , within which the rectangular regions S_1 and S_2 contain a joint distribution of M and Z that are separable (i.e. independent). Figure reproduced from the works of Johnston et. al [98].

before moving on to the modelling of peculiar velocity fields using the second property of ζ .

4.2.1 Defining T_c with Two Limits

Where Figure 4.1 was constructed with only a faint apparent limit (m_f) in mind, Figure 4.4 depicts how easily this can be extended to include the presence of a bright apparent limit as well, denoted as m_b , where a fixed range of corrected distance moduli δZ are considered.

Much like before, for a given value of δZ , every galaxy in Figure 4.4 with coordinates (M_i, Z_i) can have its regions S_1 and S_2 defined such that

$$S_1 = \left\{ (M, Z) : M_{lim}^b \le M \le M_i, \qquad Z_i - \delta Z \le Z \le Z_i \right\},$$

$$S_2 = \left\{ (M, Z) : M_i < M \le M_{lim}^f, \qquad Z_i - \delta Z \le Z_i \right\},$$

where in the case where there is no bright limit, the above region definitions will revert to their faint limit only cases [98]. In a similar fashion r_i , the expected number of galaxies in region S_1 will be defined as:

$$\frac{r_i}{N_{gal}} = \int_{Z_i - \delta Z}^{Z_i} \bar{h}(Z') dZ' \int_{M_{lim}^b}^{M_i} f(M) dM,$$
(4.12)

and the expected number of galaxies n_i in the region $S_i = S_1 \cup S_2$ being given by:

$$\frac{n_i}{N_{gal}} = \int_{Z_i - \delta Z}^{Z_i} \bar{h}(Z') \mathrm{d}Z' \int_{M^b_{lim}}^{M^f_{lim}} f(M) \mathrm{d}M \tag{4.13}$$

The integrals of Equations 4.12 and 4.13 can be rewritten such that:

$$\int_{M_{lim}^{b}}^{M_{i}} f(M) \mathrm{d}M = F[M_{i}(Z_{i})] - F\left[M_{lim}^{b}(Z_{i})\right], \qquad (4.14)$$

and

$$\int_{M_{lim}^b}^{M_{lim}^J} f(M) \mathrm{d}M = F\left[M_{lim}^f(Z_i)\right] - F\left[M_{lim}^b(Z_i)\right]$$
(4.15)

with our unbiased estimator for $\hat{\zeta}_i$ and our definition of T_c (with its expectation E_i and variance V_i) continuing to be as defined in Equations 4.10 and 4.11 respectively. With regards to $\hat{\zeta}_i$ in particular, the introduction of a bright limit has merely changed the definition of the random variable ζ itself and the membership criteria of the two regions S_1 and S_2 . Consequently, provided that both $m_*^f \leq m_{lim}^f$ and $m_*^b \geq m_{lim}^b$, $\hat{\zeta}_i$ will continue to be uniformly distributed on [0,1], and continue to be uncorrelated with (or independent from) Z_i , just as before [98].

Given that the construction of T_c is not altered or affected by the introduction of a bright limit to the data, we can expect T_c to behave as previously established, fluctuating around a plateau of 0 until the faint magnitude limit of the survey is reached, upon which the statistic exhibits a systematic 'freefall' for all magnitudes fainter than m_{lim}^f . As illustrated in Figure 4.5 where the faint limit of the SDSS-DR1 sample of early-type elliptical galaxies is computed with a bright limit applied, this is indeed the case though it is worth noting that in the case where the completeness is computed without using a bright limit, T_c has a harder time recovering the true faint limit at the 3σ confidence level. T_c however does exhibit its characteristic falloff once the test magnitudes exceed the published faint limit, indicating how the effectiveness of T_c is reliant on how best the luminosity function of the survey in question can be modelled. If it is adequately described by just a faint apparent magnitude limit, such as the SSRS2 in Figure 4.3, then



FIGURE 4.5: Performance of the T_c statistic applied to the SDSS-DR1 early-type elliptical galaxies. In the lefthand panel T_c is computed by fixing the bright magnitude limit to equal the published value of 14.55mag. The behaviour of T_c is consistent magnitude completeness up to and including the published faint limit of 17.45mag with the statistic dropping rapidly thereafter - indicating the onset of incompleteness. In the righthand panel T_c is computed using only a faint limit. T_c deviates significantly from its expected value for a complete set at magnitudes which are much brighter than the published faint magnitude limit, though T_c exhibits an even stronger freefall onc the published faint limit is exceeded. Figure reproduced from the works of Jonhston et. al [98].

the one-limit version of T_c will recover the completeness limit effectively. Conversely if the survey is properly described by both a faint and a bright apparent magnitude limit then using the two-limit version of T_c will more effectively recover the completeness limit than the one-limit variant.

Having now defined our completeness statistic T_c and demonstrated how it may be constructed and applied in both the one and two limit case, we will now proceed to make use of it in conjunction with our second property of ζ , namely its statistical independence from μ or Z to model peculiar velocity fields and constrain their cosmological parameters, focusing on β in particular.

4.3 Constraining β with ROBUST

Recalling from Equation 3.22 where we established that the peculiar velocity field $v(\mathbf{r})$ of a region of space is linearly dependent on the mass distribution of galaxies in that region of space multiplied by a factor of β , the linear redshift distortion parameter, we

have a means by which to describe a one-parameter velocity model for this field. More specifically for a β -dependent velocity field $v_{\beta}(\mathbf{r})$, there exists a solution β^* that scales the velocity field to match what is observed in reality such that

$$v_{\boldsymbol{\beta}^*}(\mathbf{r}) \equiv v(\mathbf{r}).$$

This means that for a given value of the parameter β , the model-dependent variables μ_{β} and M_{β} (which as we established in Section 4.1 will be correlated in some manner due to observational selection effects) can be calculated via the distance modulus law (refer to Section 2.3) to obtain:

$$\mu_{\beta} = 5 \log_{10} \frac{cz}{H_0} + 25 - u_{\beta}, \qquad M_{\beta} = m - \mu_{\beta}, \qquad (4.16)$$

where the β -dependent velocity u_{β} is defined as:

$$u_{\beta} = -5\log_{10}\left(1 - \frac{v_{\beta}}{cz}\right),\tag{4.17}$$

and $v_{\beta} \ll cz$. The β -dependent distance modulus and absolute magnitude μ_{β} and M_{β} are related to the 'true' M and μ via

$$\mu_{\beta} = \mu + u_{\beta^*} - u_{\beta}, \qquad M_{\beta} = u_{\beta^*} + u_{\beta}. \tag{4.18}$$

Recalling the definition of ζ from Equation 4.6 as the ratio of magnitudes in a survey up to an assumed limiting magnitude $M_{lim}(\mu_{\beta})$, we can calculate ζ from μ_{β} and M_{β} such that:

$$dP = \frac{1}{A}h(\mu)F\left[M_{lim}(\mu_{\beta})\right]d\mu C_{\beta}\theta(\zeta_{\beta})\theta(1-\zeta_{\beta})d\zeta_{\beta},$$
(4.19)

where C_{β} has the following form when $(v_{\beta^*} - v_{\beta}) \ll cz$:

$$C_{\beta} = \frac{f(M)}{f(M_{\beta})} \simeq 1 + (u_{\beta} - u_{\beta^*}) (\ln f)'(M_{\beta})$$
(4.20)

Because M_{β} (and consequently $(\ln f)'(M_{\beta})$) is correlated with μ_{β} (ala Section 4.1) and is therefore also correlated with ζ_{β} as per the first property of ζ , C_{β} essentially becomes the correlation coefficient between ζ_{β} and the modelled β -dependent velocity field u_{β} for all cases where $\beta \neq \beta^*$. Conversely for the case where $\beta = \beta^*$, then M_{β} and ζ_{β} become statistically independent and fully separable on an M-Z plot, such that $\zeta_{\beta} \equiv \zeta$ is independent from the spatial distribution of galaxies and from any β -dependent velocity field $u_{\beta}(\mathbf{r})$ [160].

Another way to think about this would be to consider either Figure 4.1 or 4.3. As we continue to alter our value of β being applied to the velocity field, it will cause the corrected distance moduli Z and the absolute magnitudes M of the galaxies in the survey to alter and move around the M-Z plot. As a result we are essentially wanting to find the optimal value of β , β^* , such that the distribution of M and Z in the plot will be fully separable due to them being statistically independent.

As a result, any test of independence between ζ_{β} and u_{β} will yield an unbiased estimate of the optimal value of β , β^* , given the unbiased nature of $\hat{\zeta}_i$ as defined earlier [160]. More specifically, when two variables are statistically independent then the coefficient of correlation between them reduces to 0, therefore for $\beta = \beta^*$ we can define:

$$\beta = \beta^* \Leftrightarrow \rho(\zeta_\beta, u_\beta) = 0, \tag{4.21}$$

where

$$\rho(\zeta_{\beta^*}, u_{\beta^*}) = \frac{n \sum_{1}^{n} \zeta_i u_i - \sum_{1}^{n} \zeta_i \sum_{1}^{n} u_i}{\sqrt{n \sum_{1}^{n} \zeta_i^2 - \left(\sum_{1}^{n} \zeta_i\right)^2} \times \sqrt{n \sum_{1}^{n} u_i^2 - \left(\sum_{1}^{n} u_i\right)^2}} = 0, \qquad (4.22)$$

and n is the number of galaxies in the survey [117].

Having now defined our correlation coefficient parameter $\rho(\zeta_{\beta}, u_{\beta})$, we can now apply it to a galaxy survey with a velocity field model such as that given in Equation 4.17 and monitor its behaviour. Considering that (much like our χ^2 -minimisation techniques of Section 3.4.5) β will be increased incrementally on the interval [0,1], we expect our logarithmic velocity model of Equation 4.17 to also exhibit a monotonic increase in value, which consequently would cause a plot of $\rho(\zeta_{\beta}, u_{\beta})$ vs. β_{trial} to also exhibit a similar monotonic increase in its value as we continue to scale over different trial values. It follows from this that the behaviour of $\rho(\zeta_{\beta}, u_{\beta})$ does not need necessarily to follow a linear progression in value over trial values of β as this will be dependent on the sort of underlying velocity field being modelled. For the sake of simplicity however we will be making use of the logarithmic model defined above in future computational chapters of this work, though the robustness of the parameter does indeed permit us to model whatever field we like, should we wish to.

Figure 4.6 illustrates the use of the correlation coefficient parameter to constrain a value of β for the IRAS 1.2Jy galaxy sample, obtaining a value of β^* around 0.1-0.15. The plot of ρ exhibits the monotonic increase we expected of it, though it must be noted that the value of β^* constrained here is inordinately low. In particular this value of $\beta^* \sim 0.1$ -0.15 is inconsistent with the published value of $\beta = 0.50 \pm 0.04$ for the IRAS sample using VELMOD and Tully-Fisher data [204]. Rauzy & Hendry however theorise



FIGURE 4.6: Observed coefficient of correlation $\rho(\zeta_{\beta}, u_{\beta})$ for the IRAS 1.2Jy survey sample as a function of the parameter β , returning a value of $\beta^*=0.1$ -0.15 as the value for which ζ and the underlying peculiar velocity field u_{β} are completely separable. Confidence levels of rejection on β at the 1, 2 and 3σ levels are denoted in dotted lines, and the means by which these confidence levels are constructed are discussed in Section 4.3.1. Figure reproduced from the works of Rauzy & Hendry [160].

that the reason behind this discrepancy lies in the effective depth of the survey being analysed. More specifically, in their work Strauss & Willick analysed the IRAS 1.2Jy galaxy sample out to a limiting redshift of $cz \leq 7500 \text{kms}^{-1}$ when applying VELMOD, while Rauzy & Hendry applied the ROBUST method to the sample on a redshift range of $1000 \leq cz \leq 12000 \text{kms}^{-1}$ [204][160]. When they repeat their analysis with a considerably truncated sample of the IRAS 1.2Jy out to $500 \leq cz \leq 5000 \text{kms}^{-1}$, the value of β^* returned by ROBUST improves to ~0.35 as shown in Figure 4.7, which is in better agreement with the published results of Strauss & Willick. What this would appear to suggest is that the predicted IRAS velocity field model, while successful in reproducing the cosmic flow locally, fails to describe the kinematics on larger scales [160].

Having now established the fundamental idea behind the ROBUST method and how it can be utilised to compute an unbiased estimate of the parameter β we now need to turn our attention to how to determine the confidence intervals on the values it computes.



FIGURE 4.7: The confidence levels of rejection for β for both the full (black) and truncated (grey) samples of the IRAS 12Jy galaxy sample, plotted as a probability function defined by $1 - 2 \operatorname{Prob} (\rho \leq -|\rho_{obs}(\beta)|)$, such that the minimum of the function indicates the optimal value of β^* . Confidence intervals at the 1, 2 and 3σ levels are denoted by dotted lines, and the means of their construction are discussed in Section 4.3.1. Figure reproduced from the works of Rauzy & Hendry [160].

4.3.1 Determining the Error on β

Determining the errors, or confidence intervals on the values of β computed by ROBUST is actually a rather simple affair, as we can exploit the characteristics of the variable ζ utilised in $\rho(\zeta_{\beta}, u_{\beta})$, namely its uniform distribution on [0,1], and its statistical independence from μ or Z to our advantage. Before we address this however let us discuss how we would determine the errors on β . We proceed by computing the normalised CDF of the values of $\rho(\zeta_{\beta}, u_{\beta})$ such that

$$F_X(\rho_i) = \frac{P(X) \le \rho_i}{N_{gal}},\tag{4.23}$$

where in effect for each value of ρ_i computed for a galaxy we count the number of galaxies that have a value of $\rho \leq \rho_i$, normalised over N_{gal} . When plotted, you obtain a CDF similar to that depicted in Figure 4.8. As per the basic rules of statistics a value of ρ_i on this plot corresponding to, say, a CDF value of 0.75 indicates that 75% of all correlation coefficient values of ρ in the distribution exist below this ρ_i . If we assume a Gaussian prior on the distribution of errors around β , we can compute either the one-sided 1,


Coefficient of correlation ρ

FIGURE 4.8: Sample cumulative distribution function of ρ with the one-sided 95% confidence level (2 σ) on β depicted (black dotted line). Figure reproduced from the work of Rauzy & Hendry [160].

2 and 3σ confidence intervals on our values of ρ by utilising the CDF values of 0.68, 0.95 and 0.997 respectively, or the two-sided confidence intervals by utilising the same values, except symmetric around the midpoint of CDF=0.5. This would correspond to CDF intervals of [0.16,0.84], [0.025,0.975] and [~0.01,~0.99] to determine the 1, 2 and 3σ confidence intervals on ρ respectively. Once we have identified these values, extrapolating the confidence intervals on β^* becomes a simple matter of referring to the points on the β - ρ plot of Figure 4.6 and identifying which trial values of β correspond to the calculated confidence values of ρ .

Having now established how we can compute our confidence intervals on β , we will now discuss our aforementioned ability to make use of the characteristics of ζ to achieve this. In particular consider the two following cases:

We know that ζ is uniform and will take a value on the interval [0,1], and that by the works of Efron & Petrosian [52], all these values are equally likely when randomly selecting a galaxy on a ζ-μ or ζ-Z plot such as depicted in the upper panel of Figure 4.2 for a magnitude complete survey. We also know that for the optimal β value of β*, the two distributions of ζ and μ/Z will be completely separable and thus the distribution of one is completely independent of the distribution of the other. Therefore, if we were to generate a random string of numbers on the interval [0,1]

and assign these values instead to the galaxies as their computed values of ζ , there should be no alteration to the uniform distribution being observed in the plot, and more importantly **no alteration to the values of** $\rho(\zeta_{\beta}, u_{\beta})$ **calculated thereafter**. We would then proceed to compute the normalised CDF for this iteration of ρ as normal and identify its confidence intervals on β as discussed earlier. If we repeat this multiple times using newly generated random number strings each time and observe the confidence intervals on ρ (and consequently β) calculated, this will give us a very good handle on the error bars on β . This is known as ζ -resampling, or Monte Carlo simulations.

2. As a variant of the above point, consider that instead of resampling values of ζ from a new uniform distribution on [0,1] and assigning them to the galaxies in our sample, we choose to take the already existing values of ζ and scramble them: reassigning them at random to different galaxies. Proceeding on what we have established above, our uniform distribution of ζ values should continue to remain unaffected by this scrambling, and therefore the values of ρ calculated thereafter will also remain unchanged. With every repetition of this we would randomly reassign our existing ζ values again to new galaxies and observe the behaviour of our normalised CDFs of ρ as normal to gain a handle on the error bars on β . This is known as ζ -scrambling, or bootstrap resampling.

We will make use of both Monte Carlo simulations and bootstrap resampling in future computational chapters of this work to observe how well both of these methods can constrain the error on β , and whether there is indeed any preference between the two.

In the case of using mock galaxy data requiring multiple mock realisations or simulations to determine β , the constraining of errors on β is even simpler to do, as one only needs to take note of the optimal value of β^* calculated in each mock, then plot a histogram of these values once all mocks are completed. By modelling a best-fit Gaussian curve to such a histogram the standard deviation of the Gaussian model σ will consequently indicate the 1σ confidence interval on β^* .

4.4 Strength of ROBUST

The strength of the ROBUST methodology lies, rather unsurprisingly, in its robustness. We have made no prior assumptions on the spatial distribution of galaxies and their luminosity function, allowing for the effects of galaxy clustering and evolution effects to be considered without introducing bias to our method, therefore eliminating the need for additional corrective procedures such as Malmquist corrections, homogeneous or otherwise. We only require that the cumulative luminosity function of the galaxies be statistically independent from their observed spatial distribution. Additionally, although ROBUST requires that the survey being used be magnitude complete in order for it to function effectively, we have illustrated how we can make use of ROBUST, and more specifically the completeness statistic derived from it, T_c , to compute both the faint and bright limits of said survey without much difficulty and proceed without further impediment. Also, in ROBUST only requiring us to make use of the survey and all the information contained within it for it to function, we have eliminated the need for secondary distance indicators such as Type Ia Sne or the Tully-Fisher relation to determine accurate distances to galaxies in the survey, as ROBUST can function without them.

With regards to the functioning of the correlation coefficient parameter ρ in particular, we have established its ability to recover a value of β for the IRAS 1.2Jy sample over small redshift scales without having to make use of the Tully-Fisher information of those galaxies (or indeed any of the priors made on the Tully-Fisher relation itself) as required by methods such as VELMOD, and still be in good agreement with those published values [204]. This in itself leads to an interesting point. The flux information of a survey such as IRAS becomes more poorly defined over larger distances and yet despite this, ROBUST was still able to recover a value for β when the entire sample was used, albeit one that was in poorer agreement with the published value of Strauss & Willick. When one considers the number of galaxies in IRAS we can infer that even with a broadly defined luminosity function that would result in poorer distance indicators on galaxies over larger distances, the sheer number of galaxies in the survey actually helps to balance this out with ROBUST being able to constrain β to a degree regardless. We will be making use of this characteristic in later sections when we experiment with altering the size of the galaxy samples we choose to apply ROBUST to, and altering the width of the luminosity functions we model to mock catalogues as we observe how well ROBUST continues to constrain β .

Having now established how ROBUST works, and the potential it presents to constrain β effectively with minimal a priori assumptions or models required, we will now proceed to apply ROBUST to the IRAS PSCz, and 2MRS surveys. With regards to the former, we will make use of mock simulations initially, assuming a 'true' value of β while experimenting with altering mock galaxy catalogue sizes, varying luminosity functions, and altering various reconstruction and distance errors applied to the galaxies to see how well ROBUST continues to constrain values of β ; while also observing to what level it remains in good agreement with our presumed 'true' value.

Once we have ascertained the viability of ROBUST from mock simulations of the IRAS PSCz, we will proceed to use the entirety of the survey up to its magnitude completeness level to seek to constrain β in its real-world setting and compare how well our value lies in agreement with published values for the IRAS PSCz. We will then proceed to apply ROBUST to the 2MRS, computing values of β for that survey as well. It is our hope that ROBUST proves itself successful in recovering β effectively from all of the aforementioned surveys, securing its potential as a powerful statistical tool for probing the velocity fields and cosmological parameters of upcoming future surveys such as the LSST, the Square Kilometre Array project (SKA) and (looking to the more immediate future) future data runs of the SDSS.

Chapter 5

Applying χ^2 to the IRAS PSCz Survey

Before we actually begin to apply ROBUST to either the IRAS PSCz or the 2MRS surveys, we will first dedicate this chapter to a real-world application of χ^2 -minimisation to the IRAS PSCz; with the goal of establishing or constraining a benchmark value for the linear redshift distortion parameter β from which we then draw our assumed 'true' value of β from for ROBUST to attempt to recover. It will also serve as the testing ground for various coordinate conversion methods that are necessary to properly embed real-world Type Ia Sne into any given survey, given the typical lack of consistency in coordinate conversion calculator provided by NASA [30] to cross-check the coordinates computed from our conversion techniques to ensure they are correct before proceeding with the χ^2 -minimisation. The results generated will be presented in Section 5.2, with a flowchart of the typical methodology we will implement here presented in Figure 5.1.

Aside - choosing the appropriate software

Before any analysis is attempted it would be prudent to decide upon a program or analysis software that would be most suited to our needs. Considering the general scope and size of the different galaxy surveys we will be analysing throughout this thesis and the statistical analyses and mathematical calculations that will be performed, it was decided that Matlab would be used for the following reasons:

• Matlab is specifically designed to work with large datasets, particularly in the matrix formats in which most of our available datasets exist, and supports a wide



FIGURE 5.1: Flowchart illustrating typical $\chi^2\text{-minimisation}$ methodology using real world data.

range of matrix manipulation techniques that can be performed with ease alongside basic and complex mathematical and statistical calculations.

- The software also provides logical libraries and the ability to perform logical operations such as searches, complex condition searching and exclusions and counts, all of which will be of great use when applying the ROBUST methodology
- In working with the software and discovering the existence of the **bsxfun** function which is specifically designed to optimise the calculation runtime on large matrices, this allows us to work with even larger datasets such as 2MASS and the SDSS in less time, further broadening our research prospects.

With our computational software now selected, we will proceed to apply χ^2 -minimisation to the IRAS PSCz, after briefly introducing the sets of real-world Type Ia Sne that we will make use of going forward.

5.1 Using Real-World Sne Sets

An Alternative to Tonry - the Radburn-Smith Sne subset

The Tonry et. al supernovae dataset, as introduced in Section 2.9.1, has served as a cornerstone for many cosmological analyses over the years, however due to it consisting primarily of high-redshift Sne, it does not lend itself naturally to our work. We are focusing primarily on the low redshift Universe, not venturing out beyond a redshift of 0.1, or equivalent distance of 30000kms^{-1} . As discussed in Section 3.3.3 this is to monitor the effect, if any, of the breakdown in isotropy and homogeneity on the value of the linear redshift distortion parameter β that we are attempting to constrain. Additionally as noted when one plots the general distribution of PSCz galaxies in the survey as a function of distance, the level of completeness of the PSCz and our confidence in the reliability of its reconstructed peculiar velocity fields starts to fall off considerably after an approximate distance limit of 15000 kms⁻¹. As a result a restricted subset of the Tonry et. al set is required.

The Type Ia Sne set developed by Radburn-Smith et. al in their seminal paper which calculated a value of β using the IRAS PSCz survey (see Section 2.10.1) of $\beta=0.55 \pm 0.06$ [158] is exactly the set we need. Specifically, it is a restricted version of the Tonry et. al set wherein which only all Sne within $150h^{-1}$ Mpc are considered, thus keeping all objects within our low redshift limit of $z\approx 0.1$. In addition, all Sne which exhibit V-band extinction values larger than 1.0 mags are excluded due to their associated errors most

likely being underestimated [158]. This produces a usable set of 98 SNe with which we can usefully constrain β with and will as such use going forward.

5.2 Constraining β with the RS Sne Set

Being aware that the PSCz survey is complete with a reliable reconstructed peculiar velocity field out to approximately a distance of 15000kms^{-1} (or conversely a redshift of 0.05), the Radburn-Smith set (henceforth named the RS set) needs to be restricted out to that same distance/redshift limit. This is to ensure that the RS Sne are fully embedded within the PSCz survey and that a velocity interpolation scheme can work as efficiently as possible. We then proceed to apply a framework of coordinate frame conversions and velocity interpolations as required, the results of which can be observed in Figure 5.2, alongside a comparison chart of our interpolated velocities for the RS Sne alongside their 'observed' velocities as calculated from the Hubble Law for our assumed 'true' value of $\beta = 0.55$. Should our assumed value of β indeed be correct, then the comparison plot should exhibit a near 1:1 ratio, which shall be illustrated by means of a red line in an attempt to fit the data. It should be noted that the area of each individual circle on these velocity comparison plots is proportional to the associated reconstruction errors for each Sne. Put another way, the smaller the circle, the smaller the associated error and the more confident we are of the true peculiar velocity of that particular Sne lying in an increasingly small range on the plot.

5.2.1 The Curiosity of Large χ^2 Values

While we obtain a β value of 0.553 ± 0.04 as demonstrated in the top panel of Figure 5.2 (the error bars have been calculated at the 1σ level by determining the β -intercept points on our χ^2 parabola at $\chi^2_{min} \pm 1$), which is in excellent agreement with the value obtained by Radburn-Smith there are a few peculiarities that must be addressed. As we established in Section 3.4.5, our expected minimum value of χ^2 should be on the order of the numbers of degrees of freedom, or nsn - 3 to be more exact considering the fixed parameters used in this analysis, with values approaching more than double this amount causing us to favour rejecting our β of 0.55 as the correct value for the IRAS PSCz [199]. Having determined in the previous section that we have 98 usable Sne from the RS set we should consequently expect to see a minimum χ^2 ideally on the order of 95 or smaller, yet this is clearly not the case as seen in the top panel of Figure 5.2, where the minimum is on the order of nearly double that. These larger values could be indicative of several things:



FIGURE 5.2: Results of utilising real-world Sne sets to constrain β for the PSCz. Upper panel depicts the χ^2 plot indicating the optimal value of β , with χ^2_{min} at $\beta = 0.553 \pm 0.04$ providing the optimal value of β at the 1 σ error level. The lower panel depicts the comparison between observed and interpolated velocities. Areas of individual plot points are proportional to the associated reconstruction errors of each individual Sne used, with the red line indicating the linear regression (or goodness-of-fit) between both sets of velocities.

- 1. Our underlying null hypothesis of β =0.55 as per the results of Radburn-Smith is incorrect in some fashion as previously mentioned,
- 2. There is some underlying irregularity in the data being used here,
- 3. Our target Sne analysis sample is too small for use

With regards to the first point this is unlikely when one considers that the Radburn-Smith value of β has ben independently verified by Zaroubi et. al using an unbiased version of the Weiner filter [211], and has also been calculated by Radburn-Smith himself using independent means via the equation:

$$\beta_I = \frac{\Omega_m^{0.6}}{b_I} = \frac{\Omega_m^{0.6} \sigma_8}{\sigma_{8,I}},$$
(5.1)

where σ_8 is the rms (root mean square) amplitude of the mass fluctuations, δ_m , averaged within a top-hat sphere of $8h^{-1}$ Mpc radius. By utilising data from WMAP and other CMB and non-CMB sources a value of $\Omega_m^{0.6}\sigma_8$ has been derived as 0.38 ± 0.04 (see the works of Spergel et. al [182] and the references therein for a more in-depth discussion and analysis of the preliminary telemetry received from the WMAP probe), and by directly integrating the PSCz power spectrum (see the works of Hamilton & Tegmark and the references therein), $\sigma_{8,I}$ was found to be 0.80 ± 0.05 [76], thus giving us a final β value of 0.48 ± 0.06 , which is also in good agreement with the value obtained via Radburn-Smith's velocity-velocity comparisons [158]. Consequently it is unlikely that our underlying null hypothesis is incorrect in this instance and is not the cause of our inordinately large χ^2 values.

With regards to the second possibility: the existence of some kind of irregularity in the raw data, the fact that the RS Sne is a subset of an already well established data set, namely the Tonry et. al set which has been used in multiple analyses of import makes this unlikely. In addition, the IRAS PSCz survey is a well established survey which continues to be used to this day without problem for all kinds of cosmological and redshift/peculiar velocity analyses due to its all-sky properties. While it is possible that the source of the irregularity exists instead in the coordinate transformation matrices that we are applying, this is also very unlikely. By making use of the online positional coordinates calculator LAMBDA provided by the Goddard Space Flight Centre [30], excellent agreement is observed with the RS Sne when their galactic latitude and longitude angles are transformed into the cartesian supergalactic coordinates necessary to properly embed them within the PSCz survey, rendering the possibility of the problem lying in those transformation matrices even less likely. With regards to the third option, it is indeed possible that utilising a target Sne sample of only 98 objects to analyse and constrain β for a redshift survey containing over 15000 objects in and of itself is not sufficient for χ^2 to be able to return the smaller values we expect of it. It has been noted by Radburn-Smith in his paper that the value of β he obtained was not affected by the applied 'cull' of Sne that exhibited V-band extinction values larger than 1.0 mags [158], thus giving us a means of increasing the size of our base analysis sample and noting whether this has any effect on the value of β returned by χ^2 . To that effect, we will reinstate the culled Sne into our sample, now giving us a sample size of 145 Sne to work with before distance restrictions are applied. Once our distance restriction of 15000kms⁻¹ is applied we have a usable sample of 105 objects. While not ideal, the addition of these few extra objects may affect the χ^2 value size enough to give us an indication as to whether we are indeed working with too small a sample or not. We reapply our developed framework and velocity interpolation scheme and proceed with our χ^2 analysis, the results of which can be seen in Figure 5.3.

There are two key things that we can take away from Figure 5.3. Firstly, we can confirm Radburn-Smith's prior observations that our addition of the culled Sne with large V-band extinction values had minimal effect on the value of β obtained by χ^2 , namely $\beta=0.541 \pm 0.04$ with our error bars once again calculated at the 1σ level using the β -intercepts to our χ^2 plot at $\chi^2_{min} \pm 1$, which continues to be in very good agreement with all previously discussed values in this section, thus suggesting that the theorised underestimation of the associated extinction errors have very little bearing on the calculation of β . However as can be noted from the Figure, the χ^2 values have not reduced, continuing to exhibit a similar deviation of nearly double what our expected minimum value of χ^2 should be (around nsn - 3 = 100 or so in this instance) when compared to the analysis when the extinction Sne were culled from the sample. This unfortunately makes it unclear as to whether the addition of these few extra Sne has had any effect (most likely not), or whether indeed having access to a much larger external Sne sample would cause our χ^2 values to worsen or improve. The latter possibility will be addressed in Section 5.3.

5.2.2 Analysing Linear Regression

The second main peculiarity in these results that needs to be addressed is the less than ideal fit of our predicted velocities against what is observed (i.e. the linear regression between the two sets of data), both in the situation where the culled version of the RS set is used or its counterpart complete with extinction Sne. As can be seen in both bottom panels of Figures 5.2 and 5.3, there exists a noticeable number of outliers, specifically Sne whose associated errors are quite large that deviate quite significantly from the 1:1



FIGURE 5.3: Results of utilising RS Sne set with the 'culled' extinction Sne reinstated into the set to constrain β for the PSCz. Upper panel depicts the χ^2 plot indicating the optimal value of β , with χ^2_{min} at $\beta = 0.541 \pm 0.04$ providing the optimal value of β at the 1σ error level. The lower panel depicts the comparison between observed and interpolated velocities. Areas of individual plot points are proportional to the associated reconstruction errors of each individual Sne used, with the red line indicating the linear regression (or goodness-of-fit) between both sets of velocities.

ratio we expect to see as denoted by the red line, and not enough Sne whose velocities have been well constrained with small errors lying along our expected ratio line. This once again lends itself to one of several possibilities, considering that we are now dealing with velocity comparisons:

- 1. There is a fundamental error in the velocity interpolation schemes being applied to calculate the β -dependent peculiar velocities of the Sne based on the underlying reconstructed peculiar velocity field of the PSCz galaxies,
- 2. The redshift reconstruction and Sne distance errors we are applying as part of the χ^2 analysis are incorrect in some fashion,
- 3. There exists some other irregularity in the data in the form of incorrectly transformed velocity reference frames or incorrect coordinate transformations, although we have already established that this is unlikely given our successful recovery of positional coordinates when checked against available calculator resources such as LAMBDA.
- 4. Once again our assumed true value of β =0.55 is incorrect in some fashion, though as previously discussed this is highly unlikely to be the case given that this value of 0.55 has been independently verified by others.

In considering the first option, it is worth briefly discussing the sort of interpolation scheme being applied to determine the peculiar velocities of the Sne once they have been embedded into the PSCz survey. We first begin by assuming a linear weighting interpolation scheme defined by the following:

Starting with the *x*-component of the velocity:

$$v_x^{pred}(SN_j) = \sum_{i=1}^n a_{ij} v_x^{pred}(PSCz, i),$$
(5.2)

where the weighting factor a_{ij} is assumed to take the form:

$$a_{ij} \propto \frac{1}{d_{ij}^2}$$

and

$$a_{ij} = \frac{\frac{1}{d_{ij}^2}}{npscz}, \qquad (5.3)$$

$$\sum_{k=1}^{k-1} \frac{1}{d_{kj}^2}$$

where k is a dummy variable summing over all galaxies, npscz is the number of galaxies in the IRAS PSCz survey and d_{ij} , the distance between each supernova and galaxy is given by:

$$d_{ij} = \sqrt{(x_{SN} - x_{gal})^2 + (y_{SN} - y_{gal})^2 + (z_{SN} - z_{gal})^2}$$
(5.4)

In effect we are calculating the distances between each supernova and each of the galaxies in the survey, and normalising our results so that each weight takes on a value between 0 and 1 and the sum of the weights is equal to unity. The more PSCz galaxies there are surrounding each supernova, the more accurately we will be able to interpolate its velocity, since the closer a galaxy is to a supernova, the more significantly will its gravitational influence affect the supernova's velocity than those which are farther away, meaning that their peculiar velocities can be expected to be more similar, if not the same due to the distribution of the mass in their vicinity. The weighting factor is designed to reflect this in the calculation: $a_{ij} \sim 0$ means d_{ij} is very large, and the galaxy is very distant, while $a_{ij} \sim 1$ means d_{ij} is very small, and the galaxy is close by. By altering the exponent of d_{ij} used in the weighting factor we are effectively restricting the sphere of interest within which we are working, and limiting our list of usable galaxies to those which are relatively closer by. We have chosen an exponent of 2 for this work to obtain a sphere of considerable and useful size, and effectively screen out whichever galaxies are too far away for our purposes.

Substituting the weights into Equation 5.2 we get:

$$v_x^{pred}(SN_j) = \sum_{i=1}^n \left[\frac{\frac{1}{d_{i_j}^2}}{\sum_{k=1}^{npscz} \frac{1}{d_{k_j}^2}} \right] v_x^{obs}(PSCz, i)$$
(5.5)

Since k is a dummy variable, we can take it out of the sum and rearrange to get the x-component of our interpolated velocity:

$$\sum_{k=1}^{npscz} \frac{1}{d_{kj}^2} \underbrace{v_x^{pred}(SN_j)}_{v_x^{int}} = \sum_{i=1}^n \frac{1}{d_{ij}^2} v_x^{obs}(PSCz, i)$$
(5.6)

The above is repeated with the y and z components, and resultant reconstructed 3D velocity is stored in a matrix V:

$$V = \begin{bmatrix} v_{xinterp} \\ v_{yinterp} \\ v_{zinterp} \end{bmatrix}$$

Since we require the line-of-sight component of this 3D velocity, we dot product it with the unit vector in the direction of the supernova from the centre of the supergalactic plane (the plane on which our version of the PSCz is defined):

$$u_{pec}^{pred}(\beta) = V' \cdot n_i, \tag{5.7}$$

where V' is the transpose of our matrix V, and the unit vector n_i is given by:

$$n_i = (\cos \theta_i \cos \phi_i, \sin \theta_i \cos \phi_i, \sin \phi_i)$$

We have now calculated the interpolated line-of-sight velocity we need for the χ^2 minimisation.

