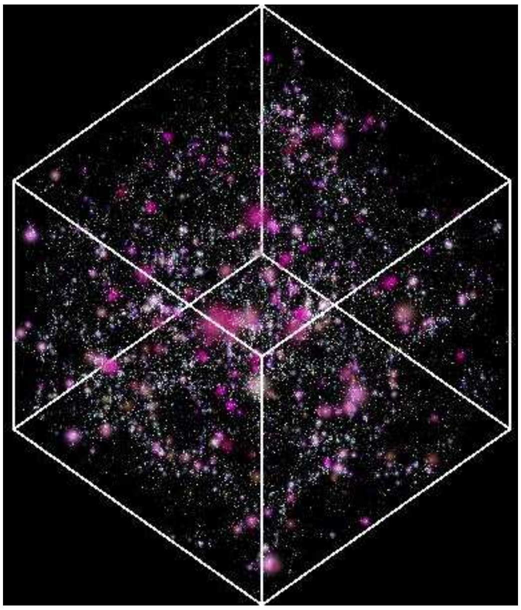
MULTISCALE PROBABILITY MAPPING

A Coarse-Grained Universe, Large-Scale Structure and a Test of the Timescape Cosmology



Anthony G. Smith

Sydney Institute for Astronomy (SIfA), School of Physics, University of Sydney

Supervisors: Andrew M. Hopkins and Richard W. Hunstead

2012

THE UNIVERSITY OF SYDNEY

A Thesis submitted for the degree of Master of Science (Research)

Title Illustration: A cube of side 160 h^{-1} Mpc at $0.06 < z < 1.12$ encompassing the scale of homogeneity, displaying groups and clusters from the MSPM catalogue as a solid particle render (Appendix B).

Abstract

We have developed a multiscale structure identification algorithm for the detection of overdensities in galaxy data that identifies structures having radii within a user-defined range. Our "multiscale probability mapping" technique can be used to construct a coarse-grained map of the underlying fine-grained galaxy distribution, from which overdense structures are then identified. Applying our approach to Data Release 7 of the Sloan Digital Sky Survey, we have compiled a catalogue of groups and clusters at $0.025 < z < 0.24$. A clear trend of increasing velocity dispersion with radius from 0.2 to 1 h^{-1} Mpc is detected, confirming the lack of a sharp division between groups and clusters.

By using our group and cluster catalogue as a coarse-grained representation of the galaxy distribution for structure sizes of $\lesssim 1\,h^{-1}$ Mpc, we demonstrate a technique to identify filaments with a false discovery rate of less than one half. We have identified 53 filaments (from an algorithmically-derived set of 100 candidates) as elongated unions of groups and clusters at $0.025 < z < 0.13$. We have identified 47 voids as 40 h^{-1} Mpc spherical volumes that contain no MSPM groups or clusters, demonstrating a unified approach to the identification of underdense and overdense features of large-scale structure. For our assumed void definition, the void volume fraction at the current epoch is $f_{\rm v0} = 0.75 \pm 0.11$. A test is proposed to distinguish the timescape cosmology from ΛCDM, based on the density of gravitationally-bound structures relative to the critical density as measured from the radius-velocity dispersion relation. An initial calculation is attempted, but significant systematic issues remain unresolved.

Publication: *Multiscale probability mapping: groups, clusters and an algorithmic search for filaments in SDSS* by Smith, A. G. Hopkins, A. M. Hunstead, R. W. Pimbblet, K. A. 2012, MNRAS.

Sections 1.2, 2.2-2.4 and 3.2-3.4 are taken from this paper with minor adaptations.

Data Products and Online Material: Group and cluster, filament and void catalogues, along with animated three-dimensional visualisations are available at:
`http://www.physics.usyd.edu.au/sifa/Main/MSPM/`
The group, cluster and filament catalogues are also available in the online version of our paper.

Contact
Anthony G. Smith: asmith@physics.usyd.edu.au
Andrew M. Hopkins: ahopkins@aao.gov.au
Richard W. Hunstead: rwh@physics.usyd.edu.au

Sydney Institute for Astronomy (SIfA): http://www.physics.usyd.edu.au/sifa/

Where necessary for distance measurements, we have adopted $H_0 = 100h$ km s^{-1} Mpc^{-1}, $\Omega_{\mathrm{M}0} = 0.3$ and $\Omega_\Lambda = 0.7$, though the exact choice of values does not significantly affect results at $0 < z < 0.25$. Except where otherwise indicated, all distances are comoving. Zeroes in subscripts (e.g. H_0) denote current epoch quantities.

Acknowledgements

In Xanadu did Kubla Khan
A stately pleasure-dome decree:
Where Alph, the sacred river, ran
Through caverns measureless to man
Down to a sunless sea.

So twice five miles of fertile ground
With walls and towers were girdled round
. . .
To such a deep delight 'twould win me,
That with music loud and long,
I would build that dome in air,
That sunny dome! those caves of ice!

– Lines 1-7 and 44-47 from Samuel Taylor Coleridge, *Kubla Khan*

Thanks to my exalted effendis, Andrew Hopkins and Dick Hunstead, for all the things supervisors do, but also for the freedom I've had to carry out some of the riskier investigations herein. Chris Miller was there at the beginning to tell us MSPM hadn't "been done". It was really handy to have Kevin Pimbblet around to help with the filament work (including an independent morphological classification of the 53 MSPM filaments in Section 3.4), which made this worthwhile in my mind. I had a real blast working on the cosmology, and found the related chat with David Wiltshire both reassuring and encouraging. Chris Blake put us in touch with Berian James, who read a near-final version of Chapter 5 and gave me some great feedback.

All told, I had useful discussions with and got handy hints from Chris Miller, Scott Croom, Ben Jelliffe, Kevin Pimbblet, Aditya Menon, Juliana Kwan, David Wiltshire, Chris Blake and Berian James. Thanks also to the anonymous referee on our paper for detailed comments that helped to improve our group, cluster and filament paper. He also deserves a hand for giving us equation 2.2 . . . well almost. Improvements to this Thesis have also been made following feedback from examiners, Scott Croom and one anonymous.

Cheers to Richard Thompson, Xenogene Gray, Kevin Aquino, Mat Guenette, Stuart Gilchrist and others for some good teaching times. And to my Overlook office-mates, Kate Randall, Ben Jelliffe and Sarah Reeves, who sent me a birthday card the night before I submitted this Thesis, on the big day before the big day! I like to think Igor, Fish Vader and Ralph among other aquatic companions wondered a little about the world outside, and that oaf at the other end of the office – blessed are the lost. Working on Dragon's remake of *Star Control 1* helped me with the minimal spanning tree and visualisation work, and I will eventually model those sand dunes!

Call me weird but I also thank the many computers who crunched the numbers. Multivac, Camellia and Daneel on the home front, siva and prominence on my desk and mammuthus for carrying the data. A lot of the heavy lifting was done by deinotherium – that hoe tusker is a trusty steed indeed!

This research has made use of NASA's Astrophysics Data System, a guide to cosmographic distance measures (Hogg 1999), the VizieR database of astronomical catalogues (Ochsenbein, Bauer & Marcout 2000) and the SIMBAD database, operated at CDS, Strasbourg, France. The .gif animations of cosmic structure (Appendix B) were encoded by images2gif from python's visvis library[1]. In Sections 2.4.4 and 5.4 we make use of group and cluster catalogues compiled by Miller et al. (2005), Berlind et al. (2006) and Yoon et al. (2008). The galaxy data on which this research is based are from the Sloan Digital Sky Survey (SDSS), Data Release 7. Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is http://www.sdss.org/.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

As a wannabe artist, I've tried to add a philosophical subtext to this Thesis and the ornate quotations are taken from these, including some of my favourites:
Samuel Taylor Coleridge, *Kubla Khan* (page iii)
Homer, *The Odyssey* (page 2)
Theodore Taylor, *The Cay* (page 3)
Jung Chang, *Wild Swans: Three Daughters of China* (page 24)
Wu Cheng-en, *Journey to the West* (page 53)
Frank Herbert, *Dune* (page 54)
Kim Stanley Robinson, *The Years of Rice and Salt* (page 66)
Isaac Asimov, *The End of Eternity* (page 77)
Theodore Taylor, *Timothy of the Cay* (page 98)
Jonathan Nolan, *Memento Mori* (page 101).
Spoilers are in *Message for the Arch Vile* on page 101. Musical quotations in the epilogue were

[1]http://code.google.com/p/visvis/

typeset with LilyPond[1].

Finally, not to disparage any of the papers in the reference list for this Thesis, but there are some papers one checks to retrieve some obscure detail. Then there are others that turn out to be intriguing, useful or cool in some other way:

Balian & Schaeffer 1989: *Scale-invariant matter distribution in the universe. I – Counts in cells*. A fascinating account of the statistics of the galaxy distribution, which I read as I was testing MSPM at the beginning of this project. Hopefully at some point I'll get around to reading its sequel on the fractal nature of the galaxy distribution.

van de Weygaert & Schaap 2009: *The Cosmic Web: Geometric Analysis*. Nice overview of large-scale structure with pretty pictures, and an account of one multiscale method to reconstruct the galaxy density field, the DTFE. The DTFE is in a way opposite to MSPM because it performs a reconstruction rather than a reductive coarse-granulation.

Hogg 1999: *Distance measures in cosmology*. This came in handy as I made the transition from contrived particle tests to The Real Thing.

Einasto et al. 1984: *Structure of superclusters and supercluster formation. III Quantitative study of the local supercluster*. A crucial choice one makes in MSPM is the range of scales to sample, and this paper statistically justified $1\,h^{-1}$ Mpc as the maximum radius of sensitivity, before the transition from groups and clusters to filaments.

Abazajian et al. 2009: *The Seventh Data Release of the Sloan Digital Sky Survey*. A good bit of data never hurts!

Miller et al. 2005 (C4): *The C4 Clustering Algorithm: Clusters of Galaxies in the Sloan Digital Sky Survey*. Having come up with a new algorithm, we discussed it with Chris Miller, who gave me a nice overview of existing techniques and wasn't aware that MSPM had "been done". And the C4 paper guided me with the parameters to use in our initial search for clusters, parameters that were later refined to suit our own purposes.

Berlind et al. 2006 (Mr20): *Percolation Galaxy Groups and Clusters in the SDSS Redshift Survey: Identification, Catalogs, and the Multiplicity Function*. Initial results from MSPM found many more structures than C4, but Mr20 was reassuring because like us it sets a relatively low threshold. The sort of tests they do with mock catalogues are pretty interesting as well, although I'm glad we didn't do anything like that with MSPM (Section 5.5.2).

Yoon et al. 2008 (Y08): *A Spectrophotometric Search for Galaxy Clusters in SDSS*. Like C4 and Mr20 this catalogue was useful for comparison.

Navarro, Frenk & White 1997: *A Universal Density Profile from Hierarchical Clustering*. I gave a talk on this but would have come across NFW profiles anyway. An interesting read to me, I knew a fair bit more about simulations after this and it contains a lot of useful little pointers.

Pimbblet 2005: *A new algorithm for the detection of intercluster galaxy filaments using galaxy orientation alignments*. There's no point in coming up with a new algorithm unless it does something existing algorithms can't, and one unsolved problem we were interested in tackling was filament finding. We had an interesting discussion with Kevin Pimbblet about this, and getting back to him years later with progress was a highlight. Aside from that I toiled away

[1] http://lilypond.org/

at the filament-finding problem for some time with little luck, and part of what led me to the success was the realisation that the scales I was looking were too small. The realisation came when I was looking at this paper, even though it isn't specifically about filament sizes. I was madly inventing new approaches to the problem when all I had to do was change one number, and therein lies an interesting lesson in knowing the background and reviewing the basics.

Pimbblet, Drinkwater & Hawkrigg 2004 (PDH): *Intercluster filaments of galaxies programme: abundance and distribution of filaments in the 2dFGRS catalogue*. The filament section of our paper (Smith et al. 2012) does something similar to what this paper did, with a sample that while smaller, was selected with an algorithm followed up by human inspection. The morphological classification we did follows the scheme in there too.

Gottlöber et al. 2003: *The structure of voids*. An intriguing picture is presented in which voids contain haloes that exhibit filaments and subvoids reminiscent of their higher-mass counterparts, a nice fractal-like characterisation. It comes from simulations and is apparently wrong, but it's still a nice image.

Wiltshire 2009b: *From Time to Timescape – Einstein's Unfinished Revolution*. I saw this paper on arXiv and read it because although it had little to do with my research interest it sounded cool, and it read that way too.

Wiltshire 2007a: *Cosmic clocks, cosmic variance and cosmic averages*. And years later the timescape was very relevant to my work, which meant reading the details and getting some gracious hints from David Wiltshire himself.

Bryan & Norman 1998 (BN98): *Statistical Properties of X-Ray Clusters: Analytic and Numerical Comparisons*. This work was important to me on a number of levels. It gave us and the referee on our paper the analytic radius-velocity dispersion relation, verified the validity of the spherical collapse model without even needing a sample of recently-formed structures and gave us the f_σ normalisation.

Voit 2005: *Tracing cosmic evolution with clusters of galaxies*. Useful review, because for half of my project I really didn't have to know very much about the physics of clusters.

Barmby & Huchra 1998: *Kinematics of the Hercules supercluster*. The *Hercules* cluster is the first structure in our group and cluster catalogue, visualised in 3D on page 25 and featured in two figures (2.1 and 5.3). In most of the work of this Thesis, I don't care much for individual cases, but this one turned out to be most illuminating. When in Chapter 5 I realised that I actually didn't have a way to infer virial radii from power-law fits, *Hercules* at least showed that we weren't far off. This paper identifies member galaxies and plots them, identifying a similar volume to what MSPM does which is vital for radius measurement. These authors also report a radius and velocity dispersion that agrees with our measurement, which is nice. If the Arch Vile is the villain of the artsy narrative in here, *Hercules* is the hero, much like its mythical namesake (page 2) – see *Message for the Arch Vile* on page 101.

Cline, Frey & Holder 2008: *Predictions of the causal entropic principle for environmental conditions of the universe*. Star Formation + Entropy + Diamonds = Awesome.

Statement of Student Contribution

The optical data were taken from the SDSS. I designed the MSPM algorithm and the elongation probability measure (Section 3.2), guided by useful discussions with my supervisors (Chapter 1), Andrew M. Hopkins and Richard W. Hunstead. I wrote and ran original matlab/octave code to implement MSPM (Chapter 2) with extensions for filament-finding (Chapter 3) and void-finding (Chapter 4) on the SDSS spectroscopic survey. Primary visual inspections were carried out by myself and an independent morphological classification of the 53 MSPM filaments (Section 3.4) was provided by Kevin A. Pimbblet. I devised and carried out the test of cosmology presented in Chapter 5.

I wrote scripts to generate the text and pdf versions of our data products (Appendix A). I wrote and ran original matlab/octave and python code to create the three-dimensional animations of groups, clusters, filaments, voids and a box the size of statistical homogeneity (Appendix B), and I wrote the SIfA web page that hosts our data products and visualisations:
http://www.physics.usyd.edu.au/sifa/Main/MSPM/ .

I wrote the text of this Thesis and produced its figures, incorporating editorial corrections and improvements as suggested by my supervisors.

I certify that this Thesis contains work carried out
by myself except where otherwise acknowledged.

Signed ....

Date:

Contents

Thesis Outline

This Thesis is divided into two main parts. Part I (Chapters 1 to 2) describes the design of "multiscale probability mapping" (MSPM) and a catalogue of galaxy groups and clusters at $0.025 < z < 0.24$. In Part II (Chapters 3 to 5) this catalogue is used as a coarse-grained representation of the galaxy distribution for structure sizes of $\lesssim 1\ h^{-1}$ Mpc to identify features of large-scale structure, and a model that considers the implications of large-scale structure for cosmology is tested. Sections 1.2, 2.2-2.4 and 3.2-3.4 are taken from Smith et al. (2012) with minor adaptations.

Chapter 1 introduces MSPM in the context of galaxy group and cluster identification, emphasising its ability to create coarse-grained representations of the galaxy distribution in Section 1.2.5. Chapter 2 reviews the history and importance of group and cluster studies before detailing the application of MSPM to the Sloan Digital Sky Survey, Data Release 7. The MSPM group and cluster selection criteria in our implementation are described in Section 2.2, and the catalogue is presented in Section 2.3. The group and cluster radius-velocity dispersion relation is analysed in Section 2.4.3 and the MSPM catalogue is compared with others in Section 2.4.4.

Chapter 3 describes a technique to identify filaments from a coarse-grained representation of the galaxy distribution such as that provided by the MSPM group and cluster catalogue. A false discovery rate of less than a half is demonstrated in Section 3.3, where a sample of MSPM filaments is presented. These filaments are morphologically classified in Section 3.4. Chapter 4 presents a sample of voids identified as spherical volumes that contain no MSPM groups or clusters in Section 4.2, and the void volume fraction at the current epoch is measured in Section 4.3. Chapter 5 revisits the issue of low group and cluster density contrasts relative to the critical density, determined from the radius-velocity dispersion relation's slope in Section 2.4.3. A test to distinguish the timescape cosmology from the standard ΛCDM model is proposed in Section 5.2.2, with an initial calculation presented in Section 5.4.

Chapter 6 summarises our results. The complete MSPM filament and void catalogues can be found in Appendix A, and visualisation of cosmic structure is described in Appendix B. Where necessary for distance measurements, we have adopted $H_0 = 100h$ km s^{-1} Mpc^{-1}, $\Omega_{\mathrm{M0}} = 0.3$ and $\Omega_\Lambda = 0.7$, though the exact choice of values does not significantly affect results at $0 < z < 0.25$. Except where otherwise indicated, all distances are comoving. Zeroes in subscripts (e.g. H_0) denote current epoch quantities.

Part I

Islands and Banks

Our ship now past the straits of th' ocean flood,
She plow'd the broad sea's billows, and made good
The isle Aeaea, where the palace stands
Of th' early riser with the rosy hands,
Active Aurora, where she loves to dance,
And where the Sun doth his prime beams advance.

When here arrived, we drew her up to land,
And trod ourselves the re-saluted sand,
Found on the shore fit resting for the night,
Slept, and expected the celestial light.

Soon as the white-and-red-mix'd-finger'd dame
Had gilt the mountains with her saffron flame,
I sent my men to Circe's house before,
To fetch deceas'd Elpenor to the shore.

Straight swell'd the high banks with felled heaps of trees,
And, full of tears, we did due exsequies
To our dead friend. Whose corse consum'd with fire
And honour'd arms, whose sepulchre entire

Alcmena next I saw, that famous wife
Was to Amphitryo, and honour'd life
Gave to the lion-hearted Hercules,
That was of Jove's embrace the great increase.
. . .
Down with these was thrust
The idol of the force of Hercules;
But his firm self did no such fate oppress,
He feasting lives amongst th' immortal states

– Homer, *The Odyssey* Book Twelve Lines 1-18, Book Eleven Lines 340-3 and 818-21
(Chapman translation)

Lost

Timothy waited a long time before answering, probably trying to choose the right words. Finally, he said, 'Young bahss, dere is, in dis part of d'sea, a few lil' cays like dis one, surround on bot' sides by hombug banks. Dey are cut off from d'res o' d'sea by dese banks . . .'

I tried to make a mental picture of that. Several small islands tucked up inside great banks of coral that made navigation dangerous was what I finally decided on.

'You think we are on one of those cays?'

'Mebbe, young bahss, mebbe.'

Fear coming back to me - I knew he'd made a mistake in bringing us ashore - I said, 'Then no ships will pass even close to us. Not even schooners! We're trapped here!' We might live here forever, I thought.

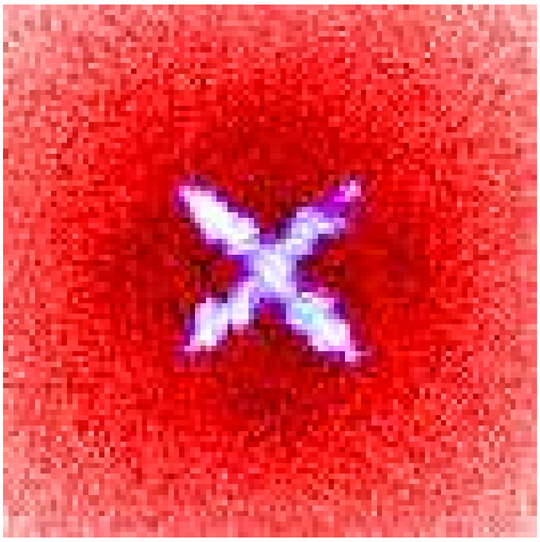
Again he did not answer directly. I was beginning to learn that he had a way of being honest while still being dishonest. He said, 'D'place I am tinking of is call Debil's Mout'. 'Tis a U-shaped ting, wit dese sharp coral banks on either side, runnin' maybe forty, fifty mile . . .'

He let that sink in. It sounded bad. But then he said, 'I do hope, young bahss, dat I am outrageous mistaken.'

– Theodore Taylor, *The Cay*

Chapter 1

Coarse-Graining the Universe



Representation of a contrived particle distribution with P (blue) and S (red).

Section 1.2 of this Chapter is taken from Smith et al. (2012) with minor adaptations.

The distribution of galaxies in the Universe is structured, comprising groups and clusters, filaments, sheets and voids. Measurements of this distribution, through statistics such as correlation functions and power spectrum analysis, have been fundamental in allowing numerous measurements of the cosmological parameters that define the geometry of our Universe. These nonlocal statistics, though, quantify global properties of this distribution, while mapping algorithms locate specific structures within it. In this Chapter we develop a multi-scale structure identification algorithm for the detection of overdensities in galaxy data that identifies structures having radii within a user-defined range. Our “multiscale probability mapping” technique combines density estimation with a shape statistic to identify local peaks in the density field. This technique takes advantage of a user-defined range of scale sizes, which are used in constructing a coarse-grained map of the underlying fine-grained galaxy distribution, from which overdense structures are then identified.

1.1 Review: Mapping Algorithms

Galaxies have long been known to preferentially reside in groups and clusters (e.g. Biviano 2000). Usually, the locations of these structures are inferred from the presence of their member galaxies, making the process of cluster detection a two-part process: first the galaxies are detected, and then the clusters are identified. The latter of these steps is a general problem that has admitted a variety of solutions. The simplest of these solutions is to manually inspect the galaxy distribution to identify overdense regions. This solution is neither objective nor precisely reproducible, and intractably time-consuming where large galaxy catalogues are concerned.

Efficient and objective algorithms are required to find and quantify groups and clusters in galaxy data, and there is great freedom in the design of these algorithms. Some are purely geometric and attempt to make a minimum of assumptions about the structures sought, while others assume properties derived from well-studied samples to optimise the search. The former approach is valuable because assumptions about the properties of clusters in the identification technique can introduce a bias toward finding only those types of structures. The latter approach is employed to optimise the extraction of structures with properties that are known to make them interesting. Many such algorithms have been developed or refined in recent times to explore the wealth of data available through galaxy surveys (e.g. Miller et al. 2005; Koester et al. 2007a; Dong et al. 2008; Robotham et al. 2011) such as the Sloan Digital Sky Survey (SDSS; York et al. 2000; Abazajian et al. 2009).

1.1.1 Astrophysically-Constrained Algorithms

In contrast to algorithms that are general enough for application to any set of particles are techniques that specifically target galaxy data.

1.1.1.1 Red Sequence

This technique takes advantage of the observation that most galaxy clusters possess a central population of early-type galaxies that exhibit a tight correlation in colour-magnitude space, known as the red sequence. All the members of a cluster will have similar histories and so will cluster in colour-magnitude space, but the red sequence clustering is said to be especially pronounced and more uniform between clusters. This is so because the massive galaxies which make up the red sequence should have all formed by $z = 2$ and are not as sensitive to relatively fleeting episodes of star formation or environmental disturbances as other cluster members.

Gladders & Yee (2000) describe a way to exploit the red sequence to identify clusters. To begin, each galaxy has position, colour and magnitude. The expected colour and slope of the red sequence as a function of redshift is used to guide the partition of the colour-magnitude diagram into colour slices. Each slice might be occupied by a red sequence, and this hypothesis is tested for each slice. The probability of a range of counts within each slice occurring at random (accounting for background densities) is found, constructing a probability density function. The probability of finding one specific count or more is found, and the task of finding clusters then becomes that of finding the high-probability slices, for the colour-magnitude diagram at each location. This technique is less useful for finding poorer groups in which the red sequence is not well defined, or for locating clusters at redshifts where the red sequence has not yet evolved.

Results (initial description by Gladders & Yee 2005) for the 100 square degree red-sequence cluster survey (RCS) give about 1000 cluster candidates, with most of them at $0.35 < z < 0.95$. Gilbank et al. (2007) have spectroscopically followed up eleven candidates with $0.6 < z_{\rm phot} < 1.2$ (as determined by the red sequence), firmly agreeing with the photometric redshifts and concluding that over this redshift range they are accurate to within 10%. Ten of the clusters were confirmed as real, where the remaining one was beyond the reach of the spectroscopic data. Working with their own sample of red sequence clusters, Muzzin et al. (2008) found that the red sequence redshift estimation is accurate to within 0.04 for $0.1 < z < 1$. Although many have used this technique, few have adapted or modified it.

1.1.1.2 Brightest Cluster Galaxy

The maxBCG algorithm (e.g. Bahcall et al. 2003; Koester et al. 2007a) simultaneously exploits three properties thought to broadly characterise galaxy clusters. These are spatial overdensity, the red sequence (also known as the "E/SO ridgeline") and the existence of a brightest cluster galaxy (BCG). The original part of maxBCG is the BCG component. Assuming that each cluster has a BCG, one can identify clusters by the presence of their BCGs. The algorithm works in this backwards fashion, so each galaxy within a survey is assigned a probability that it is a BCG. This is found as a joint probability, for which there are two components, because there are two conditions that must be satisfied: for a galaxy to be a BCG, it needs to have both environmental and intrinsic properties to match. So the likelihood $\mathcal{L}_{\rm total}$ of a galaxy being a BCG is

$$\mathcal{L}_{\rm total} = \mathcal{L}_{\rm BCG} \times \mathcal{L}_{\rm R} \tag{1.1}$$

where

- $\mathcal{L}_{\mathrm{BCG}}$ = probability that it is a BCG irrespective of environment (from expected BCG properties in redshift, colour and magnitude)
- $\mathcal{L}_{\mathrm{R}}$ = probability that it is a BCG based on its environment (density, red sequence)

Each of these terms is a function of redshift, and the redshift eventually assigned to a cluster is the one that maximises $\mathcal{L}_{\mathrm{total}}$, hence the "max" in maxBCG. $\mathcal{L}_{\mathrm{R}}$ is found by following the procedure of Postman et al. (1996), described in Section 1.1.1.4. A minor difference is that Koester et al. use different filters, one of which models the red sequence. $\mathcal{L}_{\mathrm{R}}$ is *like* the probability found by Postman et al. but not exactly the same because this time they associate their probabilities with each galaxy rather than with each position in a grid; here there is no interest in creating an image. $\mathcal{L}_{\mathrm{BCG}}$ relies on training from a set of known BCGs; an empirical relation between BCG magnitude and redshift (e.g. Brough et al. 2002; Loh & Strauss 2006) is used to select galaxies that have expected BCG properties in redshift, colour and magnitude.

The above steps are carried out on every galaxy in a catalogue, and redshifts that maximise $\mathcal{L}_{\mathrm{total}}$ are assigned to them. All the galaxies are ranked according to the probabilities they get, and the most likely galaxy is selected as a BCG. All galaxies within some arbitrary distance of this BCG are eliminated from the list of potential BCGs, and this process repeats until all galaxies are either BCGs or otherwise eliminated from the list. From the resulting BCGs, all that have probability below a threshold are rejected to keep the false discovery rate down.

Bahcall et al. (2003) found clusters using the hybrid matched filter (HMF; Section 1.1.1.4) of Kim et al. (2002) and maxBCG to produce two independent catalogues. The two catalogues were then merged, the motivation being that the two techniques would select clusters of different types: the HMF would supposedly find clusters with similar properties to those of the filter, and maxBCG would supposedly find clusters dominated by luminous red ellipticals. The BCG likelihood for any given galaxy is based on both its intrinsic and environmental properties. Bahcall et al. don't use the MF for evaluating the environmental BCG likelihood. Environment influences their BCG likelihood by weighting the intrinsic BCG likelihood by the count of galaxies that are fainter than the prospective BCG and within the appropriate E/SO ridgeline consistent with a red sequence.

In SDSS commissioning data, the combined maxBCG/HMF (BH) catalogue contains 799 clusters, 436 from HMF and 524 from maxBCG with 161 simultaneously detected by both techniques. Both techniques were used to estimate redshifts and maxBCG's estimates were closer to the spectroscopic measurements. Richness estimates were also compared (in maxBCG they come from counting galaxies within a set distance of the BCG) for clusters common to both samples, and a large scatter in the comparison was found. Bahcall et al. attribute part of the relatively low matching rate between the BCG and HMF catalogues to uncertainties in the richness estimates. The richness estimates have a dramatic effect because for both samples, a minimum richness is enforced as part of the selection. Causes of uncertainties in richness estimates include cluster substructure and uncertainties in redshift and transverse cluster centre position. Cluster selection functions (% recovery as functions of redshift and richness) for the BCG and HMF catalogues are derived from simulations, and show no obvious qualitative differences, but they are difficult to compare since the richness estimates are not quite the same for the two techniques.

To compare this with the work of Koester et al. (2007a for the algorithm description and 2007b for the catalogue presentation) four years later, the first obvious difference is that maxBCG was then applied to a later data release of the SDSS, resulting in a catalogue of clusters numbering 13 823 (the largest at the time). The clusters are at $0.1 < z < 0.3$ and rich members of this sample feature many earlier known X-ray-selected clusters.

1.1.1.3 C4

The C4 algorithm (Miller et al. 2005) is based on the premise that galaxies in any given cluster have similar star-formation histories, so along with clustering on the sky and in redshift, they should also cluster in colour and magnitude space (e.g. Goto et al. 2002).

C4 starts with a galaxy catalogue in which the galaxies each have seven properties: RA, dec, redshift and four colours. A series of steps is then carried out on each "target" galaxy. An arbitrary distance is chosen and this defines the radius of a sphere in RA, dec, redshift and colour (scaled relative to physical distance according to a cluster colour-magnitude relation) around the target galaxy. Although the size of the physical aperture (RA, dec and redshift) is fixed, C4 is multiscale in the sense that the use of colours allows the detection of structures with a range of sizes. The number of neighbouring galaxies within this sphere is counted; thus, to contribute to this count, a neighbouring galaxy must have a position sufficiently close to the target galaxy, and must have colours sufficiently close to those of the target galaxy.

In this way, a "number count" may be assigned to each galaxy. This number count is then compared with what would be expected if the target galaxy was put at a different, random location to find out how likely the obtained number count is. For each galaxy, this results in the probability of obtaining the observed count of neighbours, in position and colour space. The probabilities are then ranked and candidates rejected according to a threshold applied to minimise false discoveries.

1.1.1.4 Matched Filter

If the properties of the clusters sought are well known, these properties may be used to optimise the smoothing procedure to highlight them. Postman et al. (1996) introduced the matched filter technique, which performs a test of a model distribution of particles in position and parameter space, where the particles are galaxies and the parameters are colour and magnitude. A model cluster distribution is assumed and the probability of obtaining the observed data is calculated. A high probability indicates a high probability of the presence of the structure one was looking for under the assumed model. Postman et al. used a model in which for any given position the observed distribution is the sum of background and cluster contributions, $D = b + c$. This is modeled in position and magnitude, so the background is the expected counts as a function of magnitude and the cluster model must incorporate an expectation of the cluster's distribution in both position and magnitude, so the cluster c can be modeled as

$$c = \Lambda P \phi \tag{1.2}$$

with cluster richness Λ, radial profile P and cluster luminosity function ϕ.

In equation 12 of Postman et al. the model and observed distributions D_{model} and D_{observed} are combined as

$$\frac{(D_{\text{model}} - D_{\text{observed}})^2}{\sigma^2} \tag{1.3}$$

with background σ to quantify the deviation of the data from the model expectation in multiples of the background (which is assumed to be Poisson), so that the probability obtained in the end is a proxy for statistical significance. For a given position on the sky,

$$\ln(\mathcal{L}, \text{cluster likelihood}) \propto \int P\frac{\phi}{b}D_{\text{observed}}d^2r dm \tag{1.4}$$

The integral over d^2r in performed by integrating over all positions of galaxies throughout the survey area. Since these are just particles, D is the union of many delta functions, so the integrand can be rewritten with operators on those delta functions as in equation (17) of Postman et al.:

$$S(i,j) = \sum_{k=1}^{N_g} P[r_k(i,j)]L(m_k) \tag{1.5}$$

The integrand of the previous expression is intended to be evaluated for a single position to give the cluster likelihood at that position, and is repeated for each position. Here (i, j) corresponds to a single position and may be thought of as the pixel at (i, j). S is evaluated for each position and a map of cluster likelihood is thus produced. N_g is the total number of galaxies enumerated by k, r_k is the distance between (i, j) and the kth galaxy. P and L are the radial and flux filters, respectively, matched to the model for the structure being sought, which is where the algorithm gets its name. In this way, a smoothed map is created that highlights regions that conform to an assumed model.

The matched filter (MF) algorithm (Postman et al. 1996) has been improved much since its original implementation. MF finds the likelihood of a cluster occurring at a given position. Kawasaki et al. (1998) vary the technique to have it give not only positions and redshifts but also richness N estimates for cluster candidates. Each position in a uniform grid is tested against a model, just as in Postman et al., generating a cluster likelihood for each position. Kawasaki et al. then effectively (but not explicitly) repeat this test for models that assume different richnesses, which for each position produces a set of cluster likelihoods, one for each richness. For each position, they remember the highest of these cluster likelihoods, and the corresponding richness. Maps of these pairs of values are created, and if peaks appear at the same locations in both maps, those peaks become cluster candidates. For each of these candidates, the whole process is repeated for other redshifts. The redshift that maximises the cluster likelihood for a candidate becomes the estimated redshift, and the richness that corresponds to that maximum likelihood is the richness estimate. In this way, the redshift and richness pair that maximises the cluster likelihood is adopted, but not for each position, only for those positions which, at some redshift, exhibited peaks in the likelihood and richness maps simultaneously.

Schuecker & Böhringer (1998) examine in detail the foundations of MF, noting its reliance on the assumption of dominating background counts, and confirming that this assumption makes it best for distant cluster detection. They eliminate this problem by avoiding large number

Gaussian approximations in deriving a likelihood expression which, they show, is equivalent to that found by Postman et al. (1996) in the case of dominating background counts. In their approach, "inhomogeneous Poisson point processes" may be used to model distributions of particles that have random locations subject to a non-random structure. This is found in the presence of a cluster: the cluster imposes a non-random structure, within which the positions of the individual particles are random. The set then has both deterministic and random elements on different scales, and they can both be made random by randomising the position of the cluster, resulting in a distribution with multiscale properties. Schuecker & Böhringer construct a process that reproduces the model distribution, and they then find the probability of each possible result of this process. In this way, they produce a more general version of the MF, applying it to redshift and richness estimates as in Kawasaki et al.

Up until this point, the MF algorithm evaluates the probability of the data fitting the model of a cluster, at every location, which results in a cluster likelihood image for a given survey region. This entails sampling along a uniform grid. The adaptive matched filter (AMF) of Kepner et al. (1999) instead samples only at the locations of galaxies, in an effort to increase computational efficiency. As noted before, the original MF relies on a Gaussian approximation, which is valid when background counts dominate. For low counts resulting from poor clusters, this Gaussian approximation, the "coarse" filter, causes a bias in the redshift and richness estimates. To do the job properly, Kepner et al. replace the Gaussian approximation with a Poisson likelihood, a "fine" filter. This is more expensive computationally, so Kepner et al. apply the coarse filter for a "first pass", and follow up likely candidates with the fine filter. Finally, the AMF only "fits" a redshift for each cluster candidate from a range of values allowed by errors in redshift estimates for individual galaxies, be they spectroscopic, photometric or non-existent, so accuracy and efficiency are improved. The AMF mostly only improves technical aspects; how different kinds of redshift information in the same catalogue are dealt with, saving time by only sampling at the locations of galaxies and by running in two steps, coarse followed by fine.

Kim et al. (2002) test the AMF, comparing its results with those of the MF. This comparison is carried out using the SDSS Early Data Release, in which the MF is found to be more sensitive while the AMF estimates cluster properties more efficiently. Kim et al. attempt to combine the best attributes of both with a Hybrid Matched Filter (HMF), in which the cluster detection is done by the MF, and cluster properties are estimated by the AMF.

White & Kochanek (2002) test the AMF of Kepner et al. (1999) on a simulation to find what fraction of real clusters it may be expected to detect. They make some changes, not using the coarse filter at all, along with minimising the computational time. Their model is explicitly trained on previously found clusters, and instead of confining the redshift fitting to a range defined simply by the interval allowed by the errors, the individual galaxies are "smeared out" in redshift space according to the redshift uncertainties. The ensuing test finds that the AMF can produce cluster catalogues with high completeness and low false discovery rates.

The scale of the cluster modelled becomes a variable in Dong et al. (2008). For each candidate, a scale is chosen to maximise the cluster likelihood in much the same way as with the parameters redshift and richness.

1.1.2 Single-Scale Algorithms

If observed colour-magnitude and morphological properties are ignored or not available, galaxy positions alone must be used.

1.1.2.1 Gaussian Smoothing, Counts In Cells

Smoothing aids the interpretation of a set of particles, but is only a preliminary step toward identifying discrete structures, which may for example entail the systematic recognition of “bright” regions in the smoothed map. A simple smoothing procedure might involve the division of a survey area into tesselating cells, and the counts in those cells can be used to display the local densities, hence making it easy to identify dense regions. Similarly, the galaxy distribution can be smoothed with a Gaussian of fixed width (e.g. Balogh et al. 2004a; Yoon et al. 2008). The peaks or extended regions in the smoothed map can then be used to identify overdensities.

This smoothing performed on the input galaxies makes the galaxy distribution easier to interpret visually, but in such single-scale smoothing, a scale must be chosen, and different overdensity catalogues or properties are obtained with each possible choice. If the chosen scale is too small, no structures are identified; but if it is too large, all structures are joined together and become indistinguishable.

1.1.3 Multiscale Algorithms

Multiscale algorithms are needed to interpret the information gathered on different scales and to make a choice about which scales are most important at which locations, removing the need for manual inspection of output for many scales. Such multiscale algorithms have already been implemented.

1.1.3.1 Wavelets

Slezak, Bijaoui & Mars (1990) introduced wavelets for finding overdensities in galaxy data. Wavelets partly compensate for a shortcoming of Fourier decompositions by *locating* structures as well as identifying their presence, making them a mapping tool. In time series analysis, for example, a Fourier transform identifies the frequencies present in the data without locating them in time, like correlation analysis. Separate parts or windows of the signal can be Fourier transformed, but in doing this an arbitrary choice must be made, of what scale windows to try and of which intervals. Not all choices are equal, and one of them, where infinitesimally small windows are sampled, contends with the uncertainty principle and admits perfect positional information with no frequency information, which cannot be inferred from a sample of zero length.

By folding this choice into wavelet analysis, location information can be obtained without the need for windowing. The root problem is that the sinusoids of Fourier analysis are not localised, with each sinusoid extending throughout the whole length of the signal. Wavelets are different because different wavelets correspond to different places (τ, translation) and scales

(s). A one-dimensional signal may be composed of wavelets:

$$f(t) = \int\int \gamma(s,\tau)\psi_{s,\tau}(t)d\tau ds \qquad \text{(inverse wavelet transform)} \tag{1.6}$$

in which one makes the signal out of wavelets ψ of various scales and translations. The form of ψ is arbitrary to an extent. Its basic form is specified along with how it scales up and down; this form is constrained to ensure that the $\psi_{s,\tau}$ form a basis set. For the purposes of mapping, the coefficients $\gamma(s,\tau)$ from the wavelet transform are required so that a scale measurement can be obtained for each scale at each position τ. This process results in a smoothing for each of a set of scales.

