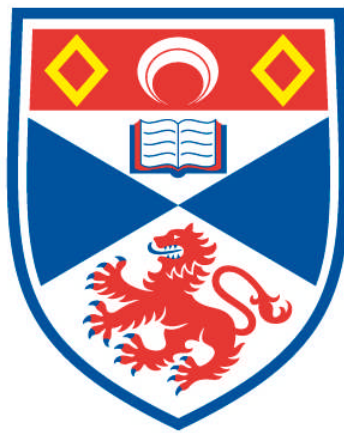


**THE COSMIC WEB UNRAVELLED
A STUDY OF FILAMENTARY STRUCTURE IN THE GALAXY
AND MASS ASSEMBLY SURVEY**

Mehmet Alpaslan

**A Thesis Submitted for the Degree of PhD
at the
University of St Andrews**



2014

**Full metadata for this item is available in
Research@StAndrews:FullText
at:**

<http://research-repository.st-andrews.ac.uk/>

Please use this identifier to cite or link to this item:

<http://hdl.handle.net/10023/4906>

This item is protected by original copyright

**This item is licensed under a
Creative Commons License**

The Cosmic Web Unravelled

A study of filamentary structure in the Galaxy and Mass Assembly survey

by

Mehmet Alpaslan

Submitted for the degree of Doctor of Philosophy in Astrophysics

May 2014



University
of
St Andrews

*I work for that, short man, large dream
I send my rockets forth between my ears
Hoping an inch of good is worth a pound of years
Aching to hear a voice cry back along the universal mall
We've reached Alpha Centauri!
We're tall!
O God! We're tall!*

An excerpt from **If Only Taller We Had Been**, by Ray Bradbury.

To my supervisors, friends, and family. Thank you.

Declaration

I, Mehmet Alpaslan, hereby certify that this thesis, which is approximately 35,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

Date May 20th 2014 Signature of candidate

I was admitted as a research student in September 2010 and as a candidate for the degree of PhD in September 2010; the higher study for which this is a record was carried out in the University of St Andrews and the International Centre for Radio Astronomy Research between 2010 and 2013.

Date May 20th 2014 Signature of candidate

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date May 20th 2014 Signature of supervisor

SIMON DRIVER

.....

Copyright Agreement

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. We have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis: Access to Printed copy and electronic publication of thesis through the University of St Andrews.

Date May 20th 2014 Signature of candidate

Date May 20th 2014 Signature of supervisor

SIMON DRIVER 

Collaboration statement

The work described in this thesis uses a number of data products available within the Galaxy and Mass Assembly database. They are listed as follows, with the primary author of every catalogue given in the parentheses: the GAMA galaxy group catalogue (Aaron S. G. Robotham); the Sérsic photometry catalogue (Lee Kelvin); the GAMA stellar mass catalogue (Edward Taylor); the matched aperture photometry catalogue (Mehmet Alpaslan); the spectral lines catalogue (Madusha Gunawardhana); the gas metallicity catalogue (Maritza Lara-Lopez); the 18-band photometry catalogue (Simon Driver); the visual morphology catalogue (Rebecca Lange). The appropriate reference for each catalogue is given in the text.

The `walk`, `findbranch`, and `makefilament` algorithms used in chapter 3 were originally written by Aaron S. G. Robotham. The line correlation function calculations in chapter 4 were performed by Danail Obreschkow. Section 4.3.1 has been written by Danail Obreschkow. The MAGPHYS outputs used to generate Figure ?? is generated by Simon Driver and Figures ?? and ?? use data generated by Simon Driver. The IOTA software described in the appendix was written by Lee Kelvin and myself.

Chapters 2 and 3 are based on expanded versions of the following papers:

- “GAMA: estimating galaxy group masses via caustic analysis”: **Mehmet Alpaslan**, Aaron S. G. Robotham, Simon Driver, Peder Norberg, John A. Peacock, Ivan Baldry, Joss Bland-Hawthorn, Sarah Brough, Andrew M. Hopkins, Lee S. Kelvin, Jochen Liske, Jon Loveday, Alexander Merson, Robert C. Nichol, and Kevin Pimbblet; 2012, Monthly Notices of the Royal Astronomical Society, Volume 426, Issue 4, pp. 2832-2846.
- “Galaxy and Mass Assembly (GAMA): The large scale structure of galaxies and comparison to mock universes”: **Mehmet Alpaslan**, Aaron S. G. Robotham, Simon Driver,

Peder Norberg, Ivan Baldry, Amanda E. Bauer, Joss Bland-Hawthorn, Michael Brown, Michelle Cluver, Matthew Colless, Caroline Foster, Andrew Hopkins, Eelco Van Kampen, Lee Kelvin, Maritza A. Lara-Lopez, Jochen Liske, Angel R. Lopez-Sanchez, Jon Loveday, Tamsyn McNaught-Roberts, Alexander Merson, Kevin Pimbblet; 2014, Monthly Notices of the Royal Astronomical Society, Volume 438, Issue 1, p.177-194.