The above linear weighting scheme has been tested rigorously throughout the development of this work; where we have generated mock galaxy positions that are coincident in position with PSCz galaxies and utilised the aforementioned scheme to calculate the line-of-sight velocity at that position and compare it with the value stored on file for the survey. In all instances the scheme has demonstrated its ability to accurately recover values for the line-of-sight peculiar velocity that are within 4-5 decimal places of the values stored on file, making it unlikely that the interpolation scheme is to blame for the deviations exhibited between observed and predicted values.

This now leads to the second possibility: that the estimated errors being associated with the Sne as part of the χ^2 analysis are incorrect in some manner. As touched upon in Section 3.4.5, the two main sources of error included in this analysis are a redshift reconstruction error, whose value is meant to describe the scale length over which we can safely assume that the underlying reconstructed peculiar velocity field of the IRAS PSCz can be deemed reliable, and a radial Sne distance scatter which is meant to encapsulate our uncertainty on the true distances to Type Ia Sne. With regards to the former, we can make use of the works of Branchini et. al wherein they determined the redshift reconstruction 1σ error of the peculiar velocity fields of surveys such as the IRAS PSCz to be approximately ~ 150kms^{-1} [19], while with respect to the latter, we have already discussed in Section 2.5 that the distances of Type Ia Sne are known to within 8%. Since we can take the redshift reconstruction error described here to be global with regards to the PSCz, we decided on formulating the errors associated with each Sne such that $\sigma_i^2 = 150^2 + N[0, \sigma_d(i)]^2$, where the latter term indicates a randomly generated Gaussian error on the order of 8% that is then added to our available Sne distances where $\sigma_d = d_{SN} \times 0.08$, d_{SN} being the Sne distance.

In experimenting with varying both the global reconstruction error applied to the data, and the level of uncertainty in our Sne distances it was found that this current setup for error estimation was the one that most successfully reduced the χ^2 residuals to the levels currently exhibited in the top panels of Figures 5.2 and 5.3; put another way, these error values are the ones that cause the value of χ^2 to minimise the most despite its still significantly overlarge value compared to what is expected. All other attempts to vary the errors applied during our χ^2 analysis only served to considerably worsen the residuals calculated and consequently caused even more significant deviations and outliers to be exhibited in the linear regressions presented in the bottom panels of Figures 5.2 and 5.3 when both predicted and observed peculiar velocities were plotted.

While having been unable to adequately determine what is causing our χ^2 analyses to return minimum values large enough to favour rejecting our assumed true β value of 0.55 as per the rules of thumb established by Wall [199] despite all independent evidence to the contrary, or what is causing the somewhat poor fit between predicted and observed peculiar velocities of the Sne and the numerous outliers and discrepancies in velocity values when one applies a goodness-of-fit test in the form of linear regression, there is one other possibility that is worth considering: namely that perhaps our β statistic by its very nature and construction is not as sensitive to the χ^2 -minimisation methodology as one would like. This manifests itself as a weaker signal being recovered by χ^2 that is represented by the larger range of χ^2 values observed, eventhough the parabolic minimum is at a value for β that is in good agreement with what we are looking for; although this does nothing to explain the observed outliers and velocity discrepancies when one attempts to fit the data with linear regression.

If we choose to proceed on the assumption that perhaps our β parameter is indeed not as sensitive to the χ^2 methodology as one would like, we can still continue to make use of the RS Sne set moving forward due to it returning values of β that are in good agreement with our assumed true value of $\beta=0.55$; just making a point of keeping an eye on χ^2 as and when we use it to see whether we continue to observe weak recovered signals and somewhat anomalous goodness-of-fit testing with linear regression.

A More Recent Alternative - the Unified Supernova Catalogue

While reaffirming the results of Radburn-Smith et. al and our consequent making use of their value of β as a benchmark for any future computations in this work has been useful, it also makes sense for us to try to confirm that value on our own, using Sne catalogues either of our own making or just simply using more recent surveys that are readily available. Looking up more recent Sne catalogues for use with χ^2 to generate results of our own that are not solely 'replicating or re-proving the works of those that came before us' was challenging however. When searching through VizieR finding catalogues of Type Ia Sne that contained both position and redshift-distance information returned fewer results than one would have hoped. While extensive catalogues exist that are older than the Tonry et. al and Union compilations that have been previously discussed, more recent surveys either contain fewer than the 100 or so Sne that we have been using up to this point (which beats the purpose), or do not possess enough Sne that have both the positional and redshift-distance information necessary for us to populate them successfully amongst the PSCz galaxies. We eventually managed to locate one survey that met our purposes, the unified supernova catalog of Lennarz et. al [120] (hereafter called the USC) which draws from three primary online supernovae catalogues: the Central Bureau for Astronomical Telegrams (CBAT), the Asiago Supernova Catalog (ASC) and the (now obsolete) SAI Supernova Catalog described in Section 2.9.3 (SSC), with its most recent drawing of data being downloaded in June of 2011 [120]. The catalogue consists of over 5000 Sne, though of those not all have redshift-distance and directional coordinates listed, requiring us to filter out all those that do. We were able to locate 444 Sne whose redshifts are within $15000 \mathrm{km s}^{-1}$, have clearly established right ascension and declination information (and indeed galactic latitude and longitude information as calculated and provided by VizieR) and distances in Mpc, and we will therefore make use of this set of Sne going forward to constrain a value of β of our own, while also making use of the considerably larger number of available usable Sne to determine whether the increased number will have any effect on the values of χ^2 returned, as discussed at the end of Section 5.2.1.

5.3 Constraining β with the USC Sne Set

Once embedded into the supergalactic frame of the PSCz utilising the relevant coordinate conversion equations, we can proceed with our velocity interpolation scheme to determine the β -dependent predicted peculiar velocities of these 444 Sne and compare those with their observed peculiar velocities as established via the Hubble Law and apply χ^2 and linear regression as normal. The results are presented in Figure 5.4.

These results are curious for two main reasons. Firstly, our value of β utilising this more recent set of Sne suggests a smaller optimal value of β of 0.44 ± 0.04 which, while smaller than the value obtained by Radburn-Smith et al, is interestingly more in line with more recent estimations of β from utilising χ^2 with the 2MRS (as shall be seen in future sections of this work) or indeed some of the more recent results calculated for β using alternate means as discussed in Section 3.4.7 and its accompanying tables of recent values, and this is by no means a bad thing.



FIGURE 5.4: Results of utilising the USC Sne set to constrain β for the PSCz. Upper panel depicts the χ^2 plot indicating the optimal value of β , with χ^2_{min} at $\beta = 0.444 \pm$ 0.04 providing the optimal value of β at the 1 σ error level. The lower panel depicts the comparison between observed and interpolated velocities. Areas of individual plot points are proportional to the associated reconstruction errors of each individual Sne used, with the red line indicating the linear regression (or goodness-of-fit) between both sets of velocities.

However the more pertinent anomaly here is two-fold: firstly is the appearance of χ^2 values so large that our assumed value of β =0.55 is rejected out of hand, let alone a value of β that is more in line with more recent values obtained by alternate means. This would appear to suggest that utilising a larger sample of Sne to perform this analysis has only served to considerably worsen and weaken the signal of β we are attempting to recover, though it is equally likely that there is an inherent fault to this subset of the USC that, unbeknownst to us, makes it a poor candidate for use with the IRAS PSCz unlike its RS counterpart. The second is the absolutely abysmal fit observed between both the predicted and observed peculiar velocities when linear regression is applied, where only a handful of the Sne with low associated errors lie along the predicted 1:1 ratio red regression line depicted in the lower panel of Figure 5.4, with discrepant velocities and outliers with large associated errors dominating throughout the plot.

It is worth pointing out that, as per the investigations carried out in Sections 5.2.1 and 5.2.2, and having already ruled out the possibility that problems reside in either the velocity interpolation schemes or the coordinate transformation matrices we are using that would cause such wildly discrepant pairs of velocities to be observed, we have continued our use of a global redshift reconstruction error of 150kms^{-1} as recommended in the works of Branchini et. al with this particular survey [19], and kept the Sne distance errors at 8% as we established in previous sections. It should be noted once again that any attempt to alter the reconstruction errors applied from their established form of $150^2 + N[0, \sigma_d]^2$ (the latter term being a Gaussian-drawn term of mean 0 and standard deviation defined by the Sne distance error σ_d) only caused the calculated parabolas generated by the USC set to worsen considerably, both in the range of χ^2 values exhibited and the value of β_{opt} returned, suggesting that the initial errors we applied would appear to be the most ideal. As was discussed previously, we expect the minimum value of this particular χ^2 parabola to be on the order of 440 or so (continuing to subtract nsn - 3 degrees of freedom to represent our main sources of error being applied during our analysis and other fixed variables) yet this clearly is not the case, reinforcing the idea that perhaps the USC Sne should be discontinued from further use, irrespective of the arguably weak signal of β it may be recovering and as such, we shall do so from this point forward.

Having now somewhat successfully recovered previously published values of β for the IRAS PSCz using χ^2 -minimisation and the RS Sne set, although with a potential caveat that β itself may not be very sensitive to such a methodology, we shall now move on to analysing the PSCz survey with an alternative technique called ROBUST and explore its effectiveness in recovering β .

Chapter 6

Probing the Peculiar Velocity Field of the PSCz with ROBUST

Having successfully applied χ^2 -minimisation to the IRAS PSCz survey and confirmed the Radburn-Smith value of $\beta = 0.55 \pm 0.06$ for that survey, we now have a benchmark value with which we can rigorously test our ROBUST methodology from Chapter 4 in both a mock and real-world environment. Having already established the software we will be using for our work we will proceed in this chapter as follows: Section 6.1 will detail the various implementations of ROBUST available to us, and Sections 6.2 through 6.5 will detail the results of each implementation. We will continue to make use of any mock methodologies discussed to explore the effects of altering different variables such as mock galaxy sample sizes and luminosity function widths etc. on the accuracy and precision with which ROBUST can estimate the value of β . Anomalies noted in our results will be explored and their root causes determined and eliminated to the best of our abilities.

6.1 Implementing ROBUST with the PSCz Survey

As discussed in Chapter 4, for a magnitude complete survey with known limits, the functionality of ROBUST lies in the assumed independence of the distribution of absolute magnitudes of galaxies from their spatial positions, such that a correlation will be introduced between these two variables should the wrong value of β be applied to the data. Considering the reliance of ROBUST on the luminosity function information of a survey, this gives us several avenues of experimentation to consider, in order to evaluate the effectiveness of ROBUST:

- 1. Mock Method 1: We adopt β_{true} of 0.55 as before and an assumed known luminosity function model, and utilise Matlab to assign 'mock' apparent magnitudes to the PSCz galaxies for an assumed, known faint limit. We then proceed to apply ROBUST, calculating the correlation coefficient $\rho(\zeta, \beta)$ over trial values of β on [0,1], correcting distances and peculiar velocities as appropriate and identify the value of β_{trial} for which $\rho(\zeta, \beta) = 0$. This is repeated over multiple mocks, generating new mock magnitudes to be assigned to the galaxies each time. We will proceed to experiment with assigning mock magnitudes to the PSCz galaxies drawn from the following:
 - (a) A Gaussian luminosity function, whose mean and standard deviation will be in keeping with the identified luminosity function of early-type galaxies of the SDSS, namely N[-21,1] [9],
 - (b) A Schechter luminosity function, whose definition and description of its various parameters will be presented in Section 6.3.1.
- 2. Mock Method 2: We proceed to repeat the above while also establishing a bright limit to the mock magnitudes being generated by Matlab and rerun RO-BUST, taking note of any alterations to the values of β recovered. Much like Method 1, this will also be repeated with both a Gaussian and Schechter luminosity function over multiple mock trials, while also serving as the testing ground for our completeness statistic T_c and its viability in identifying the proper magnitude limits of a survey. With these limits identified we can also explore the viability of ROBUST and its effectiveness in recovering β when various correct and incorrect magnitude limits are applied to the data.

While the fundamental methodology behind these methods is inherently the same, some key differences remain. In particular, Methods 1 and 2 are dealing exclusively with mock magnitudes that are generated by Matlab to be assigned to the PSCz galaxies with a given faint (and bright) limit being considered during generation, thus negating any need for us to apply T_c to determine any limits. This does however give us the means to experiment with varying the number of PSCz galaxies used during the RO-BUST analysis, as well as altering the widths and parameters of both our Gaussian and Schechter LFs and observing how well ROBUST continues to recover our assumed true value of β . With regards to Method 2 specifically, the introduction of a bright limit to the data is particularly important as it will serve as our means of testing T_c and its ability to recover the actual magnitude limits of our data, mock or otherwise, while also giving us the ability to monitor how well ROBUST functions when both the correct and incorrect limits are applied and how well it continues to recover β . We will also continue to experiment with varying mock sample and sizes and altering LF parameters and widths when a bright limit is applied and take note of how the recovered value of β is affected. Flowcharts indicating a typical application of our mock methodologies is presented in Figure 6.1.

6.2 Applying Mock Method 1a) - 'Faint Limit Only Gaussian' Mock Magnitudes

It must be noted that the peculiar velocities reconstructed in Branchini et. al from the IRAS PSCz require a scaling correction from $\beta = 1$ to $\beta = 0.55$ before any mock analyses can be performed, which is easily achieved by means of the modified Hubble Law reiterated here for ease of access:

$$cz = H_0 d + \left[v_{pec}(\beta = 1) \times \beta_{true} \right],$$

where we continue to take the Hubble constant as before as

$$H_0 = 100 \mathrm{km s^{-1} Mpc^{-1}}$$

This now only leaves the matter of generating mock magnitudes to be assigned to the positions of the PSCz galaxies. For each galaxy of given real-space distance d we can make use of the **normrnd** function in Matlab to generate a random number (in this case a mock absolute magnitude) from a normal distribution with mean and standard deviation given by μ and σ respectively. We then proceed to apply the distance modulus law of Equation 2.14 such that:

$$m = M + 5\log d + 25, (6.1)$$

where m and M represent apparent and generated absolute magnitude respectively. Provided that the computed apparent magnitude m for a given galaxy is brighter than a specified faint limit (or for later on also fainter than a given bright limit) then it is assigned to the galaxy, otherwise **normrnd** is used to generate a new mock absolute magnitude. A flowchart depicting a typical mock magnitude generation scheme for the PSCz is given in Figure 6.2.

Having previously established that we would be making use of the identified Gaussian luminosity function of early-type galaxies of the SDSS of the form N[-21,1] to generate mock magnitudes at the positions of the PSCz galaxies [9]; and knowing that we are working with all such galaxies that are within a distance restriction of 15000kms^{-1} to ensure the reliability of the reconstructed peculiar velocity field, we can apply the



FIGURE 6.1: Flowchart illustrating typical ROBUST methodology for use with generated mock magnitudes assigned to galaxies utilising either just a faint, or faint+bright magnitude limits during generation. Note that m_f and m_b stands for faint and bright apparent magnitude limit respectively [Methods 1a), 1b), 2a) and 2b)].



FIGURE 6.2: Flowchart illustrating typical mock magnitude generation and assignment to PSCz galaxies.

distance modulus law such that:

$$m = M + 5 \log d + 25$$

= -21 + 5 log $\left(\frac{15000}{H_0}\right) + 25$
= -21 + (~ 10) + 25 = 14,

giving us a general ballpark value for a mock faint limit for the PSCz galaxies of 14. Despite us making use of the luminosity function of an optical survey (the SDSS) to generate our mock magnitudes for the PSCz galaxies (taken from an infrared survey), this ballpark value is actually in good agreement with the published faint limit of the IRAS 1.2Jy sample as identified in the works of Rauzy & Hendry; given as $m_{lim} = 14.3$ [160], therefore we will make use of this faint apparent magnitude limit in all analyses going forward. The initial results of Mock Method 1 using the entirety of the PSCz (over 12000 galaxies) with a mock Gaussian luminosity function are depicted in Figure 6.3.

It is immediately clear that ROBUST has successfully recovered our assumed true value of β with a very well constrained confidence interval, and the mock $\rho(\zeta, \beta)$ plots exhibit the monotonic increase in value over [0,1] that we expected to see given the logarithmic velocity model that is being applied as per Equation 4.17, reiterated here once again for ease of access:

$$u_{\beta} = -5 \log_{10} \left(1 - \frac{v_{\beta}}{cz} \right),$$

Mock Set Size	1000	2000	5000	All (12000+)
Value of β	0.57 ± 0.37	0.56 ± 0.16	0.55 ± 0.09	0.55 ± 0.06

TABLE 6.1: Summary of estimated values of β detailing the effect of varying mock galaxy sample sizes on computed β using ROBUST during Mock Method 1a) implementation.

where v_{β} is the β -dependent radial peculiar velocity of the survey galaxies.

It is worth noting that this initial run was done making use of only 50 mocks, primarily due to the concern of computational time constraints, with a run utilising 50 mocks taking approximately 12 minutes to complete. Yet despite the somewhat 'coarse' nature of the underlying histogram when modelled to the smooth Gaussian (red line), the error bar returned on β is shown to be very well constrained at ± 0.06 . For the sake of argument we will proceed to redo this particular run utilising 200 mocks to compare and contrast the error bars on β generated in each. Ideally with the larger number of mocks used we should see a much smoother and Gaussian-shaped underlying histogram depicting β_{opt} that models far better than its 50 mock counterpart. The results of this higher mock run is presented in Figure 6.4, being completed in 90 minutes.

As expected the larger number of mocks utilised in this particular run does lend itself to a 'smoother' looking histogram when compared to its 50 mock counterpart that models better with the best-fit Gaussian curve, however there is no noticeable change in either the mean value of β_{opt} returned by ROBUST nor the confidence intervals on β , remaining the same at $\beta=0.55\pm0.06$. Given the considerably longer run times involved with using 200 mocks we will continue to move forward with our analyses making use of 50 mocks only, now confident that despite the underlying coarseness of the histograms presented we are using a sufficient number of mocks to get an accurate handle on not just the value of β but its associated error bars as well.

This now opens the door for us to experiment with varying the number of survey galaxies ROBUST utilises for its analysis and take note of any changes in the value of β recovered. To that effect we rerun our ROBUST analysis using 1000, 2000, 5000 and all survey galaxies of the PSCz with the results illustrated in Figure 6.5. A summary of the estimated values of β of each rerun is provided in Table 6.1.

A few interesting things can be taken away from Figure 6.5, namely that ROBUST continues to recover a mean estimated β that is equal to the true value of β even when a relatively small number of survey galaxies are being used, however the confidence interval on our computed β values broadens considerably as a result. When one considers the varied breadths of the luminosity functions modelled to real-world galaxy surveys



FIGURE 6.3: Initial results of Method 1a). Upper panel depicts the ρ plots for all 50 mocks generated, lower panel depicts the zero-intercept of each mock plot as a function of trial value of β plotted as a histogram with best-fit Gaussian (red line) modelled to the data. Optimal value of β returned for Method 1a) is $\beta = 0.55 \pm 0.06$, as determined from the Gaussian model. Luminosity function used during data generation is Gaussian of the form N[-21,1].



FIGURE 6.4: Results of Method 1a) redone with 200 mocks instead of 50. Upper panel depicts the ρ plots for all 200 mocks generated, lower panel depicts the zero-intercept of each mock plot as a function of trial value of β plotted as a histogram with best-fit Gaussian (red line) modelled to the data. Optimal value of β returned for Method 1a) is $\beta = 0.55 \pm 0.06$, as determined from the Gaussian model. Luminosity function used during data generation is Gaussian of the form N[-21,1].



FIGURE 6.5: Method 1a) results with varied mock set sizes. Left-hand panels illustrate the ρ plots for all 50 mocks, right-hand panels depict the histograms of optimal value of β with best-fit Gaussian (red line). Set sizes used are noted on each plot. Luminosity function generated continues to be Gaussian of form N[-21,1].

Gaussian LF Width	0.5	1	2	3
Value of β	0.56 ± 0.03	0.55 ± 0.06	0.55 ± 0.14	0.57 ± 0.21

TABLE 6.2: Summary of estimated values of β detailing the effect of varying mock generated Gaussian LF width σ on computed β using ROBUST during Mock Method 1a) implementation. Entirety of PSCz survey is utilised for this analysis.

in general (and which we will explore in more detail next) if we are to continue to use ROBUST in the future then we must ensure that the surveys we use have a bare minimum of at least 2000 viable galaxies if we wish to constrain parameters such as β with an acceptable level of accuracy. Given the expected size of future surveys such as the LSST [124], and even more recent data runs of the SDSS this should not present an issue.

We can now proceed with varying the value of σ utilised by **normrnd** when generating Gaussian magnitudes to assign to the PSCz galaxies and observe what changes are observed in the values of β recovered by ROBUST. The results of altering the LF width are presented in Figure 6.6, with a summary of the estimated values of β presented in Table 6.2.

Much like our experimentation with varying the number of galaxies used by ROBUST in its analysis, we can take note of a few interesting things. ROBUST successfully continues to recover a mean estimate for the value of β that is consistent with our assumed true value of β , though our confidence intervals on our values continue to worsen considerably as we make our luminosity function broader. It is important to mention that the entirety of the PSCz survey was utilised during this analysis, which will also have a limiting effect on the broadness of our confidence intervals. Put another way, were we to use a smaller number of galaxies with an LF whose σ is as broad as 2 or 3, we should expect the confidence intervals computed on β to be even broader than those exhibited here. To that end, provided that we have a survey that is sufficiently large (minimum of 2000 galaxies as established earlier), this would appear to suggest that ROBUST will be capable of reliably estimating the value of β provided that the luminosity function is not poorly defined or overly broad. The broader the modelled luminosity function appears to be, the greater the number of galaxies that ROBUST will need to make use of in order for it to estimate β reliably. This will be both underiably useful and important to bear in mind when we eventually move on to using ROBUST with large real-world datasets whose luminosities are described by a luminosity function that may be significantly broader than what is found for, say, the SDSS data.



FIGURE 6.6: Method 1a) results with varied mock generated Gaussian LF widths. Left-hand panels illustrate the ρ plots for all 50 mocks, right-hand panels depict the histograms of optimal value of β with best-fit Gaussian (red line). LF widths generated are noted on each plot. Entirety of the PSCz survey is utilised for this analysis.

6.3 Applying Mock Method 1b) - 'Faint Limit Only Schechter' Mock Magnitudes

To truly determine the usefulness of ROBUST however, we need to determine the extent of its functionality with luminosity functions beyond those of a simple Gaussian. To that effect making use of another commonly utilised function, namely a Schechter function and generating mock magnitudes from such a distribution to assign to the PSCz galaxies is the next logical step for our experimentations with ROBUST. Unfortunately, Matlab does not possess a function of its own similar to **normrnd** that can automatically generate random values from a Schechter distribution, primarily because such a distribution is not a typical statistical distribution used by the program. Consequently generating mock Schechter magnitudes will require a little bit of thought.

6.3.1 Defining the Schechter Luminosity Function

In brief, a Schechter luminosity function provides a parametric description of the number density of galaxies as a function of their luminosity such that:

$$n(x)dx = \phi^* x^{\alpha} \exp^{-x} dx, \quad x = \frac{L}{L^*},$$
 (6.2)

where ϕ^* is a normalisation constant. While this constant is not universal and varies with different populations and environments, one measurement from field galaxies for this constant is given as $\phi^* = 1.2 \times 10^{-2} h^3 \text{Mpc}^{-3}$ [122]. The form of the function is described with an exponential law at bright magnitudes, and a power law defined with a slope of α at fainter magnitudes, with the 'knee' of the function being indicative of L^* , the characteristic galaxy luminosity where the power-law form of the function cuts off, as seen in the first panel of Figure 6.7. This equation can also be written more conveniently in terms of absolute magnitudes instead of luminosities, giving:

$$\Phi(M) = 0.4 \ln 10\phi^* [10^{0.4(M^* - M)}]^{(1+\alpha)} \exp\left(-10^{0.4(M^* - M)}\right),\tag{6.3}$$

where M^* is the characteristic absolute magnitude or 'knee' of the function. In order to be able to randomly assign magnitudes drawn from such a luminosity function, we need to first assign specific assumed true values to the different parameters of the function and then generate a CDF, or cumulative distribution function, for it. This CDF takes values on the interval [0,1] as seen in the second panel of Figure 6.7, allowing us to



FIGURE 6.7: A sample Schechter distribution with M^* =-21 and slope α =1.09, with the CDF generated from such a distribution normalised on [0,1]

Mock Set Size	1000	2000	5000	All (12000+)
Value of β	0.54 ± 0.07	0.55 ± 0.03	0.55 ± 0.02	0.55 ± 0.01

TABLE 6.3: Summary of estimated values of β detailing the effect of varying mock galaxy sizes on computed β using ROBUST during Mock Method 1b) implementation.

exploit the **randperm** function of Matlab which allows us to generate as many unique random numbers between [0,1] as we choose. These numbers can then be mapped to the CDF we have generated to infer the absolute Schechter magnitudes that we need. The parametric values chosen to be assigned to the different components of the function are as follows:

$$\phi^* = 1.2 \times 10^{-2} h^3 \text{Mpc}^{-3}$$

$$\alpha = 1.09, \quad M^* = -21 + 5 \log(h), \quad h = \frac{H_0}{100},$$

where H_0 is the Hubble constant as before [122].

Having now established the means by which we will assign mock Schechter magnitudes to the PSCz galaxies, we can now proceed to repeat our above mock runs with ROBUST and see how well it continues to recover our assumed true value of β . While we will also rerun these analyses with varying mock galaxy sample sizes for the sake of completion, we will also experiment with altering the exponent of the modelled power law α of the Schechter function (analogous to us altering the Gaussian width σ earlier) and see what changes are observed.

Our initial results with ROBUST utilising mock Schechter magnitudes are presented in Figure 6.8. Our results continue to be very promising. ROBUST once again successfully recovers a mean estimate for β that is consistent with the assumed true value of β , and manages to do so within an even more tightly constrained confidence interval than when we used a Gaussian LF. The slight 'S-shape' curve exhibited by the mock $\rho(\zeta, \beta)$ plots is most likely due to the nature of the LF we are using, yet the typical monotonic increase in the plots across increasing values of β_{trial} on [0,1] is still generally observed as expected. Consequently we expect to see similar behaviour from ROBUST when we begin varying the sample galaxy sizes and altering the value of the power-law slope α . The results of both these analyses are presented in Figures 6.9 and 6.10, with summaries of the estimated values of β of each presented in Tables 6.3 and 6.4.



FIGURE 6.8: Initial results of Method 1b). Upper panel depicts the ρ plots for all 50 mocks generated, lower panel depicts the zero-intercept of each mock plot as a function of trial value of β plotted as a histogram with best-fit Gaussian (red line) modelled to the data. Optimal value of β returned for Method 1b) is $\beta = 0.55 \pm 0.01$, as determined from the Gaussian model. Luminosity function used during data generation is Schechter of the form estimated S[-21,1.09].


FIGURE 6.9: Method 1b) results with varied mock set sizes. Left-hand panels illustrate the ρ plots for all 50 mocks, right-hand panels depict the histograms of optimal value of β with best-fit Gaussian (red line). Set sizes used are noted on each plot. Luminosity function generated continues to be Schechter of form estimated S[-21,1.09].



FIGURE 6.10: Method 1b) results with varied mock Schechter power law slopes α . Left-hand panels illustrate the ρ plots for all 50 mocks, right-hand panels depict the histograms of optimal value of β with best-fit Gaussian (red line). α values generated are noted on each plot. Entirety of the PSCz survey is utilised for this analysis.

Schechter Power Law Slope α	0.5	1	2	3
Value of β	$\left \begin{array}{c} 0.55 \pm 0.01 \end{array} \right $	0.55 ± 0.01	0.55 ± 0.01	0.55 ± 0.02

TABLE 6.4: Summary of estimated values of β detailing the effect of varying Schechter power law slope α on computed β using ROBUST during Mock Method 1b) implementation. Entirety of PSCz survey is utilised for this analysis.

Once again there are a few points of interest that can be taken away from Figures 6.9and 6.10. Unlike the Gaussian luminosity function we applied earlier we continue to see a very tightly constrained confidence interval on our recovered values of β , quite a marked improvement in accuracy when one also considers the number of galaxies we allow ROBUST to use during its analysis that would also lend to an improved confidence interval. In particular even when we use a comparatively small number of galaxies our constrained value of β has a confidence interval that is comparable to the one computed by ROBUST when the entirety of the PSCz is used with mock Gaussian magnitudes. Also unlike Method 1a) where the confidence interval on β was heavily reliant on the Gaussian LF width used, there does not appear to be any noticeable alteration in the value of β and confidence intervals recovered by ROBUST when the power law slope α of the Schechter function is varied. This continues to suggest that ROBUST is more than capable of effectively recovering a mean estimate of β consistent with our assumed true value of β irrespective of the number of galaxies used or the kind of luminosity function that is applied, although the quality of the results returned by ROBUST improves considerably the more galaxies are available for analysis, and the better defined the luminosity function of the underlying galaxies is.

All of the above however has been established for a mock luminosity function that is well defined by a faint limit only. Considering the existence of surveys like the SDSS that are modelled by a luminosity function that is defined by both a faint and a bright limit, it makes sense to explore ROBUST's ability to work with mock surveys where both such limits are in play, and observe how well it continues to recover our assumed true value of β . This provides us with the means of testing the extent of ROBUST's usefulness as a statistical tool to constrain cosmological parameters for a broader range of real-world surveys in future chapters.



Random variable ζ

FIGURE 6.11: Sample scatter plots of the ROBUST variable ζ plotting against distance moduli μ for the case where the correct magnitude limits have been defined (top panel) and where the incorrect limits have been defined (bottom panel).

6.4 Applying Mock Method 2a) - 'Faint+Bright Limit Gaussian' Mock Magnitudes

While it is possible that any mock (or indeed true) luminosity function that is determined for the PSCz is best described by a faint limit alone, in the case where the data is indeed best described by the existence of both a faint and bright limit, then this will have a significant impact on the functioning of ROBUST. This goes back to Section 4.1, where we discussed one of the key characteristics of the variable ζ : namely that a scatter plot of ζ vs. distance moduli μ should be uniformly distributed on [0,1] IF the correct magnitude limits are applied. Any incompleteness in the survey will manifest in such a scatter plot as either an under or oversampling, such as that seen in Figure 6.11, and such a deviation in the values of ζ will carry over into the calculations of ρ , affecting the final value of β recovered.

As such it would be prudent to explore the effects adding a bright limit to a survey would have on the functioning of ROBUST, though this does come with an important caveat. It is worth pointing out that our T_c methodology as it is currently designed is not set up for the identification of a bright limit in a survey. In and of itself this is not necessarily surprising when one considers a typical M-Z plot such as that in Figure 4.1. There, a naturally occurring faint limit is to be expected due to instrumental limitations or distance restrictions, while there really is no such thing as a naturally occurring bright limit as such, which would make it difficult for T_c as it stands to be able to accurately identify it. This is even more relevant here since we are dealing with mock generated magnitudes assigned to the PSCz galaxies with a predetermined faint limit already applied, but no imposed bright limit. As such if we are wishing to explore the effects of a bright limit on the functioning of ROBUST it would be wiser to instead artificially impose an arbitrary bright limit of some value, and observe the results that ROBUST returns instead of utilising T_c . We can also continue to make use of the underlying theory from Section 4.1 that dictates that a scatter plot of ζ vs. distance moduli μ should continue to be uniformly distributed provided that the correct bright and faint limits are applied, and see how well this continues to hold true as we alter the arbitrary bright limit value that we choose to apply.

If we are to arbitrarily apply a bright limit to the data this does pose an important question however: what value of ΔZ should be used.

6.4.1 A question of ΔZ

Recall the schematic from Section 4.2.1 illustrating how we construct the S_1 and S_2 sets for use with T_c (or indeed ROBUST) when a bright limit is applied as shown again in Figure 6.12. Whereas before the S_1 and S_2 sets necessary for our analyses would only require the distance moduli and magnitude information for the particular galaxy we are defining the regions around in addition to M_L , here more variables are in play to define S_1 and S_2 correctly. Specifically not only do we need to know the magnitude and distance of our target galaxy in addition to M_L and M_B , but we also need to define a 'height' for these regions, i.e. the range of distance moduli over which we wish to define these sets. If we make ΔZ too small, then we risk throwing away a significant amount of the survey, affecting the final values of ρ calculated. In the first instance we will apply a large $\Delta Z=5$ to our region definitions and run ROBUST with an arbitrary bright limit of $m_b=8$ to see what values for ρ and optimal β are returned.

The results are presented in Figure 6.13, alongside scatter and histogram plots for the (ζ, μ) distribution of galaxy points and ζ values respectively in Figures 6.15 and 6.14 to analyse whether any deviations from uniformity have occurred.

The results returned by ROBUST when we apply an arbitrary bright limit of $m_b=8$ are interesting for a few reasons. They indicate that ROBUST continues to recover a mean estimate of β consistent with β_{true} with no noticeable change in the confidence interval,



FIGURE 6.12: Schematic illustrating the construction of regions S_1 and S_2 when a bright limit is introduced. These regions are uniquely defined for a slice of fixed width, δZ , in corrected distance modulus, and for 'trial' bright and faint apparent magnitude limits m_*^b and m_*^f , respectively. Also shown are the true bright and faint apparent magnitude limits m_{lim}^b and m_{lim}^f , within which the rectangular regions S_1 and S_2 contain a joint distribution of M and Z that are separable (i.e. independent). Figure reproduced from the works of Johnston et. al [98].

but it isn't clear whether the introduction of a bright limit had any significant impact on the analysis. On the one hand this could indicate as mentioned previously that the PSCz is more than adequately defined by the existence of a faint limit alone, however Figures 6.15 and 6.14 show something curious.

At a glance it appears that the presented scatter of (ζ, μ) values is uniformly distributed, but on closer inspection of the associated histogram one can begin to take note of a slight undersampling of data points occurring near the ζ value of 0. If our distribution is indeed uniform then for the PSCz survey which contains a total of 12087 galaxies that ROBUST can analyse given our redshift restriction, we should expect to see ~1209 galaxies in each bin, as signified by the red line. Given ROBUST's success in recovering our assumed true value of β this could once again suggest that the addition of a bright limit has had no material impact on its analysis despite the slight deviation from uniformity observed.

In order for us to determine whether this is indeed the case we will proceed to vary the arbitrary bright limit applied to the data. In particular we will make use of some bright



FIGURE 6.13: Initial results of Method 2a). Upper panel depicts the ρ plots for all 50 mocks generated, lower panel depicts the zero-intercept of each mock plot as a function of trial value of β plotted as a histogram with best-fit Gaussian (red line) modelled to the data. Optimal value of β returned for Method 2a) is $\beta = 0.56 \pm 0.07$, as determined from the Gaussian model. Luminosity function used during data generation is Gaussian of the form N[-21,1], with m_f and m_b given as 14.3 and 8 respectively. Bright limit was applied arbitrarily during S_1 , S_2 set generation.



FIGURE 6.14: Scatter plot of ζ vs. μ for a faint limit of 14.3 and an arbitrary bright limit of 8.



FIGURE 6.15: Histogram of ζ scatter points generated from a fixed faint limit of 14.3, an assumed bright limit of 8 and ΔZ =5, binned into 10 intervals on [0,1]. The thick red line denotes our expected number of galaxies per histogram bin for the distribution to be uniform.

Bright Limit	0	5	8	10
Value of β	0.55 ± 0.06	0.55 ± 0.06	0.56 ± 0.07	0.64 ± 0.07

TABLE 6.5: Summary of estimated values of β detailing the effect of varying applied apparent bright magnitude limit on computed β using ROBUST during Mock Method 2a) implementation. Entirety of PSCz is utilised for this analysis with $\Delta Z=5$.

limits that are brighter or 'closer' to our predetermined faint limit of 14.3 and see if this introduces any bias whatsoever in the values of β recovered by ROBUST, as well as use values that are considerably fainter and take note of any changes observed. It should be noted that we will continue to keep our region widths fixed at $\Delta Z=5$. Considering that we also made use of histogram plots to indicate the level of uniformity of our (ζ, μ) distributions for our bright limit of 8, we will also present similar histograms and scatter plots for the different bright limits applied and see if any useful patterns can be determined. The results of varying the bright limit retroactively applied to the PSCz are presented in Figure 6.16, our monitoring of (ζ, μ) scatter plots and histograms are presented in Figure 6.17 and a summary of the estimated values of β are given in Table 6.5.