In 1990, wavelets were relatively new. Slezak, Bijaoui & Mars (1990) begin by describing Dressler's (Dressler 1980) Nth Nearest Neighbour Distance (NNND) approach to identifying groups in galaxy data. For each position within a survey area, one can find the distance of the nth galaxy from that position, and this may be used to define a radius r_n, which encloses an nth nearest neighbour area πr_n^2. The nth nearest neighbour density of galaxies is then (Dressler 1980) $D_n = n/\pi r_n^2$. One supposes that this is a slight improvement over simple counts in cells smoothing with cells of fixed scale; this time the cell sizes vary according to the local environment. But Slezak, Bijaoui & Mars then point out flaws with this approach: the densities obtained from this depend on n (an arbitrary choice), and provide no clear way of visualising or detecting substructure.

Introducing the wavelet transform as a solution, the "mexican hat" wavelet is employed on galaxy data, recovering all of the galaxy groups identified earlier with Dressler's technique. A method to assign properties to the groups is outlined, producing a catalogue of groups with measured properties (ellipticity and position angle). A similar void-finding exercise is carried out and the wavelet transform is characterised as a mathematical microscope.

Escalera & Mazure (1992) demonstrate the usefulness of wavelets for identifying substructures in contrived tests, and then for the Abell cluster A754. They quantify the significance of detections by applying their analysis to mock catalogues, which had been done earlier, only this time they recognise that they should use clustered rather than uniform random catalogues. In all of these investigations, they smooth the various distribtions at various scales, but never outline a technique for selecting a specific scale. Finally, in their analysis of A754, they weight the individual galaxies of the distribution with their line-of-sight velocities, effectively considering this as a third dimension, and identify velocity substructure which prefigures more recent investigations such as Fleenor et al. (2005). Escalera, Slezak & Mazure (1992) carried out a similar study on the Coma cluster, which had hitherto been considered a fairly bland structure. Using 1400 galaxies, they detect structures within the Coma cluster at high significance. To combat edge effects, they find it necessary to only perform their analysis on a sub-area of all the available data.

Martìnez, Paredes & Saar (1993) fit wavelets to structures of various scales in the galaxy distribution. Escalera & MacGillivray (1995) simultaneously identify structures from groups to clusters to voids, quantifying their significance, over an area of 3000 square degrees down to a depth of $B_j = 16.4$.

Scale-scale correlations (Pando et al. 1998; Feng, Deng & Fang 2000; Yang et al. 2002; Ueda & Takeuchi 2004) can be used to discriminate between different models of the hierarchical evolution of large scale structure, and to study the power spectrum. These scale-scale correlations are based on using the coefficients derived from the wavelet transform on a given scale range statistically rather than to map. This is analogous to using counts in cells statistically rather than to map the positions of density contrasts directly.

Vikhlinin et al. (1998), discuss motivations for wavelet scale selection; see their section 3 and figure 2. More recently, Eisenhardt et al. (2008) employ wavelets in a simpler fashion, performing a simple counts-in-cells smoothing of a particle distribution before convolving this density map with

$$\frac{e^{-r^2/(2\sigma_1^2)}}{\sigma_1^2} - \frac{e^{-r^2/(2\sigma_2^2)}}{\sigma_2^2} \tag{1.7}$$

where the scales are not variable, but are fixed with $\sigma_1 = 400\,\mathrm{kpc}$ and $\sigma_2 = 1600\,\mathrm{kpc}$.

1.1.3.2 Minimal Spanning Tree and Friends Of Friends

There are many graphs capable of joining the particles of any given set, and two in particular have been used to locate overdensities, the minimal spanning tree (MST) and the Delaunay tessellation.

The MST was introduced to astrophysics by Barrow, Bhavsar, & Sonoda (1985), and was one of the first filament finders applied to galaxy catalogues. An MST is a unique tree connecting a given set of particles such that the sum of the distances is minimised. The particles are connected in such a way that there are no circuits (this is what makes it a tree), or equivalently, there is only one way of getting from one particle to another. Many such trees are possible, but the minimal spanning one is that one which minimises the sum of lengths of connecting lines. This tree is supposed to visually identify the underlying features of the distribution, and is thought to be an apt tool to this end because the MST is the least complicated way of connecting the particles, so it extracts the least complicated representation of the distribution.

Even so, the MST can still be difficult to interpret, so it may be "pruned" by cutting away branches shorter than an arbitrary length, or separated by eliminating connections longer than some arbitrary length. Pruning is intended to highlight features that are otherwise hidden by small-scale structure that doesn't reveal anything about the underlying pattern, and separating can be used to isolate groups. These can be seen as ways of varying the scale range the MST is sensitive to. The discrete groups identified in an MST without pruning are also known as percolation, or friends-of-friends (FoF) groups, and can be identified without an MST (Huchra & Geller 1982), by linking together all particles that are within an arbitrary distance of each other.

A lot of the interest in MSTs concerns the statistics they enable. Krzewina & Saslaw (1996) describe a technique for smoothing the potentially confusing jagged lines which one sees in MSTs, a process which they call "unwrinkling". This involves iteratively moving the particles of a set such that the angles at nodes of the tree are increased (closer to 180°). They find that the overall structure of the tree is unchanged by the process, though this assumes not unreasonably many iterations of the process; the number of iterations needed appears to require guidance by visual inspection. Bhavsar & Splinter (1996) show that FoF is actually an MST technique. If one finds the MST for a set of particles and removes all edges of length greater than r (separation), the groups connected by the edges that remain are exactly the same as the groups that would be found by FoF using r as the neighbourhood radius. Bhavsar & Splinter also find that this way of finding the groups involves about an order of magnitude less computational time. Of course, this doesn't make the output groups the result of an any-less-arbitrary process; just as one had to arbitrarily set the neighbourhood radius in FoF, so too one has to arbitrarily set a length to use in the MST separation process.

Huchra & Geller's (1982) work is the most popular FoF reference, but it was not the first. Turner & Gott's (1976) algorithm is equivalent, even though they describe it differently; they are amongst the first to use an objective technique of any kind to identify clusters. Their description of the FoF technique is an interesting one. Taking individual data particles as centres, they find the density within successive enclosed radii. At $r = 0$ the density ρ is infinite, but then falls with increasing r. If an arbitrary enhanced density $f\bar{\rho}$ (where $\bar{\rho}$ is the background density) is specified, the enclosed density as a function of r will eventually fall below it, for $f > 1$, since the enclosed density is destined to approach the background for large r. The radius at which this occurs defines a circle, and if this is done for each particle in a data set and all overlapping circles are joined together, groups will be identified. This is equivalent to FoF, since the condition for two particles to be friends (to be within some arbitrary separation) is equivalent to requiring a minimum local density.

Huchra & Geller (1982) introduce the more familiar description of the technique, using distances instead of densities to define a "neighbourhood" within which two data particles are considered "friends". Implementing this in three dimensions, different distances are used in the transverse and line-of-sight (redshift) dimensions to define cylindrical neighbourhoods. The transverse and line-of-sight distances are known as linking lengths. Constant linking lengths for all redshifts could be used, but Huchra & Geller urge caution about selection effects arising from incomplete sampling of the luminosity function. In a magnitude-limited survey, density will decrease with redshift because only the brightest galaxies are seen at great distances. If linking lengths that don't change with redshift are used, clusters may not be detected at great distances because of the reduced density of galaxies. Huchra & Geller counter this by using linking lengths that increase with increasing redshift. FoF is a percolative technique, which is to say that it can be seen as a construction that may be started off somewhere within a set of particles, and applied to spread throughout the area. Huchra & Geller note that FoF is commutative in the sense that two particles will be considered friends (or not) irrespective of where in the survey the percolation is begun. This means that the same groups will be selected no matter where the process starts. Press & Davis (1982) are responsible for the technique's popular name, by saying (on page 5) "any friend of a friend is a friend".

Einasto et al. (1984) use FoF to characterise structures of different scales within the local supercluster. An easy criticism to make of FoF is that any individual set of groups that comes out of the process for a given set of data particles requires an arbitrary set of linking lengths (also known as a "neighbourhood radius" which is a corresponding physical distance). But Einasto et al. counter this by trying *all* scales at once. This doesn't indicate whether a given particle should be identified as a member of a given structure on a given scale, but it does allow examination of how the properties of groups vary as the neighbourhood radius is varied. Each choice of neighbourhood radius quantifies structure on a specific scale.

If a sufficiently large neighbourhood radius is chosen, every single particle in a distribution will be assigned to one all-encompassing group. If a sufficiently small distance is chosen, no groups will be identified. At distances between these extremes, the number of groups identified and the numbers of particles in those groups may be quantified by multiplicity functions. There is a multiplicity function for each choice of neighbourhood radius, and they can be compared with models. But the qualitative behaviour of the obtained groups was of greatest interest here. Einasto et al. (1984) found that as the neighbourhood radius was increased and the identified structures changed from groups to clusters of galaxies, there was no substantial geometric difference between the two classes: clusters are simply larger. Increasing the neighbourhood radius further, the distinction of clusters from larger structures can be qualified. Most clusters are quite distinct, with few additional cluster members identified with larger neighbourhood radii. Clusters that were not very distinct appeared to be mere density enhancements in larger structures that Einasto et al. called "strings". As the neighbourhood radius was increased further, larger, geometrically distinct structures were found which Einasto et al. called strings and sheets, with larger neighbourhood radii identifying string networks, or superclusters.

DBSCAN or "Cool By Association", introduced by Ester et al. (1996) is applied to stellar associations by Caballero & Dinis (2008), and is a slightly more complicated version of FoF in which the first step is to identify data particles that are "cool". Cool particles are those that have not just one particle within their neighbourhoods, but an arbitrary number greater than one. These cool particles are linked to form groups. Cool particles are then used to identify additional group members that are "cool by association" using the usual FoF approach. The difference this has with the standard FoF approach is to identify denser groups, and to identify group members on their outskirts that are cool by association. The cool particles are called core particles, and the particles that are cool by association are called border particles. FoF is similar except that for a particle to be cool it only needs one particle within its neighbourhood.

Yang et al. (2005) define clusters as associations of galaxies that share the same parent dark matter halo. To find these clusters, they first use FoF with small linking lengths to find cluster centres and to tentatively assign cluster properties. With these tentative properties, clusters are assigned specific halo models, that are used to refine membership of the cluster.

Tago et al. (2006) discuss selection effects and how to counter them with appropriate choices for the transverse and line-of-sight linking lengths. Berlind et al. (2006) find that no combination of linking lengths can be used to recover all galaxies that inhabit the same dark matter halo.

1.1.3.3 Delaunay Tessellation Field Estimator

The Delaunay tessellation is the dual graph of the Voronoi tessellation (VT; Voronoi 1908), applied to the galaxy distribution by Icke & van de Weygaert (1987). This technique assigns a cell to each particle in a given set; this cell also encloses all positions around its particle for which that particle is the closest one. The set of locations for which particle X, say, is the closest, forms a contiguous region or cell. The set of locations for which particle Y, say, is the closest, forms another such Voronoi cell. In this way, each particle of a distribution may be assigned a Voronoi cell, and the distribution spawns a unique set of Voronoi cells that tessellate and cover the total area covered by the distribution. This results in a tessellation free of any arbitrary assumptions, and a similarly pure way of dividing a patch of sky as a prelude to calculating local densities.

This geometrical construction is of statistical interest, but for mapping, each Voronoi cell has a number of properties that may then be associated with the location of the cell. In sparsely populated regions, the Voronoi cell areas will be larger, and in the presence of filamentary structures the Voronoi cells will be elongated, indicated by the ratio of the longest cell side length to the shortest, or by the angles at the vertices.

Kiang (1966) predates Icke & van de Weygaert (1987) although the latter is often treated as the original astrophysical reference to VT. Kiang (1966) shows that VT involves few arbitrary choices. In a mostly geometric discussion the basic problem of random fragmentation is described, the partition of any region into fragments of random dimensions. The one-dimensional version of the problem is the partition of a line into segments of randomly assorted lengths and the trivial solution is to place particles at random locations along the line, constructing segments with VT.

For the extension to the two- and three-dimensional versions of the problem, the boundaries of the fragments become lines and surfaces, respectively. Kiang notes that the solutions of the day for finding these lines and surfaces involve arbitrary choices. But since a random set of particles involves no arbitrary choice, if the fragment boundaries are uniquely constructed from the particles, the boundaries are similarly not arbitrarily defined. Kiang then introduces Voronoi and Delaunay figures as natural solutions to the problem at hand, constructions that are uniquely defined by the distribution of particles.

Icke & van de Weygaert (1987) describe a simple scenario for structure formation in which, as well as the observed expansion of the Universe, the initial overdensities are accentuated by local gravitation, such that the centres of underdensities – voids – become expansion sites, with galaxies moving away from them to join overdense regions. Icke & van de Weygaert invoke a “bubble theorem” which states that the voids should become more spherical as this evolution proceeds, so that the whole scenario may be viewed as the expansion of a set of bubbles. These bubbles are locally two-dimensional walls (or sheets) that form between voids as components of a foam-like geometry. One-dimensional filaments form where the walls meet and are components of a filamentary network, in which filaments channel mass onto the (zero-dimensional) nodes of the distribution, the clusters. Icke & van de Weygaert suggest that the Voronoi foam produced by VT about a set of particles is a way to reconstruct such features of the galaxy distribution. Because a VT follows uniquely from a set of particles, the observed VT can be used to make inferences about the distribution of the particles that are the expansion centres in the expanding bubble picture. Icke & van de Weygaert also speculate about the formation rates of the topologically distinct structures of the VT – sheets, filaments and nodes – wondering whether, respectively, they would form before or after galaxies, for example, with implications for what stage of the bubble evolution we would observe.

While Icke & van de Weygaert (1987) concentrated on using VT as a model for the galaxy distribution and the statistics therein, Ebeling & Wiedenmann (1993) apply VT to mapping explicitly, recognising that the non-arbitrary nature of VT removes effects that might otherwise be imposed upon the map. Specifically, VT removes the need for conventional "bins" or "cells", so that the sizes, positions and *geometries* (this is where the "tessellation" of VT becomes important - the shapes are all different but they fit together) of these sampling spaces, which are arbitrary, don't impose their arbitrariness on the resulting map. Geometries of the sampling bins or cells makes a difference because there may be structures that are compact in one direction while extended in another, such as filaments, as noted by Schaap & van de Weygaert (2000). The densities within the cells of VT are free of that arbitrariness, but these authors note that VT is computationally demanding because it works with the given set of particles globally rather than piecewise.

The Delaunay tessellation (DT) is a dual graph of the VT such that the DT's vertices occur at the original particle positions from which the VT is constructed. Schaap & van de Weygaert (2000) begin with the example of simple binned one-dimensional data in which the value of a sampled function (local density, for example) is taken to be the same value throughout the defined length or interval of the bin. At the edges of bins, the density changes discontinuously from the value in one bin to the value in a neighbouring bin. To smooth this, one may simply draw a line between the bin centres and values, giving the one-dimensional density a "sawtooth" appearance. Where the centres of the bins are Voronoi nuclei, and the bins are Voronoi cells, this may generalised to any number of dimensions by having the density vary along the edges of the Delaunay tessellation; see their discussion in section 2. They show that this technique is able to recover filamentary structure. This technique is called the Delaunay tessellation field estimator (DTFE), and van de Weygaert & Schaap (2009) describe it in great depth.

1.2 Multiscale Probability Mapping

MST and FoF methods provide a way to identify structures on a range of scales, but the linking length is not directly tied to the scale of structure sought. Instead, the linking length effectively sets a threshold in *density*. Moreover, a threshold in density applied to a DTFE reconstruction is not directly tied to the scale of structure sought. Wavelet approaches require the choice of an analysing wavelet, introducing shape-dependence.

Our objective is to be able to detect structures having radii within a user-defined range (e.g. for finding clusters rather than features of large-scale structure), and to do this without over-specific assumptions about the properties of the target structures. We have developed a multiscale algorithm that may be directly tuned to be most sensitive to any given range of scales. By limiting the arbitrariness of our assumptions where possible, we aim for generality similar to that of statistical correlation function approaches (e.g. Balian & Schaeffer 1989; Infante 1994). While probability and scale values may be used to describe *statistical* properties of the galaxy distribution, we use these quantities to *map* the galaxy distribution by locating overdensities.

Our algorithm, multiscale probability mapping (MSPM), is able to locate overdensities in galaxy positional data, where overdensities are defined as regions that are more dense than average, more dense than surrounding locations, or both. As a multiscale algorithm, MSPM is

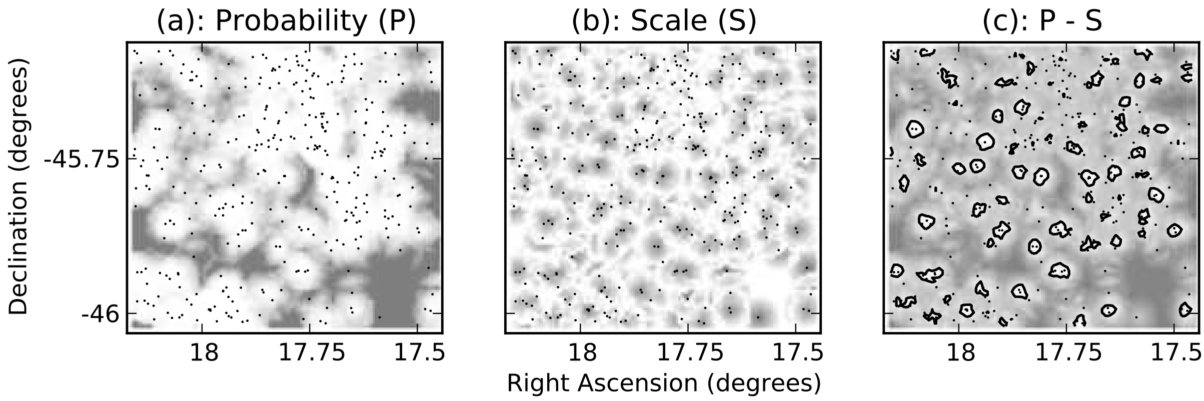


Figure 1.1: A demonstration of MSPM in two dimensions, on a sample of Extremely Red Galaxies (ERGs) studied by Smith et al. (2008). The sampled scale range is 20 arcseconds to 2 arcminutes in steps of 10 arcseconds. Each pixel corresponds to a sampling location on a uniform grid with a spacing of 20 arcseconds. Black dots are ERG positions. Sensitivity to isolated galaxies has been reduced. **(a)** A probability map showing overdensity probabilities calculated using a Poisson probability density. High probabilities are displayed as white. Probability maps are sensitive to overdensities throughout the sampled scale range. **(b)** A scale map with high scale values displayed as white. Scale maps are sensitive to density gradients within high-density islands. **(c)** A map showing $P - S$, sensitive to structures that are simultaneously more dense than average and more dense than surrounding locations, subject to the sampled scale range. Contours enclose structures that would be identified with a threshold $P - S > 0.5$. Some large structures are missed because they are beyond the sampled scale range.

sensitive to both high-density small-scale features and extended regions of intermediate density. Unlike many previous multiscale approaches, this sensitivity may be *directly* constrained to lie within a user-defined scale range. Aside from a sampled scale range and resolution, additional selection choices in our implementation are listed in Section 2.2.4.

MSPM comprises two distinct parts.

- To retain as much useful information about the galaxy distribution as possible while at the same time minimising false detections, a threshold is set in probability rather than density, such that statistically-significant regions are retained, shown in Figure 1.1(a).

- To identify structures having radii within a user-defined range, a basic shape statistic is used to identify local peaks in the density field, shown in Figure 1.1(b).

Together, these two parts provide a method by which to produce a coarse-grained map of the galaxy distribution, that encompasses a range of user-defined scale lengths, shown in Figure 1.1(c) and discussed in Section 1.2.5. This output is distinct from smoothing, because the input data are divided into separate regions. This method of deriving coarse-grained representations can be applied to any distribution of data, not only galaxy positions, and is a technique that MSPM is well-suited to in its implementation.

A precursor to MSPM has been defined and applied to a sample of Extremely Red Galaxies (ERGs) in the Phoenix Deep Survey by Smith et al. (2008). We have enhanced this algorithm by automating structure identification and assignment of scale sizes. Our complete approach is detailed here.

1.2.1 Densities and Probabilities

The input to MSPM comprises the celestial positions of a set of input galaxies (including redshifts, if available; Section 2.2), a set of user-defined spatial distances defining a scale range and resolution, and a set of sampling locations. The scale range should be chosen to extend beyond the largest structures sought if structures are to be selected on the basis of being more dense than their environments (Section 1.2.3). The sampling locations may be the galaxy positions themselves (e.g. Kepner et al. 1999), a natural choice that guarantees sufficient and economical sampling of the survey volume.

Our first step is to obtain counts of galaxies around each sampling location. Each of these counts (densities) c_i is the number of nearby galaxies within a radius equal to one of the user-defined distances r_i. Our use of redshift information in the context of these distances is described in Section 2.2. The densities obtained are then converted to probabilities because:

1. we want to detect structures that are statistically significant,

2. density alone cannot be used to set a threshold throughout a survey that exhibits a varying density (such as the typical variation with redshift resulting from a magnitude-limited survey), and

3. density *contrasts* (e.g. $\rho_{\text{inner}}/\rho_{\text{outer}}$) may be undefined in low-density regions (where $\rho_{\text{outer}} = 0$).

To compute probabilities, counts are compared with probability densities. Each probability density (e.g. Poisson or Gaussian) is the probability of obtaining each possible count of galaxies. For meaningful probabilities that take into account the statistics of the galaxy distribution, a natural choice (referred to as "empirical") is derived from a statistically-comparable ensemble of counts (for example, those obtained at a similar redshift). Then the probability of obtaining a count c at some individual location, within a specific radius r_i, is:

$$P_i(c) = \frac{\text{sampling locations with a count } c \text{ within radius } r_i}{\text{number of sampling locations in the ensemble}}. \tag{1.8}$$

With a cumulative probability density P_i' for each radius r_i containing a count c_i, we have:

$$P_i' = \sum_{c=0}^{c_i-1} [P_i(c)] + \frac{1}{2} P_i(c_i). \tag{1.9}$$

This is approximately the fraction of the probability density with $c < c_i$, and quantifies the probability of obtaining c_i or less within a radius r_i.

P_i' includes $\frac{1}{2}P_i(c_i)$ so that in the case where the i-th count is n and all the counts within its ensemble are also n, the probability is 0.5, since $P_i(c_i) = P_i(n) = 1$ (equation 1.8). In this way, a probability is associated with each sampling location, creating the i-th single-scale probability "map". This process is repeated for each of the input scales r_i, resulting in a probability map for each scale.

1.2.2 Probability and Scale Maps

To obtain our final probability map, we assign to each sampling location the maximum value of probability from all of the single-scale measurements at that location: This gives an *overdensity probability*, P, at each sampling location:

$$P = \max(P'_i). \tag{1.10}$$

If an overdensity is present on any of the sampled scales, it will be evident in this probability map.

The other half of MSPM is the shape statistic, allowing for scale selection. The scale, S, at each sampling location is defined to be the radius hosting the highest probability (P') at that location:

$$S = r(P')|_{P'=P}. \tag{1.11}$$

On this "scale map", low scale values are usually associated with local density peaks. For sampling locations close to a peak in the local density field, higher densities and hence higher probabilities P'_i will be obtained at small radii, corresponding to the close proximity of the density peak. Thus, the highest probability will be found at small r, resulting in a low scale value S.

Equivalently, if the single-scale probability values (P'_i) at a given sampling location are a probability function (of radius), that function's maximum is the overdensity probability. The scale value at that location is the radius at which the maximum occurs. To prevent probabilities rising where the count of objects does not, probability functions are defined to be zero in the absence of additional counts enclosed with increasing radius.

A probability map and a scale map are shown in Figure 1.1, demonstrating their different but complementary functions. A probability map by itself may join all structures together since most locations are overdense on at least one scale within a large scale range, or identify structures with boundaries that do not correspond to actual density variations. The scale map compensates for this behaviour by comparing local densities with surrounding locations, guaranteeing contrasts within the sampled scale range. More complex distributions than that shown in Figure 1.1 cause the scale map by itself to identify structures that may not be denser than average. Our work with SDSS does not readily produce images because an adaptive grid is used to reduce computational effort (Section 2.2). Hence, the demonstration in Figure 1.1 uses data from the Phoenix Deep Survey (Hopkins et al. 2003; Smith et al. 2008) to create two-dimensional images.

1.2.3 Thresholding with P and S

The next step is to interpret the probability and scale information assigned to each sampling location in the form of a (P, S) pair. Different combinations of the two parameters can be used to locate various features of the galaxy distribution.

The probability map highlights regions that are overdense when compared to the average density, as measured within the sampled scale range. Assuming an empirical probability density

(Section 1.2.1), regions above a threshold in probability P contain a densest fraction of the galaxy distribution by number if the sampling locations are the input galaxy positions or by volume if the sampling locations are distributed uniformly throughout the volume. For example, $P > 0.9$ contains at least the densest 10 per cent in comparison with the mean density. The fraction selected by a threshold increases as the scale range is increased. The amount that such a fraction increases also rises as the correlation between large- and small- scale structure decreases. For instance, $P > 0.9$ selects precisely the densest 10 per cent if only one scale is sampled, and more than 10 per cent as more scales are included in the probability estimator.

The scale map highlights regions that are more dense than surrounding locations over the sampled scale range. A threshold in scale S will remove sensitivity to an upper portion of the sampled scale range that depends on the threshold. For example, $S < 0.5$ removes all sampling locations where there is a peak in the probability function P_i' in the upper half of the sampled scale range. This situation tends to occur when the highest densities are found at large radii, meaning that the sampling location in question is in a relatively underdense region for radii up to the peak in the probability function. Assuming an empirical probability density, regions below a threshold in S contain a densest given fraction of the galaxy distribution in a manner similar to a threshold in probability.

P and S do not always correlate. A structure may, for example, be more dense than surrounding locations but underdense relative to the mean density. To guarantee the detection of structures that are overdense compared with both the mean density and surrounding locations, both maps are required.

P and S are interpreted jointly by subtracting S from P (where S is normalised to lie between zero and one). Subtraction is used because high densities are associated with low values on the scale map (S). Subtraction is preferable to division because it gives the two attributes of being denser than average and more dense than surrounding locations roughly equal weight: identical intervals in P and S usually contain equivalent fractions of the input galaxy distribution. A $P - S$ map is shown in Figure 1.1(c).

While various thresholds in P, S or $P - S$ may be motivated by reasoning based on known properties of the target subset of the galaxy distribution, natural choices include:

- $P > 0.5$ – denser than average,
- $S < 0.5$ – denser than surrounding locations, and
- $P - S > 0.5$ – denser than average *and* denser than surrounding locations, guaranteeing $P > 0.5$ and $S < 0.5$.

1.2.4 Structure Identification

Extended structures are identified as unions of sampling locations (galaxies) above a chosen threshold using a friends-of-friends approach to linking, and are not allowed to contain galaxies below the threshold (e.g. Section 2.2.2). The linking length used is the maximum of the sampled scale range. If the galaxy locations are used as an adaptive grid, the linked locations are identified as member galaxies. The centre of a structure may be associated with the peak of the region above the threshold, and the distance from this peak to the furthest member galaxy is a measure of radius (Section 2.3).

1.2.5 Creating Coarse-Grained Distributions

MSPM is well-suited to constructing a coarse-grained representation of the galaxy distribution in ways that previous algorithms are not. This is because S is a basic shape statistic, and P allows us to set a low threshold that considers the statistical significance of structures.

Choosing a limit of $S < 0.5$ (and similarly, $P - S > 0.5$) selects regions that are local peaks in the density field, limiting the "grain" size of the coarse-grained distribution to the sampled scale range. This feature is demonstrated in Figure 1.1(b), in which the grains (dark patches) are not allowed to merge on large scales. A threshold in density (as approximated by $P > 0.5$; Figure 1.1(a)) has the potential to join high-density islands together such that the grain-size is not uniform, and not directly controlled by the user. Since the real galaxy distribution contains structures with a variety of densities, an effective coarse-grained map should be sensitive to shape (as realised by $S < 0.5$), such that densities relative to surrounding locations (as well as to the mean density) are considered.

Such an approach should attempt to retain as much useful information about the galaxy distribution as possible, while reducing the level of noise. We would therefore like to set a low threshold while at the same time minimising false detections. In MSPM, this is attempted by thresholding with P rather than density, such that statistically-significant regions are retained. Figure 1.1(a) shows that this has an effect reminiscent of smoothing. However, the regions identified by $P - S$, shown in Figure 1.1(c), have a characteristic grain size, a result that is not guaranteed by smoothing.

The process of coarse-graining we describe is "reductive", reducing the galaxy distribution to a set of particles. This approach is fundamentally different to reconstruction of the density field at all positions within a volume such as that produced by the DTFE (Section 1.1.3.3).

Chapter 2 focuses on the use of MSPM in identifying galaxy groups and clusters, although the coarse-grained representation of the galaxy distribution produced by MSPM has additional functionality. In Chapters 3 and 4 we outline previous approaches to, and our use of MSPM in, producing such a representation for the purposes of identifying filaments of galaxies and voids, respectively.

Apart

A famous legend gave the Flatland its name. The Goddess Weaver, daughter of the Celestial Queen Mother, used to descend from the Celestial Court to bathe in a lake there. (The meteor which fell on Meteor Street is supposed to have been a stone that propped up her loom.) A boy living by the lake who looks after buffaloes sees the goddess, and they fall in love. They marry, and have a son and a daughter. The Celestial Queen Mother is jealous of their happiness, and sends some gods down to kidnap the goddess. They carry her off, and the buffalo boy rushes after them. Just as he is about to catch them, the Celestial Queen Mother pulls a hairpin from her coil and draws a huge river between them. The Silver River separates the couple permanently, except on the seventh day of the seventh moon, when magpies fly from all over China to form a bridge for the family to meet.

The Silver River is the Chinese name for the Milky Way. Over Xichang it looks vast, with a mass of stars, the bright Vega, the Goddess Weaver, on one side, and Altair, the buffalo boy, with his two children, on the other.

While she was away the truck came. I looked toward the camp and saw her running toward me, the white-golden grass surging around her blue scarf. In her right hand she carried a big colourful enamel bowl. She was running with the kind of carefulness that told me she did not want the soup with the dumplings to spill. She was still a good way off, and I could see she would not reach me for another twenty minutes or so. I did not feel I could ask the driver to wait that long, as he was already doing me a big favor. I clambered onto the back of the truck. I could see my mother still running toward me in the distance. But she no longer seemed to be carrying the bowl.

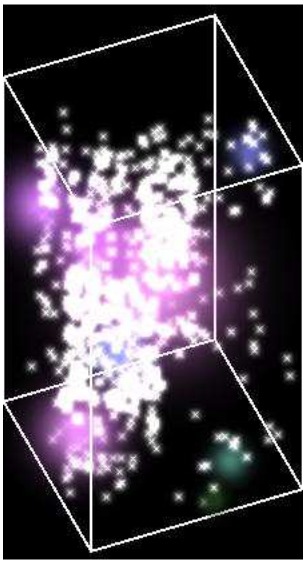
Years later, she told me the bowl had fallen from her hand when she saw me climbing onto the truck. But she still ran to the spot where we had been sitting, just to make sure I had really gone, although it could not have been anyone else getting onto the truck. There was not a single person around in that vast yellowness. For the next few days she walked around the camp as though in a trance, feeling blank and lost.

I was in the camp with my father when a letter came from Mother with the news. Immediately my father went to ask for permssion to go home and see her. Young was very sympathetic, but the camp authorities refused. My father burst out crying in front of a whole courtyard of inmates. The people from his department were taken aback. They knew him as a 'man of iron'. Early the next morning, he went to the post office and waited outside for hours until it opened. He sent a three-page telegram to my mother. It began: 'Please accept my apologies that come a lifetime too late. It is for my guilt toward you that I am happy for any punishment. I have not been a decent husband. Please get well and give me another chance.'

– Jung Chang, *Wild Swans: Three Daughters of China*. 24. 'Please Accept My Apologies That Come a Lifetime Too Late'

Chapter 2

Groups and Clusters



A solid particle render (Appendix B) of the first structure in the MSPM group and cluster catalogue, Abell 2151 (Abell 1958; Corwin 1974) in the *Hercules* supercluster (Tarenghi et al. 1979).

Sections 2.2-2.4 of this Chapter are taken from Smith et al. (2012) with minor adaptations.

In this Chapter we compile a catalogue of groups and clusters at $0.025 < z < 0.24$ (Section 2.3) based on the Sloan Digital Sky Survey, Data Release 7, quantifying their significance and comparing with other catalogues (Section 2.4.4). Most measured velocity dispersions lie between 50 and 400 km s^{-1}. A clear trend of increasing velocity dispersion with radius from 0.2 to 1 h^{-1} Mpc is detected (Section 2.4.3), confirming the lack of a sharp division between groups and clusters.

2.1 Review: Towers of Babel

2.1.1 History

2.1.1.1 "Antiquity"

The early history of cluster science is reviewed by Biviano (2000). During this time, threads of investigation were started which remain sources of intrigue to this day. Zwicky (1937) calculated the mass of the Coma cluster using the virial theorem, deriving a mass much greater than that implied by the visible light, recognising the missing (or, dark) mass problem. Zwicky also suggested, well before such studies became possible, resolving the mass determination issue by treating clusters as gravitational lenses. One early use of clusters for observational cosmology (e.g., Gunn & Oke 1975; Sandage, Kristian & Westphal 1976) proceeded via Hubble diagrams based on cluster galaxies and estimation of the deceleration parameter q_0. Gunn & Oke (1975) presented an early use of clusters for observational cosmology, via a Hubble diagram (redshift versus magnitude) based on brightest cluster galaxies.

Important early catalogues were the Abell (1958), Zwicky (Zwicky, Herzog & Wild 1961) and ACO (Abell, Corwin & Olowin 1989) catalogues. Their numerical size allowed authoritative exploration of the large-scale structure of the Universe, through analysis of clusters' spatial distribution (e.g. Bahcall & Soneira 1983; Postman, Huchra & Geller 1992; Miller et al. 1999). Even so, the Abell catalogue, for example, is only moderately complete to $z \approx 0.19$ and has a high false discovery rate (Lucey 1983), with a lack of redshift data resulting in projection effects arising from the inclusion of non-cluster foreground and background galaxies (Sutherland 1988). These shortcomings demanded the improvements which have driven developments in cluster science since.

2.1.1.2 Modern Optical

Automation and objectivity became standard in subsequent cluster catalogues (e.g. Lumsden et al. 1992; Dalton et al. 1994; Collins et al. 1995), with redshifts obtained to minimise projection effects. Progressively more distant clusters have been found, with, for example, the Palomar Distant Clusters Survey (Postman et al. 1996) containing 79 clusters at 0.2 $\lesssim$ z $\lesssim$ 1.2, and about 20 at at z $\gtrsim$ 0.9 in the Hubble Space Telescope Medium Deep Survey (Ostrander et al. 1998). A novel cluster detection technique involving the use of surface brightness fluctuations in images to detect clusters without the need for resolution of individual galaxies (Dalcanton

1996, Zaritsky et al. 1997) was employed in the Las Campanas Distant Clusters Survey to select 1073 cluster and group candidates at $0.3 \lesssim z \lesssim 1$ (Gonzalez et al. 2001).

At lower redshifts, very large cluster samples (e.g. Gal et al. 2003) have been constructed using a variety of different algorithms which exploit the wealth of photometric data in different ways. Multicolour digital photometry is now available to the extent that clusters are no longer two-dimensional distributions, but entities which can be hunted down with great precision in colour-magnitude space. As the amount of available data grows, cluster catalogues are increasingly defined by algorithms, rather than survey properties. Examples of these are the maxBCG catalogues (Bahcall et al. 2003; Koester et al. 2007b), the Red Sequence Cluster Survey (Gladders & Yee 2005) and the C4 catalogue (Miller et al. 2005).

2.1.1.3 X Ray

Another way to circumvent projection effects is to identify clusters via X-ray emission from intracluster gas, avoiding the need to distinguish an overdensity from a confusing background and foreground distribution. As well as large samples at low redshift (e.g. Ebeling et al. 1996; Böhringer et al. 2000; Böhringer et al. 2004), this method has also demonstrated the ability to select clusters at moderately high redshifts (e.g. Gioia & Luppino 1994; Rosati et al. 1998; Romer et al. 2000; Mullis et al. 2003); Sadat et al. (2004) presented a database of X-ray clusters and groups from the literature. Romer et al. (2001) expect that the XMM-*Newton* satellite will detect ≈ 750 clusters at $z > 1$ throughout its lifetime. As discussed and tested by Vikhlinin et al. (2006) and Kravtsov, Vikhlinin & Nagai (2006), mass determination from X-ray measurements involves modelling of the radial gas density and temperature: the mass is a function of each.

2.1.1.4 Weak Lensing

Luminous matter is a minor component of cluster mass, and so a mass determination method which doesn't rely on direct observation of the luminous matter directly is useful. Clusters may be detected and their masses quantified through their gravitational effects on background galaxies. Weak lensing promises "mass-selected" cluster samples that would have great utility in observational cosmology. The first such detection of a cluster without a clear optical counterpart was reported by Erben et al. (2000), which was soon followed by a $z \approx 1$ (Clowe, Trentham & Tonry 2001) candidate, the first spectroscopically confirmed candidate (Wittman et al. 2001) and the first sample selected from weak lensing shear (Wittman et al. 2006).

Unfortunately, White, van Waerbeke & Mackey (2002) find that completeness in weak lensing cluster searches can only be obtained after conceding a high false discovery rate, caused by structures along the line-of-sight, and may require complementary observations with other techniques. De Putter & White (2005) find that these uncertainties are difficult to correct for, but Johnston et al. (2007) present a possible solution to this problem using stacked samples of clusters. In any case, weak lensing is important for its sensitivity to cluster mass, and Diaferio, Geller & Rines (2005) find that it agrees well with masses derived from redshift-space caustics. Sheldon et al. (2001) used weak lensing to measure the masses of SDSS/RASS (ROSAT All-Sky Survey; X-ray detected) clusters, and, continuing this work with maxBCG clusters, Sheldon et al. (2007) find evidence that luminosity is a better indicator of mass than galaxy count. Clusters detected by weak lensing are important in calibrating the mass measurements from other techniques, even though weak lensing is observationally expensive since it requires sensitive and high-resolution observations of large fields.

2.1.1.5 Sunyaev-Zel'dovich effect

The Sunyaev-Zel'dovich (SZ) effect is a distortion of the CMB by clusters, depending on electron temperature and mass such that all clusters above a certain mass will be detected regardless of redshift. Future SZ surveys promise large catalogues of high redshift clusters, and their potential is discussed by, for example, Carlstrom, Holder & Reese (2002). The *Planck* satellite has produced an all-sky catalogue of 189 clusters, including 30 new candidates (Planck collaboration 2011).

2.1.2 Groups, Clusters and Galaxy Evolution

2.1.2.1 Morphology

Galaxy groups and clusters are important in studies of galaxy evolution and cosmology, with large samples necessary to draw robust conclusions about the role played by environment. The concentration of mass in a galaxy's environment or host halo (Haas, Schaye & Jeeson-Daniel 2012) may trigger local physical processes that influence evolution. Such processes include gas stripping (Gunn & Gott 1972), shocks in the intracluster medium (Moran et al. 2005), harassment (Moore et al. 1996) and galaxy-galaxy interaction (Ostriker & Tremaine 1975). The extent of these influences, contrasted against differences in the evolution of galaxies with specific masses, has the potential to discriminate between models of galaxy evolution.

Dressler (1980) found from a study of rich clusters that ellipticals and lenticulars were more prevalent in denser environments, with a corresponding deficit of spiral galaxies. Postman & Geller (1984) found this same morphology-density relation (MDR) in two complete redshift surveys, speculating that stripping mechanisms might be responsible for the stellar populations of lenticulars, and that the collapse time in high density environments might be too short for the formation of disks. The former of these speculations invokes "nurture", as an explanation, while the latter invokes "nature"; this duality concerns the precise role that environment plays in determining the properties of resident galaxy populations.