- “Galaxy and Mass Assembly (GAMA): Fine filaments of galaxies detected within voids”: **Mehmet Alpaslan**, Aaron S.G. Robotham, Danail Obreschkow, Samantha Penny, Simon Driver, Peder Norberg, Michael Brown, Michelle Cluver, Benne Holwerda, Andrew M. Hopkins, Eelco van Kampen, Lee S. Kelvin, Maritza A. Lara-Lopez, Jon Loveday, Smriti Mahajan, Kevin Pimbblet; 2014, Monthly Notices of the Royal Astronomical Society: Letters, Volume 440, Issue 1, p.L106-L110.

As the first author on each of these papers I was responsible for all of the research and writing, with co-authors contributing observational data products and textual wording comments.

Contents

Declaration	iii
Copyright Agreement	v
Collaboration statement	vii
Abstract	xxv
1 Filaments, voids, and the growth of large scale structure	1
1.1 Large scale structure	3
1.1.1 Constraining cosmology	4
1.1.2 How does large scale structure impact galaxies?	5
1.2 Filament finders	8
1.2.1 Density based filament finders	9
1.2.2 Friends-of-friends filament finders	10
2 The Galaxy and Mass Assembly survey	13
2.1 Galaxy and Mass Assembly	16
2.2 The GAMA Group Catalogue	18
2.2.1 The mock GAMA group catalogue	26
3 Verification of the GAMA group catalogue masses and groupfinder effectiveness	29
3.1 The caustic mass method - background	31
3.1.1 The caustic mass method - algorithm	34
3.2 Caustic mass estimates of groups	40
3.2.1 Sensitivity to definition of group centroid	41
3.2.2 Sensitivity to definition of group radius	41
3.2.3 Caustic mass estimation in mock groups	42
3.2.4 Caustic mass estimation in observed groups	52

3.3	Summary and discussion	54
4	The GAMA Large Scale Structure Catalogue	59
4.1	Data	60
4.1.1	GAMA Group Catalogue	60
4.1.2	Sample selection	62
4.2	Filaments and large scale structure	67
4.2.1	Minimal spanning trees and Scooper	67
4.2.2	walk	74
4.2.3	makebranch	76
4.2.4	Filament catalogue	81
4.3	Tendrils	87
4.3.1	The excess line correlation of LSS	88
4.3.2	Shrinking voids	93
4.4	Comparison of LSS to mocks	95
4.5	Discussion and summary	105
5	The properties of galaxies as a function of environment	111
5.0.1	GAMA catalogues	116
5.1	Sub-sample selection and mass normalisation	117
5.2	The impact of large scale structure on galaxy properties	122
5.2.1	Broad 2D distributions	123
5.2.2	Metallicity	124
5.2.3	1D distributions	126
5.2.4	Elliptical fraction	128
5.2.5	Total energy (CSED)	128
5.3	Summary of results	130
6	Conclusions and future work	133
6.1	Future work	135
A	IOTA and the aperture matched photometry catalogue	139
A.1	Introduction	139
A.2	Data processing and SWarps	140

A.3	IOTA	144
A.3.1	Source Extractor	145
A.3.2	Image preparation and CUTTERPIPE	146
A.3.3	Source extraction and photometry	147
A.4	Matched aperture photometry catalogue	150
A.4.1	Errors	153
Acknowledgements		155
Bibliography		156