The result of applying the incorrect bright limit to the data is striking, as can be seen in the bottom panels of Figure 6.16, where for $m_b=10$ ROBUST returns a mean estimate for β that is inconsistent with our assumed true value of β as the plot of $\rho(\zeta, \beta)$ has shifted significantly to the right; returning a β value of 0.64±0.07. When one considers the arbitrary bright limit of 10 and a ΔZ value of 5 being applied in this instance this is actually not surprising. Our ΔZ of 5 is larger than the difference between our faint and bright limits of 10 and 14, meaning that the shape of the S_1 and S_2 sets will be affected as a result, as can be seen in a sample M-Z plot in Figure 6.18.

For a random galaxy (highlighted in green) on the M-Z plane where a faint and bright limit of 14 and 10 respectively are applied to the data (represented by the blue lines restricting the usable distribution), the applied ΔZ of 5 causes the shapes of S_1 and S_2 to alter from square/rectangular to triangular or trapezoidal when the bright limit comes into play. In the absence of the bright limit we are effectively creating volume-restricted subsets of the galaxy data we are working with (represented by the rectangular regions outlined in black) wherein we can make their absolute magnitudes M and their corrected distance moduli Z completely separable within those regions when we generate our ζ statistic for a clearly defined set of distance moduli and absolute magnitudes. With the bright limit in play however its interference and changing of the shapes of the regions introduces potential 'imbalancing' and incompleteness that may render the separability of our parameters less valid, which would consequently have a negative knock-on effect



FIGURE 6.16: Method 2a) results with varied bright limits applied. Left-hand panels illustrate the ρ plots for all 50 mocks, right-hand panels depict the histograms of optimal value of β with best-fit Gaussian (red line). LF is of the form N[-21,1], with m_f and m_b noted on each plot. Entirety of the PSCz is utilised for this analysis with $\Delta Z=5$.



FIGURE 6.17: Monitoring of ζ behaviour with varied Gaussian bright limit. Left-hand panels illustrate the (ζ, μ) scatter plot for all galaxies used, right-hand panels depict the histograms of ζ value distribution over [0,1]. The thick red line denotes our expected number galaxies per histogram bin for a uniform distribution.



FIGURE 6.18: Sample M-Z plot depicting the trapezoidal nature of the S_1 and S_2 sets (denoted in dark red and orange respectively) when an overly large ΔZ is applied in the presence of both a faint and bright limit (represented by blue lines), creating potential 'imbalance' and possibly affecting the separability of M and Z within those regions for a randomly selected galaxy on the M-Z plane, highlighted in green.

on the generation of the ζ statistics necessary for ROBUST to function correctly. While the bottom panels of Figure 6.16 indicate that this is indeed the case, what is puzzling is the bottom panels of Figure 6.17, where the associated scatter plots and histograms of the (ζ, μ) distribution show a relatively uniform distribution for that bright limit of 10. This suggests one of two possibilities: firstly, that ROBUST is being affected by the value of ΔZ being applied or secondly, that the mere presence of the bright limit is introducing an unnecessary bias that, while for brighter limits of 8 and larger was one that ROBUST was able to work with without a problem, is affecting ROBUST's ability to work effectively.

In order to determine which is indeed the case, having already experimented with various bright limits, we will now proceed to alter the values of ΔZ applied during during S_1 , S_2 region definition while continuing to utilise an arbitrary true bright limit for the PSCz, $m_b=8$ and note any change in the values of β recovered by ROBUST. The results of altering ΔZ on computed values of β are presented in Figure 6.19, with a summary of estimated values of β given in Table 6.6.



FIGURE 6.19: Method 2a) results with varied ΔZ applied. Left-hand panels illustrate the ρ plots for all 50 mocks, right-hand panels depict the histograms of optimal value of β with best-fit Gaussian (red line). LF is of the form N[-21,1], with m_f , m_b and ΔZ used noted on each plot. Entirety of the PSCz is utilised for this analysis.

$\Delta \mathbf{Z}$	1	5	10	20
Value of β	0.56 ± 0.06	0.56 ± 0.07	0.57 ± 0.06	0.57 ± 0.06

TABLE 6.6: Summary of estimated values of β detailing the effect of varying applied ΔZ on computed β using ROBUST during Mock Method 2a) implementation. Entirety of PSCz is utilised for this analysis with m_f and m_b fixed at 14.3 and 8 respectively.

The results depicted in Figure 6.19 indicate something rather interesting. Even when a relatively small value of ΔZ is applied such as 1, the mean estimates of β computed continue to be consistent with the assumed true value of β even in the presence of a fixed bright limit. Irrespective of the value of ΔZ applied, both the recovered values of β and its confidence interval remain tightly constrained over our 50 mock trials. While on the one hand this would appear to indicate that 'throwing away' considerable amounts of the survey by applying a small ΔZ is not a major issue as β continues to be recovered effectively, it further delineates the necessity for our underlying assumption; that distance moduli μ and absolute magnitude be independent of each other, that the correct magnitude limits be maintained otherwise under/oversampling will introduce bias into our calculations for ρ . This furthermore reinforces our suspicion that it is indeed the mere presence of a bright limit being arbitrarily applied to the data that is causing the bias to appear in the results returned by ROBUST more so than the value of ΔZ , lending further credence to our assumption that the PSCz is sufficiently modelled by a faint limit alone.

For the sake of completion we will now proceed to repeat all of the above with mock generated Schechter magnitudes assigned to the PSCz galaxies, and see if the introduction of a bright limit in that scenario will continue to manifest the same sort of biases we have observed so far on the values of β recovered by ROBUST.

6.5 Applying Mock Method 2b) - 'Faint+Bright Limit Schechter' Mock Magnitudes

As with Method 2a) since our T_c methodology is not designed to easily identify a bright limit in our data, we will continue to arbitrarily apply bright limit values of our own to our Schechter mock magnitudes and observe what changes occur in the results returned by ROBUST, starting with an initial arbitrary applied limit of $m_b=7$. The results of this are presented in Figure 6.20, alongside scatter and histogram plots of the (ζ, μ) distribution and ζ values of the galaxies respectively in Figures 6.21 and 6.22 to analyse any deviations from uniformity as per usual.



FIGURE 6.20: Initial results of Method 2b). Upper panel depicts the ρ plots for all 50 mocks generated, lower panel depicts the zero-intercept of each mock plot as a function of trial value of β plotted as a histogram with best-fit Gaussian (red line) modelled to the data. Optimal value of β returned for Method 2a) is $\beta = 0.53 \pm 0.01$, as determined from the Gaussian model. Luminosity function used during data generation is Schechter of the form S[-21,1.09], with m_f and m_b given as 14.3 and 7 respectively. Bright limit was applied retroactively during S_1 , S_2 set generation.



FIGURE 6.21: Scatter plot of ζ vs. μ for a faint limit of 14.3 and an assumed bright limit of 7 for mock generated Schechter magnitudes.



FIGURE 6.22: Histogram of ζ scatter points generated from a fixed faint limit of 14.3, an assumed bright limit of 7 and $\Delta Z=1$, binned into 10 intervals on [0,1]. The thick red line denotes our expected number of galaxies per histogram bin for the distribution to be uniform.

Bright Limit	0	7	8	10
Value of β	0.55 ± 0.01	0.53 ± 0.01	0.49 ± 0.03	N/A

TABLE 6.7: Summary of estimated values of β detailing the effect of varying applied apparent bright magnitude limit on computed β using ROBUST during Mock Method 2b) implementation. Entirety of PSCz is utilised for this analysis with $\Delta Z=5$.

As before, ROBUST continues to successfully recover a mean estimate of β consistent with β_{true} , albeit at a slightly smaller value of $\beta = 0.53 \pm 0.01$, but there is something very interesting to note in Figures 6.21 and 6.22 in particular. The introduction of that bright limit of 7 has introduced a very noticeable undersampling for low values of ζ followed by a significant oversampling observed at consequent values to attempt to compensate for it, resulting in a very stark deviation from uniformity for those values and yet despite this, ROBUST continues to operate effectively. This once again suggests a couple of things: firstly that our Schechter function is far more sensitive to the use of a bright limit than its Gaussian counterpart and secondly, that our arbitrary limit used in this instance is too bright, even though ROBUST manages to function despite its interference (possibly explaining the slightly lower value returned here). To determine whether the latter is indeed the case we will proceed to experiment with varying the bright limit applied to the data and note any change in the values of β returned by ROBUST. As before with its Gaussian counterpart we will also monitor the behaviour of the relevant (ζ, μ) scatter plots and the histograms of ζ value distribution to see if we continue to observe the same patterns we saw before. The results of varying our presumed applied bright limit are presented in Figure 6.23, the tracks of (ζ, μ) scatter behaviour and uniform histograms in Figure 6.24 and a summary of the estimated values of β provided in Table 10.7. It should be noted that ΔZ was fixed at a value of 5 for this analysis.

Once again the results returned by ROBUST are rather striking. While ROBUST continues to successfully return mean estimates of β consistent with the true value of β for all arbitrary bright limits brighter than 7, as we encroach beyond that limit ROBUST continues to return poorer and poorer results, with an $m_b=10$ failing to return any value at all as the mock plots of $\rho(\zeta, \beta)$ are all consistently positive and exhibit no zero-intercept whatsoever, as can be seen in the bottom panels of Figure 6.23. The almost 'flat' nature of the S-curves exhibited in the bottom panels of the figure make it difficult to accurately determine where the approximate zero-intercept would be (at an also admittedly very negative value of β which, as we have established in earlier chapters, is not possible). What continues to be puzzling however is that despite our previous supposition that the Schechter function is sensitive to the presence of a bright



FIGURE 6.23: Method 2b) results with varied bright limits applied. Left-hand panels illustrate the ρ plots for all 50 mocks, right-hand panels depict the histograms of optimal value of β with best-fit Gaussian (red line). LF is of the form S[-21,1.09], with m_f and m_b noted on each plot. Entirety of the PSCz is utilised for this analysis with $\Delta Z=5$.



FIGURE 6.24: Monitoring of ζ behaviour with varied Schechter bright limit. Left-hand panels illustrate the (ζ, μ) scatter plot for all galaxies used, right-hand panels depict the histograms of ζ value distribution over [0,1]. The thick red line denotes our expected number of galaxies per histogram bin for a uniform distribution.

$\Delta \mathbf{Z}$	1	5	10	20
Value of β	0.54 ± 0.01	0.54 ± 0.01	0.54 ± 0.02	0.55 ± 0.01

TABLE 6.8: Summary of estimated values of β detailing the effect of varying applied ΔZ on computed β using ROBUST during Mock Method 2b) implementation. Entirety of PSCz is utilised for this analysis with m_f and m_b fixed at 14.3 and 6 respectively.

limit, and which is readily apparent in Figure 6.24 as we continue to see the same under and oversampling of ζ points causing significant deviations from uniformity, ROBUST still manages to function effectively for limits brighter than 7.

It should be noted that even for the arbitrary bright limit of $m_b=7$, unlike its Gaussian counterpart, the deviation from uniformity presented in the histogram of ζ value distribution does indeed appear to be significant enough to introduce a slight biasing in the results returned by ROBUST, depicted by the fact that the corresponding $\rho(\zeta,\beta)$ plots are shifted upwards very slightly, consequently altering the zero-intercepts computed for β_{opt} . While this in itself does not completely negate the viability of $m_b=7$ as a bright limit for the PSCz it does enforce the idea that the inclusion of a bright limit is likely unnecessary especially when a Schechter function is in use, as ROBUST has shown in previous sections that it can function perfectly well without its inclusion. It also further reinforces the paramount importance of applying the correct apparent magnitude limits to our survey if we are to expect ROBUST to function correctly. If again for the sake of argument we continue to apply a bright limit of some kind, say a slightly reduced limit of $m_b=6$, we can now proceed to experiment with varying our value of applied ΔZ as before, effectively varying how much of the survey is included in the generation of the S_1 and S_2 sets and consequent ζ statistic computation. Much like in the previous section we will look to see whether the value of ΔZ applied has any noticeable effect on the value of β returned by ROBUST or not like its Gaussian counterpart. We can also monitor whether the biasing introduced by using this arbitrary limit is removed. The results of varying ΔZ with a fixed applied bright limit of 6 are presented in Figure 6.25, with a summary of estimated values of β in Table 6.8 as before.

Much like what was seen when we varied ΔZ with our Gaussian luminosity function, there continues to be no noticeable difference in the values of β returned by ROBUST, once again suggesting that even when a considerable amount of the survey is not being used in the definition of our S_1 and S_2 sets (small values of ΔZ) ROBUST is still capable of running effectively. Similarly even when overly large values of ΔZ are applied, rendering the shape of S_1 and S_2 trapezoidal in nature, this continues to have no effect on the functionality of ROBUST, which is promising behaviour. Additionally having



FIGURE 6.25: Method 2b) results with varied ΔZ applied. Left-hand panels illustrate the ρ plots for all 50 mocks, right-hand panels depict the histograms of optimal value of β with best-fit Gaussian (red line). LF is of the form S[-21,1.09], with m_f , m_b and ΔZ used noted on each plot. Entirety of the PSCz is utilised for this analysis.

brightened our bright limit to $m_b=6$ has completely removed the biasing that was exhibited in ROBUST prior to this (this in spite of deviations from uniformity in our (ζ, μ) distributions that have been observed already), further lending credence to our suspicion that the inclusion of a bright limit to our data only serves to add unnecessary bias to the results returned by ROBUST, while also serving as a reminder that it is very sensitive to the use of the **correct** faint and bright limits at all times.

In summary, throughout all of our mock realisations and experimentations with different luminosity functions and their associated parameters, we have established that ROBUST is more than capable of recovering a value of β from the PSCz, and recovering it well the more galaxies are available for it to analyse. We have also determined that even if broad luminosity functions are used this will have little effect on the functioning of ROBUST, provided that the survey is suitably large enough to account for any bias that might be introduced with the use of such a poorly defined function. Furthermore we have highlighted the paramount importance of utilising the correct magnitude limits for a survey, as any error in these limits will introduce significant bias. We have also determined that at least as far as the PSCz itself is concerned, it is more than sufficiently modelled with just the inclusion of a faint apparent magnitude limit when working with its luminosity information, though the inclusion of a bright limit does not necessarily impact the functioning of ROBUST, provided that it does not encroach too much onto the distribution of points on a typical M-Z plot. In particular should we choose to model a Schechter function to any galaxy surveys moving forward we need to be especially careful of the bright limit we choose to apply as the function has proven itself to be quite sensitive to the addition of such a limit, while conversely ROBUST has demonstrated that, when the correct limits are in play, it is more than capable of operating effectively irrespective of the value of ΔZ utilised, which will be of great use moving forward.

With all of this information in hand, we will now proceed to work directly with the PSCz and all available luminosity information for it, and apply ROBUST to determine a real world value of β_{opt} for the survey.

Chapter 7

Applying ROBUST to the IRAS PSCz Survey

Now that we are working directly with the PSCz, we can make use of known optical B-band magnitude information for the survey, assign those magnitudes to their correct galaxies then proceed to apply T_c to determine the faint (and bright, if necessary) limits of the survey and restrict it as required. We then proceed to apply ROBUST as normal over various values of β_{trial} on [0,1] to determine β_{opt} , as presented in summary format in Figure 7.1.

7.1 Making use of MC Simulations and Bootstrap Resampling

Having established our basic methodology from previous chapters and having now made the move to utilise the real world PSCz survey with ROBUST to compute an optimal value for β , we are no longer in a position where we can constrain our errors, or confidence intervals, on β by repeating our analysis multiple times and modelling a Gaussian to the values we obtain as we have been doing up to this point. However as discussed in Section 4.3.1, we can make use of the uniform distribution of our statistic ζ on the interval [0,1] to infer a confidence interval on β in the following two ways:

1. Monte Carlo simulations: making use of the random function in Matlab, we can generate a new uniform distribution of random values on the interval [0,1] and assign these numbers as ζ values to our galaxies. Due to our statistic ζ and distance moduli μ being independent of each other by their very construction, the use of a newly generated distribution of numbers should have no effect on the



FIGURE 7.1: Flowchart illustrating typical ROBUST methodology using real world data and magnitude-position galaxy matching. Note that m_B indicates B-band apparent magnitude.

values of $\rho(\zeta, \beta)$ calculated afterwards. Repeating this multiple times will give us a set of ρ values we can then use to construct a CDF from which our confidence intervals on ρ (and consequently our optimal value of β) can be determined.

2. Bootstrap resampling: making use of the randperm function in Matlab, we instead reorder the ζ values we already have computed for the PSCz galaxies and reassign them to different galaxies. Much like with Monte Carlo simulations this rearranging of values should have no effect on the values of ρ computed afterwards due to the inherent independence of ζ and μ . Repeating this rearranging over several trials will once again provide us with a set of ρ values from which a CDF (and consequently our confidence intervals) can be constructed.

Before we proceed with applying either of these techniques to the PSCz, it makes sense to determine the validity of both by applying them in a mock setting. Put another way, we already know from our mock runs that our confidence intervals on β (whether we choose to apply a Gaussian or Schechter LF) should be on the order of ~ ±0.05 or so for a Gaussian and ~ ±0.02 for a Schechter. Consequently if we were to make use of some of our mock results from previous sections and apply MC simulations and bootstrap resampling to them, if applied correctly, we should obtain similar confidence intervals. To that effect, bearing in mind the PSCz being well defined with the use of a faint limit alone, we will now make use of the results presented both in Figures 6.3 and 6.8, to determine the confidence intervals on our values of β for both a Gaussian and Schechter LF respectively. The computed CDF plots for ρ using both methods for both a mock Gaussian and Schechter LF are presented in Figures 7.2 and 7.3 respectively, while the extrapolated confidence intervals on β from the aforementioned figures are presented in Figures 7.4 and 7.5. A summary of the extrapolated confidence intervals on β for both applications of mock LFs using both methods is given in Table 7.1.

The results presented across all the Figures are very telling. Irrespective of the kind of method used to determine the confidence intervals on β , both of them are successful in returning the kind of intervals that are in excellent agreement with those we observed in our mock scenarios with both a Gaussian and Schechter LF; while any difference being determined between either method by Matlab is deemed small enough to not be significant. Additionally it is worth bringing up that there is no difference noted in computation time by Matlab whether we choose to apply MC simulations or bootstrap resampling, making both of them equally valid for use going forward as we prepare to apply ROBUST to real world data.

Having now verified both methods available to us to correctly determine the error bars on β , we can now proceed to prepare the PSCz for use with ROBUST.



FIGURE 7.2: Computed CDFs for $\rho(\zeta, \beta)$ utilising one mock trial with the symmetrical 68%, 95% and 99.7% confidence intervals marked off with blue, red and green dotted lines respectively. Top panel CDF is generated from applying MC simulations, bottom panel CDF is generated from applying bootstrap resampling. LF function utilised is Gaussian of the form N[-21,1] with no bright limit applied.



FIGURE 7.3: Computed CDFs for $\rho(\zeta, \beta)$ utilising one mock trial with the symmetrical 68%, 95% and 99.7% confidence intervals marked off with blue, red and green dotted lines respectively. Top panel CDF is generated from applying MC simulations, bottom panel CDF is generated from applying bootstrap resampling. LF function utilised is Schechter of the form S[-21,1.09] with no bright limit applied.



FIGURE 7.4: Extrapolated confidence intervals on β utilising values of $\rho(\zeta, \beta)$ generated from one mock trial with the symmetrical 68%, 95% and 99.7% confidence intervals marked off with blue, red and green dotted lines respectively. Top panel intervals are generated from applying MC simulations, bottom panel intervals are generated from applying bootstrap resampling. LF function utilised is Gaussian of the form N[-21,1] with no bright limit applied. Entirety of the PSCz was utilised for this analysis.



FIGURE 7.5: Extrapolated confidence intervals on β utilising values of $\rho(\zeta, \beta)$ generated from one mock trial with the symmetrical 68%, 95% and 99.7% confidence intervals marked off with blue, red and green dotted lines respectively. Top panel intervals are generated from applying MC simulations, bottom panel intervals are generated from applying bootstrap resampling. LF function utilised is Schechter of the form S[-21,1.09] with no bright limit applied. Entirety of the PSCz was utilised for this analysis.

	Value of β	1σ error	2σ error	3σ error
		Gaussian LF		
MC Simulations	0.57	± 0.03	$^{+0.06}_{-0.07}$	± 0.08
Bootstrap Resampling	0.57	± 0.03	$^{+0.06}_{-0.07}$	± 0.08
		Schechter LF		
MC Simulations	0.55	$^{+0.008}_{-0.007}$	$^{+0.016}_{-0.015}$	± 0.02
Bootstrap Resampling	0.55	$+0.008 \\ -0.007$	$^{+0.016}_{-0.015}$	± 0.02

TABLE 7.1: Summary table of extrapolated confidence intervals obtained on β using both MC simulations and bootstrap resampling for both an applied Gaussian and Schechter LF.

7.2 Preparing the PSCz for use with ROBUST

In preparing the PSCz catalogue for use with ROBUST, having magnitude data for the available galaxies is key. Unfortunately therein lies the first of many problems in attempting to adapt this survey with use for ROBUST, beginning with the lack of available luminosity information for the PSCz survey in any band as provided by Branchini. Further delving into available data for the original PSCz survey from which the Branchini catalogue was developed via the VizieR archives yielded lists of B-band magnitudes for the survey, as well as additional flux information in the $25\mu m$, $65\mu m$ and $100\mu m$ bands from which magnitude information could be extrapolated. However rather frustratingly another problem presented itself; namely that no marker information or cartesian positional information was made available between the Branchini PSCz survey and the original PSCz survey on VizieR for us to be able to match up which B-band magnitude or flux belonged to which galaxy. Galaxy names as assigned in the VizieR PSCz survey (which are also incomplete) were eliminated in the Branchini PSCz survey, essentially making them nameless and unmatchable, and all attempts to track down the original form of the survey as developed by Branchini himself to determine his galaxy identities were unsuccessful. However a simple Aitoff projection of the original survey galaxies and our current Branchini PSCz survey, as seen in Figure 7.6 makes it clear that several matches DO indeed exist ... the problem simply lies in figuring out which PSCz galaxy matches to which VizieR ones.



FIGURE 7.6: Aitoff projection of the Enzo Branchini PSCz data set - blue dots, against the original survey retrieved via VizieR - green crosses.

7.3 The problem of magnitude-galaxy matching

We decided upon a few different ways by which we would attempt to solve this matching issue. The first involved taking the position information of the original survey galaxies provided in galactic ascension and declination angles from VizieR, and converting them to their equivalent cartesian supergalactic coordinate counterparts, again using the necessary conversion matrices. We then attempt to match our values with the supergalactic cartesian positions already provided in the Branchini PSCz survey. Curiously no matches were returned despite experimenting with multiple tolerance levels for accuracy (allowing for the use of PSCz galaxies that are not exactly coincident with our converted positions but are instead relatively close by, say, within a few hundreds of kms⁻¹). Matches only began to appear when a tolerance error of 400kms⁻¹ or more was applied, implying that the matching Branchini galaxies were at least 400kms⁻¹ away from their original VizieR counterparts, which is in poor agreement with what is shown in the Aitoff projections.

The second technique we decided to implement was somewhat of a reverse of the first, wherein we instead converted the Branchini supergalactic positions into their equivalent galactic ascension and declination angles, and compared the values with the original VizieR survey to see if any matches appeared. With a tolerance level of 0.1° used, matches appeared for over 10000 of the Branchini galaxies, which is far more in line with the Aitoff projection and we began to populate a modified catalogue accordingly. With this success however came another curiosity.

Instances of multiple matches began to appear which is a natural, if somewhat curious, consequence of implementing this reverse technique. It is essentially implying that several of the VizieR PSCz galaxies (we observed instances of 3-4 multiple matches) lie along the same line of sight as the Branchini galaxy we are attempting to match up which, while plausible, still occurred with enough frequency to be a source of puzzlement. While we considered going with a 'first come first served' approach wherein we would take our first match in line and assign it to the galaxy we ultimately decided against pursuing this avenue due to the possible errors that would be introduced by assigning the wrong magnitude to the wrong galaxy, and consequently decided to abandon this attempt for the time being.

A final gambit that was considered for implementation was to scrap the use of the Branchini survey in its entirety, and to instead make use of the VizieR survey exclusively, which provides heliocentric redshifts and B-band magnitudes for all PSCz galaxies. In principle one would take the redshift information provided for these galaxies and extrapolate from that both a real-space position and a radial peculiar velocity by means of a distance convergence loop (this concept will be introduced in more detail in Section 8.4), and then use all of that in conjunction with the provided B-band magnitudes to first apply T_c and determine the true faint apparent magnitude limit of the B-band data and then proceed to run ROBUST and constrain a value for β . It was ultimately decided that we would not pursue this avenue due to the lack of essential peculiar velocity information necessary for ROBUST to function (in that regard the Branchini PSCz survey we have been using up to this point has been extremely useful and responsible for streamlining a lot of our velocity-related computations), and while it is possible in principle for us to recover that information ourselves by means of our aforementioned distance convergence loops, the exhaustive computation times required as well as additional mitigating time constraints in completing the rest of this work made it an unwise course to pursue at this time.

While we were unsuccessful in our attempts to constrain a real-world value of β for the PSCz due to missing position marker information, Chapter 6 has still served as a strong foundation for the usefulness of ROBUST as a whole and its versatility in recovering β irrespective of the kind of luminosity function applied to a survey or the number of galaxies that indeed exist within that survey.

Thankfully there do exist other extensive redshift velocity surveys with which we can apply ROBUST and attempt to constrain a value of β such as the 2MRS and SDSS. We will now dedicate the next chapter to our efforts in applying ROBUST to the 2MRS in particular and calculating real-world values of β , exploring any anomalies that we discover along the way.

Chapter 8

Probing the Peculiar Velocity Field of 2MASS with ROBUST

8.1 An aside: Thinking in cubes

Before exploring the intricacies of working with real data using ROBUST, it is worth considering what effect working with such large data sets will have on computational speed.

Irrespective of the considerable size and depth of the 2MASS/2MRS survey which we covered in Sections 2.10.2 and 2.10.3, upcoming galaxy surveys such as WALLABY, the SKA and LSST projects are expected to produce terabytes, if not petabytes, of information daily that can span up to billions of objects at a time. When one considers the typical weighting matrices necessary to calculate the predicted peculiar velocities of mock objects embedded into a given redshift survey, they rely on determining the separation between the object we are considering and ALL the galaxies in the survey for each individual object; despite the fact that at significantly large distances the weighting contributed by these more distant galaxies towards the peculiar velocity at that point is statistically negligible and not worth considering, consequently lending themselves towards computationally exhaustive and time consuming runs. If we can find a way to determine the velocities of the mock objects using only the closest surrounding survey objects that would be more statistically significant, then we will effectively reduce the number of computations that need to be done per object and decrease the script runtimes by a drastic amount.


FIGURE 8.1: Embedding a blank mesh grid into an existing galaxy survey to begin generating a 3D velocity field map. Basic map grid points are indicated in green, while a random selection of galaxies within $200 \rm km s^{-1}$ of the survey are indicated in blue.

As an alternative solution consider the following: imagine a large region of space that we have populated with galaxies from a redshift survey and for each of which we know their coordinates in space and the individual components of their peculiar velocities.

Now imagine placing a cube whose width is comparable to that of the redshift survey inside the survey, and that this large cube contains thousands of points spaced at regular intervals as shown in a simplified manner in Figure 8.1. This effectively creates a massive cube that contains thousands of smaller cubelets. We know nothing about these points beyond their position in space, however using previously established methods and weighting schemes we can calculate the predicted peculiar velocity of the vertices of each cubelet using the predicted peculiar velocities of the survey galaxies around them. We would now effectively have a 3D velocity grid map generated from these galaxies that we can use to efficiently determine the interpolated peculiar velocity of any object we wish to embed in it. Since this map is being generated with the same weighting schemes that were used in ROBUST and the χ^2 scripts, this step would admittedly take a long time to do depending on the size and resolution (grid spacing) of the map, but the benefit that is to be gained from this is that once this map has been generated, <u>it never has to be done again</u>, making the one-off long map generation runtime worth it in the long run.

Now we have our map. Let us now imagine that we take an object that we know nothing about beyond its observed redshift and direction on the sky. We want to calculate what its peculiar velocity is. To do this, based on its position we can place it on our map and see that it will fall within one of the potentially thousands of mini-cubelets that inhabit this map, such as shown in Figure 8.2.

The beauty of this idea is that for any trial value of β we already know the peculiar velocity components of each vertex of this mini-cubelet that the object resides in, and that these components have been calculated to very high accuracy from a massive galaxy survey already. So now we have 8 points surrounding our object that are both known to high accuracy and are also (relatively speaking) extremely close to the object meaning that when we apply our weighting schemes, the contribution of these 8 points to the peculiar velocity of the object will be more statistically significant than if we were to use every galaxy in a large survey as we were doing before. The potential here to reduce the number of computations being done and overall running times is not to be overlooked.



FIGURE 8.2: A selection of galaxies (marked in blue) that, due to their positions, reside within a certain mini-cubelet, marked in solid green.

8.2 Carrick et. al and the 2M++ velocity field

While a significant amount of development time during this work was dedicated towards developing a grid interpolation scheme that would streamline the application of RO-BUST to any survey (the PSCz and 2MASS surveys in particular), and generate the required velocity grid maps as mentioned in the previous section, many problems and debug issues manifested that began to bring into question the feasibility of using this technique for the remainder of the project. In fact such were the number of problems and coding issues that developing the velocity grid maps was ultimately shelved in favour of more standard methods and for the sake of completing this work within the allotted time.

In that regard, the work of Carrick et. al [25] as published in 2015 during our own development efforts and the creation of the 2M++ velocity survey and its associated cubic grid granted us an unexpectedly pleasant reprieve. The 2M++ survey is an amalgam of 2MASS, with additions from the SDSS and 6dFS surveys. It is a subset of the 2MRS which exhibits superior sampling to PSCz and far greater depth than its 2MASS equivalent [25]. K band magnitudes from 2MASS are supplemented with data from SDSS-D7 and 6dFS-D3 such that $K_s < 11.5$ in regions not covered by SDSS and

6dFS and $K_s < 12.5$ elsewhere (i.e. drawn from 2MRS alone). Galaxy redshift cloning techniques were used to populate empty regions such as the Zone of Avoidance due to extinction levels [25]. The luminosity function used to model this catalogue is a Schechter given by:

$$\Phi(M) = 0.4 \log(10) n^* 10^{0.4(1+\alpha)(M^*-M)} \exp\left(-10^{0.4(M^*-M)}\right)$$
(8.1)

Magnitude incompleteness was accounted for using a luminosity-density weighting scheme where the weight assigned to each galaxy's luminosity is based on the fraction of the total luminosity expected, given the magnitude limit of the survey, to the luminosity one expects to observe at a given distance:

$$w^{L}(r) = \frac{L_{average}}{L_{observed}(r)} = \frac{\int_{L_{min}}^{\infty} L\Phi(L)dL}{\int_{4\pi r^{2} f_{min}}^{\infty} L\Phi(L)dL}.$$
(8.2)

The value of β obtained by Carrick (**0.431±0.021** as calculated from using the Tully-Fisher relation and χ^2 -minimisation with Type Ia Sne) is reliant on the magnitude incompleteness modelling used here [25]. In applying the T_c completeness statistic mentioned in Section 4.2, we can do two key things: firstly confirm (or improve upon) the magnitude limits of the survey to ensure ROBUST constrains β as efficiently as possible across the various magnitude passbands, and also use the results returned by T_c to check the validity of the underlying assumption of ROBUST that the modelled Schechter luminosity function used here is truly independent of the spatial position of the galaxies. In validating this assumption we lay out the key groundwork necessary for ROBUST to operate effectively once it is applied to the 2M++. The survey in itself is limited to a distance of $125h^{-1}$ Mpc to ensure number and magnitude completeness as much as possible, which is comparable to our limiting of the PSCz survey to 15000kms⁻¹ to ensure uniform sampling.

From this, an iterative procedure was used where the now weighted galaxies have their distances reconstructed from their observed redshifts, after which said galaxies are populated onto a cubic grid (refer to [25] and the references therein for full details of the iterative procedure). The mean mass density contrast function of the galaxies is then calculated and used to determine the line of sight peculiar velocities of the galaxies as per the equation:

$$\mathbf{v}(\mathbf{r}) = \frac{\beta^*}{4\pi} \int d^3 \mathbf{r}' \delta_g^*(\mathbf{r}') \frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3}; \tag{8.3}$$

	β^*	$\chi^2/(D.O.F)$
Forward Likelihood (LW)		
A1	$0.440{\pm}0.023$	-
SFI++ Galaxy Groups	$0.429{\pm}0.022$	-
SFI++ Field Galaxies	$0.423 {\pm} 0.045$	-
All	$0.431 {\pm} 0.021$	-
Forward Likelihood (NW)	$0.439 {\pm} 0.020$	-
Inverse VELMOD (LW)	$0.387 {\pm} 0.048$	-
χ^2 (LW)	$0.444{\pm}0.026$	2194/2899
χ^2 (NW)	$0.442{\pm}0.028$	2200/2899

TABLE 8.1: Summary of best fit values of β^* using different weighting schemes, methods of analysis and peculiar velocity datasets. Results obtained using luminosity weighting are indicated by (LW), whereas those obtained using number weighting are indicated by (NW). Unless explicitly indicated, all datasets were used for the method mentioned with the exception of Inverse VELMOD which used all individual galaxies from SFI++ [25]

where β is scaled from 0 to 1 across multiple iterations. The 'true' β value for the survey was obtained/constrained using multiple techniques as seen in Table 8.1. Consult Carrick et al. and the references therein for more information on the relevant techniques used.

We can use the populated 2M++ cubic grid developed by Carrick to interpolate the velocities of the 2MASS galaxies available to us for use in ROBUST and χ^2 -minimisation and also compare the values of β that we constrain across the different magnitude passbands with their published values.

8.3 Preparing 2MASS for use with ROBUST in conjunction with 2M++

Preparing the 2MRS survey for use with ROBUST in conjunction with the Carrick velocity grid would require some thought given the limitations of the hardware available to us, not to mention the sheer size of the velocity grid itself. The grid is a cube 256Mpc wide, consisting of 'cubelets' made of grid points 1Mpc apart, where each grid point contains the interpolated peculiar velocity cartesian components that make up the velocity grid as a whole. As such the total number of grid points available in the file comes to 257³, or approximately 17 million, which is impossible for Matlab to hold in internal memory all at once without crashing. To combat this problem the velocity grid was divided into slices 16Mpc wide, reducing the amount of data being held in Matlab from nearly 17 million entries down to slightly over 1.1 million, which is at the

uppermost limit of what the program can hold at once. As a result when we come to run our interpolation schemes, we will instead load in the velocity grid one slice at a time, populate the slice with all the 2MRS galaxies that would reside within it, run the interpolation using the available grid points, save out the relevant variables, clear out the current slice file and galaxies and load in the next one and so on.

Reference frames was the next issue. The Carrick grid is in the Local Group (LG) frame and uses its own coordinate conversion equations for the individual grid points given by the following:

$$X = \frac{(i - 128) \times 400}{256}$$
$$Y = \frac{(j - 128) \times 400}{256}$$
$$Z = \frac{(k - 128) \times 400}{256},$$

where i, j and k are given in galactic coordinates as cited in the work of Carrick [25] and can be accessed in further detail at http://cosmicflows.iap.fr/download.html. We need to ensure the 2MRS galaxies (currently in the galactic frame) are converted using these equations to ensure the galaxies are embedded properly into the velocity grid and the interpolation runs correctly.