Our knowledge of the MDR has been gradually extended to higher redshifts. Dressler et al. (1997) investigated the MDR at $z \sim 0.5$, finding that the MDR in less concentrated clusters strengthens at low redshift, admitting the interpretation that the MDR strengthens as time passes and becomes apparent in higher density clusters sooner. It was also found that the fraction of lenticulars has increased significantly since then, with a simultaneous decrease in the fraction of spirals. At $z = 1$, Smith et al. (2005) examined the elliptical/lenticular fraction, finding that this fraction has been steadily increasing in the highest density regions from $z = 1$ to $z = 0.5$, while constant in intermediate and lower density regions over that time. Postman et al. (2005) found that this increase in the early-type fraction is slower at $z \sim 1$ than in the nearby Universe.

Working with a single cluster at $z \sim 0.4$, Treu et al. (2003) found that mild trends in morphology at large radii suggest a gradual morphological change, with gas starvation as a possible reason. Treu et al. identified substructure in the infall regions of the cluster, and found that at greater radii, morphology is more strongly dependent on the local density of this substructure than on cluster radius, implying that this substructure is gradually erased by interaction with the cluster potential. This example illustrates the simultaneous influence of nature and nurture, and the value of mapping in identifying substructure.

2.1.2.2 Star Formation

The effect of environment on interstellar matter was considered very early (Spitzer & Baade 1951). In the local Universe, there is a clear star formation-local density (SFD) relation (e.g. Hashimoto et al. 1998), according to which, star formation rates (SFRs) are lower in denser environments; this appears to be a continuous function of density, rather than a discrete step divided into "cluster" or "void" environments. A rough, though neither precise nor decisive, indicator of star-formation activity is colour; actively star-forming galaxies have young stars which give their host galaxies bluer colours, while galaxies with old stellar populations have red colours. Higher-redshift rich clusters contain bluer galaxies in their cores than their lower-redshift counterparts (Butcher & Oemler 1978).

Lewis et al. (2002) found that the rising SFR with radius extends beyond virial cluster radii, to the extent that local physical processes such as gas stripping are ruled out as the sole cause. Studying the same cluster as Treu et al. (2003), Moran et al. (2005) found that near the virial radius, infalling galaxies appear to experience small bursts of star formation. This pair illustrate again a nature/nurture duality. The role of nurture is constrained in the local Universe by Hogg et al. (2004), who found that the colour-magnitude relation slope varies little with density.

Influences on star formation may be time-dependent. Environment may have been a more active nurturer at high redshift, as suggested for example by Balogh et al. (2004a). Balogh et al. argue that SFRs in *individual* star-forming galaxies are not affected by environment, even though the *fraction* of star-forming galaxies varies. This is evidence that there is not an ongoing gradual decrease in the SF rates of individual galaxies, evidence that argues against the "nature" mode of evolution. Again, by seeing the SFD relation vary beyond cluster cores, ram pressure stripping is ruled out as the only reason, suggesting that any nurturing influence is not current. By finding that SFRs in low density environments aren't particularly high, the increased prevalence of low density environments at higher redshifts (where structure formation is less advanced) is ruled out as the reason for the higher global SFR in the past, providing further evidence against current nurture. By examining the dependence of SFRs on local ($\sim$ 1.1 Mpc) and large-scale ($\sim$ 5.5 Mpc) environments, and finding that both of them are influential, it is concluded that the dependence on local environment (equivalently, short term processes) is indirect, more evidence against nurture at low redshift. Finally, these authors draw the above results together to conclude that the SFD relation can be attributed mostly to a short term high-redshift process, for which starburst activity triggered by galaxy-galaxy interaction is a good candidate.

The exploration of environmental dependence on different scales is also considered by

Kauffmann et al. (2004), who find that small-scale environment is most important in the SFD relation, strikingly similar to the intermediate density MDR results of Treu et al. (2003). Like Balogh et al. (2004a), Kauffmann et al. consider time scales. By examining SF indicators that probe the SF history on different time scales, they find that the main cause of the lower SFR in dense environments is a long term process, contradicting Balogh et al. As if to emphasise the disagreement, they also find that the structures of high mass galaxies vary little with environment, concluding that they have not been the result of major galaxy-galaxy interaction.

The seemingly divergent pictures put forward by Balogh et al. (2004a) and Kauffmann et al. (2004) are accommodated by Balogh et al. (2004b), who judge from the good fit of two gaussians to the galaxy colour distribution that the transition from blue to red is either current and fast, or occurred in the distant past over a long time. Baldry et al. (2006) find that this colour bimodality varies little with local density, concluding that it is more likely a result of galaxy stellar mass than environment, but the relative fractions of galaxies on these red and blue sequences is dependent on environment. Peng et al. (2010) show that the effects of intrinsic mass and environment can be separated at $z \lesssim 1$ and suggest that they can be attributed to distinct "mass quenching" and "environment quenching" processes that suppress star formation to produce the observed colour bimodality. In this scenario, mass-quenching is dominant and proceeds at a rate proportional to galaxies' individual SF rates, while environment quenching probabably results from major galaxy mergers.

As with the MDR, the variation of the SFD relation with redshift is of relevance to the extent to which the MD and SFD relations trend similarly, and are produced by different physical processes. Clearly, there is vast scope for qualitative difference between the MD and SFD relations over cosmic time. Baryonic matter has a limited potential for star formation, integrated over time: it is a measure of *change* as distinct from the measure of *state* that morphology is. For example, the highest density regions cannot always have been the sites of lowest SF activity as they must have formed all their stars at some point, implying a necessary variation in the balance between the MD and SFD relations over time.

Wilman et al. (2005) compare the star formation properties of galaxies in groups at $z \sim 0.4$ with those of field galaxies at the same redshift, finding that the relation is qualitatively similar to that found at low redshift, and consistent with the near-infrared observations of groups at $0.1 < z < 0.6$ by Balogh et al. (2007). This suggests that groups at $z \sim 0.4$ have mostly finished their star formation. Evidence for the phenomenon of galaxies in high density environments exhausting their star-forming capacity is also found by Poggianti et al. (2006), who compare star formation as a function of environment between high redshift ($0.4 < z < 0.8$) and low redshift (SDSS) samples. Density seems to mostly determine the fraction of star-forming galaxies in the high redshift sample, a fraction that is higher than that found at low redshift. This is similar to the trend at low redshift, except that the relation exhibits a plateau for environments above a certain mass, implying that the densest environments exhaust their star-formation capacity in their densest components first ("from the inside out").

This exhaustion effect might then be expected directly in the evolution of the red sequence, and accordingly, Gilbank et al. (2008) find that red sequence galaxies at $0.35 < z < 0.95$ are deficient at early times, where the only red sequence galaxies are the most massive ones, reinforcing the view that SF is exhausted first in the most massive systems. To accompany the work of Baldry et al. (2006) in addressing the colour bimodality, Cassata et al. (2007) study colour-magnitude diagrams of the red and blue sequences at $z \sim 0.7$ (formed mostly by early- and late-type galaxies respectively), which show slopes and normalisations independent of environment, suggesting that apart from whatever mechanism separated the blue and red populations, their evolution has been a stronger function of stellar mass than environment.

Poggianti et al. (2008) study SF and morphology as functions of environment at $0.4 < z < 0.8$, comparing with low redshift data. The fraction of star-forming galaxies is found to decrease with increasing density as expected, and the SFRs in individual galaxies are found to be independent of local environment, consistent with earlier results. Interestingly, Poggianti et al. tentatively found that the mass-averaged SF rate maximises in intermediate density environments, pointing to the qualitative change in the SFD relation found by other authors at higher redshifts ($z \gtrsim 1$), where evidence is mounting that that the SFD relation may be inverted. Cucciati et al. (2006) find that the relation between colour and density weakens dramatically towards $z \sim 1.5$. Gerke et al. (2007) find that the blue fraction for both group and field populations is constant at $0.75 < z < 1$ but there is a rise in the group blue fraction such that the blue fraction is the same in groups and in the field at $z = 1.3$. Cooper et al. (2007) find that the red fraction is only weakly correlated with overdensity at $z = 1.3$. Elbaz et al. (2007) find that at $z \sim 1$ the SFD relation is inverted such that at higher redshifts, galaxies in denser environments tend to have higher SFRs.

It is also suggested, however, that this inversion is nuanced. Cooper et al. (2008) investigate the environmental dependence of SF at $z \sim 1$, making a distinction between the total SFR per galaxy, SFR, and specific SFR per unit mass, sSFR. These authors find that SFRs are higher in high density environments at high redshifts, inverting the SFD relation from low redshift, but sSFRs as a function of environment at high redshift trends similarly to what is found at low redshift. This is an area of ongoing investigation.

2.1.2.3 Relevance to Evolutionary Scenarios

A number of papers have sought to compare the MD and SFD relations. Pimbblet et al. (2002) found a colour variation greater than expected from the MDR alone, proposing a difference in star formation activity as the reason. Gómez et al. (2003) found that a break in the SFD relation at a particular local density could not be explained by the MDR. Finally, Capak et al. (2007) find that at $0 < z < 1.2$, the conversion of spiral/irregular to elliptical/lenticular morphologies proceeds fastest in highest density regions, except in the field, where this does not seem to hold; at the same time, star formation activity seems to vary continuously, suggesting that the physical bases for the MDR and SFD relations are not entirely the same.

At low redshift, it is well established that early type (elliptical/lenticular) galaxies dominate high density environments with an increasing fraction of late type (spiral/irregular) morphologies with decreasing density; this is the low redshift MDR. At the same time, star formation is now well known to anticorrelate with density, and this SFD relation, together with the MDR, dictates star formation activity as a function of morphology. There has been much speculation about the role of environment: there is evidence that a dense environment may be responsible for both effects, engendering suppression of star formation via effects such as gas stripping, and mergers that produce early type galaxies. This "nurture" scenario, is an extreme in which environment is responsible for inducing the changes that in turn lead to morphological variation. The alternative to this is "nature". Perhaps high density environments play no role, merely occurring where the more massive galaxies are, such that it is more the intrinsic masses of the galaxies that are responsible for the MD and SFD relations, an extreme in which the (anti) correlation is incidental rather than causal, with particular morphologies and particular environments having common causes.

The broader issue is the extent to which galaxies evolve independently; these nature and nurture paradigms have similarities with, though may not be precisely identified with, the broad evolutionary models PLE (passive luminosity evolution; nature) and hierarchical merging (nurture). The truth probably lies somewhere in between, and exactly where may be probed by measures of star formation activity in different environments as a function of redshift, as well as studies of stellar populations, which may reveal how they were formed (passive evolution or galaxy-galaxy interaction). Finding morphologies in different environments at varying redshifts is important in this goal, as, for example, the number densities of ellipticals with redshift are sensitive to models of hierarchical merging.

And hence the importance of mapping: the ability to characterise structures at various redshifts to identify the resident morphologies and local star formation properties is crucially important. The continuous variation in star formation activity with density emphasises the value of a mapping technique which does more than classify environments as "cluster" or "non cluster" (where the definition of a cluster is somewhat arbitrary in the first place), black and white, and the considerations of various scales emphasises the value of a mapping technique that is sensitive to structure on multiple scales.

2.1.3 Groups, Clusters and Cosmology

2.1.3.1 Cluster Counts from Simulations

The value of cluster counts as a direct test of cosmology can be questioned, with the possibility of simply using the raw galaxy counts. But clusters can be detected without finding all their constituent galaxies and, with appropriate models, their masses can be inferred (see many of the techniques we describe). Evrard et al. (2002) give two N-body simulations, one each for ΛCDM ($\Omega_{\mathrm{M}} = 0.3, \Omega_\Lambda = 0.7$) and τCDM ($\Omega_{\mathrm{M}} = 1$). Observations are simulated to produce cluster catalogues to $z \approx 1.5$ and deeper, for comparison with both current and forthcoming observations. Concentrating on the ΛCDM results, uncertainty in σ_8 is found to be a huge source of uncertainty in the predicted cluster counts (a given σ_8 value corresponds to a value of the matter density Ω_{M} from current measurements). Given infinite sensitivity, the total number of observable Coma-sized clusters on the sky is finite because massive structures are rarer at higher redshifts, but varies between 40 and 2000 within the 90% confidence interval of σ_8.

2.1.3.2 Redshift Distribution and Power Spectrum

Majumdar & Mohr (2004) argue that both the redshift distribution of clusters and their power spectrum may be used to obtain cosmological constraints, and that both should be applied simultaneously. Instructions are given on the cluster power spectrum's use in constraining dark energy's (Section 5.1.1) equation of state. Clusters can produce constraints that are competitive with those derived from galaxy statistics, with smaller samples producing similar uncertainties because clusters are more biased mass tracers than galaxies. Moreover, given an initial galaxy catalogue, a cluster sample can be obtained at no additional observational cost.

2.1.3.3 Cluster Internal Structure

The internal structure of clusters presents an important test of numerical simulations (e.g. Lewis, Buote & Stocke 2003). The formation of groups and clusters is sensitive to cosmology (e.g. Gunn & Gott 1972), influenced by the matter density, the Hubble parameter and dark energy (Eke, Cole & Frenk 1996). This leads to scaling relations between observable properties, one of which is observationally quantified in Section 2.4.3. In Chapter 5 we exploit this size-mass relation to carry out a new test of cosmology.

2.1.3.4 Large-Scale Structure Studies

As mass tracers, clusters can highlight features of large-scale structure. Groups and clusters are not isolated, but are connected and arranged in a non-trivial manner by cosmic superstructures, including filaments, walls and superclusters (e.g. de Lapparent, Geller & Huchra 1986). Filaments may be traced by groups and clusters (e.g. Connolly et al. 1996), and often occupy the spaces between massive clusters (Pimbblet, Drinkwater & Hawkrigg 2004; Colberg, Krughoff & Connolly 2005a). Similarly, superclusters have been identified as unions of smaller structures (e.g. Einasto et al. 2001), demonstrating the value of group and cluster catalogues for the identification of large-scale structure. We exploit this utility in Chapters 3 and 4.

2.2 MSPM in SDSS DR7

2.2.1 Survey Volume and Search Apertures

We apply our MSPM approach initially to SDSS data (Data Release 7; Abazajian et al. 2009) using only the SDSS main spectroscopic sample, omitting the Luminous Red Galaxy sample (Eisenstein et al. 2001). To guarantee sufficient sampling while reducing computational effort we use the input galaxy locations as an adaptive grid (e.g. Kepner et al. 1999). Galaxies less than 2 h^{-1} Mpc from the survey edges are excluded as potential structure centres to reduce edge effects and enable comparison of structure densities with neighbouring volumes (Section 2.3.5). Because our approach relies on a comparison with counts obtained at a similar redshift, our comparison volumes at all redshifts must be large enough to allow sufficient statistical strength. We guarantee this by restricting our attention to $z > 0.025$.

We define our galaxy search volumes as cylinders with variable radii on the sky and a fixed line-of-sight interval in redshift, defined as twice an empirically-obtained *redshift radius*. To locate groups and clusters rather than features of large-scale structure (e.g. Einasto et al. 1984), our transverse radii are set at 0.2 to 2 h^{-1} Mpc in steps of 0.2 h^{-1} Mpc. This defines the scale range we sample. Although we are primarily interested in structures with radii $\lesssim 1\ h^{-1}$ Mpc, we sample the larger scales so we can select structures that are more dense than their environments. *Components* of larger structures may be detected by this approach.

Intracluster peculiar velocities stretch structures along the line of sight, preventing us from interpreting redshifts as positions in depth on $\sim 1\ h^{-1}$ Mpc scales. Results obtained from a preliminary analysis have been used as a guide to how deep our cylindrical volumes should be in the line of sight. We find that a redshift radius of 10 h^{-1} Mpc ($\Delta z \sim 0.004$) is suitable throughout our range of redshift, capturing the spread of peculiar velocities present within most structures. Thus our cylindrical search volumes have fixed line-of-sight depths in redshift of 20 h^{-1} Mpc. A smaller redshift radius would probably recover most of the same structures, but their velocity dispersions might be underestimated as a result of the removal of galaxies with large peculiar velocities.

Our larger redshift radius may allow some structures to be identified as unions of physically unassociated galaxies across large distances in the line of sight, and we use sigma-clipping (described in Section 2.3.7) to mitigate this effect.

2.2.2 Redshift Slices and Inter-Galaxy Distances

Our probabilities are computed using an empirical probability density (Section 1.2.1). The probability associated with a particular count is found by comparison with all other counts within a redshift slice of width $\Delta z = 0.005$ ($\sim 14\ h^{-1}$ Mpc), centred on that redshift. This width is chosen to provide a large background comparison volume rather than search for nearby galaxies in the line of sight, and is not related to our redshift radius. For the area of sky available in Data Release 7, $\Delta z = 0.005$ allows sufficient statistical strength, while retaining sensitivity to decreasing mean density caused by incompleteness at higher redshifts. However, some down-weighting will be suffered by structures lying in slices that host structures such as the Sloan Great Wall (e.g. Einasto et al. 2010). Figure 2.2(b) shows the redshift distribution for galaxies, with counts from slices of width $\Delta z = 0.005$. The variation exhibited by this distribution is not great enough to significantly affect our results.

Since we use the galaxies as an adaptive grid, a (P, S) pair is associated with each galaxy, where S is normalised to lie between 0 and 1. The physical lengths 0.2 h^{-1} Mpc and 2 h^{-1} Mpc are transformed to 0 and 1 respectively. To locate structures that are overdense when compared with both the mean density and surrounding locations, we set a threshold of $P - S > 0.5$. This identifies 177675 of the 619234 galaxies (29 per cent) in the original SDSS sample of galaxies brighter than $r = 17.77$ as lying in overdense environments.

For the purpose of structure identification throughout the survey volume, inter-galaxy distances are defined as

$$d = \sqrt{d_{\rm t}^2 + \left(\frac{d_{\rm los}}{e_{\rm los}}\right)^2}, \tag{2.1}$$

where $d_{\rm t}$ and $d_{\rm los}$ are the transverse (sky) and line-of-sight comoving separations respectively, and a line-of-sight elongation factor $e_{\rm los} = 10$ allows a 1 h^{-1} Mpc cluster to contain galaxies apparently up to 10 h^{-1} Mpc away in the line of sight, consistent with our 10 h^{-1} Mpc redshift radius. Figure 2.1 shows the first structure identified in our catalogue, demonstrating our structure identification on real data. Galaxies above the threshold are linked together if the distance between them is less than $d = 2$ h^{-1} Mpc, the maximum of our sampled scale range, and $P - S > 0.5$ for all galaxies (if any) between them, as shown in Figure 2.1(c).

2.2.3 Thresholds

We require that each structure, as defined in Section 2.2.2, has at least four member galaxies with magnitude $r < 17.77$, where each galaxy must have $P - S > 0.5$, so that properties can be measured for each structure. Collins et al. (1995) find that velocity dispersions determined with fewer than eight radial velocities may be inaccurate, but we have set a lower minimum membership threshold to retain sensitivity to poorer structures. Our minimum membership requirement excludes 105652 galaxies (out of 177675 with $P - S > 0.5$) that, while individually falling within overdense environments, are not members of such a group. The minimum overdensity probability of any individual galaxy that is a member of such a group is 0.56. The density of our structures is compared with their environments in Section 2.3.4. Additionally, we reject structures with low or unmeasurable local density contrasts and structures with fewer than four member galaxies after line-of-sight sigma-clipping. These additional criteria are described in Section 2.3, and affect less than one per cent of our candidate structures.

2.2.4 Summary of Selection Choices

Although we have tried to minimise the assumptions made about groups and clusters when identifying them, such assumptions are impossible to eliminate entirely. We have attempted to minimise the arbitrariness of the choices we have made. Below we list justifications for our more significant choices, along with the selection effects these choices may produce in the resultant catalogue (or cite sections of this Thesis where they are given).

- **Sampling locations:** by using the input galaxy catalogue as an adaptive grid, we guarantee sufficient sampling while reducing computational effort. Greater spatial sensitivity would be achieved by use a regular lattice of sampling locations (e.g. Kim et al. 2002 for the case of the matched filter algorithm). The adoption of a regular lattice for MSPM would require changes to our calculation of probabilities to account for the inclusion of void regions. A suitable adaptation of our approach would recover a similar catalogue of structures.

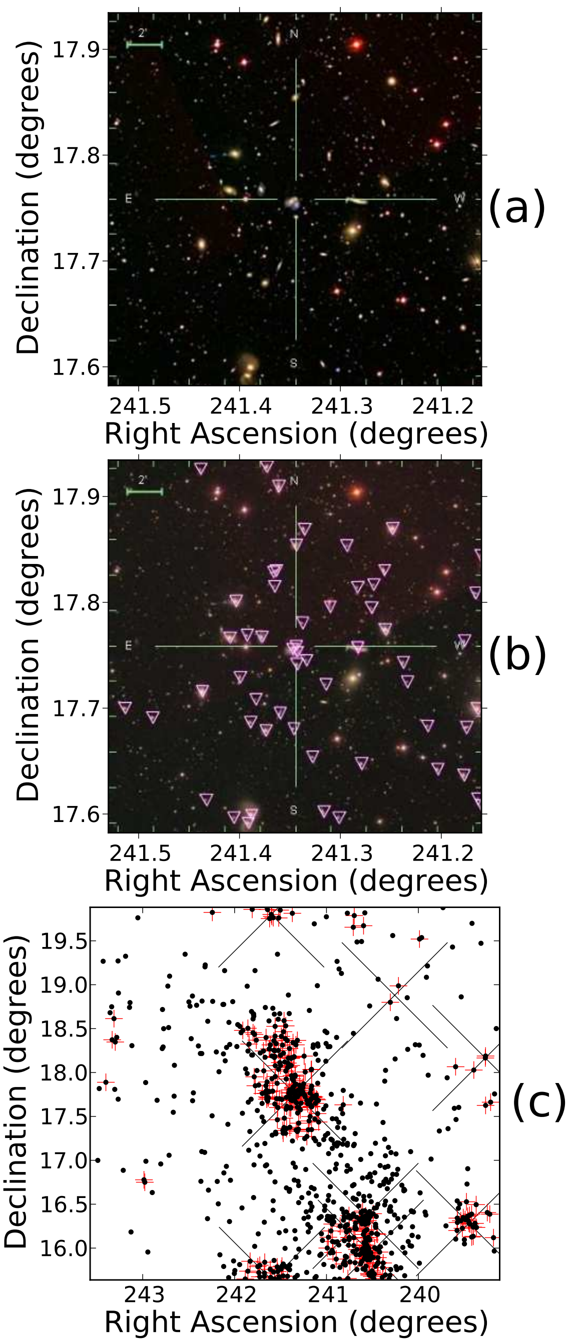


Figure 2.1: The first structure in our catalogue (Table 2.1), Abell 2151 (Abell 1958; Corwin 1974) in the Hercules supercluster (Tarenghi et al. 1979). **(a)** SDSS image centred on the cluster position, showing a transverse radius of 10.6 arcminutes (0.332 h^{-1} Mpc at $z = 0.036$). **(b)** The same image with $r < 17.77$ galaxies within a line-of-sight radius $\Delta z = 0.005$ of the cluster centre marked as triangles. **(c)** Objects within a much larger field of view, including a transverse radius of 4 h^{-1} Mpc and the same line-of-sight radius. Dots are $r < 17.77$ galaxies, which are also marked by red plusses if they are at positions with $P - S > 0.5$ and large crosses are MSPM structures, some of which are only partially visible within this redshift slice.

- **Comparison volume:** Local galaxy counts are compared with those obtained from redshift slices of width $\Delta z = 0.005$ (Section 2.2.2). From inspecting counts as a function of redshift in the SDSS volume, a narrower slice width would begin to become affected by sample variance (often referred to as cosmic variance) due to sampling an insufficient volume to infer the true local average density. A larger slice width would introduce biases in the mean density estimate, due to the Malmquist bias resulting from the magnitude limit of the survey. For example, at the nearer edge of the redshift slice, the average density would be higher than at the farther edge, merely due to the inclusion of intrinsically fainter sources only at the lower redshifts.

- **Threshold:** (Section 1.2.3) $P - S > 0.5$ selects galaxies in regions that are overdense relative to both the mean density and surrounding locations. Twenty-nine per cent of our primary galaxy sample satisfies this criterion (12 per cent remain after the minimum count threshold is enforced). A higher threshold would reduce the false discovery rate in our catalogue but, by including lower-significance detections, the MSPM catalogue retains more information about the galaxy distribution for a study of large-scale structure (Section 3). From comparisons with other catalogues, we find that our threshold affects the measured range of velocity dispersions, and hence the masses of the detected structures. Catalogues constructed from a smaller fraction of the galaxy population contain more massive groups (Section 2.3.7).

- **Projected scale range:** to locate groups and clusters rather than features of large-scale structure, our transverse sampling radii are set at 0.2 to 2 h^{-1} Mpc in steps of 0.2 h^{-1} Mpc. Our threshold includes regions with $S < 0.5$, corresponding to radii less than 1 h^{-1} Mpc. The characteristic radius of larger clusters implied by the two-point correlation function is 1 h^{-1} Mpc (e.g. Einasto et al. 1984). Sampling larger scales would potentially identify extended structures such as filaments.

- **Redshift radius r_z:** we find that $r_z = 10\,h^{-1}$ Mpc ($\Delta z \sim 0.004$) is suitable throughout our range of redshift, capturing the spread of peculiar velocities present within most structures, as discussed in Section 2.2.1.

- **Structure identification linking length:** our linking length of 2 h^{-1} Mpc corresponds to the maximum of our projected scale range. Since most of our structures have radii $< 1\,h^{-1}$ Mpc, altering this large linking length has little effect on the structures we find. Similarly, our line-of-sight elongation factor $e_{\rm los} = 10$ allows a 1 h^{-1} Mpc cluster to contain galaxies apparently up to 10 h^{-1} Mpc away in the line of sight, consistent with our 10 h^{-1} Mpc redshift radius. Altering $e_{\rm los}$ would thus produce effects similar to those of altering the redshift radius.

- **Minimum count of member galaxies:** We require that each structure has at least four member galaxies, as discussed in Section 2.2.3.

2.3 Group and Cluster Catalogue

The resultant catalogue in Table 2.1 contains 10443 structures in the redshift range $0.025 < z < 0.24$, containing a total of 72023 member galaxies, 12 per cent of the input galaxy data. This is lower than the 37 per cent identified in the Mr20 catalogue of Berlind et al. (2006), but greater than the 8 per cent contained by C4 clusters (Miller et al. 2005). Detailed comparison with these catalogues is discussed in Section 2.4.4. Structures were sought at $z > 0.24$, but none were found because of incompleteness due to the magnitude limit of the survey.

Table 2.1: Catalogue of MSPM groups and clusters in SDSS DR7.

ID	RA (J2000) (deg)	Dec (J2000) (deg)	z	P	N	R (h^{-1} Mpc)	$\mathrm{LDC}_{0.4,2}$	$\mathrm{LDC}_{1,2}$	σ_v (km s^{-1})	Galaxy $\rho/\bar{\rho}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1........	241.3437	17.7596	0.03629	1.000	157	1.696	8.6	3.9	680	513.1
2........	167.7442	28.6773	0.03270	1.000	111	0.870	10.0	4.6	645	385.0
3........	223.2302	16.6928	0.04428	1.000	62	1.555	8.8	5.9	546	659.6
4........	240.5182	15.9474	0.03341	1.000	47	0.944	5.0	3.6	408	532.4
5........	169.1441	29.2692	0.04651	1.000	66	0.941	10.0	6.7	456	583.6
6........	240.5750	16.3662	0.03889	1.000	48	0.903	4.5	3.5	272	379.7
7........	234.9231	21.7713	0.04120	1.000	71	0.782	9.0	4.2	541	475.9
8........	351.1126	14.6395	0.04003	0.999	53	0.744	4.8	3.0	715	391.8
9........	247.1607	39.5800	0.03015	0.999	94	1.187	3.6	2.3	757	351.6
10........	247.5227	40.7662	0.03047	0.999	85	1.056	4.8	2.3	580	353.2
....	....	....	....	....	....	....	....	....	....	....
1000........	195.9923	35.3664	0.03443	0.865	5	0.720	3.3	1.4	249	50.4
2000........	178.4335	22.3690	0.06570	0.987	5	0.342	9.6	9.6	185	118.2
3000........	171.6480	3.4761	0.07444	0.907	5	1.078	18.0	2.2	267	76.4
4000........	59.8634	−6.5318	0.06150	0.750	4	0.625	12.0	3.8	101	52.5
5000........	134.5177	30.3496	0.08505	0.972	8	0.691	5.8	2.7	232	199.2
6000........	122.9627	30.2907	0.07563	0.912	5	0.592	12.0	1.5	190	119.9
7000........	159.6631	23.9223	0.09476	0.676	4	1.066	16.0	4.5	117	73.0
8000........	221.1220	56.1547	0.11532	0.886	6	1.484	14.4	1.8	164	171.5
9000........	166.1805	4.2125	0.14423	0.993	4	0.608	24.0	6.0	178	334.0
10000........	155.2763	30.4010	0.15498	0.999	6	1.327	19.2	2.4	340	697.1

Locations and measured properties of MSPM structures. Entries are ordered by $P-S$ within slices of ascending redshift, where each slice has a width $\Delta z = 0.025$. This table shows only a portion of our catalogue as an indication of its content. The complete catalogue can be found in the online edition of Smith et al. (2012), or at http://www.physics.usyd.edu.au/sifa/Main/MSPM/ , along with three-dimensional visualisations.
The selection criteria are described in Section 2.2. Columns (2) to (4): position; (5): overdensity probability at peak $P-S$; (6): count of galaxies with $r < 17.77$; (7): radius enclosing region with $P-S > 0.5$; (8) to (9): local density contrasts; (10): velocity dispersion; (11): galaxy density within 0.4 h^{-1} Mpc in units of the background density. Our measurements are detailed in Section 2.3.

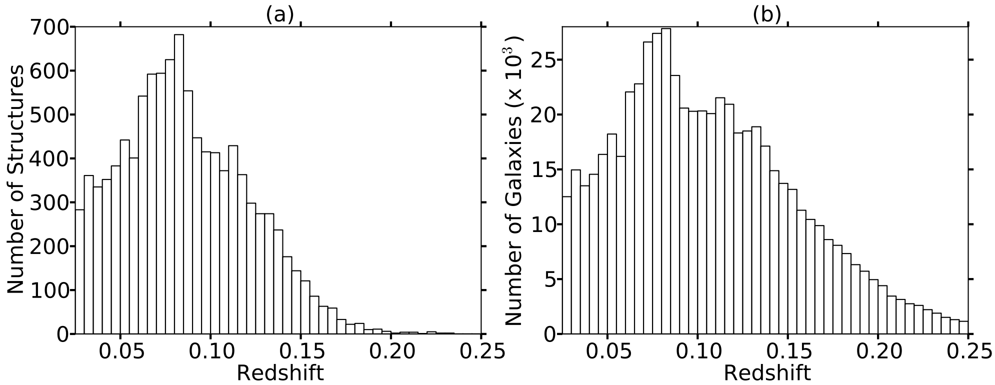


Figure 2.2: **(a)** The redshift distribution of MSPM structures in Table 2.1; mean: 0.086, median: 0.082. We restrict our attention to $z > 0.025$ (Section 2.2.1), and do not find any structures at $z > 0.24$ because of incompleteness. **(b)** The redshift distribution of galaxies from the SDSS spectroscopic survey.

For each structure, we measure a range of properties. Positions, overdensity probabilities, counts and radii result directly from our structure identification procedure. Local density contrasts are a density-based quantification of the significance of our detections and velocity dispersions measure the total mass in the systems we have found.

2.3.1 Position

Consistent with our structure selection, we quantify structures without the use of colour-magnitude information, using only our $P - S$ values and identified member galaxies. An MSPM structure is defined to have the same position on the sky as its member galaxy with the highest $P - S$ value. For structures that are elongated or asymmetric, an average position on the sky may not select the densest part of the structure. Peaks in $P - S$ most accurately identify the centres of structures containing at least eight member galaxies.

Our approach is analogous to the maximum density measure of Yoon et al. (2008). Since redshifts cannot be interpreted as precise positions on scales of $\sim 1\ h^{-1}$ Mpc, our structures have been assigned the average redshift of the member galaxies, similar to Berlind et al. (2006). Figure 2.2 shows that the redshift distribution of our catalogue peaks at $z \sim 0.08$, lower than that for the input galaxy data ($z \sim 0.1$). The difference is caused by the SDSS magnitude limit. The consequence of a magnitude limit is for the higher-redshift galaxies that enter the sample to be more luminous and massive, and typically to lie in overdense regions. However, the fainter members of such overdensities may not enter the sample, and as a result this reduces the proportion of overdensities that can be recovered at $z > 0.15$.

2.3.2 Overdensity Probability

The overdensity probability (P) reported for each structure in Table 2.1 is defined to be the value at its $P - S$ peak. Figure 2.3 shows that 71 per cent of our structures are detected with overdensity probabilities of 0.9 or greater, meaning that they are in the densest 10 per cent of

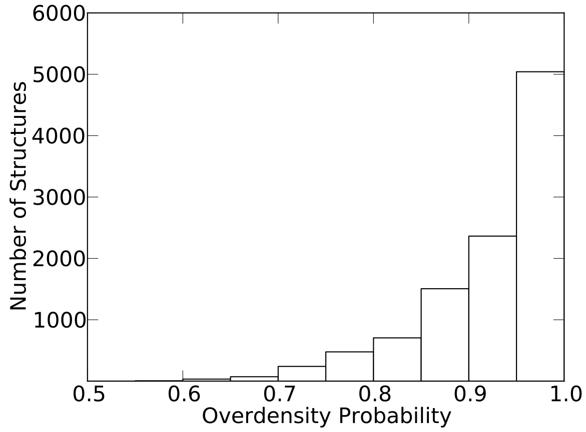


Figure 2.3: Histogram of overdensity probabilities for catalogued structures, median: 0.946. To qualify for inclusion in our catalogue, each structure must have an overdensity probability of at least 0.5, so that they are in the densest half of the galaxy distribution. Seventy-one per cent of our structures are detected with probabilities greater than 0.9.

the galaxy distribution, within the scale range sampled. Because the centres of overdensities have low S values, most (76 per cent) of these high probability values result from our chosen probability density (Section 1.2.1) over a radius of 0.2 h^{-1} Mpc, centred at the $P - S$ peak.

2.3.3 Count

The count of member galaxies is the number of galaxies associated with each structure by our structure identification process. Each member galaxy must have $P - S > 0.5$. Figure 2.4 shows that most of our structures have fewer than eight members with $r < 17.77$ above this threshold. Although we find that counts of member galaxies and overdensity probabilities are correlated, 56 per cent of structures with only four member galaxies still have probabilities greater than 0.9.

Seventy-five per cent of structures with eight or more member galaxies (23 per cent of the full sample) have probabilities greater than 0.95. These form a high-purity subset of our catalogue, with properties measured more accurately, as a higher detected number of the structure members allows a more robust estimate of their radius and velocity dispersion.

2.3.4 Radius

Because of the apparent line-of-sight elongation resulting from galaxy peculiar velocities, we cannot use redshift information to determine the total physical extent of structures on $\sim 1\ h^{-1}$ Mpc scales. We use instead the transverse (sky) distance from the $P - S$ peak to the furthest member galaxy. This measure will be sensitive to random galaxy displacements, but a more significant limitation for most of our structures is their low count of member galaxies. Figure 2.5 shows that most of our measured radii fall within our range of sampled scale values, as

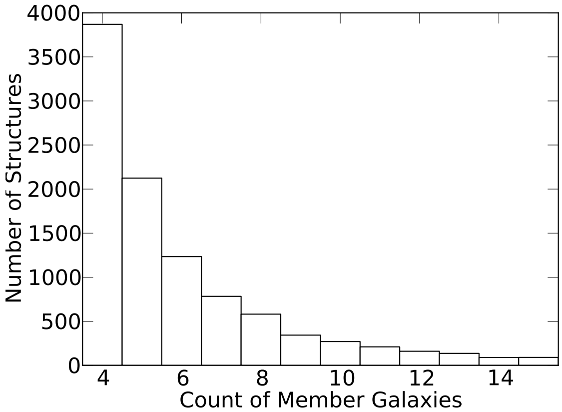


Figure 2.4: Histogram of member galaxy counts for catalogued structures The median count is 5. To qualify for our catalogue, each structure must contain at least four member galaxies (Section 2.2.3); 37 per cent of our catalogue has this count. Five per cent of our structures have 16 or more member galaxies (not shown).

expected. We have found that some of our radii are overestimated as a result of contamination by unassociated nearby galaxies.

To explore the physical significance of the radii we have measured for our structures, we have determined average galaxy density profiles as a function of radius (Figure 2.6) for structures with radii up to 1 h^{-1} Mpc. Galaxies within 10 h^{-1} Mpc in the line of sight are included in the average. The density profile for each structure is normalised to have a minimum of one, so that all structures have equal weight. Profiles from individual structures sharing similar radii are then averaged.

Figure 2.6 shows that our radii approximate boundaries containing regions twice as dense as the density at 2 h^{-1} Mpc, ρ_2, defined as the density in the annulus formed by rings of radii 1.95 and 2.0 h^{-1} Mpc. Our $P - S > 0.5$ criterion selects regions that are unusually overdense, and this corresponds approximately to regions that satisfy $\rho > 2\rho_2$. In real space this galaxy density contrast is much higher (~ 130, Section 2.3.6), without averaging over a large redshift radius. This is an empirical result, and cannot be generalised to all potential input galaxy distributions.

Our large structures tend to have radii that enclose lower densities relative to the background than those enclosed by the radii of small structures. Outlying regions of larger structures are above our $P - S$ threshold because they are recognised as extended regions of intermediate density, and thus unusual when compared with most of the galaxy distribution. The opposite effect occurs for our structures with smaller radii, since large densities at small intergalaxy separations are common, because of the intrinsic clustering of galaxies. This trend results from our decision to use the locations of the galaxy distribution itself for our probability density measurements.

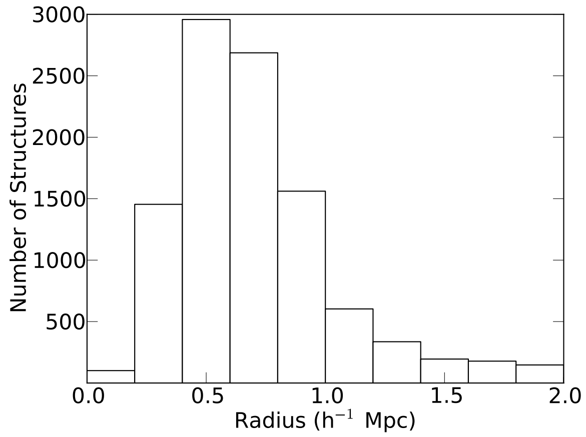


Figure 2.5: Histogram of transverse radii measured for catalogued structures within our sampled scale range; mean: 0.75, median: 0.65. Two per cent of our structures have radii greater than 2 h^{-1} Mpc (not shown).

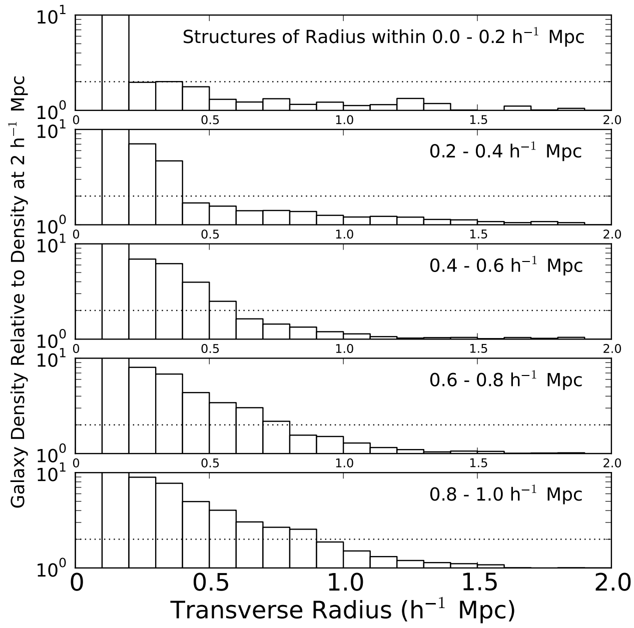


Figure 2.6: Profiles of the galaxy density relative to surrounding locations, including $P - S \leq 0.5$ galaxies, within annuli centred on $P - S$ peaks determined for our structures, averaged over $0.025 < z < 0.2$. Each panel shows an average of density profiles obtained for all structures with radii within the indicated range. Horizontal lines indicate twice the density at $2\,h^{-1}$ Mpc, defined as the density in the annulus formed by rings of radii 1.95 and 2 h^{-1} Mpc. Relative densities close to group and cluster centres are clipped at 10.