List of Figures

1.1	One of the earliest (and most puzzling images) of galaxy distribution not being completely uniform, shown in de Lapparent et al. (1986). The ‘CfA Stickman’ has been an iconic image of the startling nature of the distribution of galaxies for many years.	3
1.2	<i>Left:</i> The skeleton of a random Gaussian field. The yellow points show local minima; orange points are saddle points; and red points are local maxima. <i>Right:</i> The skeleton on the left panel overlaid onto a smoothed version of the Gaussian density field it was derived from, showing how the skeleton lines trace out filaments that connect overdensities (Novikov et al., 2006).	9
1.3	The RA- z plane of the 2dFGRS survey showing both the northern and southern galactic poles and the distribution of galaxies in connected systems colour-coded according to the multiplicity of the structures they belong to (Murphy et al., 2011).	11
2.1	A summary of the properties of modern galaxy surveys taken from Baldry et al. (2010) that shows spectral density against survey area, with a colour legend showing the primary spectrum range for each survey. Squares represent surveys that are limited by magnitude cut-offs; circles are surveys that use colour-cuts while triangles represent surveys that predominantly observe targeted objects. Table 2.1 summarises the selection criteria of these surveys.	14
2.2	An all-sky projection map of pointings and regions for a variety of galaxy surveys conducted with different instruments and telescopes, and GAMA, marked by the thick black lines. The three equatorial regions are the most complete, with the two southern fields currently being completed. Figure taken and reproduced with permission from Simon Driver (priv. comm.)	17
2.3	The GAMA G12 field shown for four different distributions, with $m_r > 19.4$ mag. The two leftmost panels show all galaxies and groups in this field (groups coloured according to total luminosity in L_\odot ; and the rightmost panels show all galaxies within groups and all galaxies outside of groups, in the field. It is very telling to see that galaxies present within groups show clear examples of redshift space distortion, whereas this is removed in the plot showing all groups. Note also how galaxies outside of groups also trace out large scale structure. . .	18

2.4	Galaxy groups are unique tracers of dark matter halos. The panel on the left shows dark matter halos generated with the Millennium catalogue with coloured circles indicating masses of groups within halos (yellow: $10^{13} - 10^{14} M_{\odot}$; cyan $10^{12} - 10^{13} M_{\odot}$; red $< 10^{12} M_{\odot}$). The right panel is the same field with only group members shown in redshift space, with prominent group member galaxies shown in white and others in magenta, with links drawn in cyan lines (Driver et al., 2009).	19
2.5	A schematic demonstration of the modified FoF algorithm of Robotham et al. (2011) applied to an idealised set of galaxies. Namely, this figure demonstrates the two step approach whereby galaxies are grouped both along the line of sight and projected on the sky. This uses both projected and radial separations and recovers the underlying group (galaxies 1, 5, and 6). This figure has been reproduced with permission from Robotham et al. (2011).	20
2.6	Top-down view (looking down in Declination) of the three equatorial GAMA fields and the groups that they contain, as per the GAMA group catalogue. Each group is denoted by a circle placed on its centre, the size of which denotes the number of galaxies in the group. The circles are coloured according to the velocity dispersion of the group. This image has been reproduced with permission from Robotham et al. (2011).	23
3.1	In this figure taken from Serra et al. (2011), a simulated galaxy cluster is plotted in redshift space, with blue points denoting galaxies within the cluster, and black points galaxies outside the cluster. The cyan lines are true caustics for this cluster, and the black lines represent the fitted caustic, with 1σ errors shown.	32
3.2	Four examples of the placement of the contour κ (black) and the caustics fitted to it (green) for four FoF mock galaxy groups (whose galaxies are shown as the black points) from the G ³ C in descending order of luminosity. It is evident here that the FFT method used to estimate the density distribution causes the final caustics to be very smooth with respect to the caustics drawn in DG97 and D99. This introduces a source of error in our caustic mass estimates. Notice how the gradient requirement $d \ln \mathcal{A} / d \ln R = 1/4$ is enforced towards the ends of the caustics in all panels, where the green lines no longer follow the black caustics.	39
3.3	Side-by-side comparison showing how the caustic mass estimation performs for a large mock galaxy group. In both panels, the black points represent the locations of galaxies in redshift space, the black lines show the contour $f_q(r, v) = \kappa$ and the green lines are the caustics drawn along $\min\{ v_u , v_l \}$. The left panel is a caustic fit only for galaxies present within this mock group and shows a clear example of the ‘trumpet’ distribution. The right panel includes nearby galaxies in redshift space out to $\pm 4R_{100}$ and $ z - z_{\text{med}} \leq 4 \times \max z - z_{\text{med}} $ whilst circling the original group members in red. In this case the trumpet distribution is lost and the caustics are artificially closed at the end of the sample.	40