8.4 An aside: Distance convergence loops

Another key issue that needs to be considered while preparing a survey of this size (or indeed any survey for that matter) for use with ROBUST, is the distinction between an object's real-space position and redshift. Recalling the Hubble Law given as:

$$cz = H_0 d + v_{pec},$$

the difference between an objects real-space position (as given by H_0d) and its redshift (cz) is its peculiar velocity. The only instance in which these two variables would be the same is when the object in question is stationary in space. While for the purpose of determining the magnitude limits of a survey (see Section 8.5 for our determination of limits for 2MRS) it can be deemed acceptable to assume that both the real-space position and redshift of a galaxy are coincident (particularly if you only have the redshift on hand for calculations) simply because the error introduced with respect to magnitude calculations is minimal in the larger scale of things, the same cannot be said when it comes to actual velocity interpolations. Consider the Carrick grid whose cubelets are

1Mpc wide, and the average peculiar velocity of a galaxy as per the PSCz catalogue which can be in the range of $\pm 400{\text{-}}600 \text{kms}^{-1}$ or possibly larger. The larger the peculiar velocity of a galaxy the greater its shift from its observed redshift meaning that it potentially resides in another cubelet than the one dictated by its observed redshift alone. This fundamentally alters the velocity values of the surrounding grid points we are using to interpolate with and introduces unnecessary errors that can be avoided. As such, distance convergence loops need to be introduced into our calculations to ensure as high a level of accuracy as possible, as demonstrated in Figure 8.3, where we initially assume that the real-space position and observed redshift are indeed coincident, identify the cubelet in which the galaxy resides, interpolate its peculiar velocity accordingly, then apply the Hubble Law to redetermine the real-space position of the galaxy, as observed redshifts are constant. Over multiple iterations the value for real-space distance will converge to a point, as will the 'true' peculiar velocity of the galaxy, which can then be used in ROBUST with confidence in their accuracy.

Three considerations for this technique need to be taken on board while applying these loops: what convergence limit should be applied, what is the maximum number of iterations that we should allow before we discount the galaxy as failing to converge, and how do we prevent errors occurring from galaxies shifting into another slice file or falling off the grid entirely?

The first point is important as it defines how long on average a convergence loop will run before the limit is achieved and has a knock-on effect on overall computation times for the entire survey. On the other hand we need a limit that is small enough to ensure as high a level of accuracy as possible. In analysing the general run-time of an individual loop we decided that a convergence limit of 10kms^{-1} was sufficient, as we observed that most galaxies converged within a maximum of 10-15 loops at worst with no considerable effect on overall run times, with those needing more iterations being few and relatively far between.

The second point arises from initial test runs of the convergence loop sequences, wherein a handful of galaxies failed to converge at all, and instead demonstrated behaviour wherein they kept bouncing back and forth between the same two values (effectively bouncing between the two nearest grid points in the Carrick velocity file) without ever converging, causing an infinite loop that Matlab would get stuck in. While we remain unsure as to what was causing these infinite bounces to occur, we suspect that the problem may reside in the cartesian velocity values of the grid points themselves for that particular cubelet being similar, such that when a new real-space position was calculated the consequent peculiar velocity that was assigned in the next loop would simply return it back to its initial position and so on and so on. It has been theorised



FIGURE 8.3: Flowchart for a typical distance convergence loop.

that the large-scale bulk flow of our Local Universe towards the constellation Virgo (also denoted as the Virgo-centric infall) may be what is preventing these particular galaxies from converging as the infall continues to 'pull' them along the flow, while the interpolation scheme attempts to have them 'bounce back' to where they should be, resulting in the infinite 'non-convergent' loops that we observe. Given that the similarity of velocity values on the Carrick grid is not something that we have active control over (nor over the Virgo-centric infall) it was instead decided to monitor the number of iterations the convergence loop was going through per galaxy. If more than 50 iterations were required without a convergence having been achieved we flagged the galaxy as 'non-convergent' and forcibly broke the loop to move onto the next galaxy. Once all galaxies were analysed we identified all those that were flagged and removed them from the survey. In multiple test runs the number of flagged galaxies came to less than 100 in a survey containing over 40000 galaxies, so their exclusion was deemed to have no measurable negative impact on the running of ROBUST as a whole.

As for the third and final point, making allowances for a galaxy to possibly shift from one slice file to another as it bounces towards a distance convergence is important to prevent Matlab from crashing out due to non-existent grid points that have not been loaded in. This was easily implemented by designing a tracker that monitors the realspace x-axis position of the galaxy in the loop relative to slice file, each of which are 16Mpc wide and increase in value in increments of 16 from 0 to 256 accordingly along the x-axis. Unfortunately should a galaxy (presumably already near one of the edges of the Carrick grid) fall off the grid entirely along any axis it was decided that such a galaxy should also be flagged as 'non-convergent' and discounted alongside the others that failed to reach convergence. In monitoring the number of galaxies that fell off the grid in this way they reached an average of 50 per test run, making their exclusion also have no harmful impact on our calculations.

8.5 Identifying the magnitude limits of the 2MRS survey

As discussed previously in Section 4.2, exploiting the characteristics of the ζ variable from which our ρ estimator is constructed in ROBUST gives us the means to test and determine the completeness of any given survey up to and including a certain magnitude limit. The completeness statistic T_c is consequently of paramount importance if we wish ROBUST to recover β effectively, even more so since we are now working with real data and no longer generating mock magnitude data for which an arbitrary faint (and/or bright) limit has already been implemented, rendering the need for such completeness statistics moot.

A cursory glance at plots of absolute magnitude vs. distance moduli in the three infrared bands of the 2MRS galaxies as shown in Figure 8.4 indicate clearly that a faint magnitude limit does exist for the survey across all three bands, so we proceed to apply our T_c statistic over a range of faint magnitude limit values in an effort to determine which one is the true limit. Recall that T_c is defined via:

$$T_{c} = \frac{\sum_{i=1}^{N_{gal}} \left(\hat{\zeta}_{i} - \frac{1}{2}\right)}{\left(\sum_{i=1}^{N_{gal}} V_{i}\right)^{\frac{1}{2}}},$$
(8.4)



FIGURE 8.4: M-Z plots for the 2MRS galaxies in the K (top panel), H (middle panel) and J infrared bands respectively (bottom panel).



FIGURE 8.5: T_c plots identifying the faint magnitude limits in the K, H and J infrared bands for the 2MRS survey. Magnitude limits at the 3σ confidence level are marked off via the dashed orange lines for all bands, given as $K_{lim}=11.55$, $H_{lim}=11.86$ and $J_{lim}=12.55$ respectively.

where V_i is the variance defined as:

$$V_i = \frac{1}{12} \cdot \frac{n_i - 1}{n_i + 1} \tag{8.5}$$

and n_i is the sum of the elements contained in sets S_1 and S_2 as defined previously in Figure 4.1 [98].

Choosing to run T_c across a range of apparent magnitude values from [5,15], the results are presented in Figure 8.5 which show a relatively clear fall-off at the 3σ level across all three bands. The returned magnitude limits, marked off via the dashed orange lines are given as $K_{lim}=11.55$, $H_{lim}=11.86$ and $J_{lim}=12.55$. The limit value returned for the K band is slightly fainter than the limit published by Carrick of 11.42 as discussed in Section 8.2, requiring us to ascertain whether this is a valid limit for ROBUST to use. Working off of the (ζ, μ) scatter plots we presented throughout Chapter 6, we will analyse the uniformity of ζ value distribution at the returned limits while making use of two additional statistical tools at our disposal to verify the validity of these values: namely χ^2 -minimisation and Kolmogorov-Smirnov testing (hereafter denoted as KS testing).



FIGURE 8.6: Monitoring of ζ behaviour at the K (top panels), H (middle panels) and J limits returned by T_c (bottom panels). Left-hand panels illustrate the (ζ, μ) scatter plot for all galaxies used, right-hand panels depict the histograms of ζ value distribution over [0,1]. The thick red line denotes our expected number of galaxies per histogram bin for a uniform distribution.



FIGURE 8.7: χ^2 track of ζ value distribution vs. expected number of values per histogram bin across varying values of apparent magnitude for the K (solid blue), H (solid red) and J (solid green) infrared bands of the 2MRS respectively. KS cut-off points after which the null hypothesis of uniformity is rejected at the 95% confidence level for all bands is represented by the associated coloured dashed vertical line. KS limits are given as $K_{lim} = 11.56$, $H_{lim} = 11.89$ and $J_{lim} = 12.55$.

For the former, consider the ζ value histograms across the various bands in Figure 8.6 where the red line denotes our expected number of ζ values per histogram bin. If one were to apply χ^2 -minimisation we would expect to find that χ^2 remains minimised for all apparent magnitude values brighter than m_{lim} for which the histogram distribution remains uniform, as the difference between the histogram bin numbers and expected number of ζ values per histogram bin would be minimal. Conversely for all apparent magnitude values fainter than m_{lim} we would expect to see χ^2 start to increase considerably as undersampling and bias starts to appear in the distribution, therefore the value of m_{lim} for which this change in χ^2 occurs would serve as indication for what the true faint limit for that particular band should be.

For the latter, the KS test serves as a statistical tool by which one can determine whether two samples differ significantly from one another. In this instance we are looking to determine whether our ζ values are drawn from a uniform distribution, so we are looking to compare them with a standard continuous uniform distribution and take note of any statistically significant difference between the two. To that effect we can make use of the **kstest2** function in Matlab which utilises the two-sample Kolmogorov-Smirnov test to return a test decision for the null hypothesis that our two sample distributions (in this case our ζ values and a test uniform distribution) are drawn from the same source. Matlab returns a logical value of 1 if the null hypothesis is rejected with 95% confidence and 0 otherwise. Therefore for all apparent magnitude values that are brighter than m_{lim} we expect the KS test to return 0, and to return 1 for all values brighter than m_{lim} , the point at which the KS test flips over signifying the true faint limit for that particular band once again. The results of applying χ^2 -minimisation and the KS test for a range of apparent magnitude values within which the limits returned by T_c are a part are presented in Figure 8.7; the χ^2 track for all three bands being represented by solid coloured lines, and the flip point for the KS test in each band being represented by the associated coloured vertical dashed line.

It is interesting to note that while the KS test returns a 'flip point' in all three bands that is in excellent agreement with the faint limits returned by T_c , $K_{lim} = 11.56$, $H_{lim} = 11.89$ and $J_{lim} = 12.55$ respectively compared to $K_{lim}=11.55$, $H_{lim}=11.86$ and $J_{lim}=12.55$ as returned by T_c , the χ^2 track returns values that are slightly fainter (by approximately 0.2 magnitudes) before it starts to increase considerably. This would appear to suggest that the KS test is a stronger statistical tool than χ^2 as it is more sensitive to changes between the two distributions, while χ^2 requires more drastic differences between the two samples to manifest before it begins to alter.

Considering the excellent agreement with T_c returned by the two-sampled KS test however we will proceed with the assumption that these limits are indeed viable for ROBUST to use, so with our magnitude limits of the 2MRS now in hand, in addition to real-space positions and peculiar velocities for the majority of the galaxies in the survey, ROBUST can now be run, and an optimal value of β generated which should hopefully be in concurrence with the published Carrick value of 0.43 ± 0.021 . A summary flowchart of the methodology that will be applied going forward is presented in Figure 8.8.



FIGURE 8.8: Flowchart for the application of data slicing, distance convergence loops and ROBUST methodology for the 2MRS survey.



FIGURE 8.9: Plots of ρ for the 2MRS survey across the K, H and J infrared bands, shown in blue, red and green respectively.

8.6 Applying ROBUST

Having all what we need to run ROBUST and constrain its errors effectively either via Monte Carlo simulations or bootstrap resampling, we proceed to run the program for the 2MRS across all three bands for 1000 trial values of β . The results returned by ROBUST are presented in Figure 8.9, and are puzzling for several reasons.

Recalling from the works of Carrick et. al [25] that the luminosity function modelled to the 2M++ is Schechter in nature we should expect to see somewhat 'S-shaped' curves in the $\rho(\zeta,\beta)$ plots across all three wavebands yet this is clearly not the case, despite the somewhat monotonic increase in values over the interval of [0,1]. Secondly, while we see clear β intercepts at $\rho(\zeta,\beta) = 0$ in all three bands, these are at inordinately low values of β that are not consistent with the value of 0.43 ± 0.021 computed by Carrick, nor with the values calculated via other methods as discussed previously in Section 3.4. Additionally while the signal recovered by ROBUST is relatively strong and clear for these low values of β there is a noticeable increase in noise in the recovered signal as we continue to increase the value of β over [0,1], consequently making it considerably harder to ascertain whether there is any sort of β -intercept at the computed Carrick value.



FIGURE 8.10: Monitoring of (ζ, μ) distribution of the 2MRS galaxies in the K band for various trial values of β . The trial values of β used are noted on each panel.

When we consider the (ζ, μ) distribution of the 2MRS galaxies in, say, the K band across several values of β as presented in Figure 8.10, the source of the noise starts to becomes apparent. The range of distance moduli presented in each panel is inordinately large, implying that several of the 2MRS galaxies continue to move implausibly closer to us as the observer as we increase the value of β . This would appear to suggest that as β increases the β -dependent peculiar velocities of the 2MRS galaxies increase by a significant enough amount that we see the 'packed' distributions presented in the various panels, consequently disrupting uniformity and causing the increased level of noise seen in Figure 8.9.

While it is not immediately clear what might be causing the 2MRS galaxies to shift so significantly as we alter the value of β it must be noted that the entirety of the 2MRS survey was utilised for this analysis without any distance restrictions applied, giving ROBUST over 43000 galaxies to work with. While once again recalling from Section 8.2 that the 2M++ survey is limited to a distance of $125h^{-1}$ Mpc to ensure number and magnitude completeness as much as possible, this restriction may not be sufficient to ensure the reliability of the underlying reconstructed peculiar velocity field of the Carrick grid for ROBUST to operate effectively. It may be wise to consider applying distance restrictions similar to those that were applied to the IRAS PSCz and observe the results that ROBUST returns and whether there is any improvement (refer to Section 8.6.1).

Before we proceed to experiment with applying our aforementioned distance restrictions to the 2MRS, it is worth attempting to identify what the confidence intervals on our current $\rho(\zeta,\beta)$ plots are in all three wavebands, if at the very least to exclude at the 3σ confidence level values of β that are unlikely to be considered for the survey. To that end we will make use of both Monte Carlo simulations and bootstrap resampling to determine the one-sided confidence intervals on $\rho(\zeta,\beta)$, the results of which are presented in Figure 8.11.

As was observed in Section 7.1 and consequently reinforced here, the use of either Monte Carlo simulations or bootstrap resampling return confidence intervals that are in excellent agreement with each other, in this instance excluding all values of $\beta > 0.64$ as denoted in the Figure via the black dashed lines representing the confidence intervals in each band. This exclusion of higher values of β from consideration for the 2MRS is at the very least consistent with the results returned by VELMOD and other methods as discussed in Section 3.4 that favour lower values of β for redshift-velocity surveys, while further rejecting the results favoured by POTENT of $\beta=1$ that, as previously discussed, is not possible. Despite establishing an 'exclusion zone' on our value of β this still does leave a lot of noisy data within which the true value of β might reside, lending credence to our suspicion that perhaps the underlying reconstructed peculiar velocity field being used with the Carrick grid requires a distance restriction of some sort in order for ROBUST to operate more effectively.

8.6.1 Redshift Restricting the 2MRS

Recalling that we restricted the IRAS PSCz survey out to 15000kms^{-1} to ensure the reliability of the underlying reconstructed peculiar velocity field for interpolation purposes, we will proceed to restrict out the number of galaxies ROBUST utilises for its analysis in a similar manner: implementing a distance restriction of $500 \leq cz \leq 15000 \text{kms}^{-1}$. The lower bound of 500kms^{-1} was selected as a precaution to ensure that no potential galaxies with anomalously large peculiar velocities and small real-space distances or redshifts near the core of the survey make it into the analysis and introduce potential sources of error or bias. Applying this redshift restriction provides ROBUST with nearly 30400 galaxies to work with, the results of that analysis being presented in Figure 8.12.

With the redshift restricted subset of the 2MRS now in play, we do observe a notable reduction in the amount of noise for all values of $\beta < 0.6$. This lends credence to our suspicion that the underlying reconstructed peculiar velocity field of the Carrick grid has



FIGURE 8.11: One-sided confidence intervals calculated on β via bootstrap resampling (top panel) and Monte Carlo simulations (bottom panel) for the K, H and J bands depicted in blue, red and green respectively. The 3σ confidence intervals for all three bands are denoted via the black dashed lines, indicating all values of $\beta > 0.64$ are to be excluded from consideration for the 2MRS survey with 99.7% confidence.



FIGURE 8.12: Plots of ρ for the redshift restricted 2MRS survey across the K, H and J infrared bands, shown in blue, red and green respectively.

been made more reliable with this restriction applied such that we now see a consistent $\rho(\zeta,\beta)$ intercept for values of $\beta \sim 0.05$ in all three wavebands, though once again this is not consistent with the value published by Carrick. Additionally the general shape of the ρ plot has altered drastically with this restriction in place. In particular for all values of $\beta \leq 0.6$ we observe parabolic behaviour as opposed to the general 'S-shaped' curves we expect to see given the Schechter function that is known to be modelled to the 2M++, consequently denoting a deviation from our expected monotonic increase in values of ρ as well.

It should be noted that the reduction in noise in the signal recovered by ROBUST is mirrored in the (ζ, μ) distribution of the 2MRS galaxies as presented in Figure 8.13 for various values of β , where the scatter plots for β values less than 0.6 are more in line with the plots we observed in earlier chapters. In particular the 2MRS galaxies are not shifting around as drastically as they were before and are exhibiting a sensible range of distance moduli and general uniformity, allowing ROBUST to operate effectively. Conversely for β values larger than 0.6 we see a return of the 'packed' distributions we saw in the previous section as the galaxies start to shift more significantly to implausibly closer distances, and the resultant overly large β -dependent peculiar velocities of those



FIGURE 8.13: Monitoring of (ζ, μ) distribution of the redshift restricted 2MRS galaxies in the K band for various trial values of β . The trial values of β used are noted on each panel.

particular galaxies cause an increase in the amount of noise in the signal recovered by ROBUST as the uniformity of the distribution is disrupted.

Given the reduced noise exhibited in Figure 8.12 it is once again worth identifying the confidence (or exclusion) intervals on β and observing whether the values returned via bootstrap resampling or Monte Carlo simulations are consistent and exclude the noisier values of β . The confidence intervals computed for this restricted subset of the 2MRS are presented in Figure 8.14, and do indeed favour the exclusion of all values of $\beta > 0.6$ from consideration for the survey due to the increased noise in the recovered signal. This is once again consistent with results returned by VELMOD and other contemporary methods that favour smaller values of β for surveys while further reinforcing the rejection of β results of unity favoured by POTENT.

8.6.2 Restricting cz_{corr} instead of cz_{obs}

Having made the decision to restrict our observable redshifts in such a manner as to ensure the reliability of the underlying reconstructed peculiar velocity field when applying ROBUST and having seen a noticeable improvement in the noise reduction in the



FIGURE 8.14: One-sided confidence intervals calculated on β via bootstrap resampling (top panel) and Monte Carlo simulations (bottom panel) for the K, H and J bands depicted in blue, red and green respectively. The 3σ confidence intervals for all three bands are denoted via the black dashed lines, indicating all values of $\beta > 0.6$ are to be excluded from consideration for the redshift restricted 2MRS survey with 99.7% confidence.



FIGURE 8.15: Plots of ρ for the restricted 2MRS survey utilising corrected redshifts across the K, H and J infrared bands, shown in blue, red and green respectively.

 $\rho(\zeta,\beta)$ plots returned by ROBUST as a result; an interesting question arises. What would happen if instead of utilising the observed redshifts of the 2MRS galaxies and holding them as constant as we have been doing so far, we choose to rather hold their real-space distances fixed instead, and utilise the modified Hubble Law to correct their redshifts accordingly, and then restrict those corrected redshifts instead? Correcting the 2MRS redshifts in such a manner and then restricting them reduces the number of available galaxies for analysis substantially (from an initial survey size of over 43000 objects we are now left with a little over 29000), though as has been established in previous chapters this is still more than enough for ROBUST to operate effectively. The results of running ROBUST utilising corrected redshifts, cz_{corr} as opposed to observed redshifts, cz_{obs} is presented in Figure 8.15.

The results are striking. On the one hand we observe the typical monotonic increase that we expect to see given the logarithmic velocity model that we are using in the construction of $\rho(\zeta, \beta)$, and yet ROBUST fails to return any value for β_{opt} whatsoever, given the consistently negative behaviour of ρ across all three bands bar at the very beginning, which can barely be seen along the left y-axis of the Figure, suggesting a dubious β_{opt} of 0 across all three bands. Indeed the 'noise' exhibited by ρ is peculiar and not like the more 'Gaussian' noise exhibited in previous sections; suggesting that



FIGURE 8.16: Monitoring of (ζ, μ) distribution of the restricted 2MRS galaxies utilising corrected redshifts in the K band for various trial values of β . The trial values of β used are noted on each panel.

the use of corrected redshifts is more of a hindrance than a useful solution with said noise being consistent throughout.

This is further mirrored by the associated (ζ, μ) distributions of the 2MRS galaxies at various values of β_{trial} as presented in Figure 8.16, where for all trial values presented, much like when the entirety of the 2MRS survey was used initially without any redshift restrictions, we see a return of the 'packed' distributions and the extreme shifting of certain galaxies to implausibly low distance moduli. This would suggest that the use of corrected redshifts has had no useful bearing on the running of ROBUST but rather, has been more counter-productive than useful. This is further delineated by the fact that any attempt to construct confidence or exclusion intervals on our value of β would be meaningless considering the consistent negative behaviour displayed and the clear lack of any sensible zero-intercept as the returned intervals would automatically exclude all values of β on the interval [0,1]. To that end we will instead proceed to explore a couple of other avenues that might be used to rectify and/or justify ROBUST's behaviour thus far.

8.6.3 Experimenting with μ instead of u_{β}

While redshift restricting the 2MRS survey has somewhat improved the results returned by ROBUST for lower values of β we are still left with no clear reason as to why we are not seeing the monotonic increase in values of ρ that we expect, nor why we are still getting inordinately small values of $\beta \sim 0.05$ returned across all three wavebands. Perhaps we would be well served to reconsider the underlying theory on which ROBUST itself is reliant, in particular our use of β -dependent peculiar velocities interpolated onto the 2MRS galaxies and our projected logarithmic velocity model such that:

$$u_{\beta} = -5\log_{10}\left(1 - \frac{v_{\beta}}{cz}\right),\,$$

and

$$\beta = \beta^* \Leftrightarrow \rho(\zeta_\beta, u_\beta) = 0.$$

When one recalls one of the key properties of the ζ statistic from Section 4.1, namely that ζ and μ are statistically independent of each other, i.e. the distribution of ζ is independent of the spatial distribution of the galaxies for a given survey provided it is magnitude complete and also independent of luminosity function, it makes sense that one should be able to calculate ρ such that

$$\beta = \beta^* \Leftrightarrow \rho(\zeta_\beta, \mu_\beta) = 0,$$

considering their statistical independence, and that ROBUST should still be able to return a valid result given how ζ is constructed. To that end we will experiment with running ROBUST on the redshift restricted subset of the 2MRS, where it calculates the correlation coefficient $\rho(\zeta, \mu_{\beta})$ using the β -rescaled distance moduli of the galaxies as opposed to their β -dependent peculiar velocities, and observe what results ROBUST returns. The results of this alternate analysis are presented in Figure 8.17.

Offhand it becomes clear that ROBUST is now returning a strong signal of monotonically increasing values of ρ over the interval [0,1], not unlike the mock Gaussian runs with the IRAS PSCz in Chapter 6 though the same sort of noise we have observed in previous runs for values of $\beta > 0.6$ continues to manifest again, albeit at a more reduced level. This is once again mirrored in the (ζ, μ) distribution plots for the 2MRS galaxies as presented for various values of β in Figure 8.18, where for values of β smaller than 0.6 we continue to see the uniform scatter and distribution of both galaxies and distance moduli that we have come to expect; while at values larger than 0.6 we once again see a returned of the 'packed' distribution that disrupts uniformity and introduces noise to the data.



FIGURE 8.17: Plots of ρ for the redshift restricted 2MRS survey where β -dependent distance moduli μ_{β} were utilised in the calculation of $\rho(\zeta, \beta)$ instead of β -dependent peculiar velocities u_{β} across the K, H and J infrared bands, shown in blue, red and green respectively.

While the monotonic increase in ρ is promising it still remains puzzling as to why the plots are not more 'S-shaped' in nature given their Schechter origins. Unfortunately however we are unable to recover any value of β for the 2MRS as no zero intercept of any kind is exhibited by $\rho(\zeta, \beta)$ in any waveband, and a backwards extrapolation of the plots available to us would return values of β for the 2MRS that are negative which, as we have established previously, is not only inconsistent with the values of β published by Carrick et. al and their contemporaries, but is also impossible given how β itself is defined. While in and of itself this does not negate the viability of using μ_{β} to calculate ρ for ROBUST, we need additional testing measures to determine whether or not its use is truly valid. Such a testing measure (relative entropy) will be explored in depth in Chapters 9 and 10.

In the meantime one possible avenue for determining the validity of using μ_{β} would be to once again calculate the confidence or exclusion intervals on β and see if they return values that are consistent with excluding all noisy values of β larger than 0.6 as we have observed in previous sections. The confidence intervals returned for $\rho(\zeta, \beta)$ using bootstrap resampling and Monte Carlo simulations are presented in Figure 8.19.



FIGURE 8.18: Monitoring of (ζ, μ) distribution of the redshift restricted 2MRS galaxies where β -dependent distance moduli μ_{β} were utilised in the calculation of $\rho(\zeta, \beta)$ instead of β -dependent peculiar velocities u_{β} in the K band for various trial values of β . The trial values of β used are noted on each panel.

Unsurprisingly the exclusion intervals returned by both methods, while consistent with each other, are not at all in keeping with the exclusion values we have calculated thus far for β , returning $\beta \sim> 0.1$ for exclusion in the K and H bands, and returning no exclusion range for the J band whatsoever. This is due majorly in part to the consistently positive nature of the ρ plots across all three wavebands returning no zero intercept of any kind, though considering this is because we are using μ_{β} to calculate $\rho(\zeta, \beta)$ to begin with, that only lends itself to the supposition that perhaps we should not be using μ_{β} to calculate ρ in the first place, despite its statistical independence from ζ . As mentioned previously, this supposition will be explored in more depth in Chapter 11.

Despite ROBUST's inability to recover an actual value of β_{opt} for the 2MRS that is in good agreement with the value published by Carrick, it is worth pointing out that (for all cases where the proper theory was applied) it at the very least has been able to confidently exclude at the 3σ level all values of $\beta \geq 0.6$ from consideration for use with the 2MRS survey which, while not an ideal result, is still very useful in validating the β results returned by others in their current works while reinforcing the general rejection of larger values of β equal to unity as returned by POTENT.



FIGURE 8.19: One-sided confidence intervals calculated on β via bootstrap resampling (top panel) and Monte Carlo simulations (bottom panel) for the K, H and J bands depicted in blue, red and green respectively. The 3σ confidence intervals for all three bands are denoted via the black dashed lines, indicating all values of $\beta \sim> 0.1$ are to be excluded from consideration for the redshift restricted 2MRS survey in the K and H band where μ_{β} is used to calculate $\rho(\zeta, \beta)$ instead of u_{β} with 99.7% confidence.

8.6.4 Confirming the χ^2 results of Carrick

Given the peculiar results returned by ROBUST thus far, it is worth exploring whether the underlying 2M++ interpolation/velocity grid of Carrick that we are utilising is indeed valid. Should there be an inherent fault in the grid then this would have a knock-on effect on the peculiar velocities that are interpolated for each galaxy and consequently have a negative effect on the values of ζ and $\rho(\zeta,\beta)$ that are computed later, and would go a long ways towards explaining ROBUST's odd behaviour. To that end we will utilise the Carrick grid and analyse it using χ^2 -minimisation once we have selected an appropriate set of Type Ia Sne to embed within the grid and interpolate and scale their β -dependent peculiar velocities.

While Carrick et. al made use of the 'First Amendment' A1 Sne set in their seminal work to compute β that consists of 245 Sne within $200h^{-1}$ Mpc and draws from 3 independent datasets [25], we were unfortunately unable to retrieve this particular dataset from ViZier to use in conjunction with χ^2 . Consequently we will instead make use of the RS subset of the Tonry et. al Sne catalogue that we utilised in Chapter 5 with consistent good results when utilised with the IRAS PSCz survey. In order to cover all our bases we will run this analysis using two different methods:

- Method 1: Embed the RS set amongst the 2MRS galaxies, whose peculiar velocities have been interpolated using the Carrick grid.
- Method 2: Embed the RS set directly into the Carrick grid, performing all necessary interpolation measures using only the one 1Mpc cube within which each individual Sne is located.

Method 1 in particular provides an additional avenue within which we can test the reliability of our velocity interpolation schemes and distance convergence loops that we have applied earlier, as a successful recovery of the Carrick value will at least exclude the possibility that the problems manifesting in ROBUST are due to some errors residing in those scripts that have been passed on onto the final interpolated peculiar velocities for the 2MRS galaxies. Method 2 conversely will serve as the means of directly identifying whether the problem does indeed reside within the Carrick grid or not. The results of applying χ^2 -minimisation utilising Method 1 are presented in Figure 8.20, alongside a comparison chart of our interpolated velocities for the Sne alongside their 'observed' velocities as calculated from the Hubble law for our assumed 'true' value of $\beta = 0.43$. Should our assumed value of β indeed be the correct value, then these comparison plots should exhibit a 1:1 ratio, as denoted by the red line. It should be noted that the area of each individual circle on these velocity comparison plots is proportional to the associated

reconstruction errors for each Sne. Put another way, the smaller the circle, the smaller the associated error and the more confident we are of the true peculiar velocity of that particular Sne lying in an increasingly smaller range on the plot.

The results shown in Figure 8.20 are very promising. We are successfully able to return a value of $\beta = 0.445 \pm 0.04$ at the 1σ error level which is in excellent agreement with the value of $\beta = 0.43 \pm 0.021$ established by Carrick et al. with a relatively decent fit with our expected linear regression 1:1 comparison ratio in velocities being observed, though certain velocity discrepancies and outliers with large associated errors continue to persist in the plot. It must be noted that this ratio is not completely in keeping with similar results generated by Radburn-Smith in his seminal paper, but this is most likely due to the fact that we are utilising the Carrick grid (or indeed in this case the 2MRS galaxies whose velocities have been interpolated from said grid) for our interpolation purposes instead of the PSCz survey utilised by him and his colleagues. Each survey has its own unique properties and indeed are tracing different velocity and mass distributions in different regions of the nearby Universe, so some deviations are to be expected.

It is also worth pointing out that the same velocity interpolation scheme described in Section 5.2.2 is what is applied throughout this Chapter with regards to the Carrick grid, albeit with us only making use of the 8 cubelet vertices closest to each galaxy in question, and that this scheme has once again been rigorously tested in a mock environment utilising mock galaxies with positions coincident with cubelet vertices in order to determine whether the scheme can successfully recover 3D peculiar velocity components for that position that are in line with the values stored in the Carrick file. As before, the scheme has proven itself more than capable of recovering the peculiar velocity components expected of it to within 4-5 decimal places, thus confirming that all deviations and outliers noted in our linear regression plot is not due to a fundamental error in the interpolation scheme being utilised, or its associated coordinate transformation matrices.

The one frustrating property that continues to persist however is the inordinately large χ^2 values, this time on the order of 250 or so instead of our expected ~100 which once again favours the value of β =0.43 for rejection, were it not for the fact that Carrick himself has independently verified this value by utilising Forward Likelihood and VELMOD [25]. This could once again lend credence to our thought that perhaps β is not very sensitive to χ^2 , though this remains unclear.

The results of our analyses using Method 2 are presented in Figure 8.21 and are just as encouraging, albeit with a couple of curiosities. To begin with we are successful in recovering a value of β in keeping with that computed by Carrick, namely $\beta =$ 0.426 ± 0.035 at the 1σ error level. This leads to a couple of important conclusions, specifically that whether we go for the indirect approach of Method 1 (using the 2MRS



FIGURE 8.20: Successful recovery of the value of $\beta = 0.445 \pm 0.04$ via χ^2 -minimisation using Method 1 (top panel), alongside a comparison of RS Sne velocities computed via interpolation in Method 1 (x-axis) with those computed using the Hubble Law (yaxis). Areas of individual plot points are proportional to the associated reconstruction errors of each individual Sne used, with the red line indicating the linear regression (or goodness-of-fit) between both sets of velocities.



FIGURE 8.21: Successful recovery of the value of $\beta = 0.426 \pm 0.035$ via χ^2 minimisation using Method 2 (top panel), alongside a comparison of RS Sne velocities computed via interpolation in Method 2 (x-axis) with those computed using the Hubble Law (y-axis). Observed 'squashing effect' is due to the existence of a singular outlier in the data. Areas of individual plot points are proportional to the associated reconstruction errors of each individual Sne used, with the red line indicating the linear regression (or goodness-of-fit) between both sets of velocities.

galaxies) or use the Carrick grid directly ala Method 2 we continue to achieve a value of β that is in excellent agreement with the Carrick value. This serves as a means of confirming that our interpolation schemes and distance convergence loops must indeed be functioning correctly and without error as the difference in value between the two Methods is almost negligible. The success of our χ^2 analysis using Method 2 also serves to indicate that there cannot be any underlying fault or errors in the Carrick grid itself, otherwise we would have failed to recover a value of β in such close agreement to their own. This unfortunately once again points to a problem in either the underlying theory or the functioning of ROBUST itself, which is where we will begin to focus our efforts in future sections.

While we have successfully recovered β using Method 2, the bottom panel of Figure 8.21 exhibits a strange 'squashing effect', which appears to be the result of a strange outlier in the data that is generated when we interpolate velocities using Method 2 as opposed to Method 1. It is worth noting however that even with the existence of this outlier in the data the observed goodness-of-fit with our linear regression model is still somewhat decent, with a reasonable number of Sne with small associated errors clustering in relative proximity to our predicted 1:1 ratio line for our assumed true β of 0.43. The reason for the existence of this outlier is not fully clear considering that we are increasingly confident in the functioning of our interpolation schemes and may be due to some unique properties of that particular Sne once it is embedded directly into the Carrick grid. When we isolate this particular outlier and remove it from the grid we get a distribution of velocities similar to that achieved earlier as depicted in Figure 8.22. Our ratio of observed to interpolated velocities continues to exhibit a somewhat 1:1 distribution that is clearer to see, further signifying that our interpolation scheme appears to be functioning correctly for the most part, though the persistence of velocity discrepancies and outliers muddling the fit to our linear regression model remains puzzling.

With the consistently less than ideal linear regression fitting and weak recovered χ^2 signals for β that have been observed throughout this work both here and in Chapter 5, it is worth considering whether our underlying assumptions as to the linearity of our reconstructed peculiar velocity fields are valid, as any deviation from said linearity would cause further sources or velocity perturbations to occur which would go some way towards explaining some of the velocity discrepancies and outliers noted in our plots so far. This possibility will be addressed in further detail as a future avenue of exploration at the tail-end of this work in Chapter 13.

In summary, we have been able to successfully recover Carrick's value of β to within acceptable agreement limits, while also utilising a different Sne set than that used by



FIGURE 8.22: Comparison of RS Sne velocities computed via interpolation in Method 2 (x-axis) with those computed using the Hubble Law (y-axis) with the outlier removed. Areas of individual plot points are proportional to the associated reconstruction errors of each individual Sne used, with the red line indicating the linear regression (or goodness-of-fit) between both sets of velocities.

them in their work. This lends credence to the validity of their published value while also confirming that our interpolation schemes are functioning correctly and that indeed no underlying fault can be observed in the Carrick grid (given the successful recovery of β using Method 2). While this unfortunately does not bring us any closer to understanding why ROBUST is continuing to return such noisy or inconsistent results we do have some additional avenues to explore that may explain its strange behaviour.

- Having now confirmed that there is no issue in either our interpolation schemes or in the Carrick grid itself, there may be an inherent fault in the 2MRS survey itself that prevents it from being fully compatible for use with ROBUST.
- There may be an error in our underlying theory upon which ROBUST is reliant, namely our assumption that the luminosity distribution of a galaxy survey and the 3D spatial positions of its component galaxies are independent of one another. This shall be explored in detail in Chapter 12.

Before we proceed to explore any possible errors in the underlying theory however, ROBUST itself offers us a secondary means by which we can test and/or confirm the validity of all of our results generated thus far. This alternative is called relative entropy, and after introducing the general theory behind its use in the following chapter, we will proceed to apply relative entropy to both the IRAS PSCz in a mock setting to fully explore its usefulness before consequently applying it to the 2MRS and seeing how well it validates the results that we have observed up to this point.