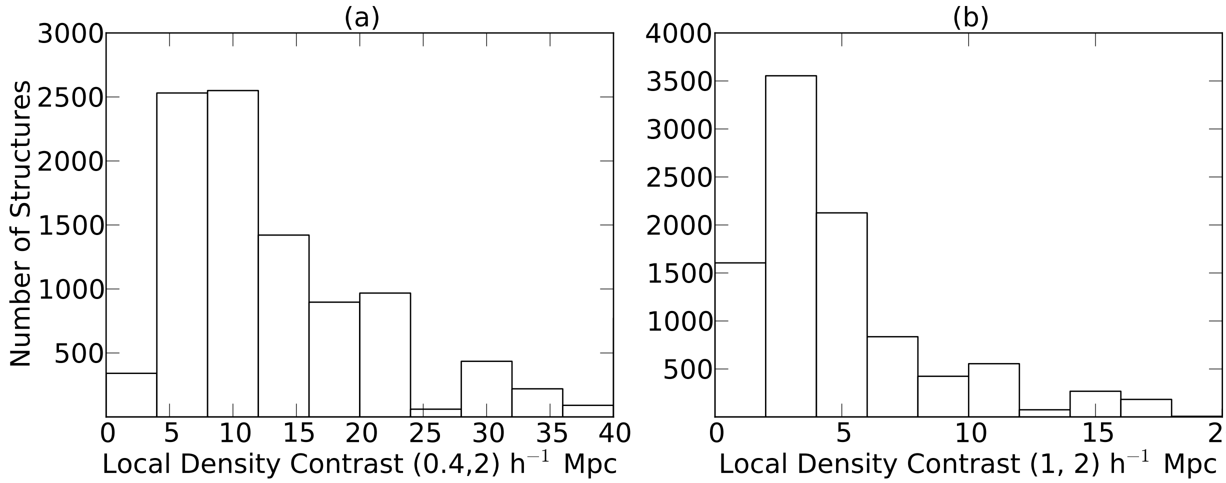


Figure 2.7: **(a)** Histogram of $\mathrm{LDC}_{0.4,2}$ for catalogued structures; mean: 17.9, median: 12. Seven per cent of our structures have $\mathrm{LDC}_{0.4,2} > 40$, excluding 107 undefined values not shown (Section 2.3.5). **(b)** Histogram of $\mathrm{LDC}_{1,2}$ for catalogued structures; mean: 5.5, median: 4. Two per cent of our structures have $\mathrm{LDC}_{1,2} > 20$, excluding 563 undefined values not shown (Section 2.3.5).

2.3.5 Local Density Contrast

We define a local density contrast (LDC) as the density within an inner radius divided by the density within the annulus formed by an outer radius projected on the sky. LDCs are found for two pairs of inner and outer radii: (0.4, 2) h^{-1} Mpc and (1, 2) h^{-1} Mpc. Galaxies further than 10 h^{-1} Mpc in the line of sight from the structure position are excluded from the galaxy density. We refer to the LDC measures with subscripts denoting the inner and outer radii in h^{-1} Mpc as $\mathrm{LDC}_{0.4,2}$ and $\mathrm{LDC}_{1,2}$. LDC measurements allow comparison with the density obtained over a large neighbouring volume.

Since small radii centred on individual galaxies are bound to yield high densities, if the count of objects within the inner radius is only one, the resulting LDC is considered unmeasurable and reported in our catalogue as -1. Structures for which neither of our LDCs can be measured are rejected from the catalogue. Of the structures that remain, where there are no galaxies in the outer annulus, the LDC is undefined, resulting from division by zero. These structures are retained in the catalogue. A larger outer radius or the inclusion of more sensitive observations would resolve this issue. Our exclusion of all structures close to the survey edges ensures that the outer radius is always within the survey.

Defining $\mathrm{LDC}_{\mathrm{max}} = \max(\mathrm{LDC}_{0.4,2}, \mathrm{LDC}_{1,2})$, we require that all of our structures have $\mathrm{LDC}_{\mathrm{max}} > 2$. This constraint excludes 9 structures from our catalogue that are less than twice as dense as surrounding locations, and all structures that do not have at least two member galaxies within 1 h^{-1} Mpc of the $P-S$ peak. On average, $\mathrm{LDC}_{\mathrm{max}} = 16.5$, excluding 563 undefined LDC values. For 90 per cent of our structures, $\mathrm{LDC}_{0.4,2} > \mathrm{LDC}_{1,2}$. The jagged appearance of the LDC histograms (Figure 2.7) is a result of integer counts dictating preferred fractions.

2.3.6 Global Density Contrast

Densities of our structures in units of the background galaxy density have also been estimated. The density of a virialised structure in units of the critical density $\rho_{\rm cr}$ is predicted by the spherical collapse model (e.g. Bryan & Norman 1998; King & Mead 2011) for $\Omega_{\rm M} = 0.3$ and our median redshift $z = 0.082$ to be $\Delta_c = 91$. This is an overdensity $\Delta = 304$ in units of the background density $\Omega_{\rm M}\rho_{\rm cr}$ (e.g. Voit 2005). Our data allow us to estimate the *galaxy overdensity* rather than the matter overdensity. Additionally, some galaxies will not be included in the SDSS spectroscopic survey because of fibre collisions, avoidance of bright stars, and other practical survey limitations, so this is not a robust measure of structure densities relative to the mean.

For each structure, we have calculated a galaxy global density contrast $\rho/\bar{\rho}$ as the density of galaxies within 0.4 h^{-1} Mpc divided by the density of galaxies within a redshift slice of width $\Delta z = 0.005$ over the entire survey area. The median global (galaxy number) density contrast for our structures is 130.6, and its average is 185.5. Because of the line-of-sight elongation caused by galaxy radial motions, in counting galaxies within a radius of 0.4 h^{-1} Mpc we have included galaxies within a line-of-sight radius of 10 h^{-1} Mpc.

2.3.7 Velocity Dispersion

Rather than use the radial velocities of all galaxies within an arbitrary transverse radius (such as 1 h^{-1} Mpc) to calculate velocity dispersions (σ_v), we use only those galaxies identified by our approach as being structure members. Equation 2.1 allows a large line-of-sight linking length that may contribute member galaxies that are far from structure centres in the line of sight. These galaxies may not be physically associated with structures, and we use line-of-sight sigma-clipping to remove them, for the purpose of calculating σ_v only. Under this procedure, σ_v is calculated using the radial velocities of all member galaxies. The initial σ_v value is used to identify outlying radial velocities, which are then removed before σ_v is then recalculated. This iterative process is also used to remove apparent groups that may result from the chance alignment of unassociated galaxies in the line of sight.

Using the biweight estimator (Beers, Flynn & Gebhardt 1990), four iterations of σ-clipping at 2σ are applied to the radial velocities. Our large redshift radius raises the possibility of multiple structures in the line of sight. To prevent an estimation of σ_v for the wrong structure, the median and mean radial velocities are fixed to prevent them from shifting during iteration. A structure is excluded from our catalogue entirely if fewer than four member galaxies remain after clipping. Less than one per cent of our candidate structures are rejected by this criterion.

The range of a catalogue's σ_v measurements is strongly dependent on the criteria used to define and select structures. Our σ_v distribution is shown in Figure 2.8 and has a median of 183 km s^{-1}; structures with eight or more members have a median of 258 km s^{-1}. The range of our σ_v measurements imply masses consistent with structures ranging from poor groups to some of the most massive clusters, with $\sigma_v > 1000$ km s^{-1}. Having allowed relatively poor structures (together containing 12 per cent of the input galaxy data) into our catalogue, we find that our median σ_v is comparable with those of samples constructed using similarly low thresholds (e.g. Berlind et al. 2006 Mr20: $\sigma_v = 128$ km s^{-1}; McConnachie et al. 2009: $\sigma_v = 227$ km s^{-1}),

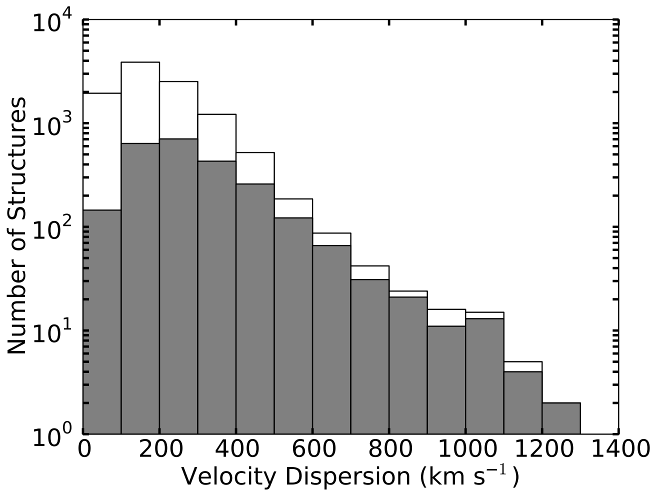


Figure 2.8: *Open histogram:* Velocity dispersions for all catalogued structures; mean: 212 km s^{-1}, median: 183 km s^{-1}. *Shaded histogram:* Structures with eight or more member galaxies; mean: 296 km s^{-1}, median: 258 km s^{-1}.

but lower than those found in studies with higher thresholds (e.g. Miller et al. 2005: σ_v = 576 km s^{-1}). This is evidence that we are correctly identifying structure members. Although the range of our results is consistent with previous measurements, most of our individual σ_v measurements are based on fewer than eight radial velocities, and may not be accurate.

2.4 Purity, Radius-Velocity Dispersion Relation and Comparisons with Other Studies

2.4.1 Average Local Density Contrast

Because the galaxy correlation function dictates that galaxies *normally* see decreasing densities as a function of radius, high local density contrast (LDC) values are only meaningful if they are also above the average LDC in the general galaxy population. We compare the LDCs in our catalogue with average LDCs obtained from the whole galaxy population in Figure 2.9. At all redshifts, LDCs obtained for our structures are more than those found for the whole galaxy population, within 1σ uncertainties. All these averages are based on LDCs where the inner count of galaxies is at least two.

2.4.2 Four-Member Detections

Over a third of our structures have exactly the minimum count of four member galaxies; the impact of this threshold is discussed in Section 2.3.3. Although this makes them marginal overdensities, they have an average $\mathrm{LDC_{max}}$ value of 18.1, *higher* than that for the remainder of the catalogue, 15.6. Figure 2.10 shows the distribution of $\mathrm{LDC_{max}}$ values obtained for this

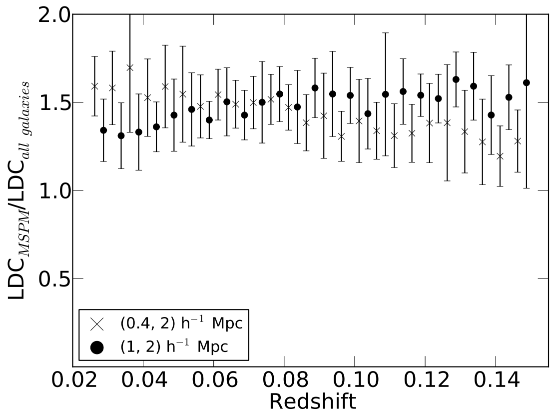


Figure 2.9: LDC values for our catalogued structures divided by the average obtained over all galaxies in the input distribution, for $0.025 < z < 0.15$. Series are offset for clarity and 1σ bootstrap uncertainties are shown. Note that "all galaxies" includes those found in our structures, which contain 72023 of the total 619234 galaxies. Average LDC values for all galaxies are overestimated since they are galaxy-weighted rather than volume-weighted, meaning that dense enviroments are sampled more than poor ones.

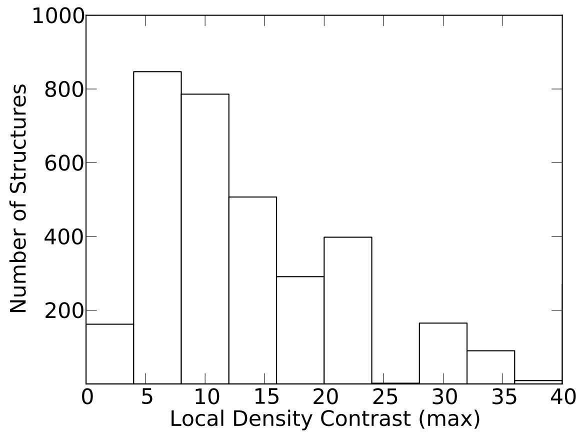


Figure 2.10: LDC$_{\text{max}}$ distribution for structures with only four member galaxies; mean 18.1, median 12. Six per cent of these four-member detections have LDC$_{\text{max}} > 40$, including 341 undefined values (Section 2.3.5).

subset. Since four-member structures tend to be surrounded by poorer environments, their high *contrast* values do not indicate higher density than more populous structures. Higher LDCs may be produced by the relative isolation of a system, its intrinsic richness, or both.

2.4.3 Radius-Velocity Dispersion Relation

Characteristic properties of groups and clusters are related to their internal dynamics and stages of evolution (Voit 2005). Older groups are more concentrated, having formed when the Universe was denser (Navarro, Frenk & White 1997), and are more relaxed and spherical than their younger counterparts (Ragone-Figueroa et al. 2010). Under simplifying assumptions, some properties can be described by scaling relations. We will treat our groups as isothermal spheres ($\rho(r) \propto r^{-2}$). More realistic mass-density profiles (e.g. Navarro, Frenk & White 1997) are shallower at small radii and steeper at large radii.

Assuming an isothermal distribution, velocity dispersion is an indicator of total enclosed virial mass, which should be proportional to σ_v^3. This mass is also proportional to the virial radius cubed (e.g. Bryan & Norman 1998; Kitayama & Suto 1996), implying a linear relation between radius (R) and σ_v. If Δ_c is the mean internal group density in units of the critical density and H is the Hubble parameter, this relation is:

$$\sigma_v = \frac{1}{2} H \Delta_c^{1/2} R. \tag{2.2}$$

With the radii we have found for the structures in our sample, we search for a relation between σ_v and radius.

Figure 2.11(a) shows σ_v against radius for all structures in the catalogue, revealing a clear trend, albeit with much scatter. Seventy-nine per cent of our structures lie within the parameter ranges shown in Figure 2.11(a), with the same data over an extended range in σ_v and R shown in Figure 2.12. The scatter can be partially attributed to uncertainties in the estimates of radius and σ_v, and to the presence of groups and clusters with various internal densities in our sample (see below). The velocity dispersions are especially uncertain for low counts of member galaxies. Radii are determined from the projected separation between the $P-S$ peak and furthest member galaxy, and so are at least uncertain by the mean separation of member galaxies. Systems in the lower-right corner of Figures 2.11(a) and 2.12 have overestimated radii as a result of contamination by nearby unassociated galaxies, and systems in the upper left may have overestimated velocity dispersions due to such contamination in the line of sight.

To quantify the trend evident in Figure 2.11(a), we have performed a least-squares fit to the data as follows. Velocity dispersions for all groups and clusters with radii between 0.2 h^{-1} Mpc and 1 h^{-1} Mpc are arranged into bins of width 0.1 h^{-1} Mpc. Clusters with radii greater than 1 h^{-1} Mpc are not included because our chosen scale range (Section 2.2.4) causes reduced sensitivity structures of this size, resulting in incompleteness that would affect our fit. Measurements that are more than one standard deviation from the mean σ_v in each bin are removed, after which 75 per cent of our data (6537 structures) at $R < 1\ h^{-1}$ Mpc remain. Our linear fit with 1σ uncertainty, shown in Figure 2.11(b), is

$$\sigma_v = (304 \pm 3)R + (8 \pm 2), \tag{2.3}$$

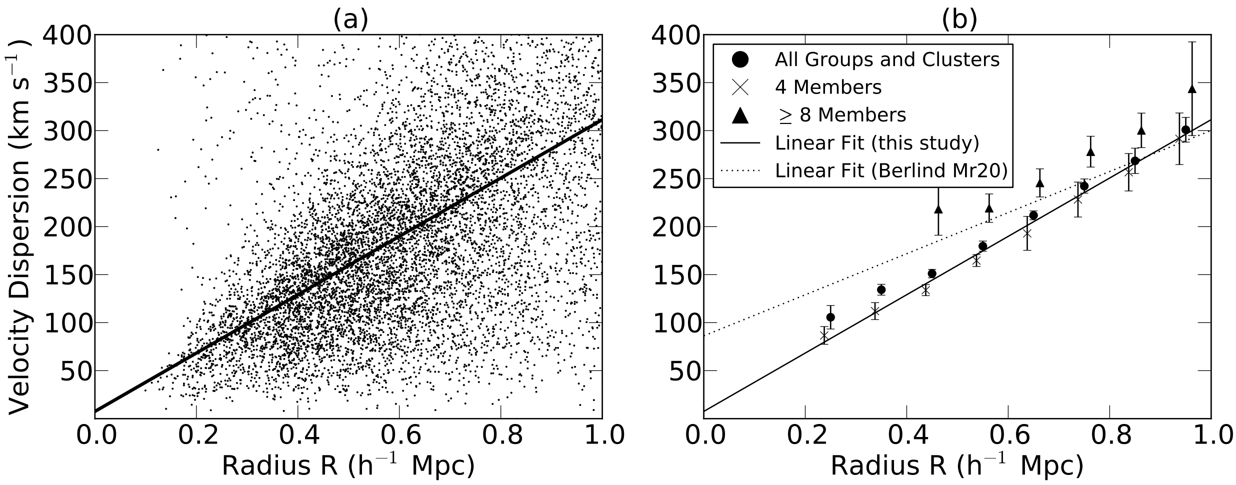


Figure 2.11: **(a)** Velocity dispersion (σ_v) against radius (R) for all MSPM groups and clusters, with our linear fit. The same data over an extended range in σ_v and R is shown in Figure 2.12. **(b)** Data points show mean group velocity dispersion against radius with 1σ bootstrap uncertainties. Series are offset for clarity. Relatively few groups and clusters with at least eight member galaxies have radii less than $0.4\,h^{-1}$ Mpc, so a reliable average cannot be obtained. Lines are fits to data with outlying velocity dispersions removed. See text for details.

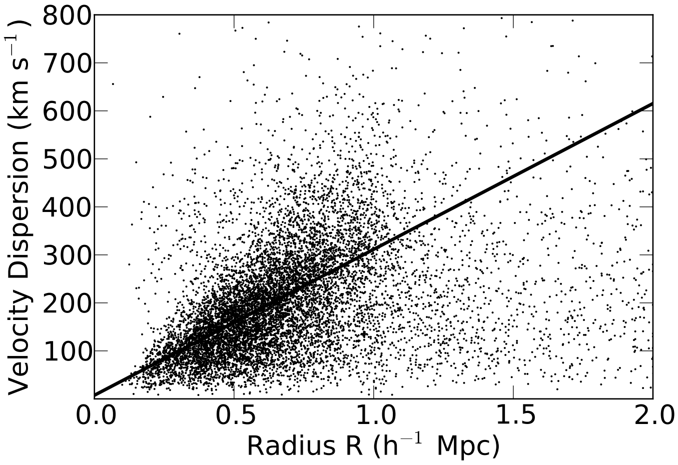


Figure 2.12: Velocity dispersion (σ_v) against radius (R) for all MSPM groups and clusters, with our linear fit as in Figure 2.11(a) but over an extended range in σ_v and R. The line is a fit to data with $0.2\,h^{-1}$ Mpc $< R < 1\,h^{-1}$ Mpc and after removal of outlying velocity dispersions. See text for details.

where R is the group or cluster radius in units of h^{-1} Mpc and σ_v is in units of km s^{-1}. This fit is to the sigma-clipped, unbinned data and is not constrained to pass through the origin. The 1σ confidence interval for the distribution of data points is ± 51 km s^{-1}.

There may also be underlying systematic uncertainties that are not reflected by the random uncertainties shown in equation 2.3. Concentrating on the slope uncertainty, a fit to the $R < 1$ h^{-1} Mpc data without sigma-clipping finds $\sigma_v/R = (283 \pm 6)$ h Mpc^{-1} km s^{-1}, suggesting a systematic uncertainty attributable to the fitting procedure of ± 21 h Mpc^{-1} km s^{-1}. Including this systematic uncertainty we find

$$\frac{\sigma_v}{R} = 304 \pm 3 \text{ (statistical)} \pm 21 \text{ (systematic)}\,, \tag{2.4}$$

where R is the group or cluster radius in units of h^{-1} Mpc and σ_v is in units of km s^{-1}.

Although an R-σ_v relation is not directly noted by Berlind et al. (2006), we have performed a similar analysis of data from their Mr20 group and cluster sample. Berlind et al. suggest that their velocity dispersions are 20 per cent too low at all multiplicities, so we apply a 20 per cent upward correction to compensate. A linear fit through their data at $R < 1$ h^{-1} Mpc has a slope of (214 ± 9) h Mpc^{-1} km s^{-1}, including only a statistical uncertainty. This slope is flatter than ours, but the fit is consistent with our data at $R > 0.5$ h^{-1} Mpc and also shown in Figure 2.11(b). We stress that close agreement cannot be expected, since R and σ_v are calculated differently by Berlind et al., who also set an effectively lower group identification threshold. This may partially account for their lower value, since the slope of the R-σ_v relation is a function of the mean group and cluster density in units of the critical density Δ_c (equation 2.2).

To apply equation 2.2 to our results, a conversion of R to units of Mpc entails division by h, removing the need for an assumed value of H_0. We also convert our radii from comoving to proper distances assuming our median redshift $z = 0.082$. The slope of our R-σ_v relation and equation 2.2 implies $\Delta_c = 43.2 \pm 7.1$. This value is significantly lower than predicted by the spherical collapse model for virialised systems (e.g. Bryan & Norman 1998; King & Mead 2011), which for $\Omega_{\mathrm{M}} = 0.3$ and our median redshift $z = 0.082$ is $\Delta_c = 91$.

We note that equation 2.2 assumes that groups have recently virialised and that R is the virial radius, assumptions we have not examined in this study. Our low Δ_c could therefore imply that many of our structures are not virialised or, alternatively, it could arise from systematic underestimation of σ_v or overestimation of R. Moreover, the value of Δ_c implied by the data of Berlind et al. (2006) is lower than ours, even though Berlind et al. optimised their linking lengths to select group-member galaxies occupying the same virialised dark matter halo. Hence, the models assumed in determining Δ_c may be flawed. We revisit this issue in Chapter 5.

A consistent variation of σ_v with increasing R to $1\,h^{-1}$ Mpc is evidence that there is no firm division between groups and clusters. An R-σ_v relation could be caused by linking criteria like equation 2.1, but we have ensured that the galaxies included in our velocity dispersion measurements are member galaxies and occupy overdense regions as demonstrated in Figure 2.6. Moreover, sigma-clipping has been used to remove galaxies spuriously included by our large redshift radius and line-of-sight elongation factor.

Table 2.2: Other Catalogues Recovered by MSPM

Other Catalogue*	Number	Fraction (%)
C4	325/466	70
Berlind Mr20	212/394	54
Y08	163/208	78

Numbers and fractions of other catalogues that are recovered by MSPM. In the case of C4, we compare at $z < 0.1$. In the case of Mr20, we consider the subset with at least four member galaxies. For Mr20 and Y08 we consider data at $0.09 < z < 0.10$. See text for details.

Table 2.3: MSPM Structures Found by Other Catalogues

Other Catalogue*	Number	Fraction (%)
C4	243/602	40
Berlind Mr20	253/362	70
Y08	162/616	26

Numbers and fractions of our MSPM catalogue recovered by other techniques, with the number of MSPM structures adjusted to reflect the varying survey areas (according to data release) and redshift intervals. In the case of C4, we consider the subset of MSPM structures with at least eight member galaxies. For Mr20 and Y08 we restrict MSPM to $0.09 < z < 0.10$. See text for details.

*C4: Miller et al. (2005); Mr20: Berlind et al. (2006); Y08: Yoon et al. (2008).

2.4.4 Comparison with Other Catalogues

Since we only use data for which spectroscopic information is available, we focus on comparisons with group and cluster catalogues that are similarly derived. When comparing any two catalogues, appropriate adjustments are made for varying survey areas and varying redshift limits at the time of the catalogue's compilation. Counterparts are identified by looking within cylinders (aligned with the line of sight) centred at group and cluster centres, with transverse and line-of-sight radii of 1 and 10 h^{-1} Mpc ($\Delta z \sim 0.004$) respectively. When determining the fraction of one catalogue recovered by another, the former catalogue's candidates centred less than 2 h^{-1} Mpc on the sky from the edges of the survey area available to both catalogues are removed so that mismatches are not caused by survey area differences. Similar adjustments are made to account for the varying redshift ranges of each catalogue. Our comparisons (with Miller et al. 2005; Berlind et al. 2006; Yoon et al. 2008) are summarised in Tables 2.2 and 2.3.

The C4 catalogue (Miller et al. 2005) is based on DR2 and offers three centroids for sky positions. In our comparison we consider the peak in the C4 density field since it is the closest analogue to our $P - S$ peaks. MSPM recovers 62 per cent (431) of the 694 C4 clusters that are more than 2 h^{-1} Mpc from the survey edges. Although the C4 catalogue is confined to the spectroscopic data, it uses the LRG sample (Eisenstein et al. 2001), which we have omitted from our input data. The C4 catalogue is thus based on approximately 1.5 times more data. At $z < 0.1$, where the LRG fraction has fallen to seven per cent, MSPM recovers 70 per cent (325) of the 466 remaining C4 candidates.

The C4 catalogue imposes a minimum galaxy membership of 8, so to find the fraction of our catalogue matched by C4, we consider the subset of MSPM structures with at least eight member galaxies. Above the minimum C4 redshift of 0.03, 40 per cent (243) of our structures with eight or more members (numbering 602 in DR2) are matched by C4. We attribute this low recovery rate to our multiscale approach and to our effectively lower threshold. Miller et al. (2005) use apertures with a fixed transverse radius of 1 h^{-1} Mpc to search for clusters, whereas we search over a range of scales. Our threshold selects groups and clusters that contain 12 per cent of the input galaxy data, whereas C4 clusters contain 8 per cent. Our median velocity dispersion for MSPM structures with at least eight members (258 km s^{-1}) is also far lower than in the C4 catalogue (576 km s^{-1}), indicating that we are finding more poor groups.

Berlind et al. (2006) use a friends-of-friends algorithm to construct a group and cluster catalogue based on DR3 data, using average member galaxy positions for their centroids. We compare with their Mr20 sample. Since Mr20 is based on volume-limited input, our comparison is carried out at $0.09 < z < 0.1$ so that our flux-limited catalogue is based on roughly equivalent data. MSPM recovers 54 per cent (212) of the 394 Mr20 groups and clusters that are more than 2 h^{-1} Mpc from the survey edges at $0.09 < z < 0.10$ and that have at least four member galaxies (to match our minimum galaxy membership). The Mr20 groups that MSPM fails to recover are judged by our approach to be less dense than surrounding locations ($S > 0.5$) within 2 h^{-1} Mpc. These groups are still local density peaks when compared with smaller-scale environments. At $0.09 < z < 0.1$, 70 per cent (253) of the MSPM catalogue (numbering 362 in DR3) is matched by Mr20.

Yoon et al. (2008; hereafter Y08) follow a Gaussian weighting scheme to measure densities and construct a cluster catalogue based on DR5 data. Like the Berlind Mr20 sample, the Y08 catalogue is based on volume-limited input, so our comparison is carried out at $0.09 < z < 0.1$. We compare MSPM positions with the sky positions of their maximum-density galaxies and their Gaussian-fitted redshifts, recovering 78 per cent (163) of the 208 Y08 clusters at $0.09 < z < 0.1$ that are more than 2 h^{-1} Mpc from the survey edges. By inspecting mismatches we find that the Y08 clusters we fail to recover are probably real systems, but are not concentrated enough for the MSPM catalogue. The Y08 catalogue allows galaxies in the photometric-only portion of the SDSS data to contribute to their detections, which may account for at least some of the mismatches. At $0.09 < z < 0.10$, 26 per cent (162) of the MSPM catalogue (numbering 616 in DR5) is matched by Y08. A much higher effective threshold is enforced by the Y08 catalogue, which contains approximately three times fewer structures at $0.09 < z < 0.1$ than the MSPM catalogue. Moreover, we have sampled a range of scales whereas Y08 follow a Gaussian weighting scheme with a fixed transverse σ of 1 Mpc (0.7 h^{-1} Mpc).

Our comparisons show that MSPM recovers most groups and clusters contained in catalogues based on similar data. However, the relatively low threshold we have set means that many candidate MSPM structures are not detected in other catalogues. Nevertheless, comparison of our structure LDCs with averages (Figure 2.9) and the radius-σ_v correlation (Section 2.4.3) are evidence for the reality of the MSPM structures. Moreover, Figure 2.6 shows that our groups and clusters are a subset of regions that have twice the density at 2 h^{-1} Mpc (averaged over a large line-of-sight distance). It remains probable that false detections and overdensities that are not gravitationally bound have been introduced by our relatively low threshold but, by including lower-significance detections, the MSPM catalogue retains more information about the galaxy distribution for a study of large-scale structure.

Part II

Large-Scale Structure

The Divine Root Conceives and the Spring Breaks Forth
As the Heart's Nature Is Cultivated, the Great Way Arises
Before Chaos was divided, Heaven and Earth were one;
All was a shapeless blur, and no men had appeared.
Once Pan Gu destroyed the Enormous Vagueness
The separation of clear and impure began.

TRIPITAKA: *Monkey, how far is it to the Western Heaven, the abode of Buddha?*

WU-KONG: *You can walk from the time of your youth till the time you grow old, and after that, till you become young again; and even after going through such a cycle a thousand times, you may still find it difficult to reach the place where you want to go. But when you perceive, by the resoluteness of your will, the Buddha-nature in all things, and when every one of your thoughts goes back to that fountain in your memory, that will be the time you arrive at Spirit Mountain.*

– Wu Cheng-en, *Journey to the West*

Desert Passage

Jessica saw the slave cribs on Bela Tegeuse down that inner corridor, saw the weeding out and the selecting that spread men to Rossak and Harmonthep. Scenes of brutal ferocity opened to her like the petals of a terrible flower. And she saw the thread of the past carried by Sayyadina after Sayyadina – first by word of mouth, hidden in the sand chanteys, then refined through their own Reverend Mothers with the discovery of the poison drug on Rossak . . . and now developed to subtle strength on Arrakis in the discovery of the Water of Life.

Far down the inner corridor, another voice screamed: "Never to forgive! Never to forget!"

But Jessica's attention was focused on the revelation of the Water of Life, seeing its source: the liquid exhalation of a dying sandworm, a maker.

Slowly, he scanned the horizon, listening, watching for the signs he had been taught.

It came from the southeast, a distant hissing, a sand-whisper. Presently he saw the faraway outline of the creature's track against the dawnlight and realized he had never before seen a maker this large, never heard of one this size. It appeared to be more than half a league long, and the rise of the sandwave at its cresting head was like the approach of a mountain.

The wild maker, the old man of the desert, loomed almost on him. Its cresting front segments threw a sandwave that would sweep across his knees.

Come up, you lovely monster, *he thought*. Up. You hear me calling. Come up. Come up.

And far to the rear along the worm's surface, Paul heard the beat of the goaders pounding the tail segments. The worm began picking up speed. Their robes flapped in the wind. The abrasive sound of their passage increased.

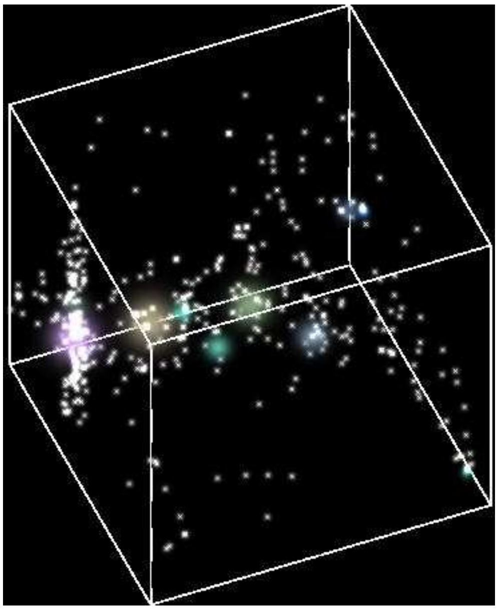
Paul looked back through the troop, found Chani's face among them. He looked at her as he spoke to Stilgar. "Then I am a sandrider, Stil?"

"Hal yawm! You are a sandrider this day."

– Frank Herbert, *Dune*

Chapter 3

Filaments



A solid particle render (Appendix B) of MSPM filament 1063. The cube has a side length of $20\, h^{-1}$ Mpc.

Sections 3.2-3.4 of this Chapter are taken from Smith et al. (2012) with minor adaptations.

In this Chapter we develop a method for quantifying elongation to measure the elongation of group and cluster environments (Section 3.2). By using our group and cluster catalogue as a coarse-grained representation of the galaxy distribution for structure sizes of $\lesssim 1\,h^{-1}$ Mpc, we demonstrate a technique to identify filaments with a false discovery rate of less than one half. We have identified 53 filaments (from an algorithmically-derived set of 100 candidates) as elongated unions of groups and clusters at $0.025 < z < 0.13$ (Section 3.3). These filaments have morphologies that are consistent with previous samples studied (Section 3.4).

3.1 Review: Spears of Light

Groups and clusters are spheroids. On larger scales, it is no longer the overdense structures which are spheroidal, but the underdense structures, the voids. In this sense, the existence of the voids one sees is not guaranteed by the existence of clusters, because overdense and underdense structures cannot simultaneously be spheroidal on the same scale. The existence of spheroidal voids on a larger scale entails the existence of overdensities that are not spheroidal: filaments and sheets.

3.1.1 Significance and Properties

Filaments are a striking feature of the galaxy distribution, and must be reproduced by any viable model of cosmic structure growth. Aragón-Calvo, van de Weygaert & Jones (2010, hereafter AWJ) find in ΛCDM simulations that more massive clusters host more filamentary connections, and this is supported in real data by Pimbblet, Drinkwater & Hawkrigg (2004), hereafter PDH. ΛCDM-simulated filaments have been morphologically classified by Colberg, Krughoff and Connolly (2005a; hereafter CKC), with results that are also supported in real data by PDH. On the other hand, Stoica, Martínez & Saar (2010) suggest that model filaments are shorter than real filaments, and do not form an extended network (see also Section 5.1.2.3). A related anomaly may be the absence of extremely large structures in simulations (Murphy, Eke & Frenk 2011; Yaryura, Baugh & Angulo 2011).

Filaments are important in mass accounting, and AWJ find from simulation that filaments contain most of the Universe's mass. Accordingly, a large fraction of the Universe's *baryonic* mass is thought to reside in filaments (Bregman 2007), but its detection is difficult because much of it exists as low-density gas, the so-called warm-hot intergalactic medium (WHIM). In galaxy clusters, the baryonic mass in the intracluster medium's gas is comparable with that contained in the galaxies and visible by X-ray emission. The WHIM has a much lower density in filaments and is ordinarily beyond detection in X-ray, but stacking has enabled a measurement of the electron density in filaments by Fraser-McKelvie, Pimbblet & Lazendic (2011).

Investigations with simulated data have helped define the properties and morphologies typical filaments may have. In simulated data, filaments are typically $2\,h^{-1}$ Mpc wide (AWJ), tend to have lengths of $\sim 15\,h^{-1}$ Mpc (Colberg 2007), with a presence that is statistically significant up to a length of $\sim 110\,h^{-1}$ Mpc (Pandey et al. 2011). PDH find that real filaments have lengths and morphologies consistent with such predictions, but further studies are required to address the possible divergence of simulated filaments on the largest scales, discussed above.

3.1.2 Filament Finders

Filaments have long been noted (e.g. Kuhn & Uson 1982) as prominent features of redshift surveys, and are apparent in deep optical images of fields containing massive clusters (e.g. Kodama et al. 2001; Ebeling, Barrett & Donovan 2004). They have a statistically significant presence (Bhavsar & Ling 1988) and become prominent when the galaxy distribution is examined on scales above $2\,h^{-1}$ Mpc (Einasto et al. 1984). However, no entirely algorithmic process has yet been employed to produce a large catalogue of filaments in real data. A range of algorithms has been suggested for the detection of filamentary structure, including minimal spanning trees (Colberg 2007), Delaunay tessellation field estimator (van de Weygaert & Schaap 2009), modeling as a marked point process (Stoica et al. 2005), skeleton (Novikov, Colombi & Doré 2006), multiscale morphology filter (Aragón-Calvo et al. 2007), “DisPerSE” (Sousbie, Pichon & Kawahara 2011) and galaxy axis orientations (Pimbblet 2005). These algorithms have mostly been applied to simulated data. Observationally, the difficulty lies in the limitations of real data (completeness, peculiar velocities and projection effects), the lower density of filaments when compared with clusters, and the possibility that simulated filaments are not accurate analogues of actual filaments. Stoica, Martínez & Saar (2010) suggest that model filaments are shorter than real filaments, and do not form an extended network.

3.1.2.1 Minimal Spanning Tree

Perhaps the first proposed filament finder was the MST (Section 1.1.3.2). Colberg (2007) introduces an algorithm that is largely derivative of the MST. Working with a set of haloes from the Millennium Run (Springel et al. 2005), FoF percolation is applied in which the haloes are treated as particles. The neighbourhood radius used in this step is the same as that required to select most of the haloes in a single group, and haloes which are not part of this group play no further role. This procedure finds haloes that are part of any possible connected structure, as opposed to haloes that are relatively isolated. The connected haloes are assigned to cells in a grid, where the cells are very large since one is only interested in large scale structure. Cells containing at least one halo are then treated as particles to be connected by an MST. This method is effective on simulated data.

3.1.2.2 Candy Model

Stoica et al. (2005) model filamentary structure as a marked point process, a Poisson process in which a variable weight or intensity associated with the process such that the result is not a random field. Instead of assigning a probability density to a count of objects within a cell (where the sum of probabilities for all possible counts is one), each possible position and configuration of a filamentary structure $\mathbf{x}$ is assigned a probability density. The probability density of a given $\mathbf{x}$ is defined with reference to an “energy”, which is a function of configuration and position. A given set of data points admits many possible sets or networks of filaments, and each of these has a total energy. The best network is the one with the least energy (high probabilities give low energies), and will depend on the arbitrary filament model that informs the definition of the probability density. The normalisation of the probability density is also arbitrary and is varied as the "temperature".

3.1.2.3 Skeleton

Novikov, Colombi & Doré (2006) discuss the "skeleton". To make a skeleton for a particle distribution, the distribution is initially smoothed into a surface, with a density ρ for each position in the survey. In this way, the skeleton is a network of lines connecting local peaks, along the ridges that run between them through saddle points, thus tracing borders of void regions. Groups may be identified by identifying regions of the skeleton above some fraction of the mean density. The equation of motion that defines the skeleton is difficult to solve, so Novikov, Colombi & Doré employ a local approximation to the skeleton that works well, and best for data that contain filaments. The numerical implementation involves a number of issues concerning the scale of initial smoothing, pixelisation scale and the identification of saddle points.

3.1.2.4 Multiscale Morphology Filter

Aragón-Calvo et al. (2007) introduce the Multiscale Morphology Filter (MMF), which automates the separation of morphologically-distinct components of structure (clusters, filaments, sheets). MMF is described as a scale space approach, and promoted as a uniform approach to detecting structures of various morphologies. Like the skeleton, MMF requires the density of the field to be defined at all points within the window of observation. To reconstruct the density field, these authors choose the Delaunay Tessellation Field Estimator (DTFE; Section 1.1.3.3) for its ability to smooth without damaging the original structure with arbitrary cell sizes and orientations.

Smoothing is performed on a variety of fixed scales to produce a family of maps. All locations on all scales are checked for the presence of structure. The structure with the best-fit morphology is assigned to each position by examining local density variations, quantified by the eigenvalues of the Hessian matrix. The strength of this technique is its ability to simultaneously search for structures with varying morphologies.

We now introduce a metric called the "elongation probability", which we then use to identify candidate filamentary structures from the MSPM catalogue.