- 3.4 The black dashed contours show the results of running the caustic algorithm on the mock catalogues and including galaxies nearby in the field, or in other groups. By comparison, the solid black contours represent the mass estimates made when considering only galaxies known to be in each group from the simulation. Finally, the red contours represent the distribution of the dynamical mass estimates from the G^3C . All three distributions have been adjusted to the same median as the dynamical mass estimates, shown in the red contours. Each contour line contains 10, 50 and 90% of the points, and the dashed green lines are 2/5/10 times away from the median. 43
- 3.5 A projection of the mock group in Figure 3.3 showing all galaxies within the mock group circled in red, and all galaxies that are detected when the algorithm includes nearby galaxies out to $\pm 4R_{100}$ and $|z - z_{\text{med}}| \leq 4 \times \max|z - z_{\text{med}}|$. The area shaded in blue represents the physical size of the simulation region at the median redshift of the group. At this range it is evident that the algorithm is including galaxies that are likely to belong to other separate groups. 44
- 3.6 Distribution of caustic (dynamical) masses for the intrinsic mock groups as a function of halo mass, drawn in black (red) for $r_{\text{AB}} \leq 19.4$. For each panel, both mass estimates have been corrected to be median-unbiased. The contours represent areas containing 10, 50 and 90% of the groups and the green lines are regions where the mass estimate is 2/5/10 times off the true mass. Of particular interest is the tendency for both distributions to follow each other very closely, particularly when over or under estimating the true halo mass. 47
- 3.7 Probability distribution functions of median-unbiased values of $\log \frac{M_{\text{est}}}{M_{\text{Halo}}}$ for the $r_{\text{AB}} < 19.4$ sample, with caustic masses drawn in black, and dynamical masses in red. The blue line is the PDF of $\log \frac{M_{\text{c}}}{M_{\text{dyn}}}$ and highlights the agreement between the two methods. The green dashed lines indicate regions that are factors of 2/5/10 away from the ‘true’ mass. The difference between scatter in the caustic and dynamical mass estimates is often minimal, with the caustic method showing less scatter for groups with $5 \leq N_{\text{FoF}} \leq 19$ 48
- 3.8 As Fig. 3.6 but for bijectively matched groups identified by the friends-of-friends algorithm instead of the true known intrinsic grouping. The coloured lines represent regions containing 10, 50 and 90 % of groups. The black contours compare the caustic mass to the FoF halo mass, and FoF dynamical mass estimate to the FoF halo mass estimate are shown by the red contours. 49
- 3.9 As Fig. 3.7 but for the bijectively matched FoF groups from the mock galaxy catalogues. The different lines show median-unbiased values of $\log \frac{M_{\text{est}}}{M_{\text{Halo}}}$ for the $r_{\text{AB}} < 19.4$ sample. The black line shows the ratio between the caustic mass and the FoF halo mass, while the red line shows the ratio between the FoF dynamical and halo masses. As before, the blue line shows the distribution of the ratio between the caustic mass and the FoF dynamical mass. 50

3.10	<i>Left:</i> A comparison of the performance of the two mass estimation methods as a function of the grouping quality parameter, Q_{Tot} . Black points correspond to caustic masses, and red ones to dynamical masses. Both methods perform worst when Q_{Tot} is close to 0, but quickly recover at approximately 0.2. The green lines show a rolling median in each 0.1 bin for the caustic mass (solid line) and for the dynamical mass (dashed line). <i>Right:</i> The ratio between the caustic and the dynamical mass estimates as a function of Q_{Tot} , with the rolling median in each 0.1 bin overplotted in green.	51
3.11	The mass ratio of caustic and dynamical masses as a function of dynamical mass. Note that the caustic masses in this figure have not been median unbiased with respect to the mocks using A_c , to show the importance of doing so. See Figure 3.12 for mass comparisons with the proper scaling factors applied.	53
3.12	Distribution of the caustic masses compared to dynamical masses for real galaxy groups from G^3C with $\sigma^2 \geq 130 \text{ km}^2 \text{ s}^{-1}$. The dashed green lines are as Figure 3.6, and the solid black lines represent regions containing 10, 50 and 90% of the data. There is a good correlation between the caustic and dynamical mass estimates as already shown in Fig 3.8 and 3.9 for the mocks.	55
3.13	As in Figure 3.7 this figure shows the PDF of the logarithm of the mass ratio between the caustic and dynamical mass estimates for all groups with $\sigma^2 \geq 130 \text{ km}^2 \text{ s}^{-2}$. The blue line shows the same distribution but for the bijectively matched FoF mock catalogues with the same velocity cut to illustrate the smaller scatter between the two methods for the real data.	56
4.1	A cutout of the G12 field, with the $0.15 \leq z \leq 0.2$ and $-1^\circ \leq \delta \leq 0^\circ$ limits. The red circles represent group centres, sized according to the number of galaxies in each group. Black points represent galaxies inside and outside of groups. . . .	61
4.2	Distribution of the total group r-band luminosity (see R11 for more details) as a function of redshift, after the sample selection process has been applied. The numbers in the top left display the number of groups kept and discarded after removing galaxies - the kept galaxies are plotted in the figure. The region shaded in highlights the region with $z > 0.213$ and is no longer volume limited, and the numbers in the bottom corners show how many groups are above and below the redshift cut. We are therefore left with 3,589 groups across all three GAMA regions, with $z \leq 0.213$ and with at least two or more galaxies with $M_r \leq -19.77$. This sample selection ensures the structures we detect are volume limited.	64
4.3	All groups in the G^3C are plotted here, with our final sample shown in red. The redshift limit of $z = 0.213$ is easily seen here. The red sample corresponds to all groups in the unshaded region in the left.	65
4.4	Three side by side cones showing the remaining galaxy sample after the selection described in Figure 4.2 for the G09, G12 and G15 regions respectively out to $z = 0.213$. All three cones span the full 5° declination range, which results in increasing projection effects at higher redshifts.	66