Chapter 9

A ROBUST Alternative -Relative Entropy

While we were successfully able to establish the usefulness of ROBUST as a statistical tool to probe the peculiar velocity fields of the PSCz in multiple mock simulations where galaxy set size, luminosity function and reconstruction errors were varied without any alteration on the 'true' value of β recovered, applying it to real world data has been problematic at best. This has included the lack of galaxy-magnitude matching information necessary for us to utilise ROBUST with the B-band magnitude information of the PSCz, and 2M++ returning a result inconsistent with that determined by Carrick using VELMOD and χ^2 -minimisation, the latter of which we confirmed for ourselves; suggesting either an inherent problem in the ROBUST methodology itself as opposed to the Carrick grid or an inherent fault within the 2MRS survey itself that causes it to not lend itself well for use with ROBUST. Various experimentation attempts have failed to bring our value of β^* any closer to ~0.43. It may also be that our correlation coefficient parameter ρ is not sensitive enough to changes in β in the 2MRS survey in order for it to be able to recover a value effectively.

The take away from the above is clear. In order for us to truly be able to determine whether the problem lies in the ROBUST methodology or elsewhere, we need an independent means of verifying the values of β ROBUST returns, one that in itself also shares some of the characteristics of our ζ statistic. If this independent means were also to return the same value of, say, $\beta \sim 0.05$ for the 2M++ then we can at least say in confidence that the problem does not lie within ROBUST itself. Thankfully, such a methodology does indeed exist that we can make use of, namely the concept of relative entropy between two variables.
9.1 Underlying Theory of Relative Entropy

While entropy by its strictest definition is a measure of the amount of disorder or chaos in a system, with all systems tending towards a state of high entropy over time, what we will be applying here is slightly different. Working from the basics of information theory we can start by defining two probability distributions μ_1 and μ_2 that are absolutely continuous with respect to one another such that

$$\mu_1 \equiv \mu_2$$

In other words, there exists no event E for which $\mu_1(E) = 0$ and $\mu_2(E) \neq 0$, or $\mu_1(E) \neq 0$ and $\mu_2(E) = 0$ [109]. We also define an additional probability measure λ such that

$$\lambda \equiv \mu_1, \lambda \equiv \mu_2,$$

meaning that for example, λ may be μ_1 , μ_2 or $(\mu_1 + \mu_2)/2$. By the Radon-Nikodym theorem (refer to the work of Kullback for the explicit definition of this theorem and its consequent derivatives [109]) we can now define two generalised probability densities for μ_1 and μ_2 , denoted $f_1(x)$ and $f_2(x)$ such that:

$$\mu_i(E) = \int_E f_i(x) d\lambda(x), \quad i = 1, 2$$
(9.1)

for all possible events, E [109]. The function $f_i(x)$ is also called the Radon-Nikodym derivative as it can be expressed as

$$d\mu_i(x) = f_i(x)d\lambda(x), \qquad f_i(x) = \frac{d\mu_i}{d\lambda}.$$

If we now choose to define our hypotheses H_i , i = 1, 2, which indicates the likelihood of a variable X being from the statistical population with probability measure μ_i then by applying Bayes' theorem it follows that:

$$P(H_i|x) = \frac{P(H_i)f_i(x)}{P(H_1)f_1(x) + P(H_2)f_2(x)} [\lambda], \quad i = 1, 2,$$
(9.2)

from which we can obtain

$$\ln \frac{f_1(x)}{f_2(x)} = \ln \frac{P(H_1|x)}{P(H_2|x)} - \ln \frac{P(H_1)}{P(H_2)} [\lambda], \qquad (9.3)$$

where $P(H_i)$, i = 1, 2 is the prior probability of H_i and $P(H_i|x)$ is the conditional probability of H_i given X = x [109]. The right hand side of Equation 9.3 is a measure of the difference between the logarithm of the odds in favour of H_1 after the observation of X = x and before the observation. This difference, which can be positive or negative, 208

may be considered as the information resulting from the observation X = x, and we define the logarithm of the likelihood ratio: $\ln [f_1(x)/f_2(x)]$ to be the information in X = x for discrimination in favour of H_1 against H_2 . It can also be thought of as the weight of evidence for H_1 given x [109]. Therefore the mean information for discrimination in favour of H_1 given H_2 for given x occurring within events E for μ_1 is given by:

$$I(1:2;E) = \frac{1}{\mu_1(E)} \int_E \ln \frac{f_1(x)}{f_2(x)} d\mu_1(x)$$

= $\frac{1}{\mu_1(E)} \int_E f_1(x) \ln \frac{f_1(x)}{f_2(x)} d\lambda(x), \qquad \mu_1(E) > 0,$ (9.4)
= $0, \qquad \mu_1(E) = 0,$

where

$$d\mu_1(x) = f_1(x)d\lambda(x).$$

If we are dealing with only one probabilistic event E, then we can rewrite our variable I as I(1:2), denoting the mean information for discrimination in favour of H_1 against H_2 per observation from μ_1 such that

$$I(1:2) = \int \ln \frac{f_1(x)}{f_2(x)} d\mu_1(x) = \left[\int f_1(x) \ln \frac{f_1(x)}{f_2(x)} d\lambda(x) \right]$$

= $\int \ln \frac{P(H_1|x)}{P(H_2|x)} d\mu_1(x) - \ln \frac{P(H_1)}{P(H_2)},$ (9.5)

where I(1:2) can also be called the information measure of μ_1 with respect to μ_2 [109]. The boxed equation in Equation 9.5 can also be considered as a *directed divergence*, and is more commonly called the **Kullback-Leibler Divergence** (as originally derived by Kullback and Leibler in their seminal work in 1951 [110]) or **relative entropy** (or Kullback-Leibler risk in some literature [34]) between two probability distributions μ_1 and μ_2 where, with regards to our purposes for this work, μ_1 represents the observed distribution of the data we have, and μ_2 represents our theoretical model for that distribution [96]. If we choose to express this integral as a summation of *n* observations over our two probability distributions we would get:

$$S\left[(\mu_1, \mu_2)\right] = \sum_{i=1}^n \mu_1(i) \ln\left(\frac{\mu_1(i)}{\mu_2(i)}\right),\tag{9.6}$$

where $S[(\mu_1, \mu_2)]$ is our relative entropy variable. More specifically, this variable will provide us with a measure of how well one distribution relates to another, and for the case where both variables are completely uncorrelated with each other, $S[(\mu_1, \mu_2)]$ will be at its minimum value, or minimum disorder or entropy. Consequently with regards to our own work the purposes for which we will want to make use of this statistic are such as to find the value of β that not only causes our correlation coefficient ρ to be 0, but also minimises $S[(\mu_1, \mu_2)]$. In order to achieve this we need to define a new variable which we can utilise to verify the correctness of our central ζ statistic while also serving as a test of independence for it.

9.2 A New Statistic, χ , and Applying Relative Entropy

Recall from Section 4.1 where we established the two key characteristics of our ζ variable for ROBUST, namely that for a magnitude complete survey:

- 1. ζ will be uniformly distributed on the interval [0,1],
- 2. ζ and μ (or Z) are statistically independent.

Now let us consider a new variable, χ , defined as the cumulative redshift distribution of galaxies in a survey such that:

$$\chi = \frac{H(z)}{H(z_u)},\tag{9.7}$$

where

$$H(z) = \int_{z}^{z_{u}} h(z')dz',$$
(9.8)

and z_u is taken to be a pre-established upper redshift limit (in the case of for example the PSCz survey this would take on the value of 15000kms⁻¹ we established earlier to ensure the reliability of the reconstructed peculiar velocity fields being used) [96]. Due to the applied normalisation over $H(z_u)$, χ now shares the same property as ζ , namely that it will be uniform on the interval [0,1]. Consequently it can be easily inferred that for a magnitude complete survey, a plot of ζ vs. χ such as that depicted in Figure 9.1 will also be perfectly uniform on a unit square provided that the correct value of β is applied during the construction of the ζ statistic. It also follows that for all values of $\beta \neq \beta^*$ a correlation will be introduced between the two variables causing a deviation from a perfect uniform distribution.

This is where we can begin to make use of relative entropy. In particular for the case where we have the correct β^* , if we were to impose a mesh grid onto our unit square such as that depicted in Figure 9.2, a perfectly uniform distribution would be such that the number of galaxies in each small square would be equal. More specifically, we can calculate the probability p_i for each cell from the observed (ζ, χ) distribution such that:

$$p_i = \frac{N(c_{mn})}{N_{gal}},\tag{9.9}$$



FIGURE 9.1: Example illustrating a typical (ζ, χ) distribution for a complete data-set for an MGC mock catalogue. The left-hand distribution shows ζ and χ estimated at the apparent magnitude limit of the survey $m_{lim} = 20.0$ and appears to be a random uniform distribution. Correlations between ζ and χ are shown on the right-hand panel where ζ and χ have been estimated at $m_* = 20.5$ (beyond the limit of the survey). Figure reproduced from the work of Johnston [96].

where $N(c_{mn})$ is the number of galaxies within the cell, c, located at (m, n), and N_{gal} is the total number of galaxies for the whole (ζ, χ) distribution [96]. It therefore follows that the theoretical model we expect to see in this case would also satisfy the condition

$$q_i = \frac{1}{C_{tot}},\tag{9.10}$$

where C_{tot} is the total number of cells that make up the imposed mesh [96]. We can now apply these definitions to compute the relative entropy of ζ and χ to obtain:

$$S\left[\left(\zeta_{\beta},\chi\right)\right] = \sum_{i=1}^{n} p_i \ln\left(\frac{p_i}{q_i}\right)$$
(9.11)

Due to the quantity q_i always being less than 1 by its very definition, $S[(\zeta_\beta, \chi)]$ will always have the property of being negative, only ever tending to 0 in the case where $p_i = q_i$. Since it is the convention that all systems tend towards increasingly positive entropy over time, we can represent this with a simple change of sign:

$$S\left[\left(\zeta_{\beta},\chi\right)\right] = -\sum_{i=1}^{n} p_i \ln\left(\frac{p_i}{q_i}\right),\tag{9.12}$$

which also gives it the property of having maximum entropy at the maximum value of $S[(\zeta_{\beta}, \chi)]$ [96].

211



FIGURE 9.2: Example of measuring the entropy of a typical (ζ, χ) distribution for a complete data-set taken at the $m_* = m_{lim} = 20.0$ of an MGC mock catalogue. We calculate the total entropy of this distribution by imposing a grid with a predetermined mesh size. In this example we have split the grid into 0.1×0.1 mesh. We then count the number of objects contained in each box, p_i , and calculate the relative entropy. We then determine the total entropy by summing each $p_1, p_2, \ldots, p_i, \ldots, p_n$. Figure reproduced from the work of Johnston [96].

An example of this relative entropy measure applied to multiple mocks of the Millennium Galaxy Catalogue (MGC) is illustrated in Figure 9.3, where it successfully constrains the mock assumed 'true' values of $\hat{\beta}$ (in this instance a measure of galaxy evolution as opposed to the linear redshift distortion parameter with which we are familiar) over multiple realisations.

It should be noted that the shape of the plot of relative entropy will be heavily reliant on the sort of parameter we are attempting to constrain, but the key point to take away is that for the correct value of $\hat{\beta}$, or β^* as we require, it should produce a clear minimum in the plot when reached. Once this minimum is established and the optimal value of β is determined, we now need to compute the confidence intervals on β . Thankfully this is simple to do.



FIGURE 9.3: Relative entropy S, vs. trial $\hat{\beta}$ (in this instance a statistic used as a measure of galaxy evolution) using MGC mocks for three different values of assumed 'true' β_{true} . For each trial $\hat{\beta}$ we have applied 50 mock realisations with successive values of $\beta_{true} = 1.0, 2.0$ and 3.0. In each case we can clearly see that the relative entropy approach minimises at the correct value. Figure reproduced from the work of Johnston [96].

9.3 Determining the Error on β

Much like before in Section 4.3.1, establishing out confidence intervals on β as determined via relative entropy is a very simple affair, due to us continuing to make use of the ζ statistic as part of its definition. In particular we can continue to make use of Monte Carlo simulations, resampling new ζ values from newly generated uniform distributions on [0,1] with each iteration, and expect to see no deviation from a perfect uniform ζ - χ distribution on a unit square, and consequently no variation in the relative entropy measurements computed thereafter. Similarly should we apply bootstrap resampling and choose to scramble our ζ values to different galaxies on said ζ - χ plot, there should continue to be no deviation in either distribution or relative entropy over multiple realisations. We will consequently make use of both when determining our error bars on β in future relative entropy chapters.

213

The same is also true where mock simulations are used, such as those depicted in Figure 9.3. As before, all that is required for us to compute our confidence intervals is to plot a histogram of the β values across all simulations that minimise $S[(\zeta_{\beta}, \chi)]$, and then model a best-fit Gaussian curve to the data where its standard deviation σ will indicate the 1σ confidence interval on β .

Having now established our derivation of relative entropy and how it can be applied in tandem with ROBUST to verify the values of β it returns, we will now proceed to apply $S[(\zeta_{\beta}, \chi)]$ to mock simulations of the PSCz using both our mock Gaussian and Schechter luminosity functions, while also experimenting with varying imposed mesh sizes to our $\zeta - \chi$ plots to observe what effect this has on the β values returned by $S[(\zeta_{\beta}, \chi)]$. Once we have established the ideal mesh size and demonstrated that we can successfully recover our assumed β_{true} of 0.55 ± 0.06 , we will then proceed to apply $S[(\zeta_{\beta}, \chi)]$ to the 2M++ and see whether the values of β returned are in good agreement with the values computed by ROBUST.

Chapter 10

Probing the Peculiar Velocity Field of the PSCz with Relative Entropy

Much like Chapter 6 where we probed the velocity field of the PSCz utilising ROBUST and the correlation coefficient $\rho(\zeta, \beta)$, we will proceed with much of the same frameworks that were developed therein, and adapt them where necessary to allow for the use of relative entropy. To that end this chapter will be structured in a very similar manner to its ROBUST counterpart, with Section 10.1 discussing the various mock methodologies at our disposal to apply to the PSCz, while Sections 10.2 through 10.5 will detail the results of each application. While we will also continue to explore the effects of varying mock survey size, Gaussian luminosity function widths and Schechter function parameters on the results returned by relative entropy, we will also dedicate Section 10.2.2 to exploring the effects that varying the number of squares (or altering the mesh size) on our $1 \times 1 \zeta - \chi$ grid (refer to Figure 9.2) will have on the accuracy and precision of the results returned by relative entropy. As before, any anomalies observed in our results will be explored and their root causes determined and eliminated to the best of our abilities.

10.1 Implementing Relative Entropy with the PSCz Survey

As discussed in Chapter 9, for a magnitude complete survey with known limits, the functionality of relative entropy lies in the assumed independence of the distribution of the β -dependent ζ statistics of galaxies as determine by ROBUST from the cumulative distribution of their redshifts as denoted by χ , such that a correlation will be introduced between these two variables should the wrong value of β be applied to the data. Once again bearing in mind the reliance of ROBUST on the luminosity function information of a survey, this gives us several avenues of experimentation to consider, in order to evaluate the effectiveness of relative entropy when it is used in conjunction with ROBUST:

- 1. Mock Method 1: We adopt β_{true} of 0.55 as before and an assumed known luminosity function model, and utilise Matlab to assign 'mock' apparent magnitudes to the PSCz galaxies for an assumed, known faint limit. We then proceed to utilise the available redshift information for the galaxies from the survey to calculate the statistic χ before then applying ROBUST as far as calculating the β -dependent ζ values for the galaxies over trial values of β on [0,1], correcting distances and peculiar velocities as appropriate. With ζ and χ now calculated we can then apply relative entropy and identify the value of β_{trial} on [0,1] for which $S[(\zeta_{\beta}, \chi)]$ is minimised. This is repeated over multiple mocks, generating new mock magnitudes to be assigned to the galaxies each time. We will proceed to experiment with assigning mock magnitudes to the PSCz galaxies drawn from the following:
 - (a) A Gaussian luminosity function, whose mean and standard deviation will be in keeping with the identified luminosity function of early-type galaxies of the SDSS, namely N[-21,1] [9],
 - (b) A Schechter luminosity function, whose standard parameters will be defined as before, namely S[-21,1.09] (refer to Section 6.3.1 for a more in depth definition of the variables and parameters of a typical Schechter function).
- 2. Mock Method 2: We proceed to repeat the above while also establishing a bright limit to the mock magnitudes being generated by Matlab and rerun RO-BUST and the application of relative entropy, taking note of any alterations to the values of β recovered. Much like Method 1, this will also be repeated with both a Gaussian and Schechter luminosity function over multiple mock trials.

Flowcharts indicating a typical application of our mock methodologies is presented in Figure 10.1.



FIGURE 10.1: Flowchart illustrating typical ROBUST/relative entropy methodology for use with generated mock magnitudes assigned to galaxies utilising either just a faint, or faint+bright magnitude limits during generation. Note that m_f and m_b stands for faint and bright apparent magnitude limit respectively [Methods 1a), 1b), 2a) and 2b)].

10.2 Applying Mock Method 1a) - 'Faint Limit Only Gaussian' Mock Magnitudes

Having already previously established in Chapter 6 the means by which we will randomly assign mock magnitudes to the PSCz galaxies (refer to Figure 6.2) and identified a sensible value for the faint limit that we would make use of $(m_{lim}=14.3)$ once we have adequately redshift restricted the PSCz survey to ensure the reliability of the underlying reconstructed peculiar velocity field, we will now proceed to make use of the entirety of the survey (over 12000 galaxies) to apply relative entropy with a mock Gaussian luminosity function and an arbitrary mesh size of 10 - i.e. 100 equal size squares overlain on our $1 \times 1 \zeta - \chi$ grid, the results of which are presented in Figure 10.2.

While it is immediately clear that relative entropy has successfully managed to recover a mean estimate for β that is in good agreement with our assumed value for β_{true} , there are a few points of interest to take note of. To begin with, unlike the relative entropy plots presented in Chapter 9 there appears to be a considerable amount of noise in the mock plots presented here. This could possibly be due to the sort of luminosity function being modelled here, in particular the specific parameters being applied such as the Gaussian standard deviation, though it could also be due to the grid resolution, or mesh size, that we are applying. Both of these possibilities will be explored later on in this section.

The second matter of interest is the considerably larger confidence interval returned on β especially compared with its ROBUST counterpart. This however is possibly due to the number of trial values of β being applied in this instance, namely ten values on the interval [0,1], thus chosen in consideration of computational time constraints (a 50-mock run utilising only ten values of β completes in approximately 12 minutes, while a similar 50-mock run utilising 50 values completes in over 90 minutes). Bearing in mind that when we proceed to apply relative entropy (only once) to the 2MRS in Chapter 11 we will be utilising 1000 trial values of β on the interval [0,1] instead of 10, we fully anticipate the broadness of the confidence intervals returned therein to be smaller.

10.2.1 The curiosity of small $S[(\zeta, \chi)]$ values

The third point of consideration, again when compared with the sample relative entropy plots presented in the previous chapter, is that the signal recovered by relative entropy when applied here is considerably weaker than one would expect, especially when compared to the signals returned by ROBUST when applied exclusively using $\rho(\zeta, \beta)$. The range of values exhibited by $S[(\zeta, \chi)]$ is on the order of 10×10^{-3} whereas in Chapter 9



FIGURE 10.2: Initial results of Method 1a). Upper panel depicts the $S[(\zeta_{\beta}, \chi)]$ plots for all 50 mocks generated, lower panel depicts the minimum of each mock plot as a function of trial value of β plotted as a histogram with best-fit Gaussian (red line) modelled to the data. Optimal value of β returned for Method 1a) is $\beta = 0.55 \pm 0.26$, as determined from the Gaussian model. Luminosity function used during data generation is Gaussian of the form N[-21,1] while arbitrary mesh size applied is 10.

the range exhibited in Figure 9.3 was on the order of 0-0.4, a good two orders of magnitude larger. On the one hand this is possibly due to the fact that the β parameter being constrained in that particular figure is tracing a different quantity altogether from our linear redshift distortion parameter (in actual fact the former parameter is being used as a measure of galaxy evolution), while additionally the underlying redshift surveys being utilised for the analysis are also different from what we are using here (the MGC as opposed to the PSCz). While both of these possibilities combined may indeed be contributing significantly to the weaker signals being returned by relative entropy here, when one takes a moment to observe the behaviour of randomly selected PSCz galaxies on both the M-Z and the ζ - χ plane, the primary source of the weak signal starts to become more apparent.

Figure 10.3 presents three different instances of 5 randomly selected PSCz galaxies being tracked in both the M-Z and the ζ - χ plane (the latter for an arbitrary mesh size of 10, as denoted by the grey lines) over incrementally increasing values of β on the interval [0,1], with ten values being used in total. While some galaxies across all three instances present a noticeable shift across the M-Z plane as the value of β (and consequently the value of their β -dependent peculiar velocity) is increased, this is not necessarily represented by an equivalent shift in the ζ - χ plane. Taking for example the purple coloured galaxy in the first instance (top row of Figure 10.3), it exhibits a decent track of movement across the M-Z plane, yet remains inside the same grid square on the ζ - χ plane. Considering that the value of $S[(\zeta, \chi)]$ is reliant on the galaxy in question moving across grid squares in order for any change in entropy to be noted, this minimal movement on the part of this particular galaxy would manifest as an extremely weak signal when recovered using relative entropy. Conversely if one considers the blue coloured galaxy in the third instance (bottom row of Figure 10.3), it exhibits a very stark shift across the M-Z plane, indicative of it possessing a very large peculiar velocity, and this manifests equivalently in the ζ - χ plane where the galaxy crosses over at least three grid squares as the value of β_{trial} is incrementally increased on [0,1]. Consequently this particular galaxy would return a stronger $S[(\zeta, \chi)]$ signal when relative entropy is applied.

The takeaway from the above is clear. Unless the PSCz galaxies being analysed have significantly large peculiar velocities, then they will not move around enough on the ζ - χ grid for a strong $S[(\zeta,\chi)]$ signal to be returned. When one considers that the general range of peculiar velocities noted for the PSCz galaxies are on the order of 400-600kms⁻¹ or smaller, the weaker signal we are observing begins to make more sense. It would appear that the galaxies do not generally have peculiar velocities large enough to produce a consistently stronger signal for $S[(\zeta,\chi)]$, at least not with how our linear redshift distortion parameter β and our zeta statistic ζ_{β} are currently being defined and constructed.



FIGURE 10.3: Three different instances of 5 randomly selected PSCz galaxies whose movements are being tracked in both the M-Z (left side panels) and the ζ - χ planes (right side panels) respectively. An arbitrary mesh size of 10 was applied to the ζ - χ grid, as denoted by the grey lines. Each circle denotes the position of each galaxy for a value of β_{trial} being increased incrementally on the interval [0,1], with ten values of β_{trial} being used in total.

Grid Resolution	Grid Resolution 5		20	30	
Value of β	0.57 ± 0.26	0.52 ± 0.29	0.57 ± 0.31	0.58 ± 0.32	

TABLE 10.1: Summary of estimated values of β detailing the effect of varying arbitrarily applied grid resolutions to the construction of the ζ - χ grid for use in computation of β with relative entropy. Entirety of PSCz survey is utilised for this analysis.

10.2.2 A question of grid resolution size

This does however raise an interesting question. If one were to alter the grid resolution (or mesh grid size) being arbitrarily applied when we are constructing the ζ - χ grid for use with relative entropy, would this cause any change in the range of values returned for $S[(\zeta, \chi)]$? Should the underlying plane be overlain with, say, 400 grid squares (equivalent of a grid resolution of 20) instead of 100, one would expect a stronger signal as a galaxy (even one with an inherently smaller peculiar velocity) would cross over more squares as the value of β_{trial} is altered over [0,1]. To that end, we will proceed to experiment with altering the arbitrarily applied grid resolution when relative entropy is run, and take note of any changes observed in the values of $S[(\zeta, \chi)]$. The results of running relative entropy with varying arbitrary grid resolutions for the standard mock Gaussian LF of N[-21,1] as applied before are presented in Figure 10.4 while our monitoring of ζ - χ histograms are presented in Figure 10.5, with a summary of the estimated β values obtained presented in Table 10.1.

The results of varying the applied grid resolution are interesting for several reasons. While we were correct in our assumption that increasing the applied resolution would cause a stronger signal in $S[(\zeta, \chi)]$ to appear when relative entropy was run, this appears to have come at the expense of accuracy as the amount of noise presented in the lefthand panels of Figure 10.4 increases significantly. When one considers the shrinking size of individual grid squares as we increase the resolution this is not necessarily surprising. As the grid squares become smaller and smaller, the likelihood of finding PSCz galaxies inhabiting them decreases substantially, causing the value of β_{opt} we are attempting to recover to become dominated and washed out by Poisson fluctuations and shot noise, as symbolised by the increasingly messy nature of the plots presented. This is also very noticeable in Figure 10.5 where as the resolution is increased the uniformity of the ζ - χ distribution becomes increasingly erratic as said shot noise and Poisson fluctuations start to dominate. This was also apparent during the Matlab runs themselves where on more than one occasion the script would crash out due to 'galaxies not inhabiting grid squares' causing zeroes and infinities to appear in the calculations that it cannot handle. Conversely when we make the grid resolution considerably smaller we see a decrease in



FIGURE 10.4: Results of varying grid resolution during use in relative entropy. Lefthand panels illustrate the $S[(\zeta, \chi)]$ plots for all 50 mocks, right-hand panels depict the histograms of optimal value of β with best-fit Gaussian (red line). Grid resolutions applied are noted on each plot. Entirety of the PSCz survey is utilised during analysis.



FIGURE 10.5: Monitoring of ζ - χ behaviour with varied grid resolution. Panels depict the bivariate histograms of ζ - χ value distribution over [0,1]. The thick red line denotes our expected number galaxies per histogram bin for a uniform distribution.

the noise (also mirrored in the considerably more uniform distribution of $\zeta - \chi$ values in the bivariate histogram in the upper left panel of Figure 10.5), but this once again comes at the expense of our $S[(\zeta, \chi)]$ values becoming even smaller. Again when one considers the now overly large size of the grid squares overlain on the $\zeta - \chi$ this makes sense as now, not enough of the galaxies are able to move across enough squares (due to their generally small peculiar velocities and the now larger distance that must be traversed) to cause any change in relative entropy to be observed.

With regards to the recovered β values themselves, while the mean estimates recovered by relative entropy are in general good agreement with our assumed value of β_{true} (though the mean value does gradually worsen as we increase the resolution), the broad confidence intervals returned here make it all the more apparent that our choice of arbitrary grid resolution is important if we wish for relative entropy to return as accurate a result as possible. Bearing in mind the problems of Poisson fluctuations and shot noise, and now having a better understanding of why the values of $S[(\zeta, \chi)]$ returned by relative entropy are as small as they are, we will continue to proceed with our analyses in the rest of the chapter with an applied grid resolution of 10, as this appears to be the 'happy

Mock Set Size	1000	2000	5000	All (12000+)
Value of β	0.54 ± 0.50	0.59 ± 0.44	0.58 ± 0.30	0.55 ± 0.26

TABLE 10.2: Summary of estimated values of β detailing the effect of varying mock galaxy sample sizes on computed β using ROBUST during Mock Method 1a) implementation.

medium' that we can best make use of.

Having now established the grid resolution we will utilise going forward we can now begin to experiment with varying the number of galaxies relative entropy utilises for its analysis and take note of any changes observed in the value of β_{opt} returned. As such we will rerun our analyses using 1000, 2000, 5000 and all survey galaxies of the PSCz with the results presented in Figure 10.6 alongside our monitoring of ζ - χ behaviour in Figure 10.7, with a summary of the estimated values of β returned for this analysis presented in Table 10.2.

In keeping with the ROBUST behaviours we observed in Chapter 6, relative entropy is successful in recovering β estimates that are in good agreement with our assumed value of β_{true} , with the confidence intervals returned in each instance improving as mock survey set size is increased. This is also mirrored in the decrease in data noise observed in the $S[(\zeta, \chi)]$ plots as we increase survey size, and the ζ - χ behaviour of the PSCz galaxies where the noise exhibited in the bivariate histograms of Figure 10.7 drops considerably for a set applied grid resolution of 10. It does however appear that in order for us to be able to constrain β with as sensible a confidence interval as possible with relative entropy we are required to use the entirety of the PSCz survey, unlike its ROBUST counterpart which was able to constrain β sensibly with a bare minimum of 2000 galaxies only. When one considers the strength of the $S[(\zeta, \chi)]$ signal recovered here in comparison with the stronger $\rho(\zeta,\beta)$ signals recovered in Chapter 6 this is not all that surprising. In addition when one bears in mind the size of the 2MRS (over 43000 galaxies available) when we come to apply relative entropy to it in Chapter 11 this should not present much of an issue and we expect the value of β constrained for that survey to have a sensibly tight confidence interval.

We can now proceed with varying the value of σ utilised by **normrnd** when generating Gaussian magnitudes to assign to the PSCz galaxies and observe what changes are observed in the values of β recovered by relative entropy. The results of altering the LF width are presented in Figure 10.8 alongside our standard monitoring of ζ - χ behaviour in Figure 10.9, with a summary of the estimated values of β presented in Table 10.3.



FIGURE 10.6: Method 1a) results with varied mock set sizes. Left-hand panels illustrate the $S[(\zeta, \chi)]$ plots for all 50 mocks, right-hand panels depict the histograms of optimal value of β with best-fit Gaussian (red line). Set sizes used are noted on each plot. Luminosity function generated continues to be Gaussian of form N[-21,1].



FIGURE 10.7: Monitoring of $\zeta - \chi$ behaviour with varied mock set size. Panels depict the bivariate histograms of $\zeta - \chi$ value distribution over [0,1]. The thick red line denotes our expected number galaxies per histogram bin for a uniform distribution.

Gaussian LF Width	0.5	1	2	3
Value of β	0.57 ± 0.13	0.55 ± 0.26	0.46 ± 0.42	0.52 ± 0.47

TABLE 10.3: Summary of estimated values of β detailing the effect of varying mock generated Gaussian LF width σ on computed β using relative entropy during Mock Method 1a) implementation. Entirety of PSCz survey is utilised for this analysis.

Once again as with its ROBUST counterpart, relative entropy succeeds in recovering mean estimates for β that are in good agreement with our assumed true value of β , with a Gaussian width of 0.5 resulting in the most tightly constrained and parabolic plots of $S[(\zeta, \chi)]$ seen thus far. Conversely as the Gaussian width is increased our confidence interval on β starts to broaden considerably as expected, once again symbolised by the increased noise in the $S[(\zeta, \chi)]$ plots. A slight curiosity does present itself however in Figure 10.9, particularly with regards to the bivariate histogram plots for a mock Gaussian width of 0.5 and 1. There, a slight undersampling can be noted causing a deviation from uniformity and yet, relative entropy still manages to successfully recover the true value of β that we are looking for. This would suggest that, much like ROBUST, relative entropy is somewhat forgiving with its required uniform distribution, still being



FIGURE 10.8: Method 1a) results with varied mock Gaussian LF widths. Left-hand panels illustrate the $S[(\zeta, \chi)]$ plots for all 50 mocks, right-hand panels depict the histograms of optimal value of β with best-fit Gaussian (red line). LF widths generated are noted on each plot. Entirety of the PSCz survey is utilised for this analysis.



FIGURE 10.9: Monitoring of $\zeta - \chi$ behaviour with varied mock Gaussian LF widths. Panels depict the bivariate histograms of $\zeta - \chi$ value distribution over [0,1]. The thick red line denotes our expected number galaxies per histogram bin for a uniform distribution.

able to function adequately despite the undersampling noted here. This would be of invaluable use should relative entropy be extended for use with other large-scale surveys such as the SDSS and LSST where the inherent modelled luminosity functions may be of varying definition and breadth.

10.3 Applying Mock Method 1b) - 'Faint Limit Only Schechter' Mock Magnitudes

As before, in order to truly determine the robustness of relative entropy as a statistical tool, we will now proceed to repeat all of the above while instead assigning mock generated Schechter magnitudes to the PSCz galaxies. Having already determined the optimal grid resolution to use for our analyses we will focus instead on the effects of altering mock survey set sizes as before, as well as the effects of altering the power law slope parameter α as part of the Schechter function definition and observe any changes noted in the values of β_{opt} recovered by relative entropy. Our initial results utilising a Schechter function are presented in Figure 10.10 and are interesting for a few reasons.



FIGURE 10.10: Initial results of Method 1b), depicting the $S[(\zeta, \chi)]$ plots for all 50 mocks generated. Optimal value of β returned for Method 1b) is $\beta = 0.5$, as determined from analysing plot minima. Luminosity function used during data generation is Schechter of the form estimated S[-21,1.09]. Entirety of the PSCz was used for this analysis.

To begin with, while relative entropy once again promisingly returns an estimate for β that is in good agreement with our assumed true value, the general shape of the $S[(\zeta, \chi)]$ plots exhibited here are more of a parabola with a 'kink' in it, not entirely unlike their sample relative entropy counterparts presented near the end of Chapter 9. This in itself is in keeping with the different 'S-shaped' ρ curves that we observed with ROBUST when we modelled Schechter magnitudes to the PSCz in Chapter 6, wherein the typical nature of such an LF is what is contributing to the change in shape that we see. Secondly, the signal recovered by relative entropy here is stronger than its Gaussian counterpart - at least one order of magnitude stronger. This suggests once again that, like its ROBUST counterpart, the Schechter function is more receptive and sensitive to changes in β , making it ideal for use with relative entropy going forward, particularly when one considers that the modelled LF of the 2M++ survey that we are using in conjunction with the 2MRS is Schechter in nature. The third thing of note is our inability to recover a confidence interval on the mean estimate for the value of β , although this is clearly due to the lack of variance in plot minima as seen in Figure 10.10, all of which return a minimum at $\beta = 0.5$, consequently hindering Matlab's ability

Mock Set Size	1000	2000	5000	All (12000+)
Value of β	0.51 ± 0.09	0.51 ± 0.05	0.5	0.5

TABLE 10.4: Summary of estimated values of β detailing the effect of varying mock galaxy sample sizes on computed β using relative entropy during Mock Method 1b) implementation.

Schechter Power Law Slope α	0.5	1	2	3
Value of β	0.5	0.5	0.51 ± 0.04	0.53 ± 0.08

TABLE 10.5: Summary of estimated values of β detailing the effect of varying Schechter power law slope α on computed β using relative entropy during Mock Method 1b) implementation. Entirety of PSCz survey is utilised for this analysis.

to generate a histogram of β_{opt} with a best-fit Gaussian modelled to it. Additionally if we are wanting to identify a more accurate value for β_{opt} should this behaviour continue would necessitate the use of more trial values of β on the interval [0,1] which, due to time constraints is computationally exhaustive. We will take note of this sort of plot behaviour as we proceed to experiment with varying mock survey set sizes or altering the parameters of our modelled Schechter and see what we can make of it.

Our results from varying mock survey set size and altering our modelled Schechter power law slope are presented in Figures 10.11 and 10.13 respectively, alongside our standard monitoring of ζ - χ behaviour in Figures 10.12 and 10.14 respectively, with a summary of the estimates of β returned by relative entropy in both cases presented in Tables 10.4 and 10.5.

Starting with our varying of mock sizes, a lot of peculiarities of note begin to arise. While in the first instance relative entropy still succeeds in recovering estimates for β that are consistent with our assumed β_{true} , it very clearly struggles to return any values for $S[(\zeta, \chi)]$ for values of β larger than 0.6 when a mock survey set size of only 100 galaxies is used. Why this is is unclear, although it does render the β estimate returned by relative entropy for those mocks somewhat suspect since the entire range of trial values of β was not utilised. It could very well be that the small number of galaxies utilised during the analysis may be the cause, and this suspicion is played out when one considers the estimates returned on β for larger and larger mock set sizes where, not only is the entire range used, but the general shape and 'tightness' of the kinked parabolas in Figure 10.11 become better defined and a clear strong signal with minimal noise is observed. At the very least this lends further confidence to our assertion that



FIGURE 10.11: Method 1b) results with varied mock set sizes. Left-hand panels illustrate the $S[(\zeta, \chi)]$ plots for all 50 mocks, right-hand panels depict the histograms of optimal value of β with best-fit Gaussian (red line) when available. Set sizes used are noted on each plot. Luminosity function generated continues to be Schechter of form S[-21,1.09].