3.2 Elongation Probability

To measure the elongation of a group or cluster's environment, we examine the configuration of nearby groups and clusters by computing their minimal spanning tree (MST; e.g. Barrow, Bhavsar, & Sonoda 1985). An MST is a graph that joins together N input particles with $N-1$ edges such that the total edge length is minimised, without closed circuits. Overdensities may be identified by the removal of long edges (e.g. Bhavsar & Splinter 1996), and the distribution of edge lengths may be used to estimate the Hausdorff dimension (e.g. Martínez et al. 1990). Adjacent edge angles can be used to measure linearity (Krzewina & Saslaw 1996). In our approach, the distribution of angles made by MST edges with a preferred direction is used to calculate an *elongation probability* P_e.

In two dimensions, an MST for an unelongated (isotropic) configuration of structures with n edges can be expected to produce $n/2$ angles less than $\pi/4$ (for example). Using this *expected count* of angles below the threshold in angle, the *actual count* and a Gaussian probability density, we obtain P_e, calculated in the same way as the overdensity probabilities. For each candidate direction sampled, elongation probabilities are calculated as the average over five angular thresholds linearly spaced between $\pi/20$ and $\pi/4$.

Fifteen candidate values of P_e are found under the assumption of each of fifteen directions that are the vectors between six locations:

- the average (group and cluster) position,
- the furthest structure from the average,
- the structure separated from the average by one quarter of the distance between the average and furthest positions,
- the structure separated from the average by half the distance between the average and furthest positions,
- the structure separated from the average by three quarters of the distance between the average and furthest positions, and
- the position of the structure ranked as having as many structures further away as closer to the average structure position.

This sampling of a limited number of directions reduces computational effort and is analogous to our use of galaxy positions as an adaptive sampling grid when identifying groups and clusters. In our implementation we are generating MSTs on group and cluster positions rather than individual galaxies, so our MSTs only have six nodes on average. For these sparse MSTs, the optimal direction will usually align with one of the edges, which will in almost all cases be one of the fifteen directions we sample. However, a sampling of all directions would allow a better estimate of P_e, and should be considered in any future implementation of this approach.

P_e is calculated for each of these directions, and the maximum value is adopted. This maximum P_e is always greater than 0.5, and a P_e distribution for MSTs comprising many nodes typically peaks at ~ 0.75. We consider a distribution to be significantly elongated if $P_e > 0.875$. This is an arbitrary threshold and we have not investigated alternative values. In the current study it demonstrates the utility of coarse-grained mapping approaches such as MSPM for filament finding. Any further work should explore the effect of a P_e threshold on purity and completeness, as well as other measures of elongation such as the inertia tensor (e.g. Ragone-Figueroa et al. 2010).

Table 3.1: Size and purity of filament samples.

$n_{\rm min}$	$F_{\rm algorithm}$	$F_{\rm likely}$	Purity
3	100	53	53%
4	25	19	76%

Numbers of algorithmically identified ($F_{\rm algorithm}$) and likely ($F_{\rm likely}$) filaments contained within, determined by the minimum number of scales $n_{\rm min}$ with elongation probabilities greater than 0.875.

3.3 MSPM Filaments

We can treat the MSPM group and cluster catalogue (Table 2.1) as a coarse-grained representation of the galaxy distribution (Section 1.2.5) with structure sizes of $\lesssim 1h^{-1}$ Mpc, made possible by our range of sampled scales and threshold in $P-S$. This demonstrates a use for MSPM's sensitivity to a user-defined scale range. Treating groups and clusters as particles improves the numerical and computational tractability of large-scale structure studies, suppresses noise contributed by isolated galaxies and reduces the prominence of apparent structures formed by line-of-sight peculiar motions. Similar approaches have previously been adopted by Colberg (2007) and Zhang et al. (2009) on simulated data. Reconstruction of the underlying matter density field from a sample of haloes can also be used to identify and classify features of large-scale structure (Wang et al. 2009; Wang et al. 2012).

Our relatively low threshold for group and cluster identification retains enough information about the galaxy distribution to identify components of filamentary structure. Our elongation probabilities, introduced above, are a measurement we have devised based on minimal spanning trees to identify filaments as elongated unions of groups and clusters.

Elongation probabilities are calculated around each of our groups and clusters, using the positions of neighbouring groups and clusters. An elongation probability is found using structures within each of a set of five radii on the sky: 2 to 10 h^{-1} Mpc in steps of 2 h^{-1} Mpc. Only the sky positions of structures within 10 h^{-1} Mpc in the line of sight are included, meaning that we are less sensitive to filaments with axes aligned close to the line of sight. We are not sensitive to thin bridges between groups and clusters less than ~ 2 h^{-1} Mpc long. An example of a filament traced by MSPM groups and clusters is shown in Figure 3.1.

Filaments are identified as unions of MSPM groups and clusters that are configured such that their elongation probability is greater than 0.875 for a minimum number $n_{\rm min}$ of the five sampled scales. The size (after removal of overlapping volumes) and purity of the resultant filament sample is determined by $n_{\rm min}$ (Table 3.1). Numbers of likely filaments contained by these samples are determined by visual inspection.

If four of the elongation probabilities are required to exceed 0.875, an algorithmically-selected sample of filaments is created with a purity of 76%, but with a sample size of only 25. For our catalogue and filament morphology work, we have used the 53 apparent filaments identified by visual inspection from the $n_{\rm min} = 3$ sample (which includes the $n_{\rm min} = 4$ sample). The 47 fields that remain were judged to be chance alignments of groups and clusters that do not appear on closer visual inspection to be subunits of filamentary structure.

There are many more filaments manifest in the data that we do not detect, so our filament

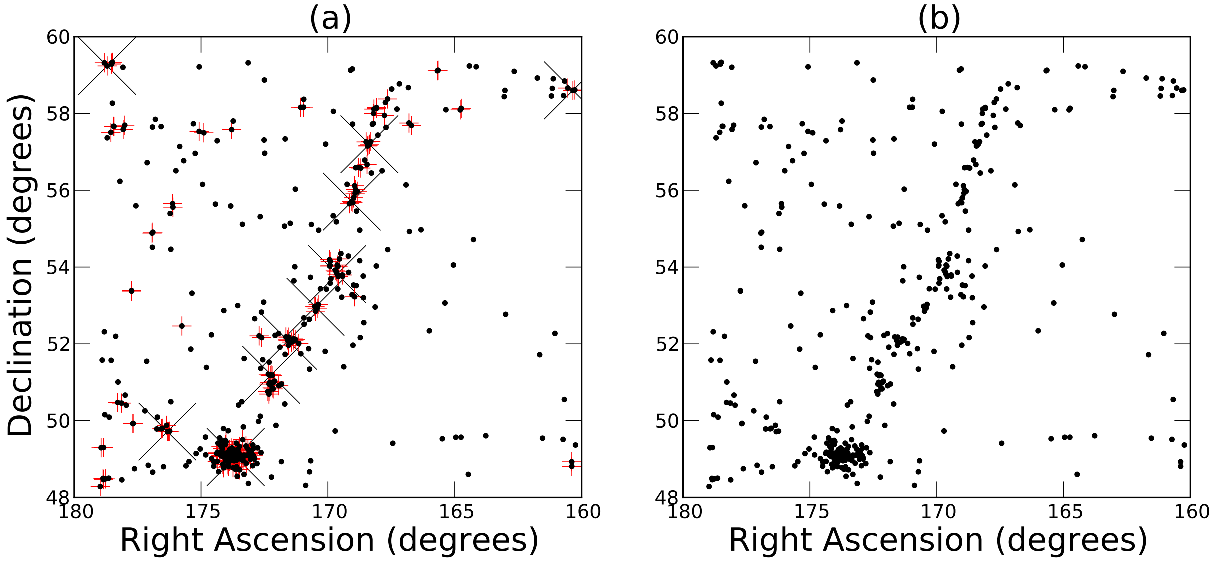


Figure 3.1: A demonstration of our filament detection method, with a filament identified in a field centred on MSPM structure 1063. Filled circles are $r < 17.77$ galaxies, which are also marked by red plusses if they are at positions with $P - S > 0.5$ and large crosses are MSPM groups and clusters. **(a)** Objects within a transverse radius of $10\,h^{-1}$ Mpc at $z = 0.03470$ and within a line-of-sight radius $\Delta z = 0.005$. **(b)** The same field of view, but with plusses and large crosses omitted.

catalogue's completeness is poor. Many filaments may fail our $P > 0.5$ requirement since they are not as dense as clusters. In other cases, the presence of neighbouring structures within 10 h^{-1} Mpc may lower filament elongation probabilities below our threshold.

Our catalogue of 53 filaments is presented in Table 3.2. Each filament is identified by the ID number of the MSPM group or cluster that lies at the centre of the field hosting the filament. No filaments are detected at $z > 0.13$ because of incompleteness.

3.4 Filament Morphologies

Our subjective morphological classification is based on the scheme introduced by PDH and CKC. For each filamentary field, projections of galaxy positions onto two orthogonal planes are inspected. In our work, one of these planes is always the sky, since we have only used sky positions in the calculation of elongation probabilities, favouring filaments that are oriented perpendicular to the line of sight. The other plane is perpendicular to the sky and parallel with the length of the filament. Our inspections are limited to fields with transverse and line-of-sight radii of $10\,h^{-1}$ Mpc that may not always capture the endpoints of the filament. A selection of filaments with different morphologies is shown in Figure 3.2.

In either plane, the projected configuration of galaxies is classified as straight, curved, uniform or irregular.

- **Straight:** galaxies form either a line or lines that are not curved.

Table 3.2: Catalogue of MSPM filaments in SDSS DR7.

ID	RA (J2000) (deg)	Dec (J2000) (deg)	z	P_e(max)	N	Morphology
(1)	(2)	(3)	(4)	(5)	(6)	(7)
68........	222.1034	18.3560	0.03999	0.924	10	V
84........	123.1876	16.8883	0.04459	0.904	10	II
299........	235.6901	8.2411	0.04039	0.894	11	I
404........	165.3420	9.3080	0.03641	0.908	10	I
663........	250.8133	24.0732	0.04696	0.905	10	II
1063........	169.6346	53.8145	0.03470	0.947	7	I
1082........	238.3466	18.3778	0.03280	0.910	15	V
1088........	149.5960	8.9077	0.04900	0.921	5	I
1494........	196.6982	60.3546	0.02784	0.895	6	I
1511........	203.9377	27.8676	0.02641	0.890	8	V
....	....	....	....	....	....	....
2547........	119.4981	40.0377	0.06630	0.908	7	I
3094........	162.2009	4.3068	0.06952	0.889	9	V
3750........	233.4721	6.8433	0.06580	0.878	3	II
4647........	175.4488	56.7553	0.09737	0.878	3	V
5149........	210.5606	5.9266	0.07834	0.878	6	V
5832........	228.2023	20.8819	0.07962	0.920	8	II
5984........	183.5435	17.7961	0.07697	0.905	4	I
6574........	149.8899	3.1282	0.08179	0.936	6	II
6846........	227.7999	5.6538	0.08455	0.905	4	I
7642........	139.4112	36.5945	0.11032	0.878	3	V

Locations, properties and morphologies of MSPM filaments. This table shows only a portion of our catalogue as an indication of its content. The complete catalogue can be found in Appendix A, Smith et al. (2012), or at http://www.physics.usyd.edu.au/sifa/Main/MSPM/ , along with three-dimensional visualisations and figures showing each filament.

Column (1): ID of central MSPM group or cluster; (2) to (4): position; (5): highest elongation probability within 10 h^{-1} Mpc; (6): count of groups and clusters within a 10 h^{-1} Mpc radius on the sky and 10 h^{-1} Mpc in the line of sight (cylindrical aperture); (7): morphological type (Section 3.4).

Table 3.3: Filament numbers and fractions by morphology.

Type	This Study (number)	This Study (per cent)	PDH (per cent)
I	26	49 ± 10	37 ± 3
II	16	30 ± 8	34 ± 3
III	0	0	4 ± 1
IV	0	0	0.8 ± 0.5
V	11	21 ± 6	26 ± 3

The abundance of each filament type in our study compared with PDH. All uncertainties are Poissonian.

- **Curved:** a line that is *continuously* bent (not simply crooked) into either a "C"- or "S"-shape.
- **Uniform:** uniformly distributed galaxies that do not form a clear line.
- **Irregular:** irregular distribution of galaxies containing large density fluctuations that obscure the linear structure of the filament.

Following PDH, each filament is assigned a morphological type based on its appearance in two orthogonal planes.

- **Type I (straight):** both are straight (e.g., Figures 3.2a, d, g).
- **Type II (warped):** at least one is curved, with neither being uniform or irregular (e.g., Figures 3.2b, e, h).
- **Type III (sheet):** one (and only one) is uniform, with the other being either straight or curved.
- **Type IV (uniform):** both are uniform.
- **Type V (irregular):** both are irregular (e.g., Figures 3.2c, f, i).

Some fields contain multiple filaments, and in these fields we classify the filament containing the groups or clusters that cause a high elongation probability. Fields are inspected independently by AGS and KAP[1] with a 4 per cent disagreement rate.

The division of our filament sample by morphology is shown in Table 3.3, and a selection of filaments with different morphologies is shown in Figure 3.2. Our results regarding the relative abundance of filament types are consistent with PDH and CKC within uncertainties, finding that most of our filaments are Type I (straight) or II (curved), with the remainder classified as Type V (irregular). The PDH sample is derived by visual inspection, so our technique is sensitive to the prominent filament morphologies apparent in real data.

Our Type I fraction of 49 per cent is marginally higher than that found by PDH (37 per cent), who assess volumes that allow greater filament curvature. PDH record filaments up to a

[1]Kevin A. Pimbblet

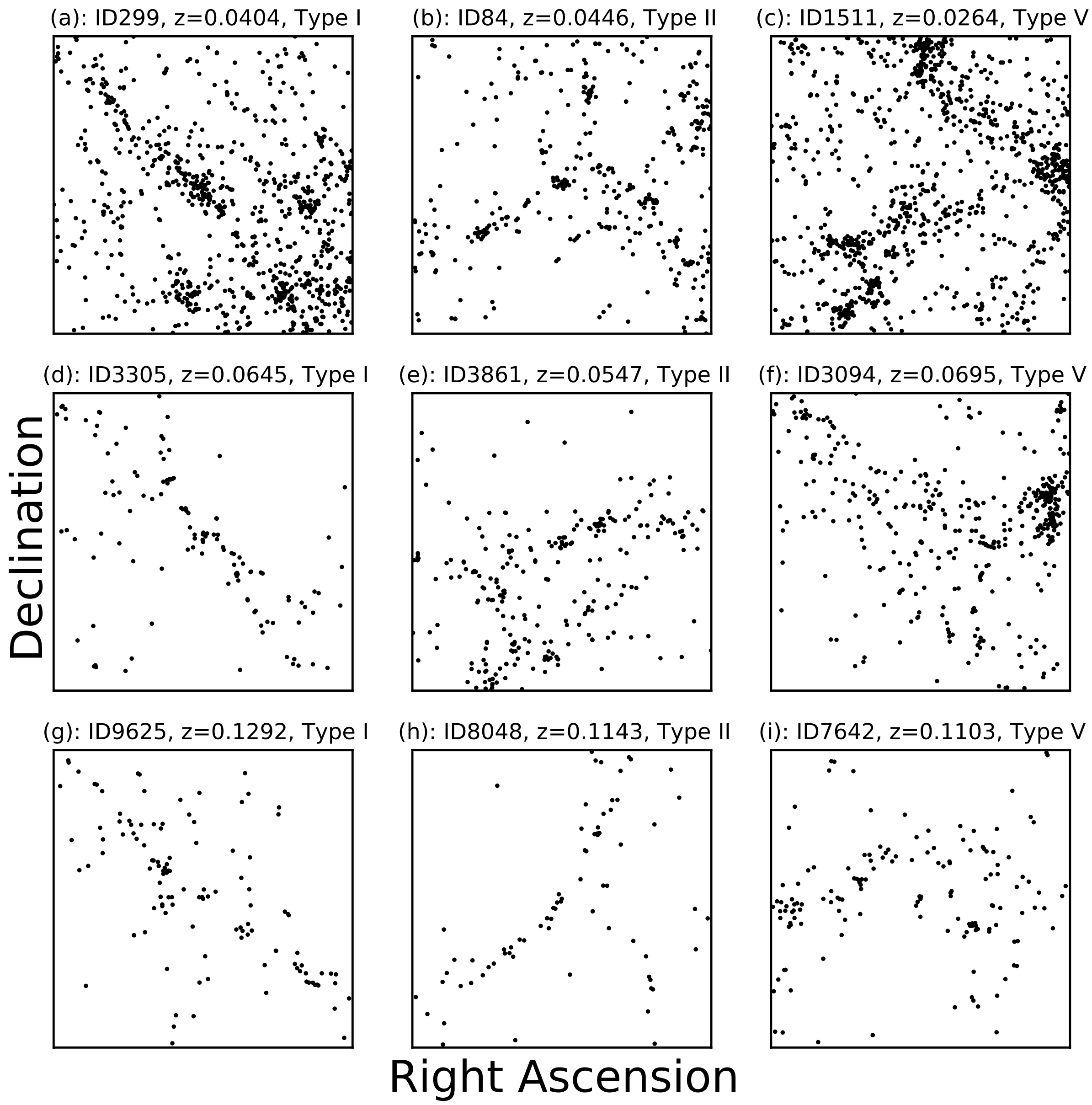


Figure 3.2: Selected fields containing filaments, centred on MSPM groups and clusters. Each panel shows $r < 17.77$ galaxies within a $20\,h^{-1}$ Mpc $\times$ $20\,h^{-1}$ Mpc square on the sky and within a line-of-sight radius $\Delta z = 0.005$ ($\approx 14\,h^{-1}$ Mpc). **(a)-(c)** Examples of types I, II and V at $z < 0.05$. **(d)-(f)** Types I, II and V at $z \gtrsim 0.05$. Field 3861 shows an example of "S-shaped" curvature. **(g)-(i)** Types I, II and V at $z > 0.1$. Figures showing all 53 filaments can be found in the online edition of Smith et al. (2012) or at http://www.physics.usyd.edu.au/sifa/Main/MSPM/ , along with three-dimensional visualisations.

length of $\approx 45\,h^{-1}$ Mpc, and find that Type I is the dominant morphology for short ($< 10\,h^{-1}$ Mpc) filaments. Since our search is restricted to fields with radii of $10\,h^{-1}$ Mpc, we do not detect filaments (or filament segments) longer than $20\,h^{-1}$ Mpc. Our elongation probability approach is also more efficient at detecting straight filaments than filaments with more complex morphologies.

Awake To Emptiness

It was disturbing, but there was nothing to be done. He too was in some kind of posthumous existence, a very hungry ghost indeed. Living on from one found bite to the next, with no name or fellows, he began to close in on himself, as during the hardest campaigns on the steppes, becoming more and more an animal, his mind shrinking in like the horns of a touched snail. For many watches at a time he thought little but the Heart Sutra. Form is emptiness, emptiness form. Not for nothing had he been named Sun Wu-Kong, Awake to Emptiness, in an earlier incarnation. Monkey in the void.

At the moment of death Kyu saw the clear white light. It was everywhere, it bathed the void in itself, and he was part of it, and sang it out into the void.

Some eternity later he thought: This is what you strive for.

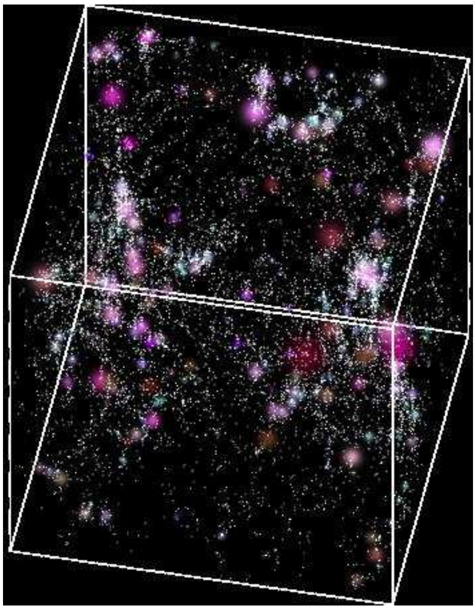
And so he fell out of it, into awareness of himself. His thoughts were continuing in their tumbling monologue reverie, even after death. Incredible but true. Perhaps he wasn't dead yet. But there was his body, hacked to pieces on the sand of the Forbidden City.

Jati, subcaste, family, village. It manifests differently. We all came into the cosmos together. New souls are born out of the void, but infrequently, for we are in the Kali-yuga, the Age of Destruction. When new souls do appear it happens like a dandelion pod, souls like seeds, floating away on the dharma wind. We are all seeds of what we could be. But the new seeds float together and never separate by much, that's my point. We have gone through many lives together already. Our jati has been particularly tight since the avalanche. That fate bound us together. We rise or fall together.

– Kim Stanley Robinson, *The Years of Rice and Salt*. Book One: Awake to Emptiness

Chapter 4

Voids



A solid particle render (Appendix B) of the nearest MSPM void, at lattice position 16110. The cube has a side length of 80 h^{-1} Mpc.

In this Chapter we identify 47 voids as 40 h^{-1} Mpc spherical volumes that contain no MSPM groups or clusters (Section 4.2), demonstrating a unified approach to the identification of underdense and overdense features of large-scale structure. This size is within the range of characteristic void sizes found in more detailed studies. The density contrast of bound structures is found to provide an estimate of the void fraction that is unbiased with respect to redshift, where filaments are resolved in the same magnitude-limited data. For our assumed void definition, the void volume fraction at the current epoch is $f_{\mathrm{v0}} = 0.75 \pm 0.11$ (Section 4.3).

4.1 Review: Sunless Seas

As with overdense structures, there is no universal void definition, and various void definitions motivate different strategies for finding them.

4.1.1 Significance

Voids fill most of the current epoch Universe by volume, with their sizes and shapes presenting important tests of numerical simulations of cosmic structure growth. The fraction of the Universe occupied by voids, the void volume fraction, is a test any viable cosmology must pass. ΛCDM correctly predicts the void fraction but may underestimate the number of void galaxies (Pan et al. 2011). Because of their extremely low density, the shapes of void regions are tidally distorted by surrounding matter. Simulated voids are aspherical (Shandarin, Sheth & Sahni 2004), and Park & Lee (2007) have shown that the distribution of void ellipticities is sensitive to cosmological parameters. In a ΛCDM cosmology, voids are more spherical given a higher dark energy density Ω_Λ, as a lower matter density ($\Omega_{\mathrm{M}} = 1 - \Omega_\Lambda$) reduces tidal effects.

Simulated voids are not completely empty, but are populated by many low-mass haloes. These haloes exhibit filaments and subvoids reminiscent of their higher-mass counterparts (Gottlöber et al. 2003). However, the expected abundance of dwarf galaxies expected to reside in these haloes has not been found in observations. Tikhonov & Klypin (2009) find that in the Local Volume, ΛCDM overpredicts the number of dwarf galaxies by a factor of ten, emphasising that ‘disagreement with the theory is staggering’. The Universe is apparently far emptier than it should be, in the low-mass regime (see also Section 5.1.2.3).

Void studies, both theoretical and observational, are required to resolve the emptiness issue and more general questions. The number of dwarf galaxies in voids is sensitive to the physics of galaxy formation and star formation in low-density environments. For example, Tinker & Conroy (2009) suggest a higher mass-to-light ratio for faint galaxies as an explanation.

4.1.2 Void Finders

4.1.2.1 Cubes and Spheres

Kauffmann & Fairall (1991) divide a volume into cubes, and find all cubes containing no galaxies. Voids are identified as cubical aggregations of these empty cubes. To better approximate void volumes, voids are extended outward by the addition of empty cubes to the initial "base voids" such that finger-like extensions of voids are avoided. One problem with this approach is that the identification of a void may be thwarted by the presence of a single galaxy, and different results will be obtained depending on the orientation of the cube axes. Kauffmann & Fairall assess the significance of their voids by comparison with random fields and publish a catalogue of 129 voids out to $z = 0.05$.

Patiri et al. (2006) define voids as the largest spheres containing no galaxies above a certain brightness or mass, and suggest two algorithms. One of these is based on distances between cells that are empty of galaxies below a certain brightness or mass. For a given void comprising many such cells, the centre of the void is found as the centre of the largest sphere that contains all of the empty cells. The other algorithm checks within trial spheres for emptiness. An empty sphere is then "grown" until its surface meets objects. This is then a potential maximal sphere, and then there is a search in close proximity for a larger sphere that might fit in the void. If a larger sphere is found that overlaps the first one, that first sphere is rejected. These techniques are applied to the 2dFGRS.

4.1.2.2 Wall Exclusion

El-Ad & Piran (1997) define a void as a contiguous region nowhere thinner than a certain diameter and free of wall galaxies. The first step is to find walls so that they may be excluded. Wall galaxies are defined as those with at least n galaxies within a radius defined with reference to the average nth nearest neighbour distance, the standard deviation thereof and an arbitrary constant. Voids are then found by identifying spherical regions containing no wall galaxies. After starting with large spheres, smaller spheres are added to compose contiguous voids of any shape. For spheres that are small enough, all the voids within the volume may join to become one large void. To avoid this problem, voids are not connected if they overlap above a defined threshold.

El-Ad & Piran (1997) find that large voids fill 50 per cent of the available space and that the sparse population of field galaxies within them tends to be faint. This approach is used by Hoyle & Vogeley (2004) to find 289 voids in the 2dFGRS, and by Pan et al. 2011 to find 1054 voids in the SDSS.

4.1.2.3 Subvoids and Merging

Aikio & Maehoenen (1998) define a distance field DF for all points within the survey volume. For any point **x**, $D(\mathbf{x})$ is the distance to the closest galaxy, which is equivalent to nth nearest neighbour distance with $n = 1$. DF is a scalar field, and its local maxima define the centres of voids. Defining field lines with direction ∇D in a manner reminiscent of the skeleton and its dual (Section 3.1.2), the size of a void associated with a local maximum is then the union of all positions from which field lines pass through that maximum. For almost every position within a survey volume, the local maximum is unique. Each position within a void must also be above an arbitrary distance from any galaxies. Aikio & Maehoenen recognise, however, that a void may contain more than one peak of the DF, and two voids are merged if the distance separating their maxima is less than or equal to both of the DF values at those maxima.

Colberg et al. (2005b) use a variant of Aikio & Maehoenen's technique, which uses density instead of nearest neighbour distance. The density field is produced by multiscale smoothing with a Gaussian filter such that the Gaussian width is varied according to the local density. Detail is lost in heavily-clustered regions but the large-scale void structure is preserved. Preliminary voids are identified in a way similar to Aikio & Maehoenen except that the nearest neighbour criterion for voidship is replaced by a maximum density threshold, and the procedure by which voids are merged is more elaborate.

4.1.2.4 Gaussian Smoothing

Just as overdensities can be found by peak-finding on a Gaussian-smoothed map, underdensities can be similarly identified as troughs in such a map. Plionis & Basilakos (2002) follow this approach to locate 22 voids out to $130h^{-1}$ Mpc.

4.1.2.5 Watershed

Platen, van de Weygaert & Jones (2007) present the watershed void finder (WVF), a technique named for a geophysical analogy. If a landscape containing hills and valleys is being flooded with water, the water collects in puddles, then pools, and then the valleys become basins. The water in two adjacent valleys will be separated by a ridge. If the ridges are built up by dams such that the bodies of water don't merge as the water level rises, they preserve a segmentation of the landscape in which each segment contains a basin. Astrophysically, the density resconstructed by the Delaunay tessellation field estimator (DTFE) plays the role of height and voids play the role of basins, with filaments and walls as the ridges. The level at which bodies of water are no longer merged but are separated by dams is arbitrary. For the purpose of finding voids, Platen et al. (2011) find that the DTFE outperforms higher-order reconstruction techniques.

4.2 MSPM Voids

Voids have been identified by treating MSPM groups and clusters (Smith et al. 2012; Chapter 2) as particles and searching for underdense regions in the SDSS volume. Filaments are resolved in the SDSS spectroscopic data up to $z \sim 0.125$ (Table 3.2). To prevent the identification of voids as a result of incompleteness at higher redshifts, we restrict our search to data in which filaments are resolved, at $z < 0.1$. Since the MSPM filaments are identified with lengths up to 20 h^{-1} Mpc, features of large-scale structure with this diameter should be resolved at $z < 0.1$.

The SDSS spectroscopic plate centres provide a convenient set of sampling locations from which to measure local densities. Since we only require that the survey volume is sampled with a minimum resolution, oversampling that may result from adjustment of plate centres for the presence of large-scale structure is not expected to bias our results. Each plate has a radius of 1.5°, which is 1.95 h^{-1} Mpc at $z = 0.025$ and 7.67 h^{-1} Mpc at $z = 0.1$. These sky positions

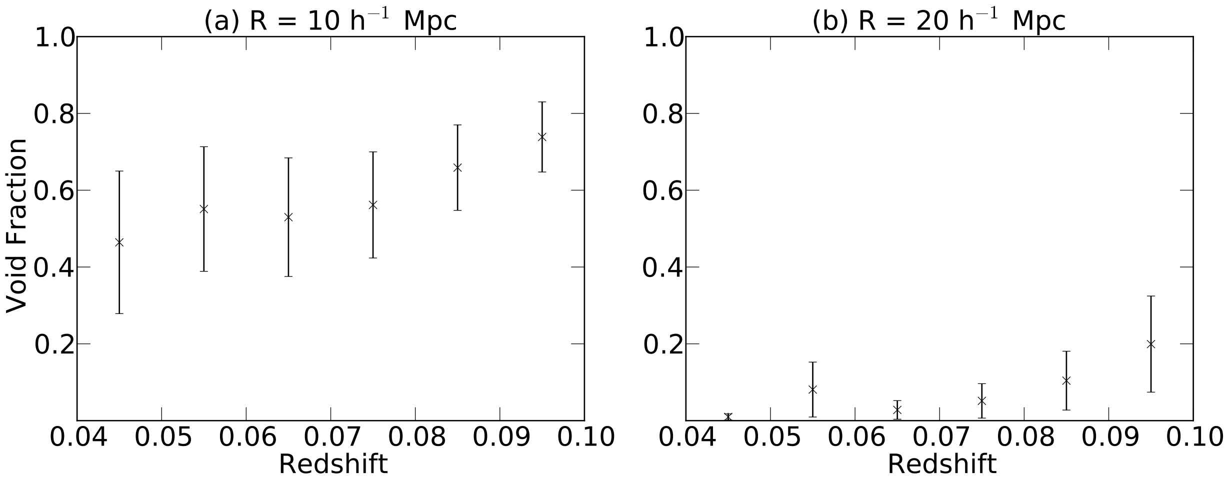


Figure 4.1: The volume occupied by voids that contain no groups or clusters at $0.04 < z < 0.10$. Lower bounds are fractions of lattice positions that have no groups or clusters within a distance R. Upper bounds are fractions of lattice positions within spheres of radius R that contain no groups or clusters. **(a)** $R = 10$ h^{-1} Mpc. **(b)** $R = 20$ h^{-1} Mpc.

are combined with sampling in redshift space with an interval of $\Delta z = 0.0025$ (~ 7 h^{-1} Mpc at $z < 0.1$) to produce a three-dimensional sampling lattice. The resolution of this lattice allows a search for voids as small as 20 h^{-1} Mpc across, coinciding with the resolution-scale suggested by the identification of filaments.

To be considered as the centre of a void of radius R h^{-1} Mpc, a plate centre must be at least $2R$ h^{-1} Mpc from the survey edges and $z = 0.025$, so that each void can be compared with its environment. This restricts the volume within which voids can be found. Voids 20 h^{-1} Mpc across cannot be found closer than $z = 0.03$. Voids 40 h^{-1} Mpc across cannot be found closer than $z = 0.035$, and can be found in less than half of the survey area at $z < 0.07$.

Defining a void as a spherical volume containing no MSPM groups or clusters, there are two voids that are 60 h^{-1} Mpc across, 47 that are 40 h^{-1} Mpc across and 1150 that are 20 h^{-1} Mpc across. Our void definition is simple and detects a greater density of voids as the apparent abundance of groups and clusters falls with increasing redshift, as shown in Figure 4.1. However, in the current study it is sufficient to demonstrate the utility of a coarse-graining approach to the identification of voids. The positions of our 47 voids that are 40 h^{-1} Mpc across are given in Table 4.1.

The closest ($z = 0.045$) of the 40 h^{-1} Mpc voids is shown in Figure 4.2 and further examples are shown in Figure 4.3. Our void candidates are underdense relative to surrounding locations, and in many cases the walls of the voids, threaded by filaments, are clearly evident (e.g. Figure 4.3b, d, f, i). A visual inspection of the smoothed three-dimensional galaxy distribution[1] in their environments confirms their reality, implying a low false discovery rate.

In the current study we have not addressed selection effects. Our approach to finding voids is single-scale, but $R = 20$ h^{-1} Mpc is within the range of characteristic void sizes found in

[1]http://www.physics.usyd.edu.au/sifa/Main/MSPM/

Table 4.1: Catalogue of 40 h^{-1} Mpc MSPM voids in SDSS DR7.

ID (1)	RA (J2000) (deg) (2)	Dec (J2000) (deg) (3)	z (4)
16110........	197.3890	19.6361	0.04500
22406........	215.3630	45.8330	0.05500
26786........	142.3790	17.0212	0.06000
30110........	154.7770	31.1830	0.06500
33619........	217.8010	8.4892	0.07000
33952........	135.9970	22.2747	0.07000
35795........	158.1730	22.6359	0.07250
36559........	186.5200	52.4890	0.07500
38881........	202.6630	51.1090	0.07750
39279........	234.7750	21.0329	0.07750
....	....	....	....
42939........	200.9950	26.4868	0.08250
46586........	154.2300	24.1352	0.08750
47835........	139.6980	32.7990	0.09000
48554........	186.2570	22.5313	0.09000
49316........	211.0950	52.1140	0.09250
50041........	208.7370	30.0585	0.09250
53043........	223.7300	42.8980	0.09750
54550........	195.7990	4.8680	0.10000
55716........	142.7850	15.1658	0.10000

Locations of MSPM voids that are 40 h^{-1} Mpc across. This table shows only a portion of our catalogue as an indication of its content. The complete catalogue can be found in Appendix A, or at http://www.physics.usyd.edu.au/sifa/Main/MSPM/ , along with three-dimensional visualisations of each void. Column (1): ID of position in sampling lattice; (2) to (4): centre position.

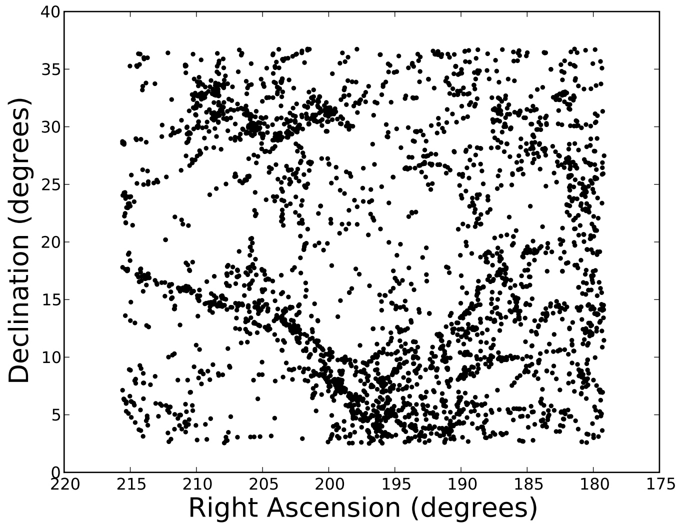


Figure 4.2: This field has a side length of 80 h^{-1} Mpc and shows a redshift slice of width $\Delta z = 0.01$ through MSPM void 16110 at $z = 0.045$. Dots are $r < 17.77$ galaxies from the SDSS spectroscopic survey. There are no MSPM structures in a sphere of diameter 40 h^{-1} Mpc at the centre of this field. We have identified 47 such non-overlapping volumes of this size.

more detailed studies (e.g. Pan et al. 2011).

4.3 The Void Volume Fraction

The void definition in Section 4.2 provides a measure of the void fraction that is sensitive to the mean density of groups and clusters. Since the MSPM group and cluster catalogue is magnitude-limited, the void fraction rises with redshift (Figure 4.1). A more robust, and more physically-meaningful, measure of the void fraction can be derived from comparison with the mean density of groups and clusters $\bar{\rho}_{\rm gc}$ at each redshift. Any measure of the void fraction requires a void definition, and any void definition divides the volume of the Universe into "void" and "wall" (filament and cluster) regions such that

$$f_{\rm v} + f_{\rm w} = 1, \tag{4.1}$$

where f_v and f_w are the void and wall volume fractions, respectively. We define voids as volumes with less than half the mean density of groups and clusters.

The measurement of the void fraction following this definition remains ambiguous, however. If the volume contained by voids of radius R is defined as the union of all regions within spheres of radius R that have densities below the threshold, an equivalent definition of the wall fraction leads to $f_v + f_w > 1$. This is because a position may be simultaneously located within spheres that are above and below the threshold. Hence, this measurement of the void fraction provides an upper bound. A lower bound is the fraction of positions at the centres of spheres with less than the mean density.

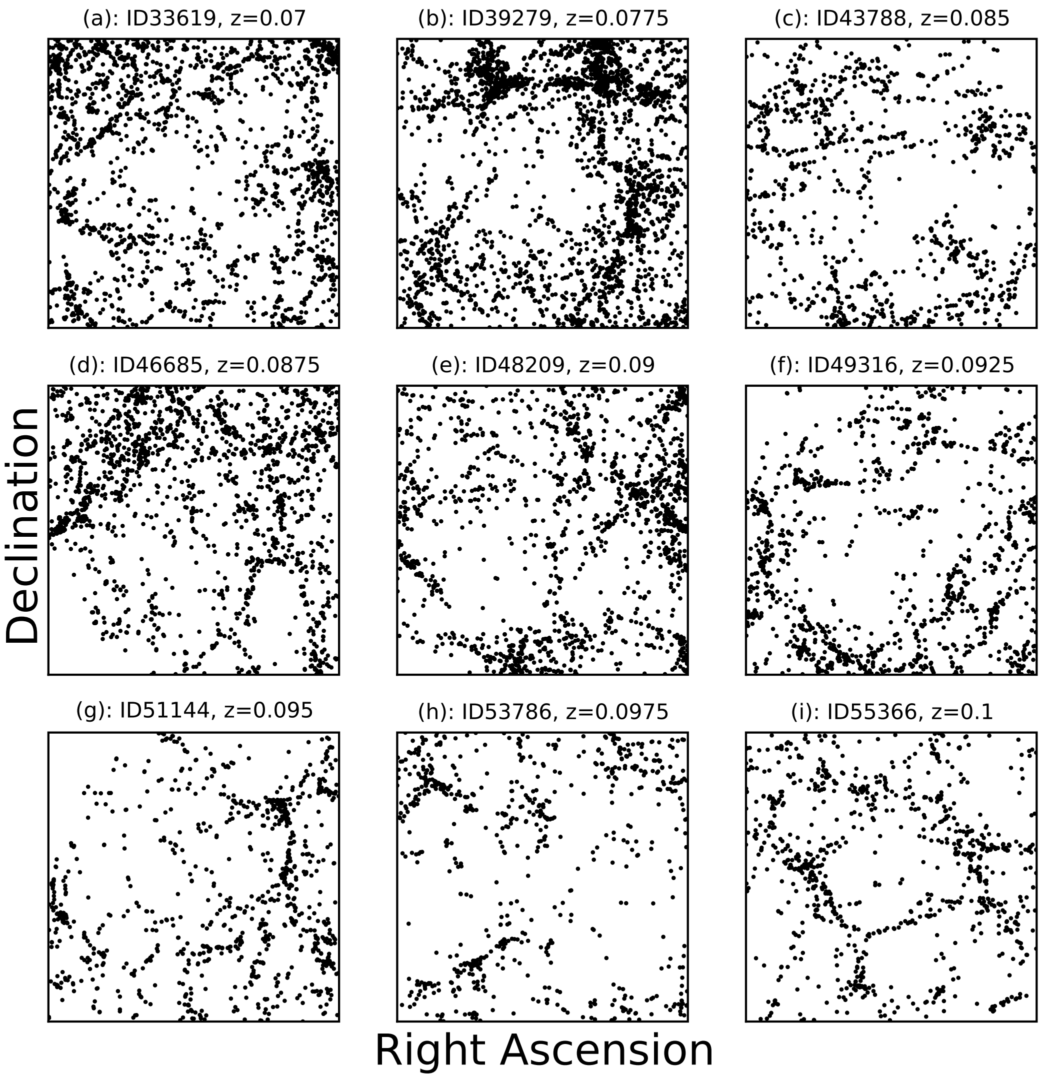


Figure 4.3: Randomly-selected fields containing voids. Each panel is centered on a sphere of diameter 40 h^{-1} Mpc that is empty of MSPM groups and clusters. Dots are $r < 17.77$ galaxies within an $80\,h^{-1}$ Mpc $\times$ $80\,h^{-1}$ Mpc square on the sky and within a line-of-sight radius $\Delta z = 0.005$ ($\approx 14\,h^{-1}$ Mpc). Figures showing all 47 voids can be found at http://www.physics.usyd.edu.au/sifa/Main/MSPM/ , along with three-dimensional visualisations.