- 4.5 Flowchart schematically describing, for one example region, all the steps taken to go from a distribution of galaxies to a network of filaments, tendrils and voids. All groups are shown on the top panel, and all ungrouped galaxies on the lower panel. The groups are then put into a minimal spanning tree and the longest edges are trimmed. Ungrouped galaxies are then scooped up around each filament, giving the large network of galaxies near filaments (shown in blue). All ungrouped galaxies are then classified as being tendril galaxies (in green) or void galaxies (shown in red). 69
- 4.6 Filaments constructed from the same minimal spanning tree, but with different maximum edge lengths, decreasing from left to right. Groups in the same filament are coloured in matching colours. As the linking length increases, all galaxies tend towards being in one single huge filament, while as it decreases, leaving only groups that are in close proximity to each other. The number of groups in filaments also drops as b decreases, as groups with no links to other groups are not considered to be filaments (a filament needs at least 2 groups). . 71
- 4.7 Cumulative fraction of groups in filaments as a function of their total r-band group luminosity, shown for different maximum edge lengths in the minimal spanning tree (given by b). As b decreases only filaments between pairs of groups that are in extreme close proximity are constructed, and the fraction of high mass groups in filaments drops to 0. Raising b to a much higher value results in all groups being in a single giant filament. I therefore select the minimum value for b at which 90% or more of galaxies with $L_{\text{Group}} \geq 10^{11} L_{\odot} h^{-2}$ are in filaments; or in other words, $f(L_{\text{Group}} \geq 10^{11} L_{\odot} h^{-2}) \geq 90\%$ 72
- 4.8 An overview of the process by which the topology of a filament is determined. The top left panel shows all nodes and links for an example filament (circles and lines respectively). In each panel, green objects represent where the algorithm is, while blue ones represent visited objects and black ones, unvisited objects. From here, walk identifies all the ends of the filament (shown in panel b) and travels along them, stopping at intersections and merging all paths that reach the same intersection (panels c and d). The algorithm associates a count value to each node (shown for the green nodes on the top right in each panel), which is the number of steps required to reach the end of the filament. The count of the centre of the filament at an intersection, the counts along each branch are summed up and assigned as the count value for that node. Therefore, the node at which all branches meet will have the highest count, and be determined to be the centre of the filament, as shown by the red node in panel e. The output of walk is fed into makebranch, which analyses this output and uses it to construct branches for the filament, and assign orders to them. These are shown in panel f, with first, second and third order filaments shown in red, orange and yellow respectively. The backbone is then defined as the single path that travels along the two first order branches. 74

4.9	Two point correlation functions as a function of comoving distance for two different galaxy populations. The black line represents galaxies in groups that are in filaments, and within $4.56 h^{-1}$ Mpc of filaments; and the red line shows the function for galaxies in voids. By setting the maximum edge length between filaments to be $5.75 h^{-1}$ Mpc and $4.13 h^{-1}$ Mpc between galaxies in tendrils, we ensure that the resulting distribution of void galaxies has no correlation signal. The grey and pink shaded areas show, for each bin, the range of values for two point correlation functions calculated by sub-sampling the GAMA regions, with a jackknife method, and serve as uncertainty estimators.	79
4.10	Galaxy distribution in the three GAMA regions, colour coded according to their environment, with groups in filaments, galaxies near filaments, galaxies in tendrils, and void galaxies shown in cyan, blue, green and red respectively. Groups, and galaxies near them form the bulky complexes of large scale structure, tendrils spreading from them in filamentary structures into voids, which seem to be populated by galaxies