Histogram count

Histogram count Histogram count 100 40 50 20 0.8 0.8 0.5 0.5 0.6 0.6 0.4 0.4 0.2 0.2 0 0 χ value 0 χ value 0 ζ value ζ value

FIGURE 10.12: Monitoring of ζ - χ behaviour with varied mock set size. Panels depict the bivariate histograms of ζ - χ value distribution over [0,1]. The thick red line denotes our expected number galaxies per histogram bin for a uniform distribution.

relative entropy (just like ROBUST) will function more adequately the more galaxies it has on hand to analyse.

What is even more puzzling however is that our assumption that a Schechter function is more receptive or sensitive to changes in β when relative entropy is applied does not appear to be entirely true. While the overall signal returned by relative entropy continues to be stronger than its Gaussian counterpart, there is a very evident discrepancy in Figure 10.12 where the introduction of the χ statistic has introduced a very noticeable undersampling for all experimented mock survey set sizes that completely disrupt uniformity at low values of ζ across the entire range of χ values on [0,1]. This undersampling actually goes some way towards explaining the 'blanks' or gaps in the associated $S[(\zeta, \chi)]$ plots for these mock survey sizes, as the lack of galaxies inhabiting those particular grid squares on a ζ - χ grid would cause zeroes to appear in the final relative entropy calculations, which Matlab proceeds to interpret as a blank or disconnect on the final plots. Why this undersampling appears when relative entropy is used in conjunction with a Schechter function is unknown, though it is possible that the very nature of the function may be playing a role in disrupting uniformity when used in conjunction with



FIGURE 10.13: Method 1b) results with varied Schechter power law slope. Left-hand panels illustrate the $S[(\zeta, \chi)]$ plots for all 50 mocks, right-hand panels depict the histograms of optimal value of β with best-fit Gaussian (red line) when available. Power law slopes used are noted on each plot.



FIGURE 10.14: Monitoring of $\zeta - \chi$ behaviour with varied Schechter power law slope. Panels depict the bivariate histograms of $\zeta - \chi$ value distribution over [0,1]. The thick red line denotes our expected number galaxies per histogram bin for a uniform distribution.

the χ statistic. It is interesting to note however that despite the significant deviation from uniformity displayed here, relative entropy still manages to recover the value of β_{true} we are seeking, at the very least lending further credence to our suspicion that relative entropy is fairly robust and forgiving in its requirement for ζ - χ uniformity in order for it to operate effectively.

Altering our Schechter power law slopes unsurprisingly does nothing to alter the deviations from uniformity noted thus far as exhibited in Figure 10.14, though the recovered estimates for the value of β_{opt} continue to be consistent with our assumed β_{true} whenever a confidence interval can be calculated. The only thing of note to be taken away, and once again in keeping with results observed in Chapter 6, is that as we continue to increase the value of the power law slope α we see a gradual broadening in the shape of the kinked parabolas as seen in Figure 10.13, though since the entirety of the PSCz survey is being utilised for that analysis, the general tightness of the plots continues to be very well defined and clear, with a strong signal of $S[(\zeta, \chi)]$ continuing to be recovered throughout.

Bright Limit	0	5	8	10
Value of β	0.58 ± 0.32	0.58 ± 0.25	0.55 ± 0.27	0.56 ± 0.27

TABLE 10.6: Summary of estimated values of β detailing the effect of varying applied apparent bright magnitude limit on computed β using relative entropy during Mock Method 2a) implementation. Entirety of PSCz is utilised for this analysis with $\Delta Z=5$.

10.4 Applying Mock Method 2a) - 'Faint+Bright Limit Gaussian' Mock Magnitudes

With all of the above analyses having been performed for the case of mock magnitudes generated with a faint limit only, we will now proceed to repeat all of the above while introducing an arbitrary bright limit to the data and taking note of any changes observed in the values of β_{opt} returned by relative entropy for both mock Gaussian and Schechter magnitudes. Given the increased sensitivity that we have observed when mock Schechter magnitudes are modelled to the PSCz galaxies in particular, it will be interesting to see how the introduction of an arbitrary bright limit will alter the results returned in the ζ - χ plane, especially when one considers how much more sensitive a Schechter LF appears to be in the presence of a bright limit when analyses were run exclusively using ROBUST, unlike its Gaussian counterpart which appeared to be more forgiving.

Having already determined in Section 6.4.1 that the value of ΔZ applied when constructing the S_1 and S_2 sets required for ROBUST to run and generate ζ in the presence of a bright limit has no bearing on the final results returned, we will proceed to fix our value of ΔZ used in our calculations at 5, and vary our arbitrarily applied bright limit; taking note of any changes observed in the value of β returned by relative entropy. Starting with our mock Gaussian magnitude case, our initial results from applying an m_b of 8 to the data and applying relative entropy are presented in Figure 10.15.

Much like its faint limit counterpart in Section 10.2, relative entropy continues to successfully recover a mean estimate for β that is in good agreement with our assumed β_{true} , and offhandedly it does not appear that the introduction of an arbitrary bright limit had any apparent effect on the results returned by relative entropy at all. Whether this is indeed the case is presented in Figure 10.16 where we begin varying the applied bright limit, with Figure 10.17 continuing our standard procedure of monitoring the $\zeta - \chi$ behaviour of the galaxies in the presence of a bright limit, with a summary of all estimates recovered for the value of β presented in Table 10.6.



FIGURE 10.15: Initial results of Method 2a). Upper panel depicts the $S[(\zeta, \chi)]$ plots for all 50 mocks generated, lower panel depicts the minimum of each mock plot as a function of trial value of β plotted as a histogram with best-fit Gaussian (red line) modelled to the data. Optimal value of β returned for Method 2a) is $\beta = 0.55 \pm 0.27$, as determined from the Gaussian model. Luminosity function used during data generation is Gaussian of the form N[-21,1], with m_f and m_b given as 14.3 and 8 respectively. Bright limit was applied arbitrarily during S_1 , S_2 set generation.



FIGURE 10.16: Method 2a) results with varied bright limits. Left-hand panels illustrate the $S[(\zeta, \chi)]$ plots for all 50 mocks, right-hand panels depict the histograms of optimal value of β with best-fit Gaussian (red line). LF is of the form N[-21,1], with m_f and m_b noted on each plot. Entirety of the PSCz is utilised during analysis with $\Delta Z=5$.



FIGURE 10.17: Monitoring of $\zeta - \chi$ behaviour with varied bright limit. Panels depict the bivariate histograms of $\zeta - \chi$ value distribution over [0,1]. The thick red line denotes our expected number galaxies per histogram bin for a uniform distribution.

The results returned in Figure 10.16 continue to be promising, with relative entropy successfully recovering mean estimates for β that are consistent with all that has come before, though there is something interesting to note with regards to the application of an $m_b=10$ in particular. Recalling from the lower panels of Figure 6.16 in Chapter 6 where an $m_b=10$ was applied to the data when we were analysing exclusively with ROBUST, we observed a noticeable shift in the $\rho(\zeta,\beta)$ plots, causing a value of β to be generated that was not consistent with β_{true} and yet here, when relative entropy is applied, we observe more of an increase in the general noise of the data than we see a noticeable shift in the range of the $S[(\zeta, \chi)]$ plots themselves. The source of this noise becomes all the more apparent in the lower panels of Figure 10.17 where, for an applied $m_b=10$, we begin to observe a clear undersampling of galaxies in the ζ - χ plane across all values of ζ for low values of χ that disrupts uniformity in that area. On the one hand this is further indication that this particular bright limit is unnecessary as it appears to be encroaching on the distribution of galaxies on the M-Z plane as discussed previously (and indeed further reinforces our belief that the PSCz is more than adequately modelled with a faint limit alone), and yet despite this disruption in uniformity relative entropy continues to show its robustness by succeeding in recovering the value of β we are looking



FIGURE 10.18: Initial results of Method 2b), depicting the $S[(\zeta, \chi)]$ plots for all 50 mocks generated. Optimal value of β returned for Method 2b) is $\beta = 0.5$, as determined from analysing plot minima. Luminosity function used during data generation is Schechter of the form estimated S[-21,1.09], with m_f and m_b given as 14.3 and 7 respectively. Bright limit was applied arbitrarily during S_1 , S_2 set generation. Entirety of the PSCz was used for this analysis.

for. This makes relative entropy as a statistical tool all the more useful for us as it has proved that it can be used in conjunction with ROBUST to not only strengthen the results that ROBUST returns on its own, but to potentially also correct some of the deviated results that it generates due to the use of incorrect limits etc..

10.5 Applying Mock Method 2b) - 'Faint+Bright Limit Schechter' Mock Magnitudes

Given the very sensitive behaviour we have observed prior to this when modelling mock Schechter magnitudes to the PSCz galaxies, introducing arbitrary bright limits to the data and running relative entropy should prove interesting. Once again having fixed our value of $\Delta Z=5$, we proceed to apply an arbitrary $m_b=7$ to our Schechter data as we did in Chapter 6, the results of which are presented in Figure 10.18.

Bright Limit	0	7	8	10
Value of β	0.5	0.5	0.5	0.5

TABLE 10.7: Summary of estimated values of β detailing the effect of varying applied apparent bright magnitude limit on computed β using relative entropy during Mock Method 2b) implementation. Entirety of PSCz is utilised for this analysis with $\Delta Z=5$.



FIGURE 10.19: Method 2b) results with varied bright limits. Panels illustrate the $S[(\zeta, \chi)]$ plots for all 50 mocks. LF is of the form S[-21,1.09], with m_f and m_b noted on each plot. Entirety of the PSCz is utilised during analysis with $\Delta Z=5$.

We continue to see much of the same behaviours that we noted previously: while relative entropy successfully recovers an estimate for β that is consistent with our assumed β_{true} , with our $S[(\zeta, \chi)]$ plots exhibiting a far stronger signal than their Gaussian counterparts, the return of the 'kinked' parabolas renders it difficult to establish a sensible confidence interval on our returned value of 0.5, as noted by the lack of histogram in Figure 10.18. What is not inherently clear however is whether the introduction of that arbitrary bright limit of $m_b=7$ did indeed have any effect on the functioning of relative entropy, which we proceed to explore in Figure 10.19 where we begin to vary our applied bright limit, with Figure 10.20 presenting our monitoring of $\zeta - \chi$ behaviour throughout, and Table 10.7 displaying a summary of the estimates of β recovered.



FIGURE 10.20: Monitoring of $\zeta - \chi$ behaviour with varied bright limit. Panels depict the bivariate histograms of $\zeta - \chi$ value distribution over [0,1]. The thick red line denotes our expected number galaxies per histogram bin for a uniform distribution.

The results returned by relative entropy are very telling. While we continue to return estimates for β that are consistent with β_{true} when one considers the number of trial values of β being utilised on the interval [0,1], the shape of the returned $S[(\zeta, \chi)]$ plots once again render it difficult to sensibly determine a confidence interval on β_{opt} . What is more interesting to note however, is that the use of arbitrary bright limits that are brighter than an m_b of 10 once again do not appear to have any appreciable impact on the functioning of relative entropy. As can be noted in Figure 10.20, while the introduction of those bright limits has not had any effect on the clear undersampling seen in the left-hand panels, their presence has not worsened the undersampling either, as relative entropy continues to function as normal. As before, the undersampling presented in these $\zeta - \chi$ plots is also responsible for the previously discussed 'blanks' or plot disconnects seen at higher values of β_{trial} in Figure 10.19. Looking at the particular case of $m_b=10$, while the relevant $S[(\zeta, \chi)]$ plots at that limit show no aberrant behaviour, the associated $\zeta - \chi$ plot at that limit (lower panels of Figure 10.20) not only exhibits the aforementioned undersampling but also starts to show indication of oversampling, noted by the diagonal clustering of points on the ζ - χ plane at low values of both ζ and χ . This would once again appear to suggest that, much like its Gaussian counterpart in both this chapter and in

all instances in Chapter 6, an arbitrary limit of $m_b=10$ is encroaching too much upon the distribution of galaxies on the M-Z plane and is consequently creating a disruption from uniformity, although it should be noted that this deviation does not appear to be significant enough in either instance to prevent relative entropy specifically from operating adequately, unlike ROBUST exclusively. This lends further credence to our belief that when used in conjunction with ROBUST, relative entropy not only provides the means to further prove upon the results that it returns on its own, but also manages to help successfully recover the value of β we are looking for even in situations wherein ROBUST would find it difficult when working alone.

In summary as before, throughout all of the mock realisations preformed in this chapter with different luminosity functions and the varying of their associated parameters, we have managed to established that relative entropy serves as a very useful ancillary statistical tool to be used in conjunction with ROBUST to recover our assumed true value of $\beta = 0.55 \pm 0.06$ for the PSCz, and recovering it well the more galaxies are available for analysis. The size of available mock survey sets continues to prove crucial particularly when broad luminosity function parameters are used as it serves to help counteract the potential biases introduced from using such poorly defined functions. Much like what was established in Chapter 6 relative entropy still requires the identification and use of the correct magnitude limits for the survey, though it has proved itself decidedly more forgiving of the use of potentially incorrect bright limits, unlike its ROBUST counterpart which is far more sensitive to the bias introduced when being utilised exclusively on its own, irrespective of the kind of luminosity function modelled to the PSCz. In addition, we have determined that the accuracy of relative entropy is not only just reliant on the number of available values of β_{trial} on the interval [0,1] for it to use, but also on the applied mesh size or grid resolution applied to the ζ - χ grid when constructing our $S[(\zeta,\chi)]$ statistic. Finding the balance between a poorer recovered signal (too large a resolution) and a signal overly dominated by shot noise and Poisson fluctuations (too fine a resolution) is paramount in order for relative entropy to operate effectively.

With all of this information in hand, we will now proceed to work directly with the 2MRS and apply relative entropy to determine whether at the very least we can confirm our ROBUST exclusion intervals on β in the KHJ bands or whether indeed by its mere inclusion, manage to successfully recover a value for β_{opt} that is in good agreement with the published Carrick et. al value of $\beta = 0.43 \pm 0.021$ for the survey.
Chapter 11

Applying Relative Entropy to 2MASS

Having now established the optimal conditions necessary for relative entropy to function adequately within a mock environment, we will now proceed to apply relative entropy to the 2MRS, and see if the results returned help to either confirm or exclude the various experimentations and restrictions we applied to the survey when we merely used ROBUST and the correlation coefficient $\rho(\zeta,\beta)$ on their own. On the basis of the results generated in the sections going forward, we will be in a better position to determine whether there indeed is an inherent fault in (or peculiar properties to) the 2MRS that prevent it from being used ideally with either ROBUST exclusively or ROBUST in conjunction with relative entropy, or whether indeed we need to go back and examine the fundamental principles upon which ROBUST itself is defined. A flowchart illustrating the methodology we will apply going forward is presented in Figure 11.1.

For each run of relative entropy completed, we will constrain our errors on β as before using Monte Carlo simulations (bootstrap resampling is to be foregone in this chapter due to the size of the survey at hand and the number of computations over the ζ - χ grid required for it to generate our required confidence intervals on $S[(\zeta, \chi)]$ being computationally exhaustive and time consuming), and monitor the ζ - χ behaviours of the galaxies over each, noting any peculiarities along the way.

11.1 Redshift Restricting the 2MRS

Having already established in Section 8.6.1 that applying an observed redshift restriction of $500 \le cz \le 15000$ substantially improves the signal returned by ROBUST and



FIGURE 11.1: Flowchart for the application of data slicing, distance convergence loops and the ROBUST methodology in conjunction with relative entropy for the 2MRS survey.



FIGURE 11.2: Plots of $S[(\zeta, \chi)]$ for the redshift restricted 2MRS survey across the K, H and J infrared bands, shown in blue, red and green respectively.

minimises the observed noise at lower levels of β , we will forego applying relative entropy to the entire unrestricted survey as we did initially in Chapter 8 and start directly with the restricted variant. The results of running relative entropy on this restricted survey with a previously established grid mesh size of 10 to be used during the construction of the ζ - χ grid is presented in Figure 11.2.

The results presented here are curious for several reasons. To start, due to the unusual generally monotonic increase in the plots of $S[(\zeta, \chi)]$ in all three wavebands, relative entropy returns a value of β_{opt} that is consistent with zero for the survey, which is neither in keeping with the low values of $\beta \sim 0.05$ returned by ROBUST exclusively nor the published results of Carrick. Furthermore it is not in keeping with established theory on how the linear redshift distortion parameter itself is defined and the value it is expected to take. In addition, considering the Schechter origin of the underlying 2M++ grid being utilised here (as discussed in Section 8.2) we were expecting to see a general plot shape not unlike the 'kinked' parabolas observed in Chapter 10. It is however interesting to note that relative entropy starts to exhibit the same strange behaviour as noted by ROBUST for values of $\beta > 0.6$, where we begin to observe a noticeable increase in the noise of the plots (though not nearly as pronounced as their ROBUST-only counterparts), and quite a significant spike in the data when that limit of 0.6 is



FIGURE 11.3: One-sided confidence intervals calculated on relative entropy produced β via Monte Carlo simulations for the K, H and J bands depicted in blue, red and green respectively. The 3σ confidence intervals for all three bands are denoted via the black dashed lines, indicating all values of $\beta > 0$ are to be excluded from consideration for the redshift restricted 2MRS survey in the K, H and J bands with 99.7% confidence.

exceeded. This behaviour is once again mirrored in the ζ - χ distributions of the galaxies for various values of β_{trial} , as presented in Figure 11.4, where for values of β_{trial} larger than 0.6 we start to see the beginnings of noticeable oversampling at very low values of both ζ and χ , which would consequently introduce noise into the signal of $S[(\zeta, \chi)]$ recovered by relative entropy.

Due to the peculiar plots returned by relative entropy, when one proceeds to apply Monte Carlo simulations to the data to determine the confidence (or exclusion) intervals on our value of β_{opt} , all values on the interval [0,1] are automatically excluded, as can be noted in Figure 11.3, which makes sense given the nature of the $S[(\zeta, \chi)]$ plots generated. While the inability of relative entropy to return results that are consistent with ROBUST does not entirely negate the values of β_{opt} returned in Section 8.6.1, particularly when one takes note of the similar behaviour between the two methods for all values of $\beta > 0.6$, it does significantly lessen their reliability, consequently making the possibility that there is an inherent fault in the 2MRS or in our ROBUST methodology all the more likely.



FIGURE 11.4: Monitoring of (ζ, χ) distribution of the redshift restricted 2MRS galaxies in the K band for various trial values of β . The trial values of β used are noted on each panel.

11.2 Restricting cz_{corr} instead of cz_{obs}

While the initial ROBUST results returned when we opted to use cz_{corr} instead of cz_{obs} were dubious at best, we can still make use of relative entropy to determine whether that avenue of experimentation should be dropped from consideration entirely from this point going forward should the program return results that are just nonsensical and dubious. The results of running relative entropy on the restricted corrected redshift variant of the 2MRS is presented in Figure 11.5, and the plots speak for themselves.

Relative entropy once again returns a value of β_{opt} for the 2MRS across all three wavebands that is consistent with a β of zero, and exhibits different behaviour to its counterpart in the previous Section. The noise in the recovered $S[(\zeta, \chi)]$ signals (while similar to its ROBUST counterpart in appearance) is constant throughout, with no alteration in behaviour as such for values of $\beta > 0.6$, manifesting instead two significant spikes in the data at β values of ~ 0 and ~ 0.3 . This is all the more apparent when one monitors the ζ - χ behaviour of the galaxies over various values of β_{trial} as presented in Figure 11.6, where consistent oversampling at low values of ζ and χ can be observed throughout.



FIGURE 11.5: Plots of $S[(\zeta, \chi)]$ for the restricted 2MRS survey utilising corrected redshifts across the K, H and J infrared bands, shown in blue, red and green respectively.



FIGURE 11.6: Monitoring of (ζ, χ) distribution of the restricted 2MRS galaxies utilising corrected redshifts in the K band for various trial values of β . The trial values of β used are noted on each panel.



FIGURE 11.7: One-sided confidence intervals calculated on relative entropy produced β via Monte Carlo simulations for the K, H and J bands depicted in blue, red and green respectively. The 3σ confidence intervals for all three bands are denoted via the black dashed lines, indicating all values of $\beta > 0$ are to be excluded from consideration for the cz_{corr} restricted 2MRS survey in the K, H and J bands with 99.7% confidence.

When one combines this once again with exclusion intervals generated via Monte Carlo simulation (as presented in Figure 11.7) which once again exclude all values of β on the interval [0,1] from consideration when cz_{corr} is used, this helps to further prove that we are better served in continuing to make use of cz_{obs} going forward as it behaves better and as such we will discontinue considering the use of cz_{corr} in any ROBUST or relative entropy analyses going forward.

11.3 Experimenting with μ instead of u_{β}

While from the get go the initial results returned by the use of cz_{corr} during analysis with ROBUST were dubious, experimenting with μ_{β} yielded some more promising results, in that we obtained the monotonically increasing plots of $\rho(\zeta, \beta)$ that we were looking for, albeit with no zero-intercepts returned to give us a value for β_{opt} . Considering the observed corrective or strengthening nature of relative entropy when it was applied to the PSCz in Chapter 10 and how it served to enhance the robustness of our methodology, it will be interesting to see the results it returns here when applied to our experimentation



FIGURE 11.8: Plots of $S[(\zeta, \chi)]$ for the redshift restricted 2MRS survey where β -dependent distance moduli μ_{β} were utilised in the calculation of $\rho(\zeta, \beta)$ instead of β -dependent peculiar velocities u_{β} across the K, H and J infrared bands, shown in blue, red and green respectively.

with μ_{β} . Figure 11.8 shows the results returned by relative entropy, and they continue to be both reassuring and puzzling.

On the one hand while the return of $S[(\zeta, \chi)]$ plots that are quite similar to those generated by relative entropy in Section 11.1 further reinforces our conclusion that we were correct to dismiss the use of cz_{corr} in any further analyses, its continued inability to recover a value of β_{opt} that is in agreement with ROBUST or the published Carrick value is puzzling, if annoyingly consistent, returning an optimal β of zero once again. We continue to observe the same spike in the data and increase in the noise of the recovered signal for values of $\beta > 0.6$ that is once again mirrored in the ζ - χ distribution of the galaxies over various values of β_{trial} on the interval [0,1] in Figure 11.9, where for values of $\beta > 0.6$ we begin to observe a very significant oversampling at low values of ζ and χ that consequently contributes to the noise observed in Figure 11.8.

The computed confidence intervals via Monte Carlo simulations are also annoyingly consistent, once again excluding all values of β on the interval [0,1] from consideration with the 2MRS survey, as seen in Figure 11.10. Considering the consistent behaviour of relative entropy across both these instances, our assumption that there is an inherent



FIGURE 11.9: Monitoring of (ζ, χ) distribution of the redshift restricted 2MRS galaxies where μ_{β} was in the calculation of $\rho(\zeta, \beta)$ instead of u_{β} in the K band for various trial values of β . The trial values of β used are noted on each panel.

fault in the survey or something fundamentally wrong with the core principles upon which ROBUST is founded is suddenly less clear.

In summary, despite the encouraging corrective and result-strengthening behaviour exhibited by relative entropy in the previous chapter where it was able to 'pick up the slack' where ROBUST on its own struggled when applied to the PSCz in multiple mock realisations, the program has consistently failed to return any value for β_{opt} that is consistent either with ROBUST in Chapter 8 or with the published value of Carrick, although the consistency of the results returned by relative entropy throughout needs to be noted, especially with regards to showing the same rate of noise in the recovered signal of $S[(\zeta, \chi)]$ for larger values of β . At the very least this lends credence to the idea that the exclusion intervals returned by ROBUST in that regard are valid, while the significantly different behaviour of relative entropy for the case of cz_{corr} further strengthens our belief that that parameter should not be used at all going forward. Unfortunately relative entropy's inability to return any value of β other than 0 for the 2MRS leaves us no closer to determining whether the fault lies in the 2MRS itself or within ROBUST somewhere.



FIGURE 11.10: One-sided confidence intervals calculated on relative entropy produced β via Monte Carlo simulations for the K, H and J bands depicted in blue, red and green respectively. The 3σ confidence intervals for all three bands are denoted via the black dashed lines, indicating all values of $\beta > 0$ are to be excluded from consideration for the restricted 2MRS survey in the K, H and J bands with 99.7% confidence when μ_{β} is utilised instead of u_{β} .

In order for us to truly be able to determine whether the latter is indeed the case we now need to turn our attention to one of the key assumptions upon which the ROBUST ζ statistic is constructed: namely that the luminosity of a galaxy is independent of its real-space position.

Chapter 12

Experimenting with Distance Dependent Luminosity Functions

In recalling the underlying theory upon which ROBUST is founded in Section 4.1 we established, for a given redshift-distance survey of galaxies that is required to be **complete** up to a given limiting magnitude m_{lim} ; or pair of limiting magnitudes m_{lim} and m_{bright} , that the distribution of absolute magnitudes M, i.e. its luminosity function F(M) is independent of the spatial positions of the galaxies. Put another way, we make the assumption that the luminosity function modelled to a given redshift-distance galaxy survey is **universal** across the whole survey. When one recalls the cosmological principle as discussed in Section 1.2 this assumption makes sense. Furthermore the advantage of applying such an assumption to a given survey leads to the construction of the ζ statistic upon which ROBUST is reliant and its two key properties, repeated here for ease of access:

- 1. ζ will be uniformly distributed on the interval [0,1] in the presence of the correct magnitude limits,
- 2. ζ and the distance moduli of galaxies μ are statistically independent, i.e. the distribution of ζ is independent of the spatial distribution of the galaxies in the survey, in keeping with the assumption made about the 'universality' of the modelled luminosity function [160].

With regards to the former property, this has granted us the ability to usefully determine the true magnitude limits of a survey by means of exploiting ζ in conjunction with the completeness statistic T_c , while the latter property has been invaluable in determining the optimal value of the linear redshift distortion parameter β when used in conjunction with the correlation coefficient $\rho(\zeta, u_{\beta})$, wherein which the value for β that corresponds to $\rho = 0$ (i.e. that ζ and the β -dependent reconstructed peculiar velocities of galaxies are independent from each other) is taken to be the optimal value of β for the survey.

An interesting question is raised from all of this however, and it relates back to the cosmological principle. While the cosmological principle holds on the macro scale with the Universe looking the same regardless of the direction one chooses to observe along, or regardless of WHERE in the Universe one chooses to observe from; over shorter scale distances, say on the order of several Mpc or smaller, deviations from homogeneity and isotropy begin to occur, primarily fuelled by variations in the local peculiar velocity and matter density fields of the Universe, causing matter to clump and cluster in filaments, voids and other cosmological structures as described briefly in Chapter 1. When one takes a moment to consider these clustering and clumpings of matter and how they would manifest in a typical survey, it is not unreasonable to wonder whether the assumption of a universal luminosity function over said survey is completely valid given the more non-uniform distribution of luminous matter.

The consequence of this is critical for the functioning of ROBUST, given the reliance of the statistic ζ on its independence from μ and consequently from the luminosity function F(M). Should F(M) not be universal, then this would introduce a very noticeable bias into the functioning of ROBUST, making it more difficult for the program to efficiently recover any value for β , mock or otherwise. To that end, and given the peculiar results returned by ROBUST for the 2MRS in previous chapters both in isolation and when used in conjunction with relative entropy, should it become apparent that the luminosity function modelled to that survey is not indeed universal then we are at least in a better position to rule out a fundamental flaw in the ROBUST methodology being the cause of the strange results and instead lend credence to the theory that the 2MRS is inherently flawed for use with ROBUST to begin with.

To determine whether this is indeed the case, we will proceed to make use of the base mock methodologies we applied to the IRAS PSCz survey in Chapter 6 and generate mock magnitudes to the positions of the 2MRS galaxies that are drawn from both a universal and a distance-dependent luminosity function and observe the results returned by ROBUST and relative entropy in both instances. Should the universal case return results similar to those presented in Chapter 8 and Chapter 11, then we can rest assured that our underlying assumption of 'universality' is valid and ROBUST is not inherently flawed as a statistical tool. Conversely, should the distance-dependent case return results for β_{opt} that are in better agreement with the published results of Carrick, then we can either at the very least rule out the use of the 2MRS survey with ROBUST and instead turn our attention to other available redshift-distance surveys that can be used, such as the SDSS.

As such, the rest of this chapter will be structured as follows: after defining both the universal and distance-dependent luminosity functions we will model, Section 12.1 will deal with applying mock distance-dependent apparent magnitudes to both the IRAS PSCz survey (to observe how well the results returned by ROBUST compare with those returned by the program in Chapter 6 where universal Gaussian and Schechter functions were modelled), and the redshift restricted 2MRS survey alongside its variant where μ_{β} was utilised in the construction of the ζ statistic instead of the β -dependent peculiar velocities u_{β} . Section 12.2 will then proceed to repeat the analyses on the 2MRS survey variants utilising mock generated universal magnitudes instead, and the results of both sets of results from both Sections will be analysed and discussed. Throughout we will also continue to make use of relative entropy and observe how its behaviour alters when a distance-dependent luminosity function is introduced.

It should be noted that due to the use of relative entropy in all of the analyses going forward and the computationally exhaustive and time consuming nature of the error bar creation for those results with regards to bootstrap resampling, all confidence and exclusion intervals generated on our $\rho(\zeta, \beta)$ and $S[(\zeta, \chi)]$ plots from this point onward will be generated using Monte Carlo simulations only. Given the extreme similarity in confidence intervals returned by both bootstrap resampling and Monte Carlo simulations in earlier chapters of this work, this is not foreseen to introduce any noticeable issues.

With regards to generating the necessary luminosity functions to draw apparent magnitudes from, we can continue to make use of the **normrnd** function that was used in Chapter 6 to generate mock magnitudes to the IRAS PSCz. Specifically **normrnd=N[mu, sigma]** allows for the use of functions for the definition of **mu** as well as integers, consequently we will continue to model a universal Gaussian to the 2MRS of the form N[-21,1] in keeping with the typical luminosity function of early-type SDSS galaxies [9]. For our distance-dependent case, we will model a simple linear relation where the brightness of the galaxies decreases with increasing distance such that, say, galaxies at a restricted redshift distance of 500kms⁻¹ have an apparent magnitude of -22, while galaxies at a restricted redshift distance of 15000kms⁻¹ have an apparent magnitude of -21. This yields a simple line equation that can be easily substituted into **mu** to generate our required apparent magnitudes while bearing in mind our pre-established faint limits for the IRAS PSCz and 2MRS surveys respectively.

12.1 Applying a Distance Dependent Luminosity Function

12.1.1 IRAS PSCz Survey

Having already established all we need in order to proceed, we will now generate mock distance-dependent apparent magnitudes to the IRAS PSCz galaxies and apply both ROBUST and relative entropy, taking note of the results for β returned by both instances. The results are presented in Figure 12.1 and are very interesting for several reasons.

Both ROBUST and relative entropy fail to return a mean estimate for the value of β_{opt} that is remotely consistent with our assumed true value of β for the PSCz (namely $\beta = 0.55 \pm 0.06$), with the former being consistently negative and parabolic in nature across all trial values of β on the interval [0,1], and the latter generally returning a value for β_{opt} that is consistent with unity which, as we have established previously, is not valid. This appears to suggest that even the introduction of a very slight distance-dependence in the form of our simple linear relation is more than enough to introduce significant bias into the functioning of both ROBUST and relative entropy. It is interesting to note however that the general minima of the ROBUST parabolas in the upper panel of Figure 12.1 are somewhat consistent with a value of $\beta = 0.5$, though considering how the parameter $\rho(\zeta, \beta)$ is constructed and its reliance on a zero-intercept to determine β_{opt} , this is likely not of any material significance.

Due to the behaviour exhibited by both ROBUST and relative entropy, getting a handle on the confidence interval for β is also equal parts difficult (given the lack of zerointercepts) and futile considering the level of bias introduced into the results, as such we will forego calculating them here and, now aware of the sensitivity of ROBUST to distance-dependent luminosity functions, cautiously proceed with applying the same function to the 2MRS galaxies and observe the changes in the values of β returned by both ROBUST and relative entropy for that survey.

12.1.2 Redshift Restricted 2MRS

The results of introducing the same linear distance dependence to the functioning of both ROBUST and relative entropy are presented in Figure 12.2 and are equal parts reassuring and puzzling.

On the one hand, the basic shape of the $\rho(\zeta, \beta)$ plots generated by ROBUST are very similar to their Chapter 8 counterparts in Section 8.6.1, exhibiting the same parabolic behaviour for low values of β and increased noise in the recovered signal for larger values



FIGURE 12.1: Results of applying mock distance-dependent apparent magnitudes to the positions of the PSCz galaxies on the value of β_{opt} as returned by ROBUST (upper panel) and relative entropy (lower panel) over several trial values of β on the interval [0,1] over 50 mock realisations. Grid mesh size applied for calculating relative entropy set at a value of 10 over all trials.



FIGURE 12.2: Results of applying mock distance-dependent apparent magnitudes to the positions of the redshift restricted 2MRS galaxies on the value of β_{opt} as returned by ROBUST (upper panel) and relative entropy (lower panel) over several trial values of β on the interval [0,1] for the K (blue), H (red) and J (green) infrared bands respectively. Grid mesh size applied for calculating relative entropy set at a value of 10.



FIGURE 12.3: Monitoring of (ζ, μ) distribution of the redshift restricted 2MRS galaxies in the K band for various trial values of β when a distance-dependent luminosity function is applied. The trial values of β used are noted on each panel.

of β , but the turnover point at which this change happens has lowered significantly from $\beta = 0.6$ to $\beta \sim 0.4$, most likely due to the introduction of our slight distance dependence. Furthermore we now have estimated values for β_{opt} in all bands returned as $\beta_{opt} \sim 0.31$, $\beta_{opt} \sim 0.34$ and $\beta_{opt} \sim 0.3$ for the K, H and J bands respectively which, while still not in good agreement with the published results of Carrick are a vast improvement over the results returned by ROBUST using the raw data in Chapter 8. Whether we can infer from this that the luminosity function of the 2MRS survey can indeed be modelled by a distance-dependent function is yet to be made clear until we ratify these results with those we shall present in Section 12.2 when we model a universal function to the data. In the meantime the observation of increased noise in the recovered $\rho(\zeta, \beta)$ signal is mirrored in the (ζ, μ) distribution of the galaxies as presented in Figure 12.3 where for all values of $\beta > 0.4$ we see the return of the 'packed' distributions that signify that the β -dependent peculiar velocities of the 2MRS galaxies are causing them to move to inordinately close distance moduli relative to the observer.



FIGURE 12.4: Monitoring of (ζ, χ) distribution of the redshift restricted 2MRS galaxies in the K band for various trial values of β when a distance-dependent luminosity function is applied. The trial values of β used are noted on each panel.

What is even more striking than the results returned by ROBUST however are the extremely noisy results returned by relative entropy across all three bands, which are completely inconsistent with their Chapter 11 counterparts in Section 11.1. While it is possible that some of this noise is due to the randomising nature of the **normrnd** function being used to generate the mock apparent magnitudes being assigned to the galaxies the fact that similar noise is not manifesting in the relative ROBUST plots renders this unlikely. It is more likely that relative entropy itself is far more sensitive to the use of a distance-dependent LF than ROBUST appears to be given the similar plots the latter has returned, and this is further reinforced in the ζ - χ behaviour of the galaxies when monitored over various trial values of β on the interval [0,1] as presented in Figure 12.4. Once again for all values of β larger than 0.4 we begin to exhibit a significant oversampling at low values of both ζ and χ , which contributes to the level of noise observed in the recovered $S[(\zeta, \chi)]$ signals.

It should be noted however that despite the very significant level of noise exhibited by relative entropy, it does show signs of a 'spike' most clearly in the J band at the same turnover point where the ROBUST plots begin to exhibit the same level of noise at around a β of 0.4 which, when cross-referenced with our calculated exclusion intervals on β using Monte Carlo simulations as presented in Figure 12.5 results in all values of $\beta > \sim 0.3$ being excluded from consideration with this variant of the survey. While on the one hand this is once again in keeping with the results presented by VELMOD and other more contemporary methods as previously discussed in earlier chapters, and helps to further reinforce the rejection of the results returned by POTENT favouring a β of unity, the exclusion of the Carrick value of $\beta = 0.43 \pm 0.021$ is concerning given how well the $\rho(\zeta, \beta)$ plots for the 2MRS appear to match.