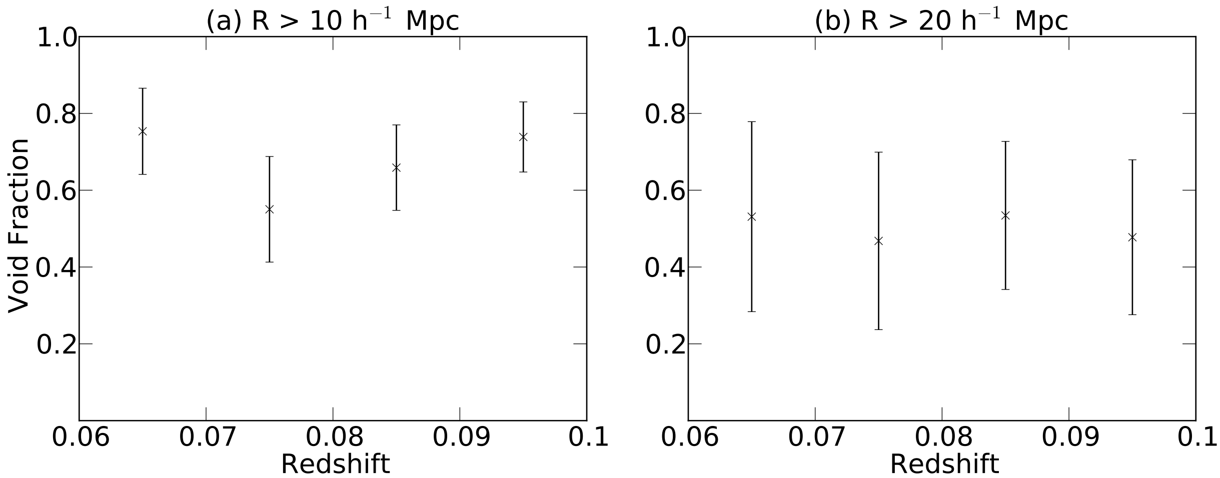


Figure 4.4: The void fraction at $0.06 < z < 0.10$, for voids with $\rho_{\rm gc} < 0.5\bar{\rho}_{\rm gc}$. **(a)** $R > 10\ h^{-1}$ Mpc. **(b)** $R > 20\ h^{-1}$ Mpc.

To remove edge effects, only volumes greater than twice the void radius from the survey edges are considered. The effect of sample (cosmic) variance must also be considered. The void fraction should be measured in a volume that encompasses the scale of homogeneity. To prevent large variance in the distances between adjacent lattice sampling positions (so that they represent equal volumes) within each redshift slice, the void fraction is determined within redshift slices no thinner than $\Delta z = 0.01$. For $R = 10\ h^{-1}$ Mpc voids, the volume available in slices of this width after exclusion of edge regions does not exceed the volume required by the scale of homogeneity found by Hogg et al. (2005) at $z < 0.06$, so we exclude these data.

Our measurements of the void fraction within redshift slices at $0.06 < z < 0.10$ are shown in Figure 4.4. The data for $R > 20\ h^{-1}$ Mpc voids do not show evidence of systematic variation with redshift, while the data for $R > 10\ h^{-1}$ Mpc voids do, as discussed below. This suggests that in a magnitude-limited survey, the density contrast of bound structures has the potential to provide an estimate of the void fraction that is unbiased with respect to redshift, where filaments are resolved in the same data. Figure 4.4 also shows that the void fraction increases when smaller voids are included, as expected.

For a final measurement of the void fraction, voids at least 20 h^{-1} Mpc across ($R > 10\ h^{-1}$ Mpc) are used to achieve the greatest accuracy. This matches the finest possible resolution suggested by our sampling lattice and detection of filaments, discussed in Section 4.2. At $z > 0.07$, the mean number of groups and clusters per $R = 10\ h^{-1}$ Mpc sphere is less than two and, because of our void definition $\rho_{\rm gc} < 0.5\bar{\rho}_{\rm gc}$, void regions identified at this depth contain no groups or clusters. This introduces a discreteness effect apparent in Figure 4.4, causing an underestimation of the void fraction at $0.07 < z < 0.08$ because the threshold is effectively rounded down to an integral number (zero) of groups and clusters per sampling sphere. The void fraction then increases with redshift, as the true density of groups and clusters approaches zero. To avoid this discreteness effect, we only use the $0.06 < z < 0.07$ slice and obtain for the current epoch void volume fraction:

$$f_{v0} = 0.75 \pm 0.11. \tag{4.2}$$

This result is derived from voids with $R > 10\ h^{-1}$ Mpc, and would increase with the inclusion of smaller voids. The mean numbers of visible groups and clusters in $R = 10\ h^{-1}$ Mpc spheres centred on MSPM filaments (Table 3.2) at $0.06 < z < 0.07$ and $0.09 < z < 0.10$ are 4 and 2.7 respectively.

Hoyle & Vogeley (2004) measure the void fraction for voids with $R > 10\ h^{-1}$ Mpc to be 0.4. However, their void definition is based on the identification of void galaxies that have fewer than three neighbours within a radius 5.6 h^{-1} Mpc. This implies a more extreme underdensity than our definition, in which voids may contain groups or clusters as long as their density is lower than half the mean.

In SDSS DR7, Pan et al. (2011) measure a void fraction of 0.62, derived from $R > 10\ h^{-1}$ Mpc voids. This value is again lower than ours, but Pan et al. define voids with a threshold in galaxy density rather than group and cluster density, of $\rho < 0.53\bar{\rho}$. Although our threshold in density contrast is almost identical, our use of groups and clusters, rather than galaxies, as mass tracers lowers the apparent density of voids, making them more extreme. Given the differences in void definition between various studies, the value of f_{v0} we have found is reasonable.

Comparison with model predictions is difficult without the application of direct tests to mock catalogues with the same void-selection threshold. However, the ΛCDM mock catalogue described by Pan et al. has a void fraction of 0.665, consistent with our value. Our measurement of the void fraction also accords with the timescape cosmology's prediction, $f_{v0} = 0.76^{+0.12}_{-0.09}$ (Leith, Ng & Wiltshire 2008). However, for the purpose of the timescape model (explored in more detail in the next Chapter), voids are defined as volumes outside disjoint "finite infinity" regions that contain gravitationally-bound structures (bound beyond the virial radius; Wiltshire 2007a), and this is not how our voids are defined. Our voids can be expected to lead to a similar void fraction if they are empty of gravitationally-bound structures, and if remaining wall volumes are dominated by finite infinity regions. The former is probably true because our voids do not contain filaments. However, because of incompleteness in our group and cluster sample it is not possible to properly explore these possibilities.

The Eternal Now

He had boarded the kettle in the 575th Century, the base of operations assigned him two years earlier. At the time the 575th had been the farthest upwhen he had ever travelled. Now he was moving upwhen to the 2456th Century.

Under ordinary circumstances he might have felt a little lost at the prospect. His native Century was in the far downwhen, the 95th Century, to be exact.

He heard her voice like a drifting wind. "Oh, you Eternals. You are so secretive. You won't share at all. Make me an Eternal."

Her voice was a sound now that didn't coalesce into separate words, just a delicately modulated sound that insinuated itself into his mind.

He must *have her, and* now. *Before any Reality Change. What was it Finge had said to him, jeering:* The now does not last, even in Eternity.

Doesn't it, though? Doesn't it?

Harlan had known exactly what he must do. Finge's angry taunting had goaded him into a frame of mind where he was ready for crime and Finge's final sneer had, at least, inspired him with the nature of the deed he must commit.

He had not wasted a moment after that. It was with excitement and even joy that he left his quarters, at all but a run, to commit a major crime against Eternity.

"Sennor," he said, "was the doubter. We reasoned with him and argued. We used mathematics and presented the results of generations of research that had preceded us in the physiotime of Eternity. He put it all to one side and presented his case by quoting the man-meets-himself paradox. You heard him talk about it. It's his favourite.

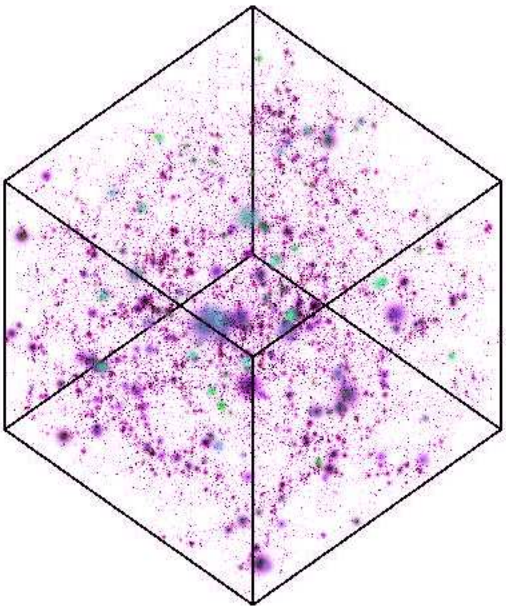
"We knew our own future, Sennor said. I, Twissell, knew, for instance, that I would survive, despite the fact that I would be quite old, until Cooper made his trip past the downwhen terminus. I knew other details of my future, the things I would do.

"Impossible, he would say. Reality must change to correct your knowledge, even if it meant the circle would never close and Eternity never be established.

– Isaac Asimov, *The End of Eternity*

Chapter 5

The Critical Density and a Test of the Timescape Cosmology



A cube of side 160 h^{-1} Mpc at $0.06 < z < 1.12$ encompassing the scale of homogeneity, showing groups and clusters from the MSPM catalogue. This is the same image as shown on the title page of this Thesis but with colours reversed and green de-emphasised: the Universe may look very different to an observer in a void.

In this Chapter we propose a test to distinguish the timescape cosmology from ΛCDM, based on the density of gravitationally-bound structures relative to the critical density as measured from the radius-velocity dispersion relation (Section 5.2.2). An initial calculation is attempted (Section 5.4), but significant systematic issues remain unresolved (Section 5.5). These issues include the measurement of virial radii and the appropriate f_σ normalisation for our sample. If these are addressed, meaningful constraints may be placed on the current epoch matter density $\Omega_{\rm M0}$.

Zeroes in subscripts (e.g. H_0) denote current epoch quantities.

5.1 A Tale of Two Models

5.1.1 ΛCDM

The current standard model for the Universe is spatially flat, containing dark energy (Λ), cold dark matter (CDM) and baryonic matter. The matter distribution is homogeneous on large scales, with current-epoch large-scale structure growing via gravity from Gaussian primordial density perturbations. Homogeneity of the Universe is evidenced by the isotropy of the cosmic microwave background (CMB), in which structure is manifest with $\Delta T/T \sim 10^{-5}$ (Smoot et al. 1992), and surveys of large-scale structure in the present Universe reinforce homogeneity on large scales (e.g. Hogg et al. 2005). This implies, to a reasonable approximation at least, that the Universe may be described by a Friedmann-Lemaître-Robertson-Walker (FLRW) geometry, with line element

$$ds^2 = -dt^2 + a^2(t)d\mathcal{X}^2, \tag{5.1}$$

where $a(t)$ is a scale factor, t is time and $\mathcal{X}$ is (comoving) position in 3-space. Our Universe appears to be spatially flat (e.g. by CMB measurements, Spergel et al. 2003) with $d\mathcal{X}^2 = dx^2 + dy^2 + dz^2$ for flatness in slices with constant t (but *spacetime* is not flat). The time evolution of the scale factor is given by the Friedmann equation (that derives from the Einstein equation for an FLRW geometry) with $k = 0$ for the flat case, and depends on the amount and form of mass-energy contained by the Universe.

The observation that light from galaxies is redshifted more at greater distances implies that the Universe is expanding and that the scale factor is increasing. The rate of expansion is quantified by the Hubble parameter $H = \dot{a}/a$ with $\dot{a} = da/dt$, which in the current epoch takes the value $H_0 = H(t_0)$. In the nearby Universe, H_0 is usually measured from a distance-redshift relation. Type Ia supernovae (SNeIa) have peak luminosities that correlate with their decay times, allowing their distances to be measured, following calibration from nearby Cepheid variable stars. SNeIa are fainter and therefore apparently more distant than expected by matter-dominated FLRW models (Riess et al. 1998; Perlmutter et al. 1999; Riess et al. 2001), implying the presence of *dark energy* that exerts a negative pressure, accelerating the expansion of the Universe.

Dark energy is poorly understood, eluding physical description and posing a foundational mystery (Frieman, Turner & Huterer 2008). However, its density Ω_Λ is a central parameter in the ΛCDM model. ΛCDM is also described by other parameters including the scalar spectral index to describe the primordial density perturbations, baryonic matter density and dark matter density, and is a good fit to CMB data (Bennett et al. 2003; Komatsu et al. 2011). ΛCDM also passes tests presented by galaxy clustering statistics (Reid et al. 2010) and the mass function of massive clusters (Vikhlinin et al. 2009; Mantz et al. 2010).

Non-standard cosmologies that do not include dark energy, while retaining an FLRW geometry, have been proposed but are disfavoured by current observations (e.g. Voit 2005). The standard cold dark matter (SCDM) model with $\Omega_{\mathrm{M0}} = 1$ implies a baryon-to-dark matter ratio that conflicts with cluster measurements (e.g. Allen, Schmidt & Fabian 2002). $\Omega_{\mathrm{M0}} = 1$ models with a "tilted" primordial density perturbation spectrum (e.g. Cen et al. 1992) are more consistent but imply values of the scalar spectral index that are lower than inferred from the CMB. Other non-standard cosmologies question the assumption of an FLRW geometry (discussed further below) and attribute cosmic acceleration to "backreaction" that results from the presence of structure in the Universe (e.g. Buchert 2008), but this effect does not appear significant enough to exclude dark energy (Ishibashi & Wald 2006). In the next Section we summarise another inhomogeneous model, the timescape cosmology, that includes backreaction but also an additional mechanism to account for cosmic acceleration, based on a reinterpretation of General Relativity. We examine this model because it is one of the few remaining observationally-viable alternatives to ΛCDM, because like much of this Thesis it is motivated by the presence of large-scale structure, and because it may resolve an anomaly encountered in Section 2.4.3.

5.1.2 Timescape Cosmology

5.1.2.1 Large-Scale Structure, Geometry and the Equivalence Principle

The timescape (TS; formerly known as "fractal bubble") cosmology (Wiltshire 2007a) is an alternative to ΛCDM that proposes a broader equivalence principle (Wiltshire 2009b) and identifies dark energy as an artifact of fitting an FLRW geometry to a universe that is not really homogeneous. Equation 5.1 is valid for a homogeneous mass-energy distribution and is conventionally assumed to provide a good approximation for our Universe. The isotropy of the CMB, with $\Delta T/T \sim 10^{-5}$, justifies this approximation at the time of last scattering, and the existence of a scale of homogeneity (e.g. Hogg et al. 2005) is usually cited as justification in the present Universe. However, a cube that encompasses this scale of homogeneity is very large (with a side length of $\sim 110\,h^{-1}$ Mpc), containing obvious structure (Figure 3.2, Figure 4.3 and Appendix B) and large voids ($\sim 40\,h^{-1}$ Mpc across) that exhibit density contrasts $|\Delta\rho|/\overline{\rho} \sim 1$, as demonstrated in Section 4.3. The extent and significance of large-scale structure has led to doubt about the adequacy of the FLRW approximation. Whether a homogeneous FLRW model can approximate well the *average* properties of an *inhomogeneous* universe remains an open question (e.g. Clarkson et al. 2012; Buchert 2008).

The averaging problem may be addressed by a generalised form of the Friedmann equations to describe the average evolution of an inhomogeneous geometry (Buchert 2000). TS interprets this result as an extension of the equivalence principle, thereby considering the effect of inhomogeneity on the local passage of time on cosmological scales. In Einstein's special relativity, a moving clock slows down relative to a stationary one, and inertial observers may be identified via Mach's principle. By general relativity's strong equivalence principle, we on Earth are accelerated upward by the ground beneath our feet, so a clock in a strong gravitational field that is not allowed to free fall undergoes acceleration, progressively slowing.

According to the strong equivalence principle, departure from the Hubble flow caused by gravity is inertial, but in TS a "cosmological equivalence principle" (Wiltshire 2008; Wiltshire 2009b), is proposed by which the expansion of particles in a static space is indistinguishable from the expansion of the space containing those particles. Observers in gravitationally-bound groups and clusters (contained in walls of large-scale structure) are decelerated from the expansion of the Universe and their clocks tick more slowly than inertial observers subject to the Hubble flow[1]. This difference in clock rates between "wall" and "void" clocks accumulates over the age of the Universe, leading to a significantly-increased age of the Universe measured by void observers and the characterisation of the Universe as a "timescape".

In TS, the time τ measured by a wall clock varies from the time t measured by a volume-average clock in a void according to $dt = \overline{\gamma} d\tau$ where the average lapse parameter $\overline{\gamma} \sim 1.4 > 1$ in the present Universe (Leith, Ng & Wiltshire 2008). The wall scale factor a is related to the volume-average scale factor $\overline{a}$ according to $a = \overline{a}/\overline{\gamma}$, reflecting the arrested inflation in walls. A geometric distinction between wall and void regions is made in which gravitationally-bound groups and clusters in walls occupy "finite infinity" regions that approach asymptotic flatness, while remaining void volumes are negatively curved. Walls and voids have FLRW-type geometries, but the average geometry is (Wiltshire 2007a):

$$ds^2 = -dt^2 + \overline{a}^2(t) d\overline{\eta}^2 + A(\overline{\eta}, t) d\Omega^2, \tag{5.2}$$

with conformal time $\overline{\eta}$ ($d\overline{\eta} = dt/\overline{a}$), particle horizon-volume averaged area A and $d\Omega^2$ as the line element on a sphere.

5.1.2.2 Dressed Cosmological Parameters and Apparent Cosmic Acceleration

Cosmological parameters may be "dressed" by the assumption of an FLRW geometry, altering their numerical values (Buchert & Carfora 2003). After relating the metric (5.2) to the wall geometry, the "bare" or "true" cosmological parameters in TS may be related to the "dressed" parameters conventionally measured by wall observers who assume an FLRW geometry. In particular, the dressed matter density and Hubble parameter relate to their bare analogues $\overline{\Omega}_{\mathrm{M}}$ and $\overline{H}$ as

$$\Omega_{\mathrm{M}} = \overline{\gamma}^3 \overline{\Omega}_{\mathrm{M}} \qquad (\text{since } a = \overline{a}/\overline{\gamma}) \tag{5.3}$$

$$H = \overline{\gamma}\overline{H} - \frac{d\overline{\gamma}}{dt},$$

[1] Gravitational energy gradients are said to be directly responsible for clock rate variance. Positive gravitational energy is associated with the negatively-curved voids.

related to the calibration of our rods and clocks, respectively. At $z \lesssim 37$ a late-time tracker solution is a good approximation (Wiltshire 2007b), giving

$$\begin{aligned}\Omega_{\mathrm{M0}} &= \frac{1}{2}(1-f_{v0})(2+f_{v0}) \\ \overline{\Omega}_{\mathrm{M0}} &= 4\frac{1-f_{v0}}{(2+f_{v0})^2} \\ \overline{H}_0 &= \frac{2(2+f_{v0})H_0}{4f_{v0}^2+f_{v0}+4}\end{aligned} \tag{5.4}$$

for current epoch parameters, where f_{v0} is the current epoch void volume fraction. At late times the model effectively has two free parameters for which H_0 and Ω_{M0} may be chosen.

In TS, the bare Hubble parameter is the underlying expansion rate (measured from local *proper* distance and *proper* time) and takes the same value in walls and voids. However, wall observers measure higher recession velocities across voids because their clocks are slower than void clocks, and therefore infer a higher Hubble parameter. The Hubble parameter is overestimated more for measurements of recession velocities in volumes with higher void fractions. Apparent cosmic acceleration, conventionally attributed to dark energy, is caused by the void-domination of the present Universe (Section 4.3). Void observers do not infer acceleration, and measure an older age for the Universe.

5.1.2.3 Tests and Anomalies

Kwan, Francis & Lewis (2009) found that TS was disfavoured by supernova data, but Smale & Wiltshire (2011) argue that TS is statistically indistinguishable from ΛCDM when a cut at the scale of statistical homogeneity is enforced and further systematic issues are considered. TS provides a slightly better fit to gamma-ray burst data than ΛCDM (Smale 2011), fitting the angular scale of the sound horizon from the CMB and the baryon acoustic oscillation scale in galaxy clustering statistics for a concordant region of the $(H_0, \Omega_{\mathrm{M0}})$ plane (Leith, Ng & Wiltshire 2008). Measurements of the dressed Hubble parameter in the local Universe vary: some are consistent with TS (e.g. Sandage et al. 2006) while others are not (e.g. Riess et al. 2009). TS addresses two puzzles related to the Hubble flow, the Sandage-de Vaucouleurs paradox and the Hubble bubble. Further observational anomalies that TS may resolve are discussed by Wiltshire (2007a) and Wiltshire (2009a), along with further tests.

Three anomalies of particular relevance to the context of this Thesis are the connectedness of the filamentary network (Stoica, Martínez & Saar 2010; Section 3.1.1), the emptiness of voids (Tikhonov & Klypin 2009; Section 4.1.1) and the low density contrasts of galaxy groups and clusters relative to the critical density (Smith et al. 2012; Section 2.4.3). The shorter ΛCDM model filaments could in principle result from dark energy breaking up the filamentary network. The emptiness of voids in real data, not reproduced by ΛCDM simulation, may (as discussed by Wiltshire 2007a) result from the variance in clock rates between walls and voids proposed by TS (but see also Tinker & Conroy 2009). If the age of the Universe is effectively greater in voids, structure formation is more advanced there than expected under an FLRW geometry in which the age of the Universe is position-independent. Although not explicitly noted in

Section 2.4.3, the density contrast of MSPM groups and clusters relative to the critical density is calculated assuming ΛCDM, and is low relative to the expectations of ΛCDM. In TS both the critical density and (bare) matter density are lower, presaging a solution to the discrepancy. In the following sections of this chapter we investigate the quantitative effect of these differences and develop a new cosmological test that is applied to both ΛCDM and TS.

5.2 Virialisation

The cosmological parameters H, Ω_{M} and Ω_Λ quantify the geometry of our Universe. Additionally, in this Section we will discuss their effect on the formation and growth of gravitationally-bound structures, motivating a new observational test.

5.2.1 Spherical Collapse

The spherical collapse model (Gunn & Gott 1972) admits scaling relations for the bulk properties of virialised structures. A spherical top-hat overdensity perturbation will defy the Hubble flow and gravitationally collapse if it is denser than the *critical density* ("overcritical"):

$$\rho_{\mathrm{cr}} = \frac{3H^2}{8\pi G}. \tag{5.5}$$

Upon collapse and virialisation, the structure has condensed such that within a sphere of radius R_{vir} it has a mean density $\Delta_c\overline{\rho}$ where $\overline{\rho}$ is the mean background density, and the mass M contained within R_{vir} is:

$$M = \frac{4\pi R_{\mathrm{vir}}^3 \rho_{\mathrm{cr}} \Delta_c}{3}. \tag{5.6}$$

If the structure is assumed to be an isothermal sphere with an internal density distribution $\rho(r) \propto r^{-2}$, the observable velocity dispersion σ_v is related to the virial mass M by (Bryan & Norman 1998):

$$\sigma_v = M^{1/3} \left(\frac{H^2(z)\Delta_c G^2}{16} \right)^{1/6}. \tag{5.7}$$

Applying equations 5.5 and 5.6 to 5.7 then leads to

$$\sigma_v = \frac{1}{2} H \Delta_c^{1/2} R_{\mathrm{vir}}, \tag{5.8}$$

similar to equation 2.2. Introducing a normalisation factor f_σ to correct for the assumption of an isothermal sphere,

$$\sigma_v = f_\sigma \frac{1}{2} H \Delta_c^{1/2} R_{\mathrm{vir}}. \tag{5.9}$$

Bryan & Norman find $f_\sigma \sim 0.9 \pm 0.05$ from simulations.

The density contrast of a density perturbation δ_c required for a structure to collapse, and hence the eventual value of Δ_c, the density of the collapsed structure in units of the critical density, is sensitive to the matter density Ω_M (Lacey & Cole 1993). Fitting functions for Δ_c are given by Bryan & Norman (1998). Defining $x = \Omega_M - 1$, these are:

$$\Delta_c = 18\pi^2 + 82x - 39x^2 \qquad \text{, for a flat universe with } \Omega_M + \Omega_\Lambda = 1\text{, and} \tag{5.10}$$

$$\Delta_c = 18\pi^2 + 60x - 32x^2 \qquad \text{, for } \Omega_M < 1 \text{ and } \Omega_\Lambda = 0. \tag{5.11}$$

The scaling relations and density contrasts predicted by this model are valid for structures that have just virialised, and are a good approximation for structures that are continuously accreting material (Eke, Cole & Frenk 1996). The boundary corresponding to Δ_c is an effective separator of virial and infall regions in an $\Omega_M = 1$ universe (Cole & Lacey 1996). Bryan & Norman (1998) use this to identify simulated halo boundaries, and find that, when appropriately normalised, the properties of simulated haloes are reasonably well-fitted by the scaling relations derived from spherical collapse, over a range of redshifts.

5.2.2 A Test of the Timescape

We now propose a test to distinguish TS from ΛCDM, based on the density of gravitationally-bound structures relative to the critical density. The critical density is substantially different between the two models at late times, and the current epoch bare parameters that directly influence group and cluster dynamics are quite different to their dressed analogues in TS.

Wiltshire (2007a) notes that the critical density of equation 5.5 is not the true critical density in TS, which should instead be $3\overline{H}^2/(8\pi G)$ with the bare Hubble parameter $\overline{H}$ related to H by equation 5.3. Consequently, the H term contributed to equation 5.8 by the critical density should be $\overline{H}$. More generally, H should be replaced by $\overline{H}$ where properties intrinsic to a bound structure are concerned, since the bare Hubble parameter pertains to the locally-measured Hubble flow, so in TS,

$$\sigma_v = f_\sigma \frac{1}{2} \overline{H} \Delta_c^{1/2} R_{\text{vir}}. \tag{5.12}$$

Δ_c for ΛCDM is related to the matter density Ω_M by equation 5.10. For TS we apply equation 5.11, but with $x = \overline{\Omega}_M - 1$ since the *bare* matter density is relevant in this case.

For ΛCDM, equations 5.9 and 5.10 relate the observable σ_v/R_{vir} to H and Ω_M, which for TS is achieved by equations 5.12, 5.11 and 5.4. Separating the observable and model parameters, we find

$$\frac{\sigma_v}{R_{\text{vir}}} = f_\sigma \frac{1}{2} H \sqrt{\Delta_c(\Omega_M)} \qquad (\Lambda\text{CDM}) \tag{5.13}$$

$$\frac{\sigma_v}{R_{\text{vir}}} = f_\sigma \frac{1}{2} \overline{H} \sqrt{\Delta_c(\overline{\Omega}_M)} \qquad (\text{TS}), \tag{5.14}$$

in which the left-hand side can be found from a linear fit through measured virial radii and velocity dispersions for a sample of groups and clusters, and the right-hand side is model-determined. In Section 2.4.3 we noted that in our implementation, H-dependence is lost in ΛCDM with the conversion from h^{-1} Mpc to Mpc, and this will remain true where the groups and clusters are observed at low redshift and $H(z) \sim H_0$. In TS the H_0 terms do not cancel out since the value of h assumed in cosmography is dressed while the Hubble parameter relevant to the spherical collapse model is bare. However, the outcome of this test is not sensitive to the value of H_0 because H_0 and $\overline{H}_0$ are related by a constant determined by the void fraction f_{v0} (equation 5.4). But the void fraction is related to the matter density, so along with the value of Δ_c, a constraint can be placed on $\Omega_{\rm M0}$. If $R_{\rm vir}$ is measured from the angular extent of groups it will be sensitive to the assumed cosmology and the parameter values therein, but at low redshift this sensitivity is slight (e.g. Wiltshire 2009a).

To invoke the spherical collapse model's scaling relations, the subject groups and clusters should have recently virialised or be continuously accreting material (Eke, Cole & Frenk 1996). Filaments channel mass onto clusters (Kodama et al. 2001; Ebeling, Barrett & Donovan 2004), so groups and clusters in filamentary environments are good candidates for the test we propose.

5.3 Refinements to the Radius-Velocity Dispersion Fit and Δ_c Determination

In this Section we describe and justify changes to the procedure outlined in Section 2.4.3.

5.3.1 Elongated Environments

Selecting recently-virialised structures is very difficult in real data, but we will attempt to address the requirement that groups and clusters should have recently virialised or be continuously accreting material, as expected by the spherical collapse model. We attempt a radius-velocity dispersion fit to groups and clusters that reside in elongated environments, alongside a refined analysis for the entire group and cluster sample. Spherical collapse is most applicable to spherical systems, and we stress that we do not attempt to select groups and clusters that are themselves elongated. The f_σ normalisation we use from Bryan & Norman (1998) is not derived from groups and clusters in elongated environments, so the results we obtain from the elongated sample may disagree with model expectations. However, this sample provides a useful comparison by indicating how sensitive our results may be to selection critieria based on environment.

We are unable to achieve a good fit by using components of the 53 MSPM filaments (Section 3.3) alone, with the removal of outlying velocity dispersions excluding over half the data. Hence, our selection criteria are relaxed to include groups and clusters in elongated environments that may not be large-scale filaments. In either case we expect to select structures with a range of masses, including nodes at intersections of filaments. Such environments are more likely sites of mass accretion, including merger events, since infall is anisotropic (Colberg et al. 1999). We use the elongation probability derived from group and cluster positions (Section 3.2), requiring that $P_e > 0.75$ within a radius of 0.2 h^{-1} Mpc. There are 1290 groups and clusters in these environments, and radii and velocity dispersion data are then analysed as in Section 2.4.3. Our linear fit with 1σ uncertainty, shown in Figure 5.1(b), is

$$\sigma_v = (256 \pm 9 \text{ (statistical)} \pm 10 \text{ (systematic)})R + (30 \pm 5), \tag{5.15}$$

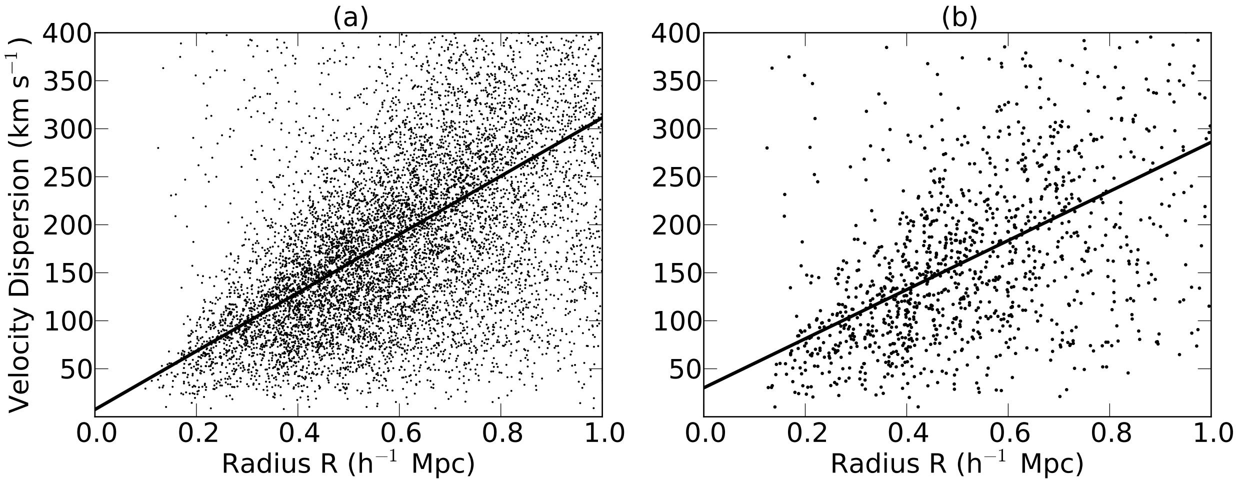


Figure 5.1: Velocity dispersion against radius (R) with linear fits to the data with outlying velocity dispersions removed. **(a)** All MSPM groups and clusters. Series for various member galaxy counts are shown in Figure 2.11. **(b)** Groups and clusters in elongated environments, containing at least four visible member galaxies. The fit may be influenced at large R by systems with radii that are overestimated because of contamination by nearby galaxies, an effect that is more likely in elongated environments. See text for details.

where R is the group or cluster radius in units of h^{-1} Mpc and σ_v is in units of km s^{-1}. This is shallower than the slope shown in Figure 5.1(a), found in Section 2.4.3 for the whole MSPM sample, consistent with the expectation that groups that have formed more recently have lower densities, having formed in a lower-density Universe (e.g. Navarro, Frenk & White 1997). Alternatively, the fit in Figure 5.1(b) may be influenced at large R by systems with radii that are overestimated because of contamination by nearby galaxies, an effect that is more likely in elongated environments.

5.3.2 Virial Radii

The radii estimated for the MSPM groups and clusters (Section 2.3.4) may not be virial radii, and should be calibrated. Ideally, such calibration would be achieved by running MSPM on simulated data, but this is beyond the present work. The virial radius R_{vir} contains an approximately isothermal mass distribution, with a projected surface galaxy distribution that can be expected to obey a power law well outside the core radius R_C (e.g. Girardi et al. 1995). Our radii R enclose regions that are approximately twice the density at $2\,h^{-1}$ Mpc, denoted by ρ_2 (Section 2.3.4). Assuming these radii as an approximation to R_{vir}, we fit power laws to density profiles at $r < R$, shown in Figure 5.2. The innermost data point ($r < 0.1\,h^{-1}$ Mpc) and innermost quarter of each profile are omitted from the fits to exclude the core. Girardi et al. (1998) find a median $R_C/R_{\text{vir}} = 0.05$.

Our power law fits show that the R values reported in our catalogue enclose approximately power-law distributions, as required of virial radii. At large distances r from the group and cluster centres the density begins to approach that of the background, becoming shallower.

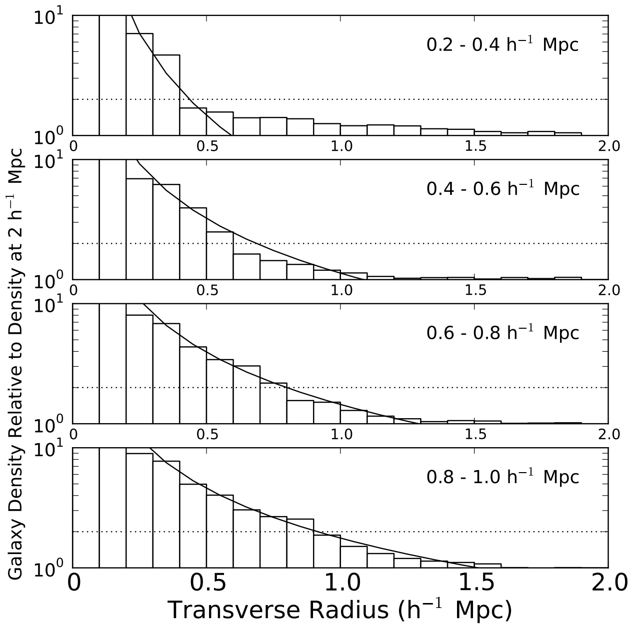


Figure 5.2: Profiles of the galaxy density relative to surrounding locations, including $P - S \leq 0.5$ galaxies, within annuli centred on $P - S$ peaks determined for our structures, averaged over $0.025 < z < 0.2$. Each panel shows an average of density profiles obtained for all structures with radii within the indicated range. Horizontal dotted lines indicate twice the density at $2\,h^{-1}$ Mpc, defined as the density in the annulus formed by rings of radii 1.95 and $2\,h^{-1}$ Mpc. Solid lines show power law fits that exclude the core. Relative densities close to group and cluster centres are clipped at 10.

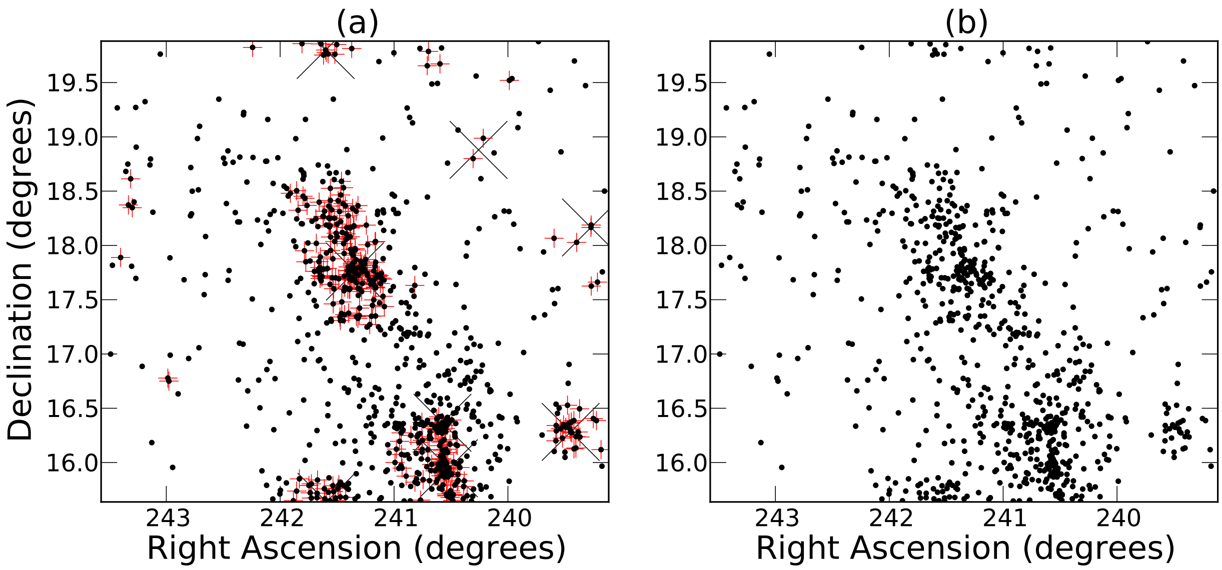


Figure 5.3: A field centred on the first structure in our group and cluster catalogue (Table 2.1), the *Hercules* cluster Abell 2151 (Abell 1958; Corwin 1974; Tarenghi et al. 1979), also shown with colour and a smaller field of view in Figure 2.1. Filled circles are $r < 17.77$ galaxies, which are also marked by red plusses if they are at positions with $P - S > 0.5$ (cluster member galaxies) and large crosses are MSPM groups and clusters. **(a)** Objects within a transverse radius of $4\,h^{-1}$ Mpc at $z = 0.03629$ and within a line-of-sight radius $\Delta z = 0.005$. **(b)** The same field of view, but with plusses and large crosses omitted.

Unfortunately, this shallowing of the density profile is expected both within and outside the virial radius (e.g. Navarro, Frenk & White 1997), and is not a feature that can be used to estimate $R_{\rm vir}$.

It is conventional to define virial radii with a threshold in density, but in the present study we cannot reliably justify any specific threshold in galaxy density. To calibrate our radii, the first structure in our group and cluster catalogue, the *Hercules* cluster Abell 2151 is a useful case to consider. For this structure we have measured $R = 1.7\,h^{-1}$ Mpc from the distance between its centre and farthest member galaxy, which is correct if member galaxies are confined to the virial component of the cluster. Visually, the infall region can be distinguished from the virial component by its anisotropy. *Hercules* is shown in Figure 5.3, and the region highlighted by the positions of member galaxies appears to pass this visual test. Moreover, in a dynamical study of *Hercules*, Bird, Dickey & Salpeter (1993) apply the virial theorem to all galaxies within a radius of $1.8\,h^{-1}$ Mpc. Barmby & Huchra (1998) exclude galaxies outside a radius of $1.6\,h^{-1}$ Mpc, and their assignment of cluster membership for the purpose of virial analysis is similar to our member identification. This case suggests that $R_{\rm vir}/R \sim 1$, although it is of course only one example, chosen only because it is the first structure in our catalogue, and *Hercules* may not be representative, exhibiting clear substructure and containing much more mass than most of our other groups and clusters. On the other hand, a perfectly spherical cluster would certainly be unrepresentative.

In the current study we assume that the virial radius of *Hercules* lies between the published estimates (1.6 h^{-1} Mpc and 1.8 h^{-1} Mpc), using this range to calibrate our radii, finding:

$$R_{\text{vir}} = (1.00 \pm 0.06)R, \tag{5.16}$$

with both R_{vir} and R in h^{-1} Mpc. This relation is applied to our σ_v/R slope to obtain σ_v/R_{vir} as required by equations 5.13.

5.3.3 Summary of Refinements to σ_v/R_{vir} and Δ_c Determination

Changes to our procedure from Section 2.4.3 are:

- use of a sample of groups and clusters in elongated environments to improve justification for the spherical collapse model's application (Section 5.3.1),
- adoption of an f_σ normalisation to correct for the assumption of an isothermal sphere (equation 5.9),
- calibration of our radii to the *Hercules* cluster Abell 2151,
- reduction of model assumptions in the expected value of Δ_c (equations 5.10 and 5.11), and
- adoption of best-fit matter densities Ω_{M0} (Table 5.1).

We will also no longer attempt to account for redshift-variation of Δ_c as this involves model assumptions, and our revised analysis will include the Mr20 (Berlind et al. 2006) and Y08 (Yoon et al. 2008) group and cluster catalogues. These are the comparison catalogues from Section 2.4.4, except that we cannot use the C4 catalogue in this way because it does not measure cluster radii. In Section 2.4.3 it was noted that conversion from h^{-1} Mpc to Mpc removes the need for an assumed value of H_0 in the calculation of Δ_c. In TS, although the difference between H_0 and $\overline{H}_0$ is of interest, the outcome of this test remains insensitive to the value of H_0. We stress that these refinements do not in any fashion make our test rigorous, but they do allow us to demonstrate a calculation and explore the issues that must be confronted when carrying out this test. Further refinements are beyond the present study.

5.4 Radius-Velocity Dispersion Relation in ΛCDM and the Timescape

5.4.1 Procedure and Parameters

In this Section we implement refinements to, and extend, our earlier radius-velocity dispersion relation analysis. Velocity dispersions σ_v are measured from the radial motions of group-member galaxies (for MSPM see Section 2.3.7). Group and cluster radii R are determined from the angular extent of groups (for our data see Section 2.3.4) on the sky and an assumed cosmology, for which we have used ΛCDM with $\Omega_{\text{M0}} = 0.3$ and $\Omega_\Lambda = 0.7$. The choice of parameter

Table 5.1: Observational and Model Parameters†

Parameter	Value	Source*
R-σ_v Slope – MSPM Whole Sample	$\sigma_v/R = 304 \pm 24\ h$ Mpc^{-1} km s^{-1}	S12; Section 2.4.3
R-σ_v Slope – MSPM Elongated	$\sigma_v/R = 256 \pm 19\ h$ Mpc^{-1} km s^{-1}	Section 5.3.1
R-σ_v Slope – Mr20	$\sigma_v/R_{\rm vir} = 214 \pm 9\ h$ Mpc^{-1} km s^{-1}	Mr20
R-σ_v Slope – Y08	$\sigma_v/R_{\rm vir} = 1008 \pm 30\ h$ Mpc^{-1} km s^{-1}	Y08
Group Radius Calibration (MSPM)	$R_{\rm vir}/R = 1.00 \pm 0.06$	Section 5.3.2
Virial Fitting σ_v Normalisation	$f_\sigma = 0.90 \pm 0.05$	BN98
Current Matter Density – ΛCDM	$\Omega_{\rm M0} = 0.274^{+0.019}_{-0.018}$	WMAP7
Current Dressed Matter Density – TS	$\Omega_{\rm M0} = 0.33^{+0.11}_{-0.16}$	LNW

†Parameter values relevant to our analysis. Apart from $R_{\rm vir}/R$, all uncertainties are random. $\sigma_v/R_{\rm vir}$ from Mr20 and Y08 are calculated by us using data from those catalogues. ΛCDM constraints are from cosmic microwave background (CMB), baryon acoustic oscillation (BAO) and nearby-Universe H_0 measurements. TS constraints are from type Ia supernovae, for which it should be noted that an arbitrary normalisation is assumed. However, these values are broadly consistent with CMB and BAO constraints on TS.

*S12: Smith et al. (2012); Mr20: Berlind et al. (2006); Y08: Yoon et al. (2008); BN98: Bryan & Norman (1998), Figure 3; WMAP7: Komatsu et al. (2011); LNW: Leith, Ng & Wiltshire (2008).

values is not important at the redshifts we work with, especially if $\Omega_{\rm M} + \Omega_\Lambda = 1$. Similarly, distance measurements are very similar in TS at low redshift (Wiltshire 2009a).

A σ_v/R slope is derived from a least-squares linear fit, and our fitting procedure is detailed in Section 2.4.3. If R is reported in Mpc with an assumed value of h we convert back to h^{-1} Mpc with the same value of h. R measurements in h^{-1} Mpc are converted to Mpc by multiplying the slope by h (of the assumed model). If R is measured in comoving distance, the slope is multiplied by $(1+z)$ to convert to proper distance, where z is the median redshift of the sample. Ideally we would apply this correction to each group and cluster before performing the fit, but since we are interested in the data's overall trend this should not significantly affect our results, especially given our narrow range of redshift. For MSPM, $\sigma_v/R_{\rm vir}$ is obtained from our slope through division by $R_{\rm vir}/R$ (Section 5.3.2).

Various observational and model parameters relevant to our analysis are listed in Table 5.1.

5.4.2 Δ_c and the Dressed Matter Density

We now revisit the calculation of Δ_c and its comparison with model expectations, initially performed in Section 2.4.3. From the ΛCDM matter density in Table 5.1 and equation 5.10, we expect $\Delta_c = 97.6^{+2.6}_{-2.5}$ in ΛCDM. From the TS dressed matter density we find its bare equivalent (equations 5.4), and from equation 5.11 we expect $\Delta_c = 100.7^{+6.8}_{-8.1}$ in TS. The expected density of bound structures relative to the critical density is similar in the two models, but the critical density is lower in TS. These expected Δ_c values are compared with measurements that derive from the observed radius-velocity dispersion relation and the assumption of parameters in Table 5.1. Agreement for the value of Δ_c is effectively a test of equality in equations 5.13. For each

Table 5.2: Measurements of Δ_c in Group and Cluster Catalogues

Catalogue*	ΛCDM	TS
Model Expectation	$97.6^{+2.6}_{-2.5}$	$100.7^{+6.8}_{-8.1}$
MSPM Whole Sample	$53.1^{+25.7}_{-17.3}$	$87.3^{+28.3}_{-18.3}$
MSPM Elongated	$37.7^{+17.4}_{-11.8}$	$61.9^{+19}_{-12.2}$
Mr20	$25.3^{+5.5}_{-4.4}$	$41.5^{+3.6}_{-1.4}$
Y08	$580.3^{+109.8}_{-90.2}$	$953.2^{+59.0}_{-10.5}$

For each catalogue, Δ_c is calculated from the observed radius-velocity dispersion relation and parameter values for the respective models from Table 5.1. Agreement is reached only for the main MSPM catalogue with TS. Uncertainties originate from the values in Table 5.1, and are greater than reported in Section 2.4.3 following the inclusion of uncertainties for f_σ and $R_{\rm vir}/R$. These results are also shown in Figure 5.4. The model expectations (top line) are not significantly different, but the estimates when measuring Δ_c depend importantly on the model assumed, as a consequence of equations 5.13 and 5.14.

*MSPM Whole Sample: Smith et al. (2012), Section 2.2; MSPM Elongated: Section 5.3.1; Mr20: Berlind et al. (2006); Y08: Yoon et al. (2008).

catalogue, two Δ_c values are calculated, one for ΛCDM and one for TS. Δ_c values found for various catalogues are listed in Table 5.2, and shown in Figure 5.4.

Agreement with ΛCDM is not obtained for any of the group samples. Measurements of Δ_c are more consistent with TS except for the Y08 sample, with agreement for the main MSPM sample. At face value, TS is favoured over ΛCDM based on the initial results of this test, but there are large systematic issues that have not been quantified, as discussed in the next Section. As we have noted, the outcome of this test is not sensitive to the value of H_0 in either model. In the present study the systematics are not understood well enough to provide meaningful constraints on $\Omega_{\rm M0}$. But as an example, if a value of $\Omega_{\rm M0}$ from Table 5.1 is no longer assumed, for TS, only the main and elongated MSPM group samples admit $0 < \Omega_{\rm M0} < 1$, with $\Omega_{\rm M0} < 0.49$ and $\Omega_{\rm M0} < 0.27$ respectively.

5.5 Discussion

5.5.1 Density Contrast Expectations

The values of Δ_c expected by ΛCDM and TS are similar, with $\Delta_c \sim 100$. This is the density of virialised structures in units of the critical density $\rho_{\rm cr}$, and should not be confused with density in units of the background density $\overline{\rho} = \Omega_M \rho_{\rm cr}$ (ΛCDM) or $\overline{\rho} = \overline{\Omega}_M \rho_{\rm cr}$ (TS). In units of the background density we therefore expect virialised structures to have densities $\sim \Delta_c/\overline{\Omega}_M$. The lower value of $\overline{\Omega}_M$ in TS (e.g. Leith, Ng & Wiltshire 2008) implies a higher expectation for this quantity. At the same time we may still find lower absolute densities in bound systems because the critical density is lower. TS expectations for bound systems are also discussed by Wiltshire (2007a, Section 8.3).

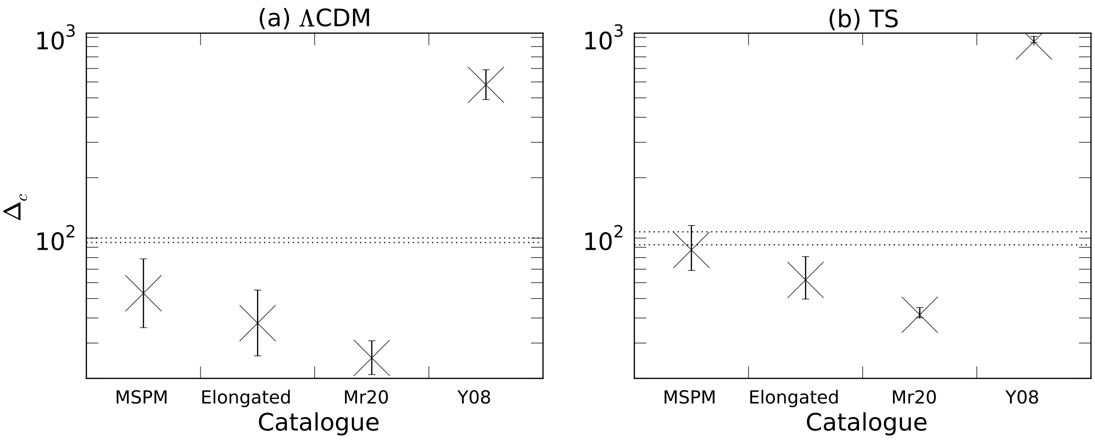


Figure 5.4: Measurements of Δ_c from each catalogue as reported in Table 5.2. Dotted horizontal lines indicate model expectations. **(a)** ΛCDM. **(b)** TS.

5.5.2 Tension with Other Catalogues

The radius-velocity dispersion relation slopes $\sigma_v/R_{\rm vir}$ determined from the various catalogues (Table 5.1) are very different. As discussed in Section 2.4.4, all of these catalogues are derived from the SDSS DR7 spectroscopic survey data (although Y08's density measurements are augmented by photometric-only data), so any differences arise from the details of their compilation and property measurements. The MSPM samples agree most closely with the expectations of TS, but they *also* approach the expectations of ΛCDM more closely than the other catalogues. This leads us to consider whether there is any reason to prefer MSPM over the others.

In the Mr20 catalogue of Berlind et al. (2006), the method of velocity dispersion estimation is comparable to our own and the radius estimation also uses the positions of member galaxies. The identification of member galaxies is different, however, based on friends-of-friends linking with a linking length of 0.2 times the mean interparticle separation. According to comparison with mock catalogues in Berlind et al., this should result in radii that encompass single virialised haloes, making them suitable for our test of cosmology. However, the mock catalogues are derived from ΛCDM simulations, which is clearly problematic for the test we have carried out. If TS is correct and the density contrast of bound systems relative to the *background* is greater than usually assumed (as discussed above in Section 5.5.1), a threshold intended to select virialised systems in ΛCDM will be too low for a TS universe, resulting in radii that are overestimated and a $\sigma_v/R_{\rm vir}$ that is too low. Hence, the Mr20 data would be expected to produce a lower Δ_c than expected by TS.

The group and cluster catalogue of Yoon et al. (2008) reports velocity dispersions calculated using the same estimator as ours, but virial radii are found from the density of member galaxies compared with the background. Yoon et al. do not state what threshold they used, but by calling their measure r_{200} they have presumably adopted the radius that encloses a mean density of 200 times the critical density ($\overline{\rho} = \Omega_M \rho_{\rm cr}$ in ΛCDM). This is motivated by the spherical collapse model's prediction that in a matter-dominated universe with $\Omega_{\rm M} = 1$, virialised structures have a mean density of $18\pi^2 \approx 200$ times the critical density (c.f. equations 5.10-5.11). However, our Universe has $0.2 \lesssim \Omega_{\rm M0} \lesssim 0.4$ and $\Delta_c \sim 100 < 18\pi^2$, so r_{200} is likely to be an underestimate (e.g. Voit 2005), resulting in a $\sigma_v/R_{\rm vir}$ that is too high. It might still be possible to carry out our test with these radii, but this would require adoption of an f_σ normalisation found from simulations in which radii are defined as r_{200}. Moreover, while the galaxy density is a reasonable approximation to the underlying dark matter density *within* clusters (Katgert, Biviano & Mazure 2004; Lin, Mohr & Stanford 2004; Hansen et al. 2005), this approximation has not been validated for comparisons with the background density.

The above considerations advance the MSPM catalogue's suitability for the test we have carried out because it is less model-dependent. Our catalogue's construction has not made use of simulations, or assumed a model-determined threshold in density. A cosmology has been assumed to determine distances, but at low redshift this does not affect our comparison of ΛCDM and TS. On the other hand, the lack of comparison with simulated data from either model prevents an independent determination of f_σ appropriate for the radii measured in our catalogue, and this is one of the main reasons we cannot place meaningful constraints on cosmology. Because we have not carried out tests on mock data, we have a poor understanding of how our measured radii relate to virial radii. Moreover, although we have compared MSPM with other approaches (Section 2.4.4), MSPM's selection function is still not well understood. If a population of structures is not detected by MSPM this potentially impacts upon the σ_v/R slope. Tests on mock data would help us to address these questions, but issues of model-dependence would need to be considered.

5.5.3 Systematic Issues within the MSPM Catalogue

The radii reported in the MSPM catalogue are distances between group centres and furthest group member galaxies. This method is sensitive both to the low member-number resolution and to member galaxy positions at the outskirts of groups, but we only require that radii are correct on average. However, it has not been established that our radii are virial and this is a significant caveat. In Section 5.3.2 we showed that the radius measured for the well-studied case of the *Hercules* cluster is close to values assumed in previous work. But this is of course only one example and *Hercules* may not be representative, exhibiting clear substructure and containing much more mass than most of our other groups and clusters. The visual separation of virial and infall regions shown in Figure 5.3 is not discernible for most of our other groups and clusters, which are not as well resolved. However, we generally find that where this distinction is clear, the member galaxies from which radii are determined appear to identify the virial component.

Another issue concerning the radius measurements is misidentification of the group centre. Group centres in the MSPM catalogue have the sky position of the member galaxy with the highest $P - S$ value, since for structures that are elongated or asymmetric, an average position on the sky may not select the densest part of the structure. However, the member-number resolution of our groups limits the accuracy of this measure. Unlike misidentification of the group edge, this will have a systematic effect, because misidentification of the group centre increases radii in a way that flattens the density profiles (Figure 5.2) and cannot be corrected by our power law fits.

The severity of this effect is expected to be reduced in groups with higher counts of visible member galaxies. Figure 2.11 shows that the σ_v/R slope for groups with at least eight member galaxies is similar to the slope obtained for four-member detections at large R, so this does not appear to be a significant issue. There does appear to be divergence at small R, but this may be a result of contamination in the line-of-sight, to which groups with higher member counts are more susceptible. Such contamination may be filamentary structure oriented with the line of sight, which has the greatest impact at small R. Some radii are overestimated because of contamination from galaxies that are not physically associated with groups, but further investigation would be required to quantify this effect.

The MSPM catalogue contains many groups with fewer than eight visible member galaxies, and group properties measured from these may thus be inaccurate. However, in our test we only require that these properties are correct *on average*, and groups with outlying velocity dispersions have been excluded from our linear fit. Moreover, our equations 2.3 and 5.16, obtained from the entire MSPM sample, admit our measurements for the *Hercules* cluster $(R, \sigma_v) = (1.7\,h^{-1}\,\text{Mpc}, 680\text{ km s}^{-1})$. Our own measurements for *Hercules* are in turn consistent with, for example, Barmby & Huchra (1998) who find $\sigma_v = 705^{+46}_{-39}$ km s^{-1} after excluding galaxies outside a radius of 1.6 h^{-1} Mpc.

The MSPM catalogue is compiled with a threshold low enough to retain sensitivity to filamentary structure (Chapter 3) that probably admits some false discoveries. Many MSPM groups are visually underwhelming and have irregular shapes, although for many of them this is because they are poorly sampled, a simple consequence of incompleteness. Particle distributions with lower member-number resolution are more likely to appear irregular. The σ_v/R slope for structures with only four visible member galaxies is similar to that obtained with more populous structures (Figure 2.11), a similarity that is difficult to explain if a large fraction of them are false discoveries. If there are false discoveries, their effect can be roughly estimated by finding σ_v/R for the MSPM catalogue constructed with a higher threshold in peak overdensity probability P (Section 2.3.2). After adopting a $P > 0.95$ cut that retains 5041 of the original MSPM sample, $\sigma_v/R = (330 \pm 5)\,h$ Mpc^{-1} km s^{-1}, which is about nine per cent higher. A commensurate 18 per cent upward correction to the measured Δ_c would improve agreement with the expectations of both ΛCDM and TS. False discoveries therefore are unlikely to have a significant impact on our test.

The MSPM catalogue is based on magnitude-limited data, and the structures it includes at higher redshifts will be more massive, with their more intrinsically-luminous member galaxies used to determine properties. At $z < 0.1$ we find $\sigma_v/R = (277 \pm 4)\,h$ Mpc^{-1} km s^{-1}, while at $z > 0.1$ we find $\sigma_v/R = (324 \pm 6)\,h$ Mpc^{-1} km s^{-1}, indicating a bias that may result from underestimation of radii at higher redshifts. The $z < 0.1$ data are probably more reliable, but a volume-limited study may be the best way to address this bias.

We can simultaneously estimate the impact of some of the abovementioned systematic issues by obtaining σ_v/R for a subset of the MSPM catalogue at $z < 0.1$ with high-significance detections ($P > 0.95$) that each have at least eight visible member galaxies and radii greater than 0.5 h^{-1} Mpc. From this set of 1289 groups and clusters we find $\sigma_v/R = (314 \pm 21)\,h$ Mpc^{-1}

km s^{-1}, consistent with our original determination. This suggests that the main systematic issue affecting the validity of our result is the calibration of radii.

5.5.4 MSPM Structures in Elongated Environments

The "elongated MSPM" subset is selected using a measure of elongation on group and cluster positions, as detailed in Section 5.3.1. However, for our test we have used Bryan & Norman's f_σ normalisation, which is not derived from a sample of groups and clusters selected to lie in elongated environments. Bryan & Norman recover clear scaling relations despite the lack of an attempt to select recently-formed structures, so this may not be an important aspect of our analysis. However, our work with the elongated MSPM sample is an indication of how sensitive our results may be to selection critieria based on environment.

It is also possible (and certain in at least some cases) that in elongated environments, galaxies are spuriously identified as members of groups, causing a systematic overestimation of radii responsible for systems shown in the lower-right of Figure 5.1(b), reducing the measured σ_v/R. It would also be interesting to look for elongation by using individual galaxy positions to select a population of recently-formed bound structures. Younger groups are more aspherical than their older counterparts (Ragone-Figueroa et al. 2010). Of course, care would need to be taken because the spherical collapse model is less relevant in the most elongated cases.

5.5.5 Validity of Spherical Collapse

An isothermal density profile is assumed for equation 5.8, but we have adopted the f_σ normalisation of Bryan & Norman (1998) to compensate, in equation 5.9. By assuming the same f_σ correction to spherical collapse regardless of the cosmology, we have assumed that bound structures in a TS universe form and grow as in an FLRW model with no dark energy and lower matter density. While there is no obvious reason to doubt this similarity, this assumption is untested. The f_σ correction is a crucial ingredient in our test, and relies on the assumption that the virial radii we have estimated are close to those derived from the Δ_c threshold applied to simulated haloes by Bryan & Norman. Ultimately, this assumption can only be validated by tests on mock catalogues. To apply the spherical collapse model, the subject groups and clusters should have recently virialised or be continuously accreting material. We have attempted to address this requirement in Section 5.3.1, finding that the value of Δ_c inferred from the elongated MSPM subset is not significantly different to that inferred from the main MSPM sample. But as we have noted, the f_σ normalisation we have used is not drawn from a similar simulated sample.

Spherical collapse is often described as statistically valid while not as reliable on a case-by-case basis, which is all we require, and is still used in halo modelling (e.g. King & Mead 2011). Most importantly, it has been tested in simulations by Bryan & Norman and others, and the fact we are applying it to real data does not invalidate its use as long as dynamics on $\sim 1\ h^{-1}$ Mpc scales remains unchanged. However, there is a more general ellipsoidal collapse model (Sheth, Mo & Tormen 2001), in which a higher initial density perturbation is generally required for collapse (explored further by Ludlow & Porciani 2011), but a detailed investigation of the implications is beyond the present study.

5.5.6 Existing Cluster Measurements

The mass function of massive clusters has been measured and is not in tension with ΛCDM (Vikhlinin et al. 2009; Mantz et al. 2010). However, the present study is concerned not specifically with the amount of mass but the volume over which it is distributed, so our results are not in direct tension with the measured mass function. Moreover, without considering these studies in detail, we note that they rely on samples of X-ray luminous clusters that are more massive than the groups in our sample, which has a median velocity dispersion $\sigma_v \sim 200$ km s^{-1}. These studies also concentrate on the inner regions of clusters where properties are more easily measured, typically within the radius r_{500} containing a mean density 500 times the critical density.

5.5.7 Summary

We have attempted to use the $\sigma_v/R_{\rm vir}$ slopes as a cosmological probe, but significant systematic issues remain unresolved. The main MSPM catalogue implies a lower value of Δ_c than expected by ΛCDM, which may be interpreted as evidence that its groups are not virialised. However, we have shown that this interpretation is model-dependent, since its Δ_c agrees with the TS expectation. There are a number of possible systematic issues within the MSPM catalogue, which must be understood before ΛCDM and TS can be truly differentiated by this test, for which group and cluster samples are now large enough. Determination of virial radii is the most significant of these, and relevant to the value of f_σ appropriate for our sample. If we have underestimated virial radii, for example, the adopted f_σ may be too low for use with the velocity dispersions measured within R.

Radial velocities could be used to find radii that contain gravitationally-bound mass (e.g. Donnelly et al. 2001). Alternatively, if a value of f_σ is obtained from groups in simulated data with radii measured by our existing approach, determination of virial radii may not be necessary. A volume-limited study may be most appropriate to eliminate biases that result from incompleteness.

By assuming the same f_σ correction to spherical collapse regardless of the cosmology, we have assumed that bound structures in a TS universe form and grow as in an FLRW model with no dark energy and lower matter density, and this assumption is untested. The variation of Δ_c with redshift has not been considered, and in the case of ΛCDM $\Delta_c = 91$ is expected at the median redshift of our sample. Accounting for this would also require consideration of redshift variation in H and $\Omega_{\rm M}$. Addressing the systematic issues to achieve the required precision is challenging, and we have also other noted ways in which our analysis could be improved. If precise measurements of large group and cluster samples become possible at much higher redshifts, the time evolution of the Hubble parameter $H(z)$ may be constrained by variation of the slope of the radius-velocity dispersion relation.

Found

I walked along east beach, remembering, remembering.

Then I went uphill, to the palm trees and our crude but strong hut, and Timothy's grave.

I looked down. The coral stones and shells were still where I'd left them.

I stood there for a little while, feeling very close to him, shut my eyes, then said, "Dis b'dat outrageous cay, eh, Timothy?"

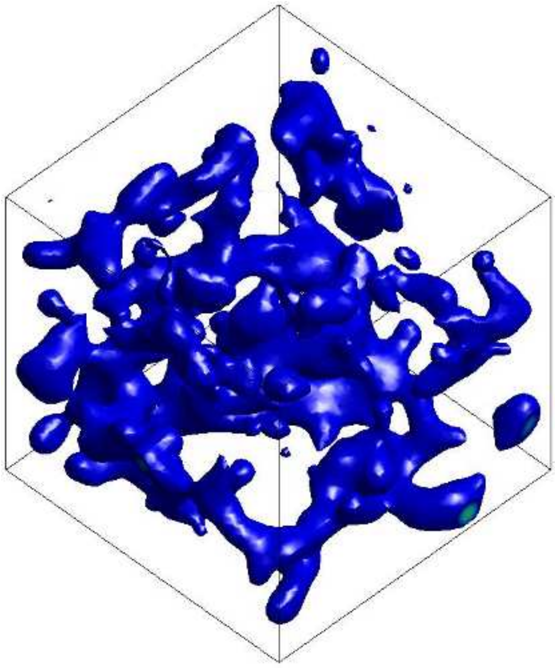
On the wind that was rustling the palms I thought I heard laughter, and a voice from above that said, "Dis be it, Phill-eep . . ."

I wasn't dreaming.

– Theodore Taylor, *Timothy of the Cay*

Chapter 6

Summary



A cube of side 160 h^{-1} Mpc at $0.06 < z < 1.12$ encompassing the scale of homogeneity, showing the Gaussian-smoothed ($\sigma = 5\ h^{-1}$ Mpc) galaxy distribution (Appendix B). The filamentary network is clearly visible, as well as a large void. This is the same view as shown on the title page of this Thesis.

We have designed and implemented a new algorithm, multiscale probability mapping, for the detection of structures in the galaxy distribution. MSPM can be made sensitive to any chosen range of scales and identifies member galaxies. Our work with SDSS DR7 data demonstrates its abilities:

- by finding groups and clusters with a range of sizes we have quantified a radius-velocity dispersion trend not highlighted in previous work,
- by identifying groups and clusters through their statistical significance we are able to set a relatively low threshold,
- MSPM's sensitivity to a user-defined scale range allows us to produce a coarse-grained representation of the galaxy distribution with a user-defined grain size, and
- using our group and cluster catalogue, we have demonstrated a technique to identify filaments algorithmically with a false discovery rate of less than 50 per cent.

We have used our group and cluster catalogue as a coarse-grained representation of the galaxy distribution for structure sizes of $\lesssim 1\,h^{-1}$ Mpc to search for large-scale structure, identifying 53 filaments and 47 voids that are $40\,h^{-1}$ Mpc across. Our approach to filament-finding is the first to achieve a false discovery rate of less than one half in real data, and our work with voids demonstrates a unified approach to identifying underdense and overdense features of large-scale structure. The morphological similarity of our filaments to those of PDH shows that algorithmic filament searches have the potential to produce results comparable to visual inspections of real data. However, our filament catalogue omits many filaments present in the data, and we lack an objective way to quantify their morphology. Future approaches will address these challenges.

As well as studying large-scale structure we have considered a cosmology motivated by its presence, the timescape model. A test to distinguish TS from ΛCDM has been outlined and attempted, but significant systematic issues remain unresolved. Resolving these issues poses a challenge, but the precision required to separate the two models is not great, and not precluded by the size of modern group and cluster catalogues. The MSPM catalogue presented in this Thesis is the most model-independent of the catalogues we have examined, but it also lacks calibration to simulated data. A volume-limited study with a well-motivated threshold to select virialised components of bound systems, or direct comparison with simulated data, has the potential to address the systematic issues we have discussed and provide meaningful cosmological constraints.

The data products made available in this work are a catalogue of 10443 groups and clusters at $0.025 < z < 0.24$, a catalogue of 53 filaments and a catalogue of 47 voids. These products, three-dimensional visualisations and further information about MSPM can be found at http://www.physics.usyd.edu.au/sifa/Main/MSPM/ .

Epilogue: Message for the Arch Vile

Here's the truth: People, even regular people, are never just any one person with one set of attributes. It's not that simple. We're all at the mercy of the limbic system, clouds of electricity drifting through the brain. Every man is broken into twenty-four-hour fractions, and then again within those twenty-four hours. It's a daily pantomime, one man yielding control to the next: a backstage crowded with old hacks clamoring for their turn in the spotlight. Every week, every day. The angry man hands the baton over to the sulking man, and in turn to the sex addict, the introvert, the conversationalist. Every man is a mob, a chain gang of idiots.

This is the tragedy of life. Because for a few minutes of every day, every man becomes a genius. Moments of clarity, insight, whatever you want to call them. The clouds part, the planets get in a neat little line, and everything becomes obvious. I should quit smoking, maybe, or here's how I could make a fast million, or such and such is the key to eternal happiness. That's the miserable truth. For a few moments, the secrets of the universe are opened to us. Life is a cheap parlor trick.

But then the genius, the savant, has to hand over the controls to the next guy down the pike, most likely the guy who just wants to eat potato chips, and insight and brilliance and salvation are all entrusted to a moron or a hedonist or a narcoleptic.

– Jonathan Nolan, *Memento Mori*

I am probably violating some rule of artistic discourse (Goethe's, I think) by giving spoilers that reveal the meaning of the florid quotations in this Thesis, but here goes. Along with music for the closing credits ;)

Kubla Khan (page iii) – *Allegro non troppo*

– Pyotr Tchaikovsky, opening theme from I: *Allegro non troppo e molto maetoso* in Piano Concerto Number One.

Kubla Khan was the Mongol barbarian king who completed the arduous project started by his grandfather Genghis Khan, the conquest of China, in 1276. He received the Venetian explorer Marco Polo, who returned to Europe telling of opulent splendour in the Far East. His tales included a description of Kublai's court in Xanadu, that inspired Coleridge's poem[1].

Xanadu's initial, X, was MSPM's original name. The river, sunless sea, walls and towers are identified respectively with a filament, void, bubble walls and clusters. When brought to life in three dimensions, these features of the galaxy distribution have all the majesty attributed to the grounds of Kubla's palace. Speaking of visualisation, the sunny dome and caves of ice refer to the two methods by which I rendered the galaxy distribution (Appendix B): the solid particle renders and Gaussian isosurfaces, which after a fashion, I did indeed build.

When the river is said to flow to the sunless sea, this is apparently at odds with filaments channelling mass onto clusters, but exactly how the Universe is to be likened to a landscape is ambiguous. The gravitational potential is higher in the voids, meaning that mass flows away from them, but rivers flow to seas, and clusters aren't reminiscent of seas at all. The voids are the seas, the sunless seas. As for what they are seas of, and what flows into them, that's another story!

[1] thanks wikipedia

Lost (page 3), Found (page 98) – *Under a Red Sky*

– Espen Gätzschmann & Tore Aune Fjellstad, *Under a Red Sky* (available online[1]) in *The Ur-Quan Masters*. Based on the original by Dan Nicholson.

In *Star Control II* an impenetrable blood red shield seals Earth off from the outside, like our two heroes on the cay. Continuing the topographical theme, let's flood our landscape, such that the towers are the only features that rise above sea level. These become *islands*, cut off from one another by the ocean just as Timothy and Phillip are cut off from civilisation. In MSPM we may not be interested in every island, but those islands that are also *banks* (Figure 1.1), towering above their surrounds. Just as not every island is a bank, not every bank is an island. These are submerged hills, "hombug banks" as Timothy calls them.

With our two heroes lost on an island out there, how do we find them? Finding all the islands is a good start, and that's what MSPM does. But like all mapping algorithms, MSPM must be guided by choices, and these choices introduce selection effects that colour what we see (Section 2.2.4). *Choice* itself is an interesting concept. There are always reasons for selecting one choice over others, implying some determinism. But the freedom to make a choice simultaneously implies some indeterminacy.

So as we set out to find the right island, we are forced to make choices. We make these choices to obtain *measurements* – choices guide MSPM to make its probability and scale measurements. MSPM is an instrument of measurement as well as of choice. We can't measure without making a choice, nor can we choose without measuring – because we can't decide without information, and we have no information without measurement. Also also, wherever am I going with all this?

[1] Available from http://www.medievalfuture.com/precursors/remix.php?remix=1

Apart (page 24) – *The Leaning Tower of Babel*

– Evil Horde, *The Leaning Tower of Babel* (available online[1]) in *The Dark Side of Phobos*. Based on Robert 'Bobby' Prince's *Nobody Told Me About id* for the *Doom* map *Tower of Babel*.

The biblical tower of Babel was intended to reach the heavens: a bridge between the finite and the infinite. Fittingly, in TS there are *finite infinity* regions that contain groups and clusters, the most massive gravitationally-bound structures in the Universe. The process by which these brutes form is much the same as how a tower is built: pile stuff upon stuff upon stuff. This is why groups and clusters are called towers of Babel in Section 2.1's title. Before the biblical tower could be finished, humanity is said to have been scattered all over the world, flung apart like the most massive clusters. Our towers of Babel seem to be just as futile, piling mass upon mass only makes a bottomless pit.

Similarly abject situations are described in the epigraph on page 24. The first illustrates that we can be seemingly separated from one another by forces beyond our control. The second illustrates a situation in which we cannot run fast enough to overcome what pulls us apart. These two cases will be revisited later.

In the third situation, one party is not only separated from the other, but also not allowed to see the other. We have noted that the determinism/indeterminacy duality appears to be a fundamental feature of our reality, and it manifests formally in quantum mechanics. The wave function evolves deterministically according to the Schrödinger equation, but since it only describes the evolution of probabilities, specific measurement outcomes are not determined (it is only the expectation values that are). When a measurement is made, the wave function probabilistically collapses: an individual observer will only encounter one specific outcome of the experiment. When we are separated from our friend and not allowed to see them, we are uncertain about their condition and they become a superposition of all the states they could be in[2].

[1] Available from http://doom.ocremix.org/main.html

[2] This situation is close to what is known in quantum mechanics as "Wigner's friend" and is an illustration of the quantum measurement problem.

Our friend becoming a superposition is a problem. When *we* make measurements, we do not become superpositions. There appears to be something about consciousness (or the measurement apparatus commandeered by a conscious observer) that collapses the wave function. If our friend becomes a superposition he is apparently not conscious and we are alone[1]. This situation is avoided in the many worlds interpretation of quantum mechanics, but in that case the probability of us ever meeting the same version of the friend we knew before quickly becomes infinitesimal, which does not really improve the situation.

The epigraph on page 24 thus illustrates three situations that I attribute to a villain I dub the *Arch Vile*.

[1] And here lies another way to pose the quantum measurement problem: if events must be agreed to by more than one observer to be "real", what is real?

Desert Passage (page 54) – *Sign of the Worm*

– Stéphane Picq, *Sign of the Worm* in *Dune: Spice Opera*[1].

Between the towers of our cosmic landscape are walls, threaded by filaments that channel mass onto clusters. These filaments often run through space that is otherwise empty, like rivers through desert (e.g. Figure 3.1), hence the desert analogy on page 54. In this analogy filaments are represented by sandworms, which in *Dune* are a means of transport through the desert (hence "desert passage") and suitably gargantuan. Left to themselves, galaxies redden and die as they exhaust their star formation potential. Mass supplied by filaments helps to keep clusters going, making filaments a source of life. Sandworms are similarly a source of life: spice derived from them has geriatric properties The related Water of Life (as mentioned in the first part of the epigraph) enables bridges between minds.

On that note, we return to our earlier problem, the Arch Vile driving us apart. We begin by making precise our idea that filaments are a source of life. There are vague notions of what life is, and also inelegantly specific definitions that pertain to what are conventionally regarded as organisms. The definition I like best is that of *observer*, identifiable by the act of measurement. It makes sense: consciousness is what seems to make measurements to collapse wave functions, and on page 103 we linked measurement and choice, which is a self-evident property of existence to most living things. Measurement requires thermodynamic free energy that produces entropy. Turning this around, entropy production is a tracer of life.

In our Universe, starlight produces the most entropy. The massive elliptical galaxies that lie at the cores of clusters are often described as "red and dead" – and because they are not forming stars, they really are! We noted that in quantum mechanics we are alone. General relativity is ambivalent in this regard because while it accommodates as many observers as we please, it doesn't actually require that any exist. So is there a comparable theoretical framework that does? It's called the anthropic principle, and the *causal entropic principle* is a variant that identifies observers in the way we've described. The causal entropic principle is successful when coupled with the observed star formation rate history (Cline, Frey & Holder 2008) and contains an awful lot of observers, so this particular challenge of the Arch Vile has been answered.

[1]http://en.wikipedia.org/wiki/Spice_Opera#Audio

Awake To Emptiness (page 66) – *Mystic Shadows*

– Kevin Palivec, *Mystic Shadows* in *Star Control II*. Remixed by Jouni Airaksinen as *Mystic Shadows* (available online[1]) in *The Ur-Quan Masters*.

Mystic Shadows is the track that plays in *Star Control II*'s QuasiSpace, a plane of existence with rather different properties to normal space. Cosmic voids might not be quite so different, but the lower matter density along with general relativity introduces a modicum of difference at least. Or more than a modicum, if the TS cosmology's treatment of the passage of time there is correct. In the cosmic landscape voids are the sunless seas, in which TS claims the age of the Universe is gigayears greater, so that we may say it is time itself that flows to them via the rivers in *Kubla Khan*, to complete our earlier analogy.

On page 66, Kyu is banished from the void by a mere thought: if there were thinking beings in the void it wouldn't be much a void now, would it? But just as filaments are a source of life for clusters, the voids are in turn a source of life for filaments. Accordingly, in the third part of the epigraph Bold speaks of souls born out of the void. When he says it happens infrequently this is also analogous to the current cosmic situation, as the voids have less to give in the current epoch. Indeed, identifying this as an Age of Destruction is apt, because when all the mass eventually flows to the clusters, the cosmic web of structure will be lost.

In the first part of the epigraph a sutra is quoted: form is emptiness, emptiness form. A structure can be marked by overdensity, but it can also be marked by the complementary underdensity. This is a prosaic observation by itself, but there is more to the emptiness that surrounds the events that directly interest us if we generalise the concept of void to *environment*. A projectile is motionless if there is no stationary background by which to gauge its motion. *Mach's principle* advocates the influence of environment on local events, suggesting that the environment is not a mere bystander but a player. Inertial frames are empirically found to be stationary or moving with constant velocity with respect to the background defined by distant objects in the Universe The relevance of Mach's principle in TS is discussed by Wiltshire (2009b). So this along with the voids sets the scene for the timescape and our next confrontation with the Arch Vile.

[1] Available from http://www.medievalfuture.com/precursors/remix.php?remix=34

The Eternal Now (page 77), Islands and Banks (page 2), Large-Scale Structure (page 53), *Memento Mori* (page 101) – *Message for the Archvile*

– Robert 'Bobby' Prince, *Message for the Archvile* in *Doom II: Hell on Earth*. Remixed by Mazedude as *Westside Archvile* (available online[1]) in *Delta-Q-Delta*.