12.1.3 2MRS: Using μ_{β} instead of u_{β}

The results of applying the same distance-dependent LF to the μ_{β} variant of the 2MRS and running both ROBUST and relative entropy are presented in Figure 12.6 and are curious for several reasons.

Unlike the previous section when compared with their counterparts in Section 11.3, the plots returned by ROBUST when a distance-dependent LF is applied are only similar to a point, specifically where the 'noise turnover' occurs when the $\rho(\zeta,\beta)$ plots stop exhibiting their monotonic increase over trial values of β on [0,1] and instead begin to fall off as the noise increases in the recovered signal. It should be noted however that the noise turnover point exhibited is the same, manifesting at a $\beta \sim 0.4$ which is at least consistent behaviour. Conversely much like their Section 11.3 counterparts we continue to recover no values for β_{opt} with the exception of the K band, which recovers a value of β_{opt} consistent with the $\beta \sim 0.05$ that we observed in earlier sections of Chapter 8. This lends itself to one of two possibilities: firstly that, much like our (now discontinued) use of cz_{corr} , perhaps the use of μ_{β} in conjunction with ROBUST and relative entropy is not conducive for our purposes despite the theory suggesting that its use should be fine or secondly, that once again the introduction of a slight distance-dependence to the performed analysis is introducing a significant bias into the calculations. The ramifications of the second alternative are interesting, considering our running hypothesis that the 2MRS survey may be best modelled by a distance-dependent function given the results of the previous section. Once again however until we model a universal function to the data for both variants we cannot say for certain what the most likely scenario is.

As before, the increased noise noted in the $\rho(\zeta, \beta)$ plots is mirrored in their observed (ζ, μ) distributions over various trial values of β as presented in Figure 12.7, where once again for all trial values of β larger than 0.4 we see the return of the 'packed' distributions indicative of significant galaxy shifting to implausibly small distance moduli as the value



FIGURE 12.5: One-sided confidence intervals calculated on β via Monte Carlo simulations for both ROBUST (upper panel) and relative entropy (lower panel) for the K, H and J bands depicted in blue, red and green respectively. The 3σ confidence intervals for all three bands are denoted via the black dashed lines, indicating all values of $\beta > \sim 0.3$ are to be excluded from consideration for the redshift restricted 2MRS survey with 99.7% confidence.



FIGURE 12.6: Results of applying mock distance-dependent apparent magnitudes to the positions of the redshift restricted 2MRS galaxies where μ_{β} is utilised instead of u_{β} on the value of β_{opt} as returned by ROBUST (upper panel) and relative entropy (lower panel) over several trial values of β on the interval [0,1] for the K (blue), H (red) and J (green) infrared bands respectively. Grid mesh size applied for calculating relative entropy set at a value of 10.



FIGURE 12.7: Monitoring of (ζ, μ) distribution of the μ_{β} reliant redshift restricted 2MRS galaxies in the K band for various trial values of β when a distance-dependent luminosity function is applied. The trial values of β used are noted on each panel.

of β is increased, causing deviations from uniformity and a subsequent increase in the noise of the signal recovered by ROBUST.

It is curious to note that once again, relative entropy struggles to return plots in all bands that are remotely consistent with their counterparts in Chapter 11 though they are quite similar to those presented in the previous section, as the $S[(\zeta, \chi)]$ plots are completely dominated by noise; exhibiting the same 'spike' at a noise turnover point of $\beta \sim 0.4$ that is consistent with that seen in ROBUST. While again it is possible that the **normrnd** function is contributing to this noise, a glance at the ζ - χ behaviour of the 2MRS galaxies across trial values of β as seen in Figure 12.8 once again makes it clear that the introduction of a slight distance-dependence is more than enough to cause a very significant oversampling to occur at low values of ζ and χ for values of $\beta > 0.4$, resulting in additional noise being exhibited in the plots returned by relative entropy.

When we compute our exclusion intervals on β using Monte Carlo simulations as presented in Figure 12.9, all values of β are excluded from consideration for the H and J bands, and all values of $\beta > 0.1$ in the K band are excluded, rendering the value returned by ROBUST for that band redundant. This inconsistency of behaviour lends





FIGURE 12.8: Monitoring of (ζ, χ) distribution of the μ_{β} reliant redshift restricted 2MRS galaxies in the K band for various trial values of β when a distance-dependent luminosity function is applied. The trial values of β used are noted on each panel.

itself further to the idea that perhaps μ_{β} , much like cz_{corr} , is not valid for use with either ROBUST or relative entropy, though once we see the results when a universal LF is modelled we will know for certain.

12.2 Applying a Universal Luminosity Function

12.2.1 Redshift Restricted 2MRS

Having chosen to model the same universal mock Gaussian LF of N[-21,1] that we applied to the IRAS PSCz in Chapter 6, the results of applying such a universal LF to the redshift restricted 2MRS are presented in Figure 12.10 and are very promising.

We continue to see general shapes for the $\rho(\zeta,\beta)$ plots returned by ROBUST that are similar not only to their counterparts in Chapter 8, but also to their distance-dependent counterparts in this chapter, although admittedly with regards to both this case and the latter the noticeable shift downwards and appearance of multiple zero-intercepts must be noted. Of these intercepts we can recover values for β_{opt} such that $\beta = 0.32$, $\beta = 0.325$



FIGURE 12.9: One-sided confidence intervals calculated on β via Monte Carlo simulations for both ROBUST (upper panel) and relative entropy (lower panel) for the K, H and J bands depicted in blue, red and green respectively. The 3σ confidence intervals for all three bands are denoted via the black dashed lines, indicating all values of $\beta > \sim 0.1$ for the K Band only are to be excluded from consideration for the μ_{β} reliant redshift restricted 2MRS survey with 99.7% confidence.



FIGURE 12.10: Results of applying mock universal apparent magnitudes to the positions of the redshift restricted 2MRS galaxies on the value of β_{opt} as returned by ROBUST (upper panel) and relative entropy (lower panel) over several trial values of β on the interval [0,1] for the K (blue), H (red) and J (green) infrared bands respectively.

Grid mesh size applied for calculating relative entropy set at a value of 10.



FIGURE 12.11: Monitoring of (ζ, μ) distribution of the redshift restricted 2MRS galaxies in the K band for various trial values of β when a universal luminosity function is applied. The trial values of β used are noted on each panel.

and $\beta = 0.3$ for the K, H and J bands respectively, which are in very good agreement with the distance-dependent results recovered earlier, though still not consistent with the published value of Carrick. We continue to observe the same 'noise turnover point' at a β of 0.4 which is likewise mirrored in the (ζ, μ) distributions of the galaxies as seen in Figure 12.11, where the 'packed' distributions of points are observed for all values of β larger than 0.4.

The behaviour of relative entropy continues to be predictable as seen in the lower panel of Figure 12.10, returning very noisy results across all three bands that fail to recover a viable value for β_{opt} and continue to exhibit the same clear spike in the returned $S[(\zeta, \chi)]$ plots (seen most clearly once again in the J band) at the same noise turnover point of $\beta = 0.4$. After this point larger values of β start to return a clear oversampling of galaxies on the ζ - χ plane as observed in Figure 12.12, which would add significant noise to an already noisy and randomised Gaussian function.

The exclusion intervals for β_{opt} for a mock universal luminosity function as calculated using Monte Carlo simulations are presented in Figure 12.13, and generally exclude all



FIGURE 12.12: Monitoring of (ζ, χ) distribution of the redshift restricted 2MRS galaxies in the K band for various trial values of β when a universal luminosity function is applied. The trial values of β used are noted on each panel.

values of $\beta \sim> 0.3$ from consideration for use with the 2MRS in all three bands at the 3σ confidence level, once again rendering the estimates for β_{opt} returned by ROBUST redundant. While this is again in keeping with results presented using VELMOD and other contemporary methods with the exception of POTENT, the repeated rejection of the published Carrick value for β_{opt} continues to be concerning when one considers the similarity between the $\rho(\zeta, \beta)$ plots observed so far with respect to the 2MRS.

12.2.2 2MRS: Using μ_{β} instead of u_{β}

The results of modelling a universal Gaussian luminosity function to the μ_{β} reliant redshift restricted variant of the 2MRS using both ROBUST and relative entropy are presented in Figure 12.15, and continue to show more of the same kind of behaviour we have come to expect thus far.

We continue to note the same deviation from monotonically increasing behaviour of the $\rho(\zeta,\beta)$ plots across all three bands, with the noise turnover point continuing to manifest at a β of 0.4; and the relative entropy plots continuing to be as noisy as ever with a very



FIGURE 12.13: One-sided confidence intervals calculated on β via Monte Carlo simulations for both ROBUST (upper panel) and relative entropy (lower panel) for the K, H and J bands depicted in blue, red and green respectively. The 3σ confidence intervals for all three bands are denoted via the black dashed lines, indicating all values of $\beta > \sim 0.3$ are to be excluded from consideration for the redshift restricted 2MRS survey with 99.7% confidence.



FIGURE 12.14: Monitoring of (ζ, μ) distribution of the μ_{β} reliant redshift restricted 2MRS galaxies in the K band for various trial values of β when a universal luminosity function is applied. The trial values of β used are noted on each panel.

clear spike in the data being observed at the same point again in all three bands. This continued deviation from monotonic behaviour remains inconsistent with the behaviour presented by ROBUST in Chapter 8 where the 'raw' luminosity information was used, lending further credence to our theory that (despite the fact that throughout this chapter both universal and distance-dependent modelled functions to the 2MRS have continued to yield very similar results), μ_{β} should not be considered as a parameter for use with ROBUST and the construction of the underlying ζ statistic due to its inability to yield consistent results throughout this work, despite promising initial signs.

As before we continue to see the same deviations from uniformity begin to manifest in the (ζ, μ) distributions of the galaxies for all trial values larger than $\beta > 0.4$ as noted in Figure 12.14, where the β -dependent peculiar velocities begin to dominate and introduce noise into the recovered signal by ROBUST. Similarly we see a familiar oversampling of $\zeta - \chi$ points for low values of ζ and χ begin to manifest in the $\zeta - \chi$ plane as seen in Figure 12.16, which would continue to contribute noticeably to the noise already present in the data given how the mock apparent magnitudes are being generated.



FIGURE 12.15: Results of applying mock universal apparent magnitudes to the positions of the μ_{β} reliant redshift restricted 2MRS galaxies on the value of β_{opt} as returned by ROBUST (upper panel) and relative entropy (lower panel) over several trial values of β on the interval [0,1] for the K (blue), H (red) and J (green) infrared bands respectively. Grid mesh size applied for calculating relative entropy set at a value of 10.



0

0

 ζ value

 χ value

FIGURE 12.16: Monitoring of (ζ, χ) distribution of the μ_{β} reliant redshift restricted 2MRS galaxies in the K band for various trial values of β when a universal luminosity function is applied. The trial values of β used are noted on each panel.

0

0

 ζ value

 χ value

When this is combined with the Monte Carlo generated exclusion intervals for β_{opt} as seen in Figure 12.17, we find that almost all values of β across the interval of [0,1] are excluded from consideration for this variant of the 2MRS in all three bands, with only values of $\beta \sim 0.05$ somewhat making it into consideration, which is once again at odds with the results generated by ROBUST using the raw luminosity data in Chapter 8, and also inconsistent with the K band only results returned by ROBUST for the distance-dependent case of this variant presented earlier. Considering the sheer lack of consistency presented by this case throughout this work, we now move to discount the use of μ_{β} in any further analyses with either ROBUST or relative entropy, due to its inability to function adequately with either program.

The key takeaways from all of the above analyses would appear to be clear. Our baseline assumption that the luminosity function of a survey is independent of the real-space spatial positions of the galaxies is critical to the functioning of ROBUST with regards to the IRAS PSCz survey, implying that it is best modelled with a universal luminosity function (either Schechter or Gaussian) and a well-defined faint limit (the latter of which we already proved in earlier chapters). The introduction of even a very slight



FIGURE 12.17: One-sided confidence intervals calculated on β via Monte Carlo simulations for both ROBUST (upper panel) and relative entropy (lower panel) for the K, H and J bands depicted in blue, red and green respectively. The 3σ confidence intervals for all three bands are denoted via the black dashed lines, indicating all values of $\beta > \sim 0.3$ are to be excluded from consideration for the μ_{β} reliant redshift restricted 2MRS survey with 99.7% confidence.

linear distance dependence was more than enough to introduce bias significant enough that neither ROBUST nor relative entropy were able to function properly and recover reliable mean estimates for the value of β_{opt} .

With regards to the 2MRS in particular, our assumption of universality would also appear to be mostly valid, as the similarity of ROBUST plots between Chapters 8 and the distance dependent and universal counterparts presented within this chapter suggest that the raw luminosity data for the 2MRS is predominantly universal in origin. It must be noted however that the similarity of the distance-dependent case of the redshift restricted 2MRS to both its universal and 'raw data' counterpart could be indicative of something interesting: namely that the slight distance dependence that we introduced was not significant enough to introduce a major change to the runnings of ROBUST in particular, though the downwards shifting of the ROBUST plots in this chapter would appear to suggest that this may not necessarily be the case. It must also be noted that the slight differences presented between the 'raw data' μ_{β} variants of the 2MRS and their universal and distance-dependent modelled counterparts would appear to suggest that, despite the promising results initially returned in Section 8.6.3, any further use of μ_{β} in the construction of the ζ statistic for use with ROBUST must be discontinued.

Regardless of the case presented however, relative entropy consistently fails to return values for β_{opt} that are in agreement with its raw data counterparts when either a universal or a distance-dependent LF is modelled, suggesting that it is extremely sensitive to both the existence of any sort of distance dependence in the modelled luminosity function; and the existence of any randomly introduced noise when modelling apparent magnitudes in a mock realisation, making it less useful as an ancillary tool for ROBUST to make use of in that scenario. Why this holds true for the 2MRS and not the PSCz where we were successfully able to make use of mock Schechter and Gaussian apparent magnitudes modelled to the positions of those galaxies and recover mean estimates on the value for β_{opt} that were consistent with our assumed true value of $\beta = 0.55 \pm 0.06$ for that survey is unclear. It could be that there is an inherent property to the 2MRS that, when made to react with our mock modelled magnitudes causes the substantial noise that we have seen throughout this chapter.

When one puts all of the above together we can more confidently say that our underlying assumption of universality, the foundation upon which ROBUST is reliant, is most likely **not** the problem causing the peculiar results returned by both ROBUST and relative entropy in Chapters 8 and 11, that indeed the programs are functioning correctly; and that there is possibly an inherent peculiar property or fault to the 2MRS survey that unfortunately renders it not very compatible with any further use with either ROBUST or relative entropy beyond the initial exclusion results presented in the aforementioned

chapters which, at the very least, help to further reinforce the rejection of values of β equal to unity returned by methods such as POTENT while allowing for the continued existence of possible values of β for the survey that are consistent with $\beta < 0.6$ and lower, such as those returned by VELMOD, χ^2 -minimisation and other constraining techniques.

Chapter 13

Conclusions and Future Avenues

While the subject of cosmology is almost as old as civilisation itself, it has taken many leaps and bounds forward in recent years, giving us a clearer view of the heavens and our place in it than we ever had before. In particular with the advent of deep and highly detailed redshift-distance galaxy surveys of various patches of the Universe we are now in the very desirable position of being able to probe the local peculiar velocity and matter density fields of the nearby Universe and make use of various statistical tools at our disposal to constrain various cosmological parameters that consequently give us a better handle on understanding just how our Universe functions and how best it can be modelled.

In attempting to constrain one such parameter, the linear redshift distortion parameter β which helps to determine the ratio of luminous to dark matter in a region of space, we have made use of two key statistical tools throughout this work: namely χ^2 -minimisation and ROBUST (the latter being eventually supplemented with analyses using relative entropy).

For its part, χ^2 -minimisation has proven itself to be a simple and effective velocityvelocity comparison tool that can be used in conjunction with a variety of secondary distance indicators, with Type Ia Sne being the most typically chosen due to their high intrinsic luminosities and highly accurate distance estimates to within 7-8%. While simple however it is not without its limitations, primarily its reliance on secondary distance information with which to perform its comparisons. Should indicators be chosen whose distances are not known to a high enough accuracy, then the functionality of χ^2 minimisation is significantly affected.

Despite these grievances however χ^2 -minimisation has proven itself capable not only of reconfirming the Radburn-Smith et. al computed value for the IRAS PSCz survey of
$\beta = 0.55 \pm 0.06$ with their own RS Sne set; it has also successfully validated the viability of the Carrick et. al velocity grid as constructed from the 2M++ survey, confirming their published value of $\beta = 0.43 \pm 0.021$ for the 2MRS survey in particular.

In confirming all of the aforementioned published values for different surveys we were able to set the benchmark values necessary with which to test the viability of the second novel statistical tool at our disposal: namely ROBUST. ROBUST has served as the main focus of this work due to its reliance on one key property for any given redshift-distance survey, specifically that for a survey that is complete up to a given limiting magnitude (or pair of faint and bright magnitudes), the luminosity function of the survey must be independent of the real-space positions of the galaxies. In creating such a key property we are consequently able to construct a statistic by which we can probe the parameters of the peculiar velocity field of a survey (β in this instance) while making as few a priori assumptions about the galaxies in the survey as possible.

We have since established that ROBUST functions effectively in a controlled mock environment when applied to mock subsets of the IRAS PSCz galaxies modelled with various luminosity functions, proving its robustness when the parameters of such functions are altered to account for poorly defined functions, or when the mock subsets being generated are reduced or increased in size potentially introducing further chance for Malmquist bias to affect the results. In particular we take note of the fact that provided the sample size being analysed is large enough (a minimum of 2000 galaxies) ROBUST is more than capable of recovering our assumed true value for β with an increasingly tightly constrained confidence interval, managing to account fairly well for the use of broad and poorly defined functions though the latter will start to affect the accuracy of the returned confidence intervals eventually. Additionally we observed that Schechter functions are far more sensitive to the use of the correct faint and bright limits than its Gaussian counterpart, manifesting significant bias in the results returned by ROBUST if the wrong limits were applied. Having said that the bias introduced into ROBUST only began to noticeably manifest for arbitrarily applied bright limits that were almost as faint as our established faint limit for the survey; meaning that that limit was encroaching upon the distribution of galaxies in the M-Z plane, causing deviations from the uniformity that ROBUST requires. This observation lends itself quite well to the theory that the IRAS PSCz is best modelled by a faint limit alone but that so long as an arbitrary bright limit is applied that is not too faint, then ROBUST will continue to function properly, further cementing its usefulness. The final key observation to take away from utilising ROBUST in a mock environment is that its functionality appears to be completely independent from the value of ΔZ chosen to define the S_1 and S_2 sets upon which the construction of the ζ statistic is reliant when a bright limit is in play. In particular whether a small value of ΔZ is applied that practically throws away a large

part of the survey from consideration for analysis, or a large value of ΔZ is applied that in the presence of a bright limit causes the underlying S_1 and S_2 sets to become trapezoidal in nature and potentially introduce deviations from uniformity; ROBUST continues to successfully recover mean estimates for the value of β that are consistent with our assumed value of β_{true} , once again enhancing its usefulness.

Unfortunately in making the move to working with real data this is where ROBUST begins to struggle, failing to return values of β for the 2MRS that are consistent with the published and independently confirmed Carrick value despite our experimentation with various parameters and distance restrictions to attempt to improve the returned results. However it must be noted that ROBUST has proven itself capable of computing one-sided exclusion intervals on β for the 2MRS that are in very good agreement with recent values determined by other means, specifically excluding all values of $\beta > 0.6$ from consideration for acceptance as β_{opt} . This is important as it serves as one of many proverbial nails in the coffin for the density-density comparison method POTENT that has continually favoured values of β close to unity; while further reinforcing the validity of results returned by χ^2 -minimisation techniques and VELMOD amongst others which favour values on the range of $\beta \sim 0.3 - 0.4$.

This behaviour is further ratified when one begins to make use of relative entropy to complement the results returned by ROBUST. While in a mock environment relative entropy behaves in a similar way to ROBUST, manifesting increased sensitivity to the use of Schechter functions it also proves that it can serve as a viable strengthening factor for the results returned by ROBUST, and can even pick up the slack so to speak where ROBUST would struggle particularly in the presence of incorrect bright limits. In particular however when applied to the 2MRS relative entropy fails to recover a value of β that is consistent with the published Carrick value, but successfully returns possible exclusion intervals that are comparable to those returned by ROBUST, once again favouring results returned by VELMOD and others while further rejecting those favoured by POTENT.

Our consequent exploration into the use of distance-dependent luminosity functions with ROBUST and relative entropy was two-fold. In determining whether the underlying LF of the 2MRS was indeed universal we could rule out whether the peculiar results being returned by the program was due to a fundamental flaw in the underlying theory of ROBUST itself, or due to an inherent peculiar property of the 2MRS that renders it not very viable for further use. Secondly we could use our modelling of distance-dependent LFs to test the limits of both ROBUST and relative entropy and what sort of potential biases would be introduced as a result. With regards to the latter it becomes clear that ROBUST is indeed sensitive to the introduction of even a mild linear dependence though this is more noticeable with the IRAS PSCz than it was with the 2MRS. Results returned for the IRAS PSCz failed to recover a mean estimate for the value of β at all, while relative entropy begins to exhibit significant noise across the board, theorised to be due to the use of the **normrnd** function in Matlab to generate mock apparent magnitudes and the consequent randomised noise that this introduces into the data. In considering the 2MRS itself, the use of both mock universal and distance-dependent apparent magnitudes being modelled to the galaxies returns consistent results from ROBUST that are in relatively good agreement with the results returned by ROBUST when analysing the raw data directly. As stated before this lends itself to one of two possibilities: firstly that the raw 2MRS data does indeed appear to be relatively universal in origin or second: that the introduced linear dependence is not significant enough to materially impact the functioning of ROBUST thought this latter possibility seems to be less likely when one considers the results more thoroughly. Relative entropy's behaviour for its part throughout this experimentation remains constant in that it is very averse to the introduction of any sort of distance dependence into its analysis, which manifests as constant noise across the board for all trial values of β on the interval [0,1], rendering it effectively useless as an ancillary tool at least where distance-dependent luminosity functions are involved.

Our final conclusion from analysing all of the information we have at hand tends us towards stating that there is nothing inherently wrong with the functionality of RO-BUST, and that its underlying assumption of luminosity function independence from spatial position is valid and sensible. There unfortunately appears to be some inherent peculiar property to the 2MRS that renders it not very viable for use with ROBUST let alone relative entropy ... it is only a shame that we were unable to determine this quicker so that we could move on to experimenting with other real world surveys and see how well ROBUST and its ancillary tools fare with them.

All of the above discoveries lend themselves to a few potential avenues of work however that may be pursued in the future:

1. Revisiting vorticity. The entirety of this work (and the functionality of RO-BUST in particular) has its root in linear perturbation theory, specifically making the assumption that the distances and scale lengths over which we are working are large enough that the peculiar velocity field we are using can be deemed **irrotational** over those lengths (i.e. $\nabla \times v = 0$). Put another way, we have assumed that the decaying vorticity mode of our linear perturbation theory as discussed in Section 3.2 has a small enough dampening scale such that it vanishes over large distances, leaving only the linear gravitational growth term to dominate over time. Should this key assumption not hold true at the relatively low redshifts within which we have been working, it would go a long way towards explaining some of the peculiar results returned not just by ROBUST and relative entropy, but by χ^2 -minimisation with its weakly recovered signals in β and the somewhat poor linear regression fitting we have been observing. This supposition lends itself to two key avenues of exploration to pursue:

- Redevelop our ROBUST framework within a perturbation theory model wherein which the vorticity modes of the peculiar velocity field are considered, and see whether the predicted additional velocities exhibited by the galaxies in a survey as a result help to improve upon the results for β returned by ROBUST and relative entropy for the 2MRS.
- A simpler alternative to the theoretically and computationally exhaustive option listed above would be to consider cutting the 2MRS at lower redshifts (i.e. excluding all galaxies that fall below a certain threshold limit from consideration for use with ROBUST and relative entropy) where the velocity field is indeed deemed to be rotational according to perturbation theory. While we have already been applying this in practice with the 2MRS, generally applying a redshift restriction limit of $500 \le cz \le 15000 \text{kms}^{-1}$ throughout all our analyses, it is possible that this lower exclusion limit of 500kms^{-1} is not enough to completely remove all effects of vorticity from the field. Conversely it is also possible that our higher exclusion limit of 15000kms^{-1} , while in keeping with previous determinations that this limit ensures the reliability of the underlying reconstructed peculiar velocity field, may also be too large; consequently meaning that the effects of v_{pec} at those distances will be less significant on our analyses. To that end, experimenting with varying our applied redshift limits more extensively (while also observing and quantifying how the luminosity function of the underlying survey is affected as a result, if at all) may help improve the results returned by ROBUST and relative entropy.
- 2. Making better use of available IRAS PSCz flux information. Throughout Chapter 7 we attempted to make use of available optical B-band magnitude information for the galaxies, despite the fact that the PSCz is at its core an infrared survey which, when one thinks about it, is rather counter-intuitive. The result returned by ROBUST in this instance would be considerably different to the value of 0.55 that we expect since optical magnitudes (and by extension an optical survey) would be tracing a different matter distribution than their infrared flux counterparts, putting it at odds with the underlying infrared survey. It would be far wiser to forego the use of the optical magnitudes in their entirety and instead focus on making use of the available infrared flux information for the PSCz

galaxies in the 25μ m, 60μ m and 100μ m ranges; converting those fluxes to their corresponding apparent magnitudes and work from there, applying ROBUST and relative entropy as normal.

- 3. Testing for memory leaks. As was discussed in Chapter 8, Matlab has demonstrated an inability to hold the entirety of the Carrick grid in memory without it being divided into smaller slices each 16Mpc wide. This is rather unusual for Matlab as it has demonstrated in other disciplines that it is more than capable of holding over 17 million data entries for analysis without issue. It is possible that there is a fundamental error in our Matlab scripts somewhere that is causing a memory leak to occur, preventing the entirety of the Carrick grid from being held in memory as normal. As such it would be worth experimenting with creating a mock galaxy catalogue of our own, consisting of over 43000 galaxies with mock positions and peculiar velocities that mimic those seen in the 2MRS (not unlike our analyses assigning mock magnitudes to the positions of the IRAS PSCz galaxies in Chapter 5) and seeing whether our velocity interpolation schemes, distance convergence loops and ROBUST/relative entropy analyses can be carried out as normal.
 - As an aside to this it is worth pointing out that, while Matlab is quite good at handling data matrices of considerable size and analysing them, other statistical programs now exist that are considerably more flexible and powerful such as Python, which may not be as susceptible to the kinds of memory leaks we experienced in this work. To that end it may be worth experimenting with rewriting our Matlab scripts in Python or similar more powerful programs moving forward to make future computational efforts easier.
- 4. Where the 2MRS failed another may succeed. We have made mention of the various data runs of the SDSS throughout this work but due to time constraints were unable to actually apply ROBUST or relative entropy to it properly and document the results they return. Perhaps the use of a survey that is defined by both a faint and a bright limit such as the SDSS will yield better results than the 2MRS. It is also worth noting that since the SDSS is an optical survey it will inherently trace a different mass distribution on the sky than its 2MRS infrared counterpart, and that perhaps this signal can be more easily recovered by ROBUST and relative entropy. It is worth noting that as of the time of writing, no value for β has yet been constrained for any DR of the SDSS, making it a very enticing target for future analysis.
- 5. Changing velocity models. We have been taking it for granted that the underlying peculiar velocity field of the 2MRS (and indeed the IRAS PSCz during our

mock realisations) is best modelled by a logarithmic equation as defined back in Section 4.3. There is no reason to suppose that this is completely accurate for either survey, and ROBUST is set up to be able to operate with any kind of velocity field model we wish to apply. It would be worth exploring the effects of altering the underlying velocity field model of the 2MRS and observing what effects it has on the values of β_{opt} returned by either ROBUST or relative entropy. Perhaps in altering said velocity field equation we may be more successful in recovering a value for β_{opt} that is in better agreement with the published Carrick value.

- 6. Experimenting more with distance dependence. Given the lack of noticeable change in ROBUST when modelling universal or distance-dependent LFs to the 2MRS it will be interesting to see how ROBUST reacts when more extreme distance-dependent relations are modelled and whether it continues to be as 'forgiving' and function adequately. Additionally it was already established by Carrick et. al that the 2M++ survey of which the 2MRS is a part is best modelled by a Schechter function, yet in our works we made use of modelling Gaussians due to the ease of their generation within Matlab. When one considers the tenets of the Central Limit Theorem - that with a large enough sample size containing independently generated variables, the computed mean and deviation of such a distribution will tend towards a Gaussian as the sample gets larger - and the fact that the 2MRS survey we were utilising consisted of over 43000 galaxies to analyse; making use of Gaussian functions did not appear to be unreasonable. For the sake of completion however it would be wise to model various universal and distance-dependent Schechter functions to the 2MRS galaxies and see how well their results as returned by ROBUST and relative entropy compare to what has been presented in this work.
- 7. Questioning Universality. Throughout the entirety of this work not only have we made use of the assumption of universality when it comes to the LF of a survey, but we have also assumed that the computed magnitude limits of a survey as returned by the completeness statistic T_c are also universal for a survey. This may not necessarily be the case. Given the breakdown of homogeneity and isotropy on smaller scale distances resulting in variations in the local peculiar velocity and matter density fields of the nearby Universe it is conceivable that depending on what direction along the sky one chooses to observe within a survey, the perceived magnitude limit along that direction may be different than if one were to choose another direction along the sky to observe along. This variance in magnitude limit as determined by a 'directional T_c ' would need to be accounted for within the functioning of ROBUST, either within iterative running of the program along various directional vectors and then summing the results, or perhaps by running

ROBUST and relative entropy and then integrating the results over shells of increasing distance, modifying the values of m_{lim} (and m_{bright} if necessary) utilised as one goes along.

13.1 More Recent All-Sky Redshift-Distance Surveys to Consider for Future Use

While we have made considerable use of the IRAS PSCz survey throughout this work, and its usefulness as an all-sky survey in the infrared band cannot be understated for observational cosmology and our ability to map the local galaxy density and velocity fields (and consequently the local gravity fields as well); its age is beginning to show. In particular the PSCz catalogue of Saunders et. al that we have been using was compiled in 2000 [170] and a lot of developments have been had in the meantime. Since 2000 many more surveys have been generated such as 2MRS (see Section 2.10.2) and 2M++, with others ongoing to this day such as the SDSS (whose most recent data runs we have already briefly covered in Section 2.10.3), each of them larger and more in-depth and detailed than those that came before it. It makes sense to briefly discuss and recommend some of the more recent all-sky surveys that are now available for us to use, spiritual successors to the IRAS PSCz that can help pave the way for more meaningful and extensive uses for ROBUST, relative entropy and various other statistical tools in future probings of the local density, peculiar velocity and gravity fields of the nearby Universe, while also honouring the foundations set by the PSCz and its predecessors.

13.1.1 Cosmic-flows3

Developed in 2016 by Brent Tully, Hélène Courtois and Jenny Sorce as an expanded compendium of galaxy distances that built upon the two prior releases of Cosmicflows developed by Tully in 2008 and 2013, unsurprisingly called *Cosmic-flows* [193] and *Cosmic-flows2* [194] (hereafter denoted CF and CF2 respectively), Cosmic-flows3 (CF3) is an all-sky infrared survey that contains over 17,600 galaxies, consequently providing the dense spatial coverage required for one to study the streams and eddies of the local cosmic flow and constrain a whole host of parameters such as β and H_0 to name a few [41]. Some of the more pertinent additions to this iteration of the project when compared to CF and CF2 include over 2200 distances that have been derived using the Tully-Fisher relation and photometry at 3.6mm, in addition to over 8800 distances calculated using the Fundamental Plane methodology from the 6dFGS collaboration [192].



FIGURE 13.1: The density of galaxies in the GLADE catalogue, presented as an Aitoff projection on the sky. Figure reproduced from the GLADE documentation works of Dálya et. al [57].

Minor augmentations to Type Ia Sne compilations included in CF3 have also been included, wherein a zero-point calibration of their luminosities has allowed for a value for H_0 to be constrained, given as $76.2 \pm 3.4 \text{kms}^{-1} \text{Mpc}^{-1}$; while conversely by instead imposing a restriction on the observed peculiar velocity monopole term representative of global velocity infall or outflow in the local velocity field has allowed for a value of $H_0=75 \pm 2 \text{kms}^{-1} \text{Mpc}^{-1}$ to be constrained [192].

For a more in-depth introduction into the Cosmic-flows project since its inception in 2011, all relevant data-products, publications and project descriptions can be accessed at https://www.ipnl.in2p3.fr/projet/cosmicflows/.

13.1.2 GLADE

Developed in 2018 by Dálya et. al, the Galaxy List for the Advanced Detector Era, otherwise known as GLADE, is an all-sky optical survey with high completeness as depicted in Figure 13.1 that is primarily meant to be used in identifying gravitational wave sources in order to support future electromagnetic follow-up projects of the LIGO/Virgo Collaboration [57]; though its formidable size and depth lends itself well to providing the input data necessary on the matter distribution of the local Universe for astrophysical or cosmological simulations [45]. In particular GLADE has been constructed (combined and matched) from four existing galaxy catalogues: GWGC, 2MPZ, 2MASS XSC and

HyperLEDA, with further extension to GLADE being provided via incorporation of the SDSS-DR12Q catalog. This results in GLADE containing over 3.2 million usable objects for analysis, which is two orders of magnitude greater than the number of galaxies in the GWGC catalogue alone (over 53,000 galaxies) [57]. Due to the sheer size of this survey, GLADE is complete up to an optical luminosity distance of $d_L = 37^{+3}_{-4}$ Mpc in terms of the cumulative B-band luminosity of galaxies within luminosity distance d_L , and contains all of the brightest galaxies giving half of the total B-band luminosity up to $d_L = 91$ Mpc [45].

The GLADE catalogue can be accessed in its entirety alongside all available documentation and previous versions of the data at http://aquarius.elte.hu/glade/index. html.

Whether one decides to more forward with any of the suggestions mentioned here with either of these newer surveys is entirely up to the reader but one way or another the truth is clear. There is still so much more out there for us to learn about and explore, and so many tools at our disposal with which to probe our Universe and further our understanding of it, ROBUST being one of them. With technology advancing at a fantastic rate and petabytes of additional information becoming available every passing day for us to analyse there is no better time or place to be a cosmologist, and it will be interesting to see where we go next as we continue to look up into the stars.

Bibliography

- F. D. Albareti et al. The thirteenth data release of the Sloan Digital Sky Survey: First spectroscopic data from the SDSS-IV Survey mapping nearby galaxies at Apache Point Observatory. *The Astrophysical Journal Supplement Series*, 233(2): 25, December 2017.
- [2] R. Amanullah et al. Spectra and Hubble Space Telescope light curves of six Type Ia supernovae at 0.511 ≤ z ≤ 1.12 and the Union2 Compilation. Astrophysical Journal, 716(1):712–738, June 2010.
- [3] M. A. Barstow et al. Gravitational redshift. In I. Ridpath, editor, Oxford Dictionary of Astronomy, page 207. Oxford University Press, London, 2nd edition, January 2007.
- [4] M. A. Barstow et al. K-correction. In I. Ridpath, editor, Oxford Dictionary of Astronomy, page 267. Oxford University Press, London, 2nd edition, January 2007.
- [5] M. A. Barstow et al. Redshift. In I. Ridpath, editor, Oxford Dictionary of Astronomy, page 412. Oxford University Press, London, 2nd edition, January 2007.
- [6] M. A. Barstow et al. Redshift survey. In I. Ridpath, editor, Oxford Dictionary of Astronomy, page 414. Oxford University Press, London, 2nd edition, January 2007.
- [7] M. A. Barstow et al. Velocity dispersion. In I. Ridpath, editor, Oxford Dictionary of Astronomy, page 524. Oxford University Press, London, 2nd edition, January 2007.
- [8] S. Basilakos and M. Plionis. Cosmological evolution of linear bias. Astrophysical Journal, 550(2):522–527, April 2001.
- M. Bernardi et al. Early-type galaxies in the Sloan Digital Sky Survey. II. correlations between observables. *The Astronomical Journal*, 125(4):1849–1865, April 2003.