The epigraph on page 77 depicts time travel via a device called a kettle. A kettle is a bit different to time travel toys that appear in other tales of chronological daring. Instead of behaving like a vehicle with an extra degree of freedom, it moves like an elevator in a kettle shaft, up and down the dimension of time. This version of time travel is closest to what we would have in TS because the shaft exists for each accessible time just as in TS, each accessible age of the universe has a corresponding location or locations in physical space. Physical space in a TS universe is like a (three-dimensional) kettle shaft.

The next part of the epigraph concerns a division between those who know the truth, and those who do not. History is constantly reviewed and reshaped in a utilitarian fashion by time travellers called Eternals who reside in Eternity, in which the kettles move. The details of this manipulation of reality are not transparent to ordinary denizens of Time who form the vast majority, the Timers. The parallel here is with the suggestion in TS that the true description of our Universe is hidden from the vast majority of us because observers live in bound systems, with slow clocks (Section 5.1.2.2). We have to recalibrate our measurements to account for this, but an observer in a volume-average location, which in the current epoch Universe is in a void, does not. These volume-average void observers are like the Eternals: they can see the truth directly, and are a vanishingly small minority.

For the third part of the epigraph, insert corny waffle about the binding forces that oppose the Arch Vile.

The last part of the epigraph is about reality and the conditions it must satisfy. Along the lines of the anthropic principle, which we've met, the fact of our consciousness and related self-evident features of our existence place boundary conditions on reality. One such feature I've mentioned is choice, which implies (at least) the determinism/indeterminacy duality that indeed turns out to be empirical fact. This reasoning from the "inside out" or from the "self outward" must of course be met from the outside by empirical data supplied by Nature.

In the quoted passage, the constraint to be satisfied by reality is that we cannot know our own future. If in TS we are effectively looking into the future when we observe voids, is this constraint violated? More years might have passed in a void so

[1] Available from http://doom2.ocremix.org/tracks/

we see *a* future, but it is the future of a region of the Universe that was underdense in the primordial density field. Since we are in a bound system that was primordially overdense, what we see is not *our* future. So it is as Sennor says in the epigraph: reality changes to correct our knowledge. It is the inhomogeneity that leads to clock rate variance and sight of the future, but it is that very same inhomogeneity that guarantees it is not our own future.

On the left is the upper-case Greek letter lambda that symbolises dark energy, and it has the shape of an arch. On the right is an Arch-vile[1] adversary from *Doom*, about to do some damage. The Arch-vile's attack sends its hapless target flying, kind of like dark energy. For these reasons I characterise (or demonise?) dark energy as the Arch Vile. On page 104, one of the situations I attributed to the Arch Vile was that of being unable to run quickly enough to reach a destination. This is analogous to recession velocities of distant objects that exceed the speed of light, according to the Hubble flow. Dark energy supposedly makes this situation worse by accelerating this recession, defying gravitational braking such that these places will never be accessible to us.

Every villain faces off against a hero, and ours turned out to be Hercules. When in Chapter 5 I realised that I actually didn't have a way to infer virial radii from power-law fits, previous measurements of the *Hercules* cluster at least showed that we weren't far off (Section 5.3.2). It's actually quite suggestive that our measurements, which involve few assumptions, coincide with values found for this well-studied cluster. So *Hercules* added weight to the result favouring TS, a model without dark energy, without the Arch Vile. Its namesake, the mythical hero thus features in the second part of the epigraph to Part I on page 2. The first part of that epigraph refers to the islands and banks (the $P > 0.5$ and $S < 0.5$ regions, respectively) that are used to produce the coarse grains with which Part I is concerned. In *The Odyssey*, Odysseus visits one after the next in his quest to return home, hindered by godly opposition with seas stirred against him as he is shipwrecked and loses all his comrades. Getting MSPM to work was also testing at times.

The epigraph to Part II on page 53 is from another legendary adventure. The first part alludes to cosmic structure formation from the very beginning, hence the "Divine

[1] http://doom.wikia.com/wiki/Arch-vile

Root" and oneness of Heaven and Earth. The "shapeless blur" refers to the surface of last scattering we see as the CMB, which is shapeless because it is very smooth, with density contrasts of order 10^{-5}. Since then, structure formation has formed overdensities and underdensities, the "separation of clear and impure" (Figure 3.2, Figure 4.3 and Appendix B). The second part of the epigraph is about TS as we go from landscape to timescape. In a fanciful trip across the Universe, one's age measured relative to the Beginning would fluctuate in a timescape.

Unfortunately, the Arch Vile is a "neurospheric" feature of reality like choice, and works to separate us internally from our own memories and thoughts, motivating the quotation from *Memento Mori* on page 101. At the end we proposed a test of the timescape to hopefully find out if we can do without dark energy. If we can, far from exerting its apparently dominating force, it was only ever our own creation. In the cosmological arena we may yet defeat it, and that is our message for the Arch Vile.

Appendices

Appendix A

Filament and Void Catalogues

These catalogues are also available at http://www.physics.usyd.edu.au/sifa/Main/MSPM/ , along with three-dimensional visualisations and figures.

Table A.1: Catalogue of MSPM filaments in SDSS DR7.

ID (1)	RA (J2000) (deg) (2)	Dec (J2000) (deg) (3)	z (4)	P_e(max) (5)	N (6)	Morphology (7)
68........	222.1034	18.3560	0.03999	0.924	10	V
84........	123.1876	16.8883	0.04459	0.904	10	II
299........	235.6901	8.2411	0.04039	0.894	11	I
404........	165.3420	9.3080	0.03641	0.908	10	I
663........	250.8133	24.0732	0.04696	0.905	10	II
1063........	169.6346	53.8145	0.03470	0.947	7	I
1082........	238.3466	18.3778	0.03280	0.910	15	V
1088........	149.5960	8.9077	0.04900	0.921	5	I
1494........	196.6982	60.3546	0.02784	0.895	6	I
1511........	203.9377	27.8676	0.02641	0.890	8	V
1541........	174.8082	35.9602	0.03973	0.904	8	V
1946........	208.3502	19.3225	0.07120	0.926	6	I
2088........	201.0014	59.0449	0.07286	0.918	8	I
2547........	119.4981	40.0377	0.06630	0.908	7	I
2770........	174.2258	44.2294	0.05886	0.894	4	II
2937........	234.3922	14.3926	0.05201	0.910	7	I
3086........	184.3658	-0.7805	0.07020	0.878	3	V
3094........	162.2009	4.3068	0.06952	0.889	9	V
3305........	241.0991	11.0421	0.06452	0.924	5	I
3502........	223.0171	17.3097	0.05819	0.920	9	II
3656........	193.7737	38.6280	0.05186	0.894	5	II
3750........	233.4721	6.8433	0.06580	0.878	3	II
3861........	195.1090	52.6712	0.05473	0.919	8	II
4074........	139.3883	53.2514	0.05782	0.924	5	II
4604........	155.2910	9.8616	0.09877	0.898	5	I
4647........	175.4488	56.7553	0.09737	0.878	3	V
4885........	148.4584	19.9723	0.08844	0.904	4	I
4976........	194.8470	29.9823	0.08425	0.878	3	I
5021........	175.1702	10.0262	0.08225	0.887	9	I
5149........	210.5606	5.9266	0.07834	0.878	6	V
5381........	211.1576	41.8530	0.09360	0.878	3	II
5527........	183.7894	36.0110	0.08906	0.878	3	I
5745........	236.1185	29.6604	0.08242	0.904	8	V
5832........	228.2023	20.8819	0.07962	0.920	8	II
5853........	226.7288	7.1890	0.07958	0.943	13	V
5935........	181.9290	23.8841	0.07748	0.919	9	I
5972........	189.9455	13.8891	0.07587	0.905	4	II
5984........	183.5435	17.7961	0.07697	0.905	4	I
6007........	119.5306	40.8289	0.07555	0.878	3	I
6252........	198.6709	19.9712	0.09046	0.919	5	II
6280........	229.2817	3.5373	0.08093	0.905	7	I
6574........	149.8899	3.1282	0.08179	0.936	6	II
6656........	203.1080	40.0689	0.08085	0.878	3	I
6695........	135.0231	53.7212	0.09165	0.894	4	I
6760........	148.4709	20.6304	0.07863	0.937	6	II
6846........	227.7999	5.6538	0.08455	0.905	4	I
6858........	170.3387	38.3315	0.08609	0.905	6	II
6920........	150.8208	18.5711	0.07888	0.924	5	I
7039........	237.9363	45.5654	0.11889	0.878	3	I
7642........	139.4112	36.5945	0.11032	0.878	3	V
8048........	118.0812	36.1567	0.11426	0.878	3	II
8555........	156.8494	11.1645	0.11731	0.921	5	I
9625........	163.9845	40.7248	0.12916	0.905	4	I

Locations, properties and morphologies of MSPM filaments. The sample is discussed in Chapter 3 and Smith et al. (2012). This catalogue is also available at http://www.physics.usyd.edu.au/sifa/Main/MSPM/ , along with three-dimensional visualisations and figures showing each filament.

Column (1): ID of central MSPM group or cluster; (2) to (4): centre position; (5): highest elongation probability within 10 h^{-1} Mpc; (6): count of groups and clusters within a 10 h^{-1} Mpc radius on the sky and 10 h^{-1} Mpc in the line of sight (cylindrical aperture); (7): morphological type (Section 3.4).

Table A.2: Catalogue of 40 h^{-1} Mpc MSPM voids in SDSS DR7.

ID (1)	RA (J2000) (deg) (2)	Dec (J2000) (deg) (3)	z (4)
16110........	197.3890	19.6361	0.04500
22406........	215.3630	45.8330	0.05500
26786........	142.3790	17.0212	0.06000
30110........	154.7770	31.1830	0.06500
33619........	217.8010	8.4892	0.07000
33952........	135.9970	22.2747	0.07000
35795........	158.1730	22.6359	0.07250
36559........	186.5200	52.4890	0.07500
38881........	202.6630	51.1090	0.07750
39279........	234.7750	21.0329	0.07750
40506........	175.4650	44.6020	0.08000
40699........	245.6250	32.5760	0.08000
42116........	233.8290	43.1790	0.08250
42939........	200.9950	26.4868	0.08250
43788........	157.7830	54.3840	0.08500
44807........	144.8920	17.6372	0.08500
45996........	191.9280	42.3300	0.08750
46586........	154.2300	24.1352	0.08750
46675........	178.5440	17.2228	0.08750
46685........	239.6280	12.8591	0.08750
47728........	217.0760	37.8180	0.09000
47835........	139.6980	32.7990	0.09000
48026........	205.9970	6.8914	0.09000
48146........	162.4230	30.3100	0.09000
48209........	136.3210	24.4600	0.09000
48554........	186.2570	22.5313	0.09000
48603........	209.5160	20.7059	0.09000
48912........	124.7200	43.7750	0.09250
49082........	144.6190	52.9490	0.09250
49316........	211.0950	52.1140	0.09250
49549........	230.8300	37.1300	0.09250
49796........	194.7540	14.7160	0.09250
50023........	174.0710	33.6840	0.09250
50041........	208.7370	30.0585	0.09250
51144........	231.2170	50.0670	0.09500
51452........	169.9550	12.0130	0.09500
52731........	138.4380	44.4390	0.09750
53043........	223.7300	42.8980	0.09750
53747........	193.6730	27.3407	0.09750
53786........	145.2230	24.1405	0.09750
53954........	201.0870	19.2293	0.09750
54550........	195.7990	4.8680	0.10000
55015........	205.4750	45.7480	0.10000
55169........	135.0510	9.8450	0.10000
55366........	203.0700	34.7950	0.10000
55716........	142.7850	15.1658	0.10000
55822........	231.2320	16.2831	0.10000

Locations of MSPM voids that are 40 h^{-1} Mpc across. The sample is discussed in Chapter 4. The complete catalogue can also be found at http://www.physics.usyd.edu.au/sifa/Main/MSPM/ , along with three-dimensional visualisations of each void.
Column (1): ID of position in sampling lattice; (2) to (4): position.

Appendix B

Visualisation: Universe in a Box

In this Thesis we have presented various figures (including Figures 3.2 and 4.3) and renders (as title images for each chapter) that show cosmic structure evidenced by the galaxy distribution. We have also made animated spins of 20 groups and clusters (from Table 2.1), 53 filaments (Table A.1) and 47 voids (Table A.2), along with a box that encompasses the rumoured scale of statistical homogeneity (shown on the title page of this Thesis and page 99), at http://www.physics.usyd.edu.au/sifa/Main/MSPM/ .

For each field, three modes of visualisation have been produced (Figure B.1): particle plots that show the raw galaxy distribution, isosurfaces based on the smoothed galaxy density field and solid particle renders.

In solid particle renders, MSPM groups and clusters are Gaussian blurs with various properties colour-coded. Overdensity probabilities between 0.5 and 1 are mapped to blue values between 0 and 1, and this is also the intensity of each group. Radius is red such that a $\sim 1\,h^{-1}$ Mpc structure has a red value of one and a $0.5\,h^{-1}$ Mpc structure has a red value of 0.5, etc. The radius (σ) of the Gaussian blur is the radius of the structure. Elongation values used for the green colours are averages of the five values obtained within $10\,h^{-1}$ Mpc. The actual elongation probabilities calculated for our paper are maximum values that are the same for many adjacent groups and not as visually interesting to use.

To see them, visit
http://www.physics.usyd.edu.au/sifa/Main/MSPM/ .

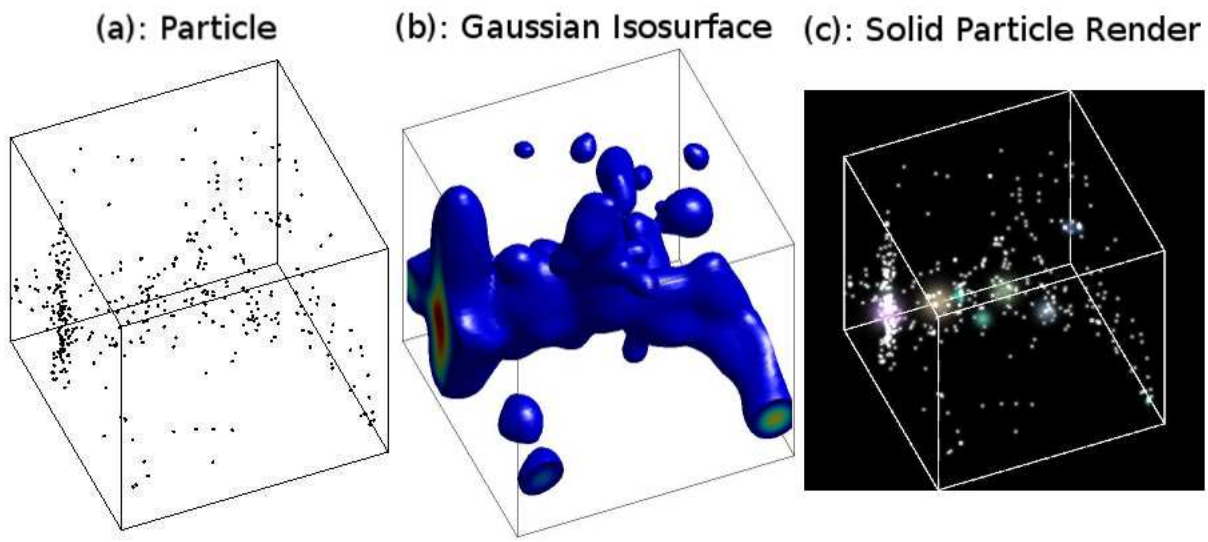


Figure B.1: Three modes of visualisation of MSPM filament 1063 (also shown in Figure 3.1), from the same viewing angle. **(a)** The galaxy distribution, with galaxies as black markers. **(b)** The filament as a contiguous structure. The raw galaxy distribution has been smoothed in three dimensions with a Gaussian filter. The surfaces shown follow contours of twice the mean density. **(c)** A solid particle render with individual galaxies plotted as spoked white dots with radii of $0.125\,h^{-1}$ Mpc. MSPM groups and clusters are Gaussian blurs with various properties colour-coded.

Bibliography

Abazajian, K. N., et al. 2009, ApJS, 182, 543

Abell, G. O. 1958, ApJS, 3, 211

Abell, G. O., Corwin, Jr., H. G., Olowin, R. P. 1989, ApJS, 70, 1

Aikio, J. Maehoenen, P. 1998, ApJ, 497, 534

Allen, S. W., Schmidt, R. W., Fabian, A. C. 2002, MNRAS, 334, L11

Aragón-Calvo, M. A., Jones, B. J. T., van de Weygaert, R., van der Hulst, J. M. 2007, A&A, 474, 315

Aragón-Calvo, M. A., van de Weygaert, R., Jones, B. J. T. 2010, MNRAS, 408, 2163 (AWJ)

Bahcall, N. A. Soneira, R. M. 1983, ApJ, 270, 20

Bahcall, N. A., et al. 2003, ApJS, 148, 243

Baldry, I. K., Balogh, M. L., Bower, R. G., Glazebrook, K., Nichol, R. C., Bamford, S. P., Budavari, T. 2006, MNRAS, 373, 469

Balian, R. Schaeffer, R. 1989, A&A, 220, 1

Balogh, M., et al. 2004a, MNRAS, 348, 1355

Balogh, M. L., Baldry, I. K., Nichol, R., Miller, C., Bower, R., Glazebrook, K. 2004b, ApJ, 615, L101

Balogh, M. L., et al. 2007, MNRAS, 374, 1169

Barmby, P. Huchra, J. P. 1998, AJ, 115, 6

Barrow, J. D., Bhavsar, S. P., Sonoda, D. H. 1985, MNRAS, 216, 17

Beers, T. C., Flynn, K., Gebhardt, K. 1990, AJ, 100, 32

Bennett, C. L., et al. 2003, ApJS, 148, 1

Berlind, A. A., et al. 2006, ApJS, 167, 1 (Mr20)

Bhavsar, S. P. Ling, E. N. 1988, ApJ, 331, L63

Bhavsar, S. P. Splinter, R. J. 1996, MNRAS, 282, 1461

Bird, C. M., Dickey, J. M., Salpeter, E. E. 1993, ApJ, 404, 81

Biviano, A. 2000, in Constructing the Universe with Clusters of Galaxies

Böhringer, H., et al. 2000, ApJS, 129, 435

Böhringer, H., et al. 2004, A&A, 425, 367

Bregman, J. N. 2007, ARA&A, 45, 221

Brough, S., Collins, C. A., Burke, D. J., Mann, R. G., Lynam, P. D. 2002, MNRAS, 329, L53

Bryan, G. L. Norman, M. L. 1998, ApJ, 495, 80 (BN98)

Buchert, T. 2000, General Relativity and Gravitation, 32, 105

Buchert, T. 2008, General Relativity and Gravitation, 40, 467

Buchert, T. Carfora, M. 2003, Physical Review Letters, 90, 031101

Butcher, H. Oemler, Jr., A. 1978, ApJ, 219, 18

Caballero, J. A. Dinis, L. 2008, ArXiv e-prints, 807

Capak, P., Abraham, R. G., Ellis, R. S., Mobasher, B., Scoville, N., Sheth, K., Koekemoer, A. 2007, ApJS, 172, 284

Carlstrom, J. E., Holder, G. P., Reese, E. D. 2002, ARA&A, 40, 643

Cassata, P., et al. 2007, ApJS, 172, 270

Cen, R., Gnedin, N. Y., Kofman, L. A., Ostriker, J. P. 1992, ApJ, 399, L11

Clarkson, C., Clifton, T., Coley, A., Sung, R. 2012, Phys. Rev. D, 85, 043506

Cline, J. M., Frey, A. R., Holder, G. 2008, Phys. Rev. D, 77, 063520

Clowe, D., Trentham, N., Tonry, J. 2001, A&A, 369, 16

Colberg, J. M. 2007, MNRAS, 375, 337

Colberg, J. M., Krughoff, K. S., Connolly, A. J. 2005a, MNRAS, 359, 272 (CKC)

Colberg, J. M., Sheth, R. K., Diaferio, A., Gao, L., Yoshida, N. 2005b, MNRAS, 360, 216

Colberg, J. M., White, S. D. M., Jenkins, A., Pearce, F. R. 1999, MNRAS, 308, 593

Cole, S. Lacey, C. 1996, MNRAS, 281, 716

Collins, C. A., Guzzo, L., Nichol, R. C., Lumsden, S. L. 1995, MNRAS, 274, 1071

Connolly, A. J., Szalay, A. S., Koo, D., Romer, A. K., Holden, B., Nichol, R. C., Miyaji, T. 1996, ApJ, 473, L67+

Cooper, M. C., et al. 2007, MNRAS, 376, 1445

Cooper, M. C., et al. 2008, MNRAS, 383, 1058

Corwin, Jr., H. G. 1974, AJ, 79, 1356

Cucciati, O., et al. 2006, A&A, 458, 39

Dalcanton, J. J. 1996, ApJ, 466, 92

Dalton, G. B., Efstathiou, G., Maddox, S. J., Sutherland, W. J. 1994, MNRAS, 269, 151

de Lapparent, V., Geller, M. J., Huchra, J. P. 1986, ApJ, 302, L1

de Putter, R. White, M. 2005, New Astronomy, 10, 676

Diaferio, A., Geller, M. J., Rines, K. J. 2005, ApJ, 628, L97

Dong, F., Pierpaoli, E., Gunn, J. E., Wechsler, R. H. 2008, ApJ, 676, 868

Donnelly, R. H., Forman, W., Jones, C., Quintana, H., Ramirez, A., Churazov, E., Gilfanov, M. 2001, ApJ, 562, 254

Dressler, A. 1980, ApJ, 236, 351

Dressler, A., et al. 1997, ApJ, 490, 577

Ebeling, H., Barrett, E., Donovan, D. 2004, ApJ, 609, L49

Ebeling, H., Voges, W., Bohringer, H., Edge, A. C., Huchra, J. P., Briel, U. G. 1996, MNRAS, 281, 799

Ebeling, H. Wiedenmann, G. 1993, Phys. Rev. E, 47, 704

Einasto, J., Klypin, A. A., Saar, E., Shandarin, S. F. 1984, MNRAS, 206, 529

Einasto, M., Einasto, J., Tago, E., Müller, V., Andernach, H. 2001, AJ, 122, 2222

Einasto, M., et al. 2010, A&A, 522, A92

Eisenhardt, P. R. M., et al. 2008, ArXiv e-prints, 804

Eisenstein, D. J., et al. 2001, AJ, 122, 2267

Eke, V. R., Cole, S., Frenk, C. S. 1996, MNRAS, 282, 263

El-Ad, H. Piran, T. 1997, ApJ, 491, 421

Elbaz, D., et al. 2007, A&A, 468, 33

Erben, T., van Waerbeke, L., Mellier, Y., Schneider, P., Cuillandre, J.-C., Castander, F. J., Dantel-Fort, M. 2000, A&A, 355, 23

Escalera, E. MacGillivray, H. T. 1995, A&A, 298, 1

Escalera, E. Mazure, A. 1992, ApJ, 388, 23

Escalera, E., Slezak, E., Mazure, A. 1992, A&A, 264, 379

Ester et al. Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining (KDD96). AAAI Press, Menlo Park, CA, pp. 226-231, August. Available from http://dns2.icar.cnr.it/manco/Teaching/2005/datamining/articoli/KDD-96.final.frame.pdf. 1996

Evrard, A. E., et al. 2002, ApJ, 573, 7

Feng, L., Deng, Z.-G., Fang, L.-Z. 2000, ApJ, 530, 53

Fleenor, M. C., Rose, J. A., Christiansen, W. A., Hunstead, R. W., Johnston-Hollitt, M., Drinkwater, M. J., Saunders, W. 2005, AJ, 130, 957

Fraser-McKelvie, A., Pimbblet, K. A., Lazendic, J. S. 2011, MNRAS, 415, 1961

Frieman, J. A., Turner, M. S., Huterer, D. 2008, ARA&A, 46, 385

Gal, R. R., de Carvalho, R. R., Lopes, P. A. A., Djorgovski, S. G., Brunner, R. J., Mahabal, A., Odewahn, S. C. 2003, AJ, 125, 2064

Gerke, B. F., et al. 2007, MNRAS, 376, 1425

Gilbank, D. G., Yee, H. K. C., Ellingson, E., Gladders, M. D., Barrientos, L. F., Blindert, K. 2007, AJ, 134, 282

Gilbank, D. G., Yee, H. K. C., Ellingson, E., Gladders, M. D., Loh, Y.-S., Barrientos, L. F., Barkhouse, W. A. 2008, ApJ, 673, 742

Gioia, I. M. Luppino, G. A. 1994, ApJS, 94, 583

Girardi, M., Biviano, A., Giuricin, G., Mardirossian, F., Mezzetti, M. 1995, ApJ, 438, 527

Girardi, M., Giuricin, G., Mardirossian, F., Mezzetti, M., Boschin, W. 1998, ApJ, 505, 74

Gladders, M. D. Yee, H. K. C. 2000, AJ, 120, 2148

Gladders, M. D. Yee, H. K. C. 2005, ApJS, 157, 1

Gómez, P. L., et al. 2003, ApJ, 584, 210

Gonzalez, A. H., Zaritsky, D., Dalcanton, J. J., Nelson, A. 2001, ApJS, 137, 117

Goto, T., et al. 2002, AJ, 123, 1807

Gottlöber, S., Łokas, E. L., Klypin, A., Hoffman, Y. 2003, MNRAS, 344, 715

Gunn, J. E. Gott, III, J. R. 1972, ApJ, 176, 1

Gunn, J. E. Oke, J. B. 1975, ApJ, 195, 255

Haas, M. R., Schaye, J., Jeeson-Daniel, A. 2012, MNRAS, 419, 2133

Hansen, S. M., McKay, T. A., Wechsler, R. H., Annis, J., Sheldon, E. S., Kimball, A. 2005, ApJ, 633, 122

Hashimoto, Y., Oemler, A. J., Lin, H., Tucker, D. L. 1998, ApJ, 499, 589

Hogg, D. W. 1999, ArXiv Astrophysics e-prints, arXiv:astro-ph/9905116

Hogg, D. W., Eisenstein, D. J., Blanton, M. R., Bahcall, N. A., Brinkmann, J., Gunn, J. E., Schneider, D. P. 2005, ApJ, 624, 54

Hogg, D. W., et al. 2004, ApJ, 601, L29

Hopkins, A. M., Afonso, J., Chan, B., Cram, L. E., Georgakakis, A., Mobasher, B. 2003, AJ, 125, 465

Hoyle, F. Vogeley, M. S. 2004, ApJ, 607, 751

Huchra, J. P. Geller, M. J. 1982, ApJ, 257, 423

Icke, V. van de Weygaert, R. 1987, A&A, 184, 16

Infante, L. 1994, A&A, 282, 353

Ishibashi, A. Wald, R. M. 2006, Classical and Quantum Gravity, 23, 235

Johnston, D. E., Sheldon, E. S., Tasitsiomi, A., Frieman, J. A., Wechsler, R. H., McKay, T. A. 2007, ApJ, 656, 27

Katgert, P., Biviano, A., Mazure, A. 2004, ApJ, 600, 657

Kauffmann, G. Fairall, A. P. 1991, MNRAS, 248, 313

Kauffmann, G., White, S. D. M., Heckman, T. M., Ménard, B., Brinchmann, J., Charlot, S., Tremonti, C., Brinkmann, J. 2004, MNRAS, 353, 713

Kawasaki, W., Shimasaku, K., Doi, M., Okamura, S. 1998, A&AS, 130, 567

Kepner, J., Fan, X., Bahcall, N., Gunn, J., Lupton, R., Xu, G. 1999, ApJ, 517, 78

Kiang, T. 1966, Zeitschrift fur Astrophysik, 64, 433

Kim, R. S. J., et al. 2002, AJ, 123, 20

King, L. J. Mead, J. M. G. 2011, MNRAS, 416, 2539

Kitayama, T. Suto, Y. 1996, ApJ, 469, 480

Kodama, T., Smail, I., Nakata, F., Okamura, S., Bower, R. G. 2001, ApJ, 562, L9

Koester, B. P., et al. 2007a, ApJ, 660, 221

Koester, B. P., et al. 2007b, ApJ, 660, 239

Komatsu, E., et al. 2011, ApJS, 192, 18 (WMAP7)

Kravtsov, A. V., Vikhlinin, A., Nagai, D. 2006, ApJ, 650, 128

Krzewina, L. G. Saslaw, W. C. 1996, MNRAS, 278, 869

Kuhn, J. R. Uson, J. M. 1982, ApJ, 263, L47

Kwan, J., Francis, M. J., Lewis, G. F. 2009, MNRAS, 399, L6

Lacey, C. Cole, S. 1993, MNRAS, 262, 627

Leith, B. M., Ng, S. C. C., Wiltshire, D. L. 2008, ApJ, 672, L91 (LNW)

Lewis, A. D., Buote, D. A., Stocke, J. T. 2003, ApJ, 586, 135

Lewis, I., et al. 2002, MNRAS, 334, 673

Lin, Y.-T., Mohr, J. J., Stanford, S. A. 2004, ApJ, 610, 745

Loh, Y. Strauss, M. A. 2006, MNRAS, 366, 373

Lucey, J. R. 1983, MNRAS, 204, 33

Ludlow, A. D. Porciani, C. 2011, ArXiv e-prints

Lumsden, S. L., Nichol, R. C., Collins, C. A., Guzzo, L. 1992, MNRAS, 258, 1

Majumdar, S. Mohr, J. J. 2004, ApJ, 613, 41

Mantz, A., Allen, S. W., Rapetti, D., Ebeling, H. 2010, MNRAS, 406, 1759

Martinez, V. J., Jones, B. J. T., Dominguez-Tenreiro, R., van de Weygaert, R. 1990, ApJ, 357, 50

Martinez, V. J., Paredes, S., Saar, E. 1993, MNRAS, 260, 365

McConnachie, A. W., Patton, D. R., Ellison, S. L., Simard, L. 2009, MNRAS, 395, 255

Miller, C. J., Batuski, D. J., Slinglend, K. A., Hill, J. M. 1999, ApJ, 523, 492

Miller, C. J., et al. 2005, AJ, 130, 968

Moore, B., Katz, N., Lake, G., Dressler, A., Oemler, A. 1996, Nature, 379, 613

Moran, S. M., Ellis, R. S., Treu, T., Smail, I., Dressler, A., Coil, A. L., Smith, G. P. 2005, ApJ, 634, 977

Mullis, C. R., et al. 2003, ApJ, 594, 154

Murphy, D. N. A., Eke, V. R., Frenk, C. S. 2011, MNRAS, 413, 2288

Muzzin, A., Wilson, G., Lacy, M., Yee, H. K. C., Stanford, S. A. 2008, ArXiv e-prints

Navarro, J. F., Frenk, C. S., White, S. D. M. 1997, ApJ, 490, 493

Novikov, D., Colombi, S., Doré, O. 2006, MNRAS, 366, 1201

Ochsenbein, F., Bauer, P., Marcout, J. 2000, A&AS, 143, 23

Ostrander, E. J., Nichol, R. C., Ratnatunga, K. U., Griffiths, R. E. 1998, AJ, 116, 2644

Ostriker, J. P. Tremaine, S. D. 1975, ApJ, 202, L113

Pan, D. C., Vogeley, M. S., Hoyle, F., Choi, Y.-Y., Park, C. 2011, ArXiv e-prints

Pandey, B., Kulkarni, G., Bharadwaj, S., Souradeep, T. 2011, MNRAS, 411, 332

Pando, J., Lipa, P., Greiner, M., Fang, L.-Z. 1998, ApJ, 496, 9

Park, D. Lee, J. 2007, Physical Review Letters, 98, 081301

Patiri, S. G., Betancort-Rijo, J. E., Prada, F., Klypin, A., Gottlöber, S. 2006, MNRAS, 369, 335

Peng, Y.-j., et al. 2010, ApJ, 721, 193

Perlmutter, S., et al. 1999, ApJ, 517, 565

Pimbblet, K. A. 2005, MNRAS, 358, 256

Pimbblet, K. A., Drinkwater, M. J., Hawkrigg, M. C. 2004, MNRAS, 354, L61 (PDH)

Pimbblet, K. A., Smail, I., Kodama, T., Couch, W. J., Edge, A. C., Zabludoff, A. I., O'Hely, E. 2002, MNRAS, 331, 333

Planck Collaboration, et al. 2011, A&A, 536, A8

Platen, E., van de Weygaert, R., Jones, B. J. T. 2007, MNRAS, 380, 551

Platen, E., van de Weygaert, R., Jones, B. J. T., Vegter, G., Calvo, M. A. A. 2011, MNRAS, 416, 2494

Plionis, M. Basilakos, S. 2002, MNRAS, 330, 399

Poggianti, B. M., et al. 2006, ApJ, 642, 188

Poggianti, B. M., et al. 2008, ApJ, 684, 888

Postman, M. Geller, M. J. 1984, ApJ, 281, 95

Postman, M., Huchra, J. P., Geller, M. J. 1992, ApJ, 384, 404

Postman, M., Lubin, L. M., Gunn, J. E., Oke, J. B., Hoessel, J. G., Schneider, D. P., Christensen, J. A. 1996, AJ, 111, 615

Postman, M., et al. 2005, ApJ, 623, 721

Press, W. H. Davis, M. 1982, ApJ, 259, 449

Ragone-Figueroa, C., Plionis, M., Merchán, M., Gottlöber, S., Yepes, G. 2010, MNRAS, 407, 581

Reid, B. A., et al. 2010, MNRAS, 404, 60

Riess, A. G., et al. 1998, AJ, 116, 1009

Riess, A. G., et al. 2001, ApJ, 560, 49

Riess, A. G., et al. 2009, ApJ, 699, 539

Robotham, A. S. G., et al. 2011, MNRAS, 416, 2640

Romer, A. K., Viana, P. T. P., Liddle, A. R., Mann, R. G. 2001, ApJ, 547, 594

Romer, A. K., et al. 2000, ApJS, 126, 209

Rosati, P., della Ceca, R., Norman, C., Giacconi, R. 1998, ApJ, 492, L21+

Sadat, R., Blanchard, A., Kneib, J.-P., Mathez, G., Madore, B., Mazzarella, J. M. 2004, A&A, 424, 1097

Sandage, A., Kristian, J., Westphal, J. A. 1976, ApJ, 205, 688

Sandage, A., Tammann, G. A., Saha, A., Reindl, B., Macchetto, F. D., Panagia, N. 2006, ApJ, 653, 843

Schaap, W. E. van de Weygaert, R. 2000, A&A, 363, L29

Schuecker, P. Boehringer, H. 1998, A&A, 339, 315

Shandarin, S. F., Sheth, J. V., Sahni, V. 2004, MNRAS, 353, 162

Sheldon, E. S., et al. 2001, ApJ, 554, 881

Sheldon, E. S., et al. 2007, ArXiv e-prints

Sheth, R. K., Mo, H. J., Tormen, G. 2001, MNRAS, 323, 1

Slezak, E., Bijaoui, A., Mars, G. 1990, A&A, 227, 301

Smale, P. R. 2011, MNRAS, 418, 2779

Smale, P. R. Wiltshire, D. L. 2011, MNRAS, 413, 367

Smith, A. G., Hopkins, A. M., Hunstead, R. W., Pimbblet, K. A. 2012, MNRAS, 422, 25 (S12)

Smith, A. G., et al. 2008, AJ, 136, 358

Smith, G. P., Treu, T., Ellis, R. S., Moran, S. M., Dressler, A. 2005, ApJ, 620, 78

Smoot, G. F., et al. 1992, ApJ, 396, L1

Sousbie, T., Pichon, C., Kawahara, H. 2011, MNRAS, 414, 384

Spergel, D. N., et al. 2003, ApJS, 148, 175

Spitzer, L. J. Baade, W. 1951, ApJ, 113, 413

Springel, V., et al. 2005, Nature, 435, 629

Stoica, R. S., Martínez, V. J., Mateu, J., Saar, E. 2005, A&A, 434, 423

Stoica, R. S., Martínez, V. J., Saar, E. 2010, A&A, 510, A38+

Sutherland, W. 1988, MNRAS, 234, 159

Tago, E., et al. 2006, Astronomische Nachrichten, 327, 365

Tarenghi, M., Tifft, W. G., Chincarini, G., Rood, H. J., Thompson, L. A. 1979, ApJ, 234, 793

Tikhonov, A. V. Klypin, A. 2009, MNRAS, 395, 1915

Tinker, J. L. Conroy, C. 2009, ApJ, 691, 633

Treu, T., Ellis, R. S., Kneib, J.-P., Dressler, A., Smail, I., Czoske, O., Oemler, A., Natarajan, P. 2003, ApJ, 591, 53

Turner, E. L. Gott, III, J. R. 1976, ApJS, 32, 409

Ueda, H. Takeuchi, T. T. 2004, PASJ, 56, 581

van de Weygaert, R. Schaap, W. 2009, in Lecture Notes in Physics, Berlin Springer Verlag, Vol. 665, Data Analysis in Cosmology, ed. V. J. Martínez, E. Saar, E. Martínez-González, & M.-J. Pons-Bordería, 291–413

Vikhlinin, A., Kravtsov, A., Forman, W., Jones, C., Markevitch, M., Murray, S. S., Van Speybroeck, L. 2006, ApJ, 640, 691

Vikhlinin, A., McNamara, B. R., Forman, W., Jones, C., Quintana, H., Hornstrup, A. 1998, ApJ, 502, 558

Vikhlinin, A., et al. 2009, ApJ, 692, 1060

Voit, G. M. 2005, Reviews of Modern Physics, 77, 207

Voronoi, G. 1908, J. reine angew. Math., 134, 198

Wang, H., Mo, H. J., Jing, Y. P., Guo, Y., van den Bosch, F. C., Yang, X. 2009, MNRAS, 394, 398

Wang, H., Mo, H. J., Yang, X., van den Bosch, F. C. 2012, MNRAS, 420, 1809

White, M. Kochanek, C. S. 2002, ApJ, 574, 24

White, M., van Waerbeke, L., Mackey, J. 2002, ApJ, 575, 640

Wilman, D. J., Balogh, M. L., Bower, R. G., Mulchaey, J. S., Oemler, A., Carlberg, R. G., Morris, S. L., Whitaker, R. J. 2005, MNRAS, 358, 71

Wiltshire, D. L. 2007a, New Journal of Physics, 9, 377

Wiltshire, D. L. 2007b, Physical Review Letters, 99, 251101

Wiltshire, D. L. 2008, Phys. Rev. D, 78, 084032

Wiltshire, D. L. 2009a, Phys. Rev. D, 80, 123512

Wiltshire, D. L. 2009b, International Journal of Modern Physics D, 18, 2121

Wittman, D., Dell'Antonio, I. P., Hughes, J. P., Margoniner, V. E., Tyson, J. A., Cohen, J. G., Norman, D. 2006, ApJ, 643, 128

Wittman, D., Tyson, J. A., Margoniner, V. E., Cohen, J. G., Dell'Antonio, I. P. 2001, ApJ, 557, L89

Yang, X., Feng, L.-L., Chu, Y., Fang, L.-Z. 2002, ApJ, 566, 630

Yang, X., Mo, H. J., van den Bosch, F. C., Jing, Y. P. 2005, MNRAS, 356, 1293

Yaryura, C. Y., Baugh, C. M., Angulo, R. E. 2011, MNRAS, 413, 1311

Yoon, J. H., Schawinski, K., Sheen, Y., Ree, C. H., Yi, S. K. 2008, ApJS, 176, 414 (Y08)

York, D. G., et al. 2000, AJ, 120, 1579

Zaritsky, D., Nelson, A. E., Dalcanton, J. J., Gonzalez, A. H. 1997, ApJ, 480, L91+

Zhang, Y., Yang, X., Faltenbacher, A., Springel, V., Lin, W., Wang, H. 2009, ApJ, 706, 747

Zwicky, F. 1937, ApJ, 86, 217

Zwicky, F., Herzog, E., Wild, P. 1961, Catalogue of galaxies and of clusters of galaxies, Vol. I (Pasadena: California Institute of Technology (CIT), —c1961)