- [10] E. Bertschinger and A. Dekel. Recovering the full velocity and density fields from large-scale redshift-distance samples. Astrophysical Journal, Part 2 - Letters, 336: L5–L8, January 1989.
- [11] M. Bilicki, M. Chodorowski, W. Hellwing, T. Jarrett, and G. Mamon. Precision cosmology with the 2MASS clustering dipole. In New developments in the study of the large scale structure of the Universe, Proceedings of the Thirteenth Marcel Grossman Meeting on General Relativity, page 3, Stokholm, Sweden, November 2012.
- [12] A. Blaauw, C. S. Gum, J. L. Pawsey, and G. Westerhout. The new I. A. U. system of galactic coordinates (1958 revision). MNRAS, 121:123, 1960.
- [13] J. P. Blakeslee et al. A first comparison of the surface brightness fluctuation survey distances with the galaxy density field: Implications for H₀ and Ω . Astrophysical Journal, 527(2):L73–L76, 1999.
- [14] M. R. Blanton et al. New York University Value-Added Galaxy Catalog: A galaxy catalog based on new public surveys. *The Astronomical Journal*, 129(6):2562–2578, June 2005.
- [15] G. R. Blumenthal, S. M. Faber, J. R. Primack, and M. J. Rees. Formation of galaxies and large-scale structure with cold dark matter. *Nature*, 311:517–525, 1984.
- [16] G. D. Bothun et al. A catalog of radio, optical, and infrared observations of spiral galaxies in clusters. Astrophysical Journal Supplement Series, 57:423–472, March 1985.
- [17] D. Branch and D. L. Miller. Type Ia supernovae as standard candles. Astrophysical Journal, Part 2 - Letters, 405(1):L5–L8, 1993.
- [18] D. Branch and G. A. Tammann. Type IA supernovae as standard candles. Annual Review of Astronomy & Astrophysics, 30:359–389, 1992.
- [19] E. Branchini, A. Eldar, and A. Nusser. Peculiar velocity reconstruction with the fast action method: Tests on mock redshift surveys. MNRAS, 335(1):53–72, September 2002.
- [20] E. Branchini et al. Comparing the SFI peculiar velocities with the PSCz gravity field: a VELMOD analysis. MNRAS, 326(3):1191–1204, September 2001.
- [21] J. S. Bullock and M. Boylan-Kolchin. Small-scale challenges to the ACDM paradigm. Annual Review of Astronomy & Astrophysics, 55(1):343–387, August 2017.

- [22] A. G. Butkevich, A. V. Berdyugin, and P. Teerikorpi. Statistical biases in stellar astronomy: The Malmquist bias revisited. MNRAS, 362(1):321–330, September 2005.
- [23] R. R. Caldwell. Dark energy. PhysicsWorld.com, 2004. Accessed 2/9/2012.
- [24] IPAC Caltech. About 2MASS. IPAC Caltech, February 2006. Accessed 13/4/2015.
- [25] J. Carrick, S. J. Turnbull, G. Lavaux, and M. J. Hudson. Cosmological parameters from the comparison of peculiar velocities with predictions from the 2M++ density field. MNRAS, 450(1):317–332, June 2015.
- [26] H. B. G. Casimir. On the attraction between two perfectly conducting plates. Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen, B51: 793–795, May 1948.
- [27] H. B. G. Casimir and D. Polder. The influence of retardation on the London-van der Waals forces. *Physical Review*, 73(4):360–372, February 1948.
- [28] S. Chandrasekhar. The highly collapsed configurations of a stellar mass (second paper). MNRAS, 95:207–225, January 1935.
- [29] C. Chuang and Y. Wang. Using multipoles of the correlation function to measure h(z), d_A(z) and β(z) from Sloan Digital Sky Survey luminous red galaxies. MNRAS, 431(3):2634–2644, May 2013.
- [30] D. T. Chuss. LAMBDA Coordinate conversions. National Aeronautics and Space Administration - Goddard Space Flight Center, 2010. Accessed 27/8/2012.
- [31] S. Clark. What were the first celestial objects? In S. Blackburn, editor, The Universe, The Big Questions, page 208. Hachette UK, October 2011.
- [32] C. Clarkson, O. Umeh, R. Maartens, and R. Durrer. What is the distance to the CMB? Journal of Cosmology & Astroparticle Physics, 2014(11):15, November 2014.
- [33] M. Colless et al. The peculiar motions of early-type galaxies in two distant regions
 VII. Peculiar velocities and bulk motions. MNRAS, 321(2):277–305, February 2001.
- [34] D. Commenges. Information theory and statistics: An overview. *eprint arXiv*, page 22, November 2015.
- [35] A. Conley et al. SiFTO: An empirical method for fitting SN Ia light curves. Astrophysical Journal, 681(1):482–498, 2008.

- [36] COSMOS The SAO Encyclopedia of Astronomy. Baryonic Acoustic Oscillations
 COSMOS. Swinburne University of Technology, 1999. Accessed 13/6/2018.
- [37] COSMOS The SAO Encyclopedia of Astronomy. Critical Density | COSMOS. Swinburne University of Technology, 1999. Accessed 14/6/2018.
- [38] COSMOS The SAO Encyclopedia of Astronomy. Gravitational Redshift | COS-MOS. Swinburne University of Technology, 1999. Accessed 9/2/2019.
- [39] COSMOS The SAO Encyclopedia of Astronomy. Supergalactic Coordinate System | COSMOS. Swinburne University of Technology, 1999. Accessed 6/6/2018.
- [40] S. Courteau, J. A. Willick, M. A. Strauss, D. Schlegel, and M. Postman. Shellflow.
 I. The convergence of the velocity field at 6000 kilometers per second. Astrophysical Journal, 544(2):636–640, December 2000.
- [41] H. Courtois, R. B. Tully, and J. R. Fisher. Cosmic Flows home. Institute de Physique Nucléaire, 2017. Accessed 11/2/2019.
- [42] L. N. da Costa et al. Redshift-distance survey of early-type galaxies: Dipole of the velocity field. Astrophysical Journal, 537(2):L81–L84, July 2000.
- [43] L. A. L. da Silva. The classification of supernovae. Astrophysics and Space Science, 202(2):215–236, April 1993.
- [44] D. A. Dale et al. Seeking the local convergence depth: The Abell cluster dipole flow to $200h^{-1}$ Mpc. Astrophysical Journal, 510(1):L11–L14, January 1999.
- [45] G. Dálya et al. GLADE: A galaxy catalogue for multimessenger searches in the advanced gravitational-wave detector era. MNRAS, 479(2):2374–2381, September 2018.
- [46] R. L. Davies, G. Efstathiou, S. M. Fall, G. Illingworth, and P. L. Schechter. The kinematic properties of faint elliptical galaxies. Astrophysical Journal, Part I, 266: 41–57, March 1983.
- [47] M. Davis, A. Nusser, and J. A. Willick. Comparison of velocity and gravity fields: The Mark III Tully-Fisher Catalog versus the IRAS 1.2 Jy Survey. Astrophysical Journal, 473:22, December 1996.
- [48] W. de Sitter. On the relativity of inertia. Remarks concerning Einstein's latest hypothesis. Koninklijke Nederlandsche Akademie van Wetenschappen Proceedings, 19(2):1217–1225, March 1917.
- [49] W. de Sitter. On the curvature of space. Koninklijke Nederlandsche Akademie van Wetenschappen Proceedings, 20(1):229–243, 1918.

- [50] A. Dekel. Dynamics of cosmic flows. Annual Review of Astronomy and Astrophysics, 32:371–418, 1994.
- [51] A. Dekel et al. POTENT reconstruction from Mark III velocities. Astrophysical Journal, 522(1):1–38, September 1999.
- [52] B. Efron and V. Petrosian. A simple test of independence for truncated data with applications to redshift surveys. Astrophysical Journal, Part I, 399(2):345–352, November 1992.
- [53] G. Efstathiou. H_0 revisited. MNRAS, 440(2):1138–1152, May 2014.
- [54] J. Einasto, A. Kaasik, and E. Saar. Dynamic evidence on massive coronas of galaxies. *Nature*, 250:309–310, 1974.
- [55] A. Einstein. Zur Elektrodynamik bewegter Körper. Annalen der Physik, 332(10): 891–921, 1905.
- [56] A. Einstein. The foundation of the general theory of relativity. In *The Principle of Relativity*, pages 109–164. Dover Books on Physics, June 1952.
- [57] Eötvös Gravity Research Group. GLADE catalog. Eötvös Gravity Research Group, 2018. Accessed 11/2/2019.
- [58] P. Erdoğdu et al. Reconstructed density and velocity fields from the 2MASS Redshift Survey. MNRAS, 373(1):45–64, November 2006.
- [59] M. Erlewine. Matrix Software: Learning Astrology The local supergalaxy. Matrix Software, 2017. Accessed 6/6/2018.
- [60] ESA. Planck and the Cosmic Microwave Background / Planck / Space ESA. European Space Agency, 2013. Accessed 30/5/2018.
- [61] S. M. Faber and R. E. Jackson. Velocity dispersions and mass-to-light ratios for elliptical galaxies. Astrophysical Journal, 204:668–683, March 1976.
- [62] U. Feindt et al. Measuring cosmic bulk flows with Type Ia supernovae from the Nearby Supernova Factory. Astronomy & Astrophysics, 560:12, December 2013.
- [63] A. V. Fillipenko. Evidence from Type Ia supernovae for an accelerating universe and dark energy. In W. L. Freedman, editor, *Measuring and Modeling the Universe*, Carnegie Observatories Astrophysics Series, page 270. Carnegie Observatories Centennial Symposia, Cambridge University Press, 2004.
- [64] K. B. Fisher, O. Lahav, Y. Hoffman, D. Lynden-Bell, and S. Zaroubi. Wiener reconstruction of density, velocity and potential fields from all-sky galaxy redshift surveys. MNRAS, 272(4):885–908, February 1995.

- [65] D. Fleisch. A Student's Guide to Maxwell's Equations. Cambridge University Press, Cambridge, UK, January 2008.
- [66] P. Fosalba, E. Gaztañaga, and F. J. Castander. Detection of the Integrated Sachs-Wolfe and Sunyaev-Zeldovich effects from the Cosmic Microwave Backgroundgalaxy correlation. Astrophysical Journal, 597(2):L89–L92, November 2003.
- [67] W. Freudling et al. Large-scale power spectrum and cosmological parameters from SFI peculiar velocities. Astrophysical Journal, 523(1):1–15, September 1999.
- [68] A. Friedmann. Über die Krümmung des Raumes. Zeitschrift für Physik, 10:377– 386, 1922.
- [69] A. Friedmann. On the curvature of space. General Relativity and Gravitation, 31: 1991, December 1999.
- [70] J. A. Frieman, M. S. Turner, and D. Huterer. Dark energy and the accelerating universe. Annual Review of Astronomy & Astrophysics, 46(1):385–432, September 2008.
- [71] G. Gentile et al. The cored distribution of dark matter in spiral galaxies. MNRAS, 351(3):903–922, July 2004.
- [72] R. Giovanelli et al. The I Band Tully-Fisher relation for cluster galaxies: a template relation, its scatter and bias corrections. *The Astronomical Journal*, 113: 53–79, January 1997.
- [73] J. Guy, P. Astier, S. Nobili, N. Regnault, and R. Pain. SALT: a spectral adaptive light curve template for Type Ia supernovae. Astronomy & Astrophysics, 443(3): 781–791, December 2005.
- [74] J. Guy et al. SALT2: using distant supernovae to improve the use of Type Ia supernovae as distance indicators. Astronomy & Astrophysics, 466(1):11–21, April 2007.
- [75] N. Hamaus et al. Multipole analysis of redshift-space distortions around cosmic voids. Journal of Cosmology & Astroparticle Physics, 07:20, July 2017.
- [76] A. J. S. Hamilton and M. Tegmark. The real-space power spectrum of the PSCz survey from 0.01 to 300hMpc⁻¹. MNRAS, 330(3):506–530, March 2002.
- [77] A. J. Hawken et al. The VIMOS Public Extragalactic Redshift Survey. Measuring the growth rate of structure around cosmic voids. Astronomy & Astrophysics, 607: 14, November 2017.

- [78] M. Hendry. Galaxy kinematics. Taken from the graduate Galaxies course taught at the University of Glasgow, April 2012.
- [79] M. A. Hendry, J. F. L. Simmons, and A. M. Newsam. What do we mean by "Malmquist Bias"? In F. R. Bouchet and M. Lachièze-Ray, editors, *Cosmic Velocity Fields*, Proceedings of the 9th IAP Astrophysics Meeting, page 23, Institut d'Astrophysique, Paris, July 1993. Gif-sur-Yvette: Editions Frontieres.
- [80] Y. Hoffman, H. M. Courtois, and R. B. Tully. Cosmic bulk flow and the local motion from Cosmicflows-2. MNRAS, 449(4):4494–4505, June 2015.
- [81] D. W. Hogg. Distance measures in cosmology. Institute for Advanced Study, 2000.
- [82] W. Hu. CMB Introduction. Department of Astronomy & Astrophysics, University of Chicago, 2007. Accessed 30/5/2018.
- [83] W. Hu and M. White. Acoustic signatures in the Cosmic Microwave Background. Astrophysical Journal, 471:30, November 1996.
- [84] E. Hubble. A relation between distance and radial velocity among extra-galactic nebulae. Proceedings of the National Academy of Sciences of the United States of America, 15(3):168–173, March 1929.
- [85] HubbleSite. Hubblesite Dark Energy Type Ia Supernovae. National Aeronatics and Space Administration, 1999. Accessed 31/5/2018.
- [86] J. Huchra, V. de Lapparent, and M. Geller. The CfA redshift survey and catalog. Smithsonian Astrophysical Observatory, 1985. Accessed 18/8/2012.
- [87] J. P. Huchra. The Hubble constant. Harvard-Smithsonian Center for Astrophysics, 2008. Accessed 30/8/2012.
- [88] J. P. Huchra. Published values of H_0 . Harvard-Smithsonian Center for Astrophysics, October 2010. Accessed 31/5/2018.
- [89] M. J. Hudson. Optical galaxies within 8000kms⁻¹ IV. The peculiar velocity field. MNRAS, 266(2):475, 1994.
- [90] M. J. Hudson and R. W. Pike. Cosmological parameters from the comparison of the 2MASS gravity field with peculiar velocity surveys. Astrophysical Journal, 635 (1):11–21, 2005.
- [91] M. J. Hudson, R. J. Smith, J. R. Lucey, and E. Branchini. Streaming motions of galaxy clusters within 12000kms⁻¹ - V. The peculiar velocity field. MNRAS, 352 (1):61–75, 2004.

- [92] M. J. Hudson, R. J. Smith, J. R. Lucey, D. J. Schlegel, and R. L. Davies. A large-scale bulk flow of galaxy clusters. Astrophysical Journal, 512(2):L79–L82, February 1999.
- [93] M. L. Humason, N. U. Mayall, and A. R. Sandage. Redshifts and magnitudes of extragalactic nebulae. The Astronomical Journal, 61:97–162, 1956.
- [94] Optimizing Constraints on Primordial Non-Gaussianity from Large-Scale Structure, Institute for the Early Universe Workshop, EWHA Women's University, Seoul, South Korea, May 2010. IEU, University Observatory Munich.
- [95] S. Islam. Probing the peculiar velocity field of galaxies. Technical report, University of Glasgow, Scotland, September 2012.
- [96] R. Johnston. The Complete Universe. PhD thesis, School of Physics & Astronomy, University of Glasgow, Scotland, July 2009.
- [97] R. Johnston. Shedding light on the galaxy luminosity function. The Astronomy & Astrophysics Review, 19:41, February 2011.
- [98] R. Johnston, L. Teodoro, and M. Hendry. Completeness I. Revisited, reviewed and revived. MNRAS, 376(4):1757–1766, April 2007.
- [99] R. Johnston, L. Teodoro, and M. Hendry. Completeness II. A signal-to-noise ratio approach for completeness estimators applied to galaxy magnitude-redshift surveys. MNRAS, 405(2):1187–1194, June 2010.
- [100] R. Johnston, L. Teodoro, and M. Hendry. Completeness III. Identifying characteristic systematics and evolution in galaxy redshift surveys. MNRAS, 421(1): 270–283, March 2012.
- [101] F. D. Kahn and L. Woltjer. Intergalactic matter and the galaxy. Astrophysical Journal, 130:705, 1959.
- [102] N. Kaiser. On the spatial correlations of Abell clusters. Astrophysical Journal, Part 2 - Letters to the Editor, 284:L9–L12, 1984.
- [103] A. Klypin et al. Where are the missing galactic satellites? Astrophysical Journal, 522(1):82–92, September 1999.
- [104] T. Kolatt and A. Dekel. Large-scale power spectrum from peculiar velocities. Astrophysical Journal, 479(2):592–605, April 1997.
- [105] E. Komatsu et al. Five-year Wilkinson Microwave Anisotropy Probe observations: Cosmological interpretation. Astrophysical Journal Supplement Series, 180(2): 330–376, February 2009.

- [106] E. Komatsu et al. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Cosmological interpretation. Astrophysical Journal Supplement Series, 192(2):47, February 2011.
- [107] J. Kormendy, N. Drory, R. Bender, and M. E. Cornell. Bulgeless giant galaxies challenge our picture of galaxy formation by hierarchical clustering. Astrophysical Journal, 723(1):54–80, November 2010.
- [108] M. Kowalski et al. Improved cosmological constraints from new, old, and combined supernova data sets. Astrophysical Journal, 686(2):749–778, October 2008.
- [109] S. Kullback. Information Theory and Statistics. Dover Publications, Gloucester, Massuchusetts, 1968.
- [110] S. Kullback and R. A. Leibler. On information and sufficiency. The Annals of Mathematical Statistics, 22(1):79–86, 1951.
- [111] O. Lahav et al. Dynamical effects of the cosmological constant. MNRAS, 251: 128–136, 1991.
- [112] O. Lahav et al. The supergalactic plane revisited with the Optical Redshift Survey. MNRAS, 312(1):166–176, 2000.
- [113] O. Lahav and Y. Suto. Measuring our universe from galaxy redshift surveys. Living Reviews in Relativity, 7(1):81, July 2004.
- [114] S. K. Lamoreaux. Demonstration of the Casimir Force in the 0.6 to 6µm Range. Physical Review Letters, 78(1):5–8, January 1997.
- [115] S. D. Landy and A. S. Szalay. A general analytical solution to the problem of Malmquist bias due to lognormal distance errors. Astrophysical Journal, Part I, 391(2):494–501, June 1992.
- [116] T. R. Lauer and M. Postman. The motion of the Local Group with respect to the 15,000 kilometer per second Abell cluster inertial frame. Astrophysical Journal, Part 1, 425(2):418–438, April 1994.
- [117] S. Leighton. Estimation of the linear redshift distortion parameter, β , from current and future galaxy surveys. Master's thesis, School of Physics & Astronomy, September 2011.
- [118] G. Lemaître. Un univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques. Annales de la Société Scientifique de Bruxelles, A47:49–59, 1927.

- [119] G. Lemaître. Expansion of the universe: A homogeneous universe of constant mass and increasing radius accounting for the radial velocity of extra-galactic nebulae. MNRAS, 91:483–490, March 1931.
- [120] D. Lennarz, D. Altmann, and C. Wiebusch. A unified supernova catalogue. Astronomy & Astrophysics, 538:9, February 2012.
- [121] A. Liddle. An Introduction to Modern Cosmology. Wiley, 2nd edition, 2003.
- [122] M. Longair. Galaxy Formation. Springer-Verlag Berlin Heidelberg, 2nd edition, 2008.
- [123] M. López-Corredoira. Alcock-Paczyński cosmological test. Astrophysical Journal, 781(2):15, February 2014.
- [124] LSST Corporation. About LSST | The Large Synoptic Survey Telescope. LSST Corporation, 2018. Accessed 14/10/2018.
- [125] C. Maccone. Deep Space Flight and Communications: Exploiting the Sun as a Gravitational Lens, chapter A simple introduction to the KLT (Karhunen-Loève Transform), pages 151–179. Springer, Berlin, Heidelberg, Praxis Publishing Ltd, Chichester, UK, 2009.
- [126] C. Magoulas et al. Measuring the cosmic bulk flow with 6dFGSv. In The Zel'dovich Universe: Genesis and Growth of the Cosmic Web, volume 308 of Proceedings of the International Astronomical Union, pages 336–339. IAU Symposium, October 2016.
- [127] K. G. Malmquist. On some relations in stellar statistics. Meddelanden fran Lunds Astronomiska Observatorium Series I, 100:1–52, March 1922.
- [128] K. G. Malmquist. A contribution to the problem of determining the distribution in space of the stars. Meddelanden fran Lunds Astronomiska Observatorium, Series I, 106:1–12, February 1925.
- [129] S. S. McGaugh. Baryonic Tully-Fisher Relations. In G. José, S. J. Funes, and E. M. Corsini, editors, *Galaxy Disks and Disk Galaxies*, volume 230 of ASP Conference Series, pages 541–544, Pontifical Gregorian University, June 2001. San Francisco: Astronomical Society of the Pacific.
- [130] S. S. McGaugh, J. M. Schombert, G. D. Bothun, and W. J. G. de Blok. The Baryonic Tully-Fisher Relation. Astrophysical Journal, 533(2):L99–L102, April 2000.

- [131] M. McQuinn and M. White. Cosmological perturbation theory in 1+1 dimensions. Journal of Cosmology and Astroparticle Physics, 2016(1):9–10, January 2016.
- [132] N. Metcalfe and T. Shanks, editors. The 2dF-SDSS LRG and QSO (2SLAQ) Survey: LRG Clustering and Redshift-Space Distortions, volume 379 of Cosmic Frontiers ASP Conference Series, Durham University, Durham, UK, December 2005. ASP.
- [133] R. Minkowski. Spectra of supernovae. Publications of the Astronomical Society of the Pacific, 53(314):224, August 1941.
- [134] R. Minkowski. Internal dispersion of velocities in other galaxies. In G. C. McVittie, editor, *Problems of Extra-Galactic Research*, Proceedings from IAU Symposium no. 15, page 115, New York, 1962. International Astronomical Union Symposium, Macmillan Press.
- [135] A. Mosley. The starry messenger: The celestial sphere. Department of History and Philosophy of Science - University of Cambridge, 1999. Accessed 17/8/2012.
- [136] R. L. Murphy. Nature of dark matter. AstronomyOnline.org, 2012. Accessed 2/9/2012.
- [137] NASA. NASA Hubble goes to the eXtreme to assemble farthest-ever view of the Universe. National Aeronautics and Space Administration, September 2012. Accessed 8/6/2018.
- [138] NASA. WMAP Shape of the Universe. National Aeronatics and Space Administration, January 2014. Accessed 14/6/2018.
- [139] J. D. Neill, M. J. Hudson, and A. Conley. The peculiar velocities of local Type Ia supernovae and their impact on cosmology. Astrophysical Journal, 661(2):L123– L126, 2007.
- [140] A. Nusser and M. Davis. On the prediction of velocity fields from redshift space galaxy samples. Astrophysical Journal, Part 2, 421(1):L1–L4, January 1994.
- [141] F. Ochsenbein, P. Bauer, and J. Marcout. The VizieR database of astronomical catalogues. Astronomy & Astrophysics Supplement, 143:23–32, April 2000.
- [142] I. Odderskov, S. Hannestad, and T. Haugbølle. On the local variation of the Hubble constant. Journal of Cosmology & Astroparticle Physics, 2014(10):15, October 2014.
- [143] C. G. Park and C. Park. Power spectrum of cosmic momentum field measured from the SFI galaxy sample. Astrophysical Journal, 637(1):1–11, 2006.

- [144] M. S. Pawlowski et al. Co-orbiting satellite galaxy structures are still in conflict with the distribution of primordial dwarf galaxies. MNRAS, 442(3):2362–2380, August 2014.
- [145] J. Peacock. MPhys advanced cosmology. Taken from the SUPA graduate Advanced Cosmology course taught at the University of Edinburgh, March 2013.
- [146] J. A. Peacock. Large-scale surveys and cosmic structure. Taken from the lectures delivered at the 2002 Tenerife Winter School, "Dark matter and dark energy in the universe", September 2003.
- [147] P. J. E. Peebles. The Large-Scale Structure of the Universe. Princeton Series in Physics. Princeton University Press, Princeton, New Jersey, November 1980.
- [148] W. J. Percival et al. The 2dF Galaxy Redshift Survey: The power spectrum and the matter content of the Universe. MNRAS, 327(4):1297–1306, November 2001.
- [149] S. Perlmutter et al. Measurements of Ω and Λ from 42 high-redshift supernovae. The Astrophysical Journal, 517(2):565–586, September 1999.
- [150] R. A. Pildis, J. M. Schombert, and A. Eder. Gas-rich dwarf galaxies from the Second Palomar Sky Survey. II. Optical properties. Astrophysical Journal, 481(1): 157–168, May 1997.
- [151] Planck. The cosmic microwave background Planck mission. European Space Agency, 2013. Accessed 3/1/2015.
- [152] Planck Collaboration et al. Planck 2013 results. XV. CMB power spectra and likelihood. Astronomy & Astophysics, 571:60, November 2014.
- [153] Planck Collaboration et al. Planck 2013 results. XVI. Cosmological parameters. Astronomy & Astrophysics, 571:66, November 2014.
- [154] Planck Collaboration et al. Planck 2015 results. XI. CMB power spectra, likelihoods, and robustness of parameters. Astronomy & Astrophysics, 594:99, September 2016.
- [155] A. Pope et al. Cosmological parameters from eigenmode analysis of Sloan Digital Sky Survey galaxy redshifts. In THE NEW COSMOLOGY: Conference on Strings and Cosmology, volume 743 of The Mitchell Symposium on Observational Cosmology. AIP Conference Proceedings, pages 120–128. AIP, December 2004.
- [156] A. Del Popolo and M. Le Delliou. Small scale problems of the ACDM model: A short review. Galaxies, 5(1):17, February 2017.

- [157] J. Primack. A brief history of dark matter. University of California: Santa Cruz Physics Department, January 2012. Accessed 30/7/2012.
- [158] D. J. Radburn-Smith, J. R. Lucey, and M. J. Hudson. A comparison of local Type Ia supernovae with the IRAS PSCz gravity field. MNRAS, 355(4):1378–1382, 2004.
- [159] S. Rauzy. A simple tool for assessing the completeness in apparent magnitude of magnitude-redshift samples. MNRAS, 324(1):51–56, June 2001.
- [160] S. Rauzy and M. Hendry. A ROBUST method for fitting peculiar velocity field models. MNRAS, 316(3):621–630, August 2000.
- [161] E. Regos and A. S. Szalay. Multipole expansion of the large-scale velocity field
 Using the tensor window function. Astrophysical Journal, Part I, 345:627–636, October 1989.
- [162] A. G. Riess, M. Davis, J. Baker, and R. P. Kirshner. The velocity field from Type Ia supernovae matches the gravity field from galaxy surveys. Astrophysical Journal Letters, 488(1):L1, 1997.
- [163] A. G. Riess et al. Observational evidence from supernovae for an accelerating universe and a cosmological constant. The Astronomical Journal, 116(3):1009– 1038, September 1998.
- [164] A. G. Riess et al. A 3% solution: Determination of the Hubble Constant with the Hubble Space Telescope and Wide Field Camera 3. Astrophysical Journal, 730(2): 18, April 2011.
- [165] V. C. Rubin and W.K. Ford. Rotation of the Andromeda Nebula from a spectroscopic survey of emission regions. Astrophysical Journal, 159:379, 1970.
- [166] B. Salmon. FRS110: Dark matter. Princeton University WordPress Service, 2005. Accessed 28/8/2012.
- [167] J. J. Salzer and M. P. Haynes. HI redshift surveys and large scale structure. In E. D. Skillman, editor, *The Minnesota Lectures on Extragalactic Neutral Hydro*gen, volume 106 of Astronomical Society of the Pacific Conference Series, page 357, University of Minnesota, Minneapolis, Minnesota, 1996. San Francisco: Astronomical Society of the Pacific, ASP.
- [168] A. Sandage. Current problems in the extragalactic distance scale. Astrophysical Journal, 127:513, May 1958.
- [169] A. Sandage and G. A. Tammann. Steps toward the Hubble constant. III The distance and stellar content of the M101 group of galaxies. Astrophysical Journal, 194:223–243, December 1974.

- [170] W. Saunders et al. The PSCz catalogue. MNRAS, 317(1):55–63, 2000.
- [171] P. L. Schechter. Mass-to-light ratios for elliptical galaxies. The Astronomical Journal, 85:801–811, July 1980.
- [172] I. M. Schmoldt et al. On density and velocity fields and β from the IRAS PSCz survey. The Astronomical Journal, 118(3):1146–1160, September 1999.
- [173] SDSS. SDSS filters. Astrophysical Research Consortium, 2002. Accessed 7/6/2018.
- [174] SDSS. Skyserver: About SDSS. Astrophysical Research Consortium, 2005. Accessed 6/6/2017.
- [175] SDSS. Scope | SDSS. Astrophysical Research Consortium, 2015. Accessed 6/6/2017.
- [176] SDSS. Scope | SDSS. Astrophysical Research Consortium, 2018. Accessed 10/6/2018.
- [177] B. Senevirathne. Probing the large-scale galaxy density and peculiar velocity fields. Master's thesis, University of Glasgow, 2010.
- [178] S. F. Shandarin and Y. B. Zeldovich. The large-scale structure of the Universe: Turbulence, intermittency, structures in a self-gravitating medium. *Reviews of Modern Physics*, 61(2):185–220, April 1989.
- [179] J. Silk. A Short History of the Universe. Scientific American Library, 1997.
- [180] M.F. Skrutskie et al. The Two Micron All Sky Survey. The Astronomical Journal, 131(2):1163–1183, February 2006.
- [181] V. M. Slipher. Nebulae. Proceedings of the American Philosophical Society, 56: 403–409, 1917.
- [182] D. N. Spergel et al. First-Year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Preliminary maps and basic results. The Astrophysical Journal Supplement Series, 148(1):1–27, September 2003.
- [183] D. N. Spergel et al. Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Implications for cosmology. Astrophysical Journal Supplement Series, 170(2):377–408, June 2007.
- [184] M. A. Strauss and J. A. Willick. The density and peculiar velocity fields of nearby galaxies. *Physics Reports*, 261:271–431, February 1995.
- [185] N. Suzuki et al. Supernova Cosmology Project. Lawrence Berkeley National Lab, April 2010. Accessed 3/9/2012.

- [186] N. Suzuki et al. The Hubble Space Telescope Cluster Supernova Survey. V. Improving the dark-energy constraints above z > 1 and building an early-type-hosted supernova sample. Astrophysical Journal, 746(1):1–24, 2012.
- [187] R. Tojeiro. Understanding the Cosmic Microwave Background Temperature power spectrum. Taken from the Pedagogical Seminars given at the Institute for Astronomy at the University of Edinburgh, March 2006.
- [188] J. L. Tonry, J. P. Blakeslee, E. A. Ajhar, and A. Dressler. The surface brightness fluctuation survey of galaxy distances. II. Local and large-scale flows. Astrophysical Journal, 530(2):625–651, February 2000.
- [189] J. L. Tonry et al. Cosmological results from high-z supernovae. Astrophysical Journal, 594(1):1–24, 2003.
- [190] S. Torres-Flores, B. Epinat, P. Amram, H. Plana, and C. Mendes de Oliveira. GHASP: an Hα kinematic survey of spiral and irregular galaxies - IX. The nearinfrared, stellar and baryonic Tully-Fisher relations. MNRAS, 416(3):1936–1948, September 2011.
- [191] D. Yu Tsetkov et al. SAI Supernova Catalog. Sternberg Astronomical Institute, 2012. Accessed 18/8/2012.
- [192] R. B. Tully, H. M. Courtois, and J. G. Sorce. Cosmicflows-3. The Astronomical Journal, 152(2):21, August 2016.
- [193] R. B. Tully et al. Our peculiar motion away from the local void. The Astrophysical Journal, 676(1):184–205, March 2008.
- [194] R. B. Tully et al. Cosmicflows-2: The Data. The Astronomical Journal, 146(4): 25, October 2013.
- [195] R. B. Tully and J. R. Fisher. A new method of determining distances to galaxies. Astronomy & Astrophysics, 54(3):661–673, February 1977.
- [196] S. J. Turnbull et al. Cosmic flows in the nearby universe from Type Ia supernovae. MNRAS, 420(1):447–454, February 2012.
- [197] G. Vaucouleurs. Evidence for a local super galaxy. The Astronomical Journal, 58: 30, 1953.
- [198] F. Vincent. Positional astronomy: Galactic coordinates. Astronomy Group: University of St. Andrews, 2011. Accessed 17/8/2012.

- [199] J. V. Wall. Practical Statistics for Astronomers II. Correlation, Data-modelling and Sample Comparison. Quarterly Journal of the Royal Astronomical Society, 37:519, December 1996.
- [200] S. Watson. An exposition on inflationary cosmology. eprint arXiv:astroph/0005003, pages 18–19, April 2000.
- [201] M. White. The Zel'dovich Approximation. MNRAS, 439(4):3630–3640, April 2014.
- [202] N. Wiener. Extrapolation, interpolation, and smoothing of stationary time series. Technology Press of the Massachusetts Institute of Technology, 1949.
- [203] J. A. Willick. The Las Campanas Observatory/Palomar 10,000 kilometer per second cluster survey. II. Constraints on large-scale streaming. Astrophysical Journal, 522(2):647–660, September 1999.
- [204] J. A. Willick and M. A. Strauss. Maximum likelihood comparison of Tully-Fisher and redshift data. II. Results from an expanded sample. Astrophysical Journal, 507(1):64–83, November 1998.
- [205] J. A. Willick, M. A. Strauss, A. Dekel, and T. Kolatt. Maximum likelihood comparisons of Tully-Fisher and redshift data: Constraints on Ω and biasing. Astrophysical Journal, 486(2):629–664, September 1997.
- [206] WMAP. WMAP Content of the universe. National Aeronautics and Space Administration, 2011. Accessed 18/8/2012.
- [207] N. Yoshida, A. Sokasian, L. Hernquist, and V. Springel. Early structure formation and reionization in a warm dark matter cosmology. Astrophysical Journal, 591(1): L1–L4, 2003.
- [208] D. Zaritsky et al. The Baryonic Tully-Fisher relationship for S⁴G galaxies and the "condensed" baryon fraction of galaxies. *The Astronomical Journal*, 147(6): 11, June 2014.
- [209] S. Zaroubi. Unbiased reconstruction of the large-scale structure. MNRAS, 331(4): 901–908, April 2002.
- [210] S. Zaroubi. Cosmic Flows: Review of Recent Developments. In L. Celnikier and J. T. Thanh Van, editors, *Frontiers of the Universe*, Proceedings of the XIIIrd Rencontres de Blois, page 65, Vietnam, 2004. Tha Giai Publishers.
- [211] S. Zaroubi, E. Branchini, Y. Hoffman, and L. N. da Costa. Consistent β values from density-density and velocity-velocity comparisons. MNRAS, 336(4):1234– 1246, November 2002.

- [212] S. Zaroubi et al. Large-scale power spectrum and structures from the ENEAR galaxy peculiar velocity catalogue. *MNRAS*, 326(1):375–386, September 2001.
- [213] S. Zaroubi, Y. Hoffman, K. B. Fisher, and O. Lahav. Wiener reconstruction of the large-scale structure. Astrophysical Journal, 449:446, August 1995.
- [214] S. Zaroubi, I. Zehavi, A. Dekel, Y. Hoffman, and T. Kolatt. Large-scale power spectrum from peculiar velocities via likelihood analysis. Astrophysical Journal, 486(1):21–31, September 1997.
- [215] I. B. Zel'dovich, J. Einasto, and S. F. Shandarin. Giant voids in the universe. *Nature*, 300:407–413, 1982.
- [216] Y. B. Zel'dovich. Gravitational instability: An approximate theory for large density perturbations. Astronomy & Astrophysics, 5:84–89, March 1970.
- [217] F. Zwicky. On the masses of nebulae and of clusters of nebulae. Astrophysical Journal, 86:217, 1937.
- [218] F. Zwicky. Basic results of the international search for Supernovae. Annales d'Astrophysique, 27(300):300, February 1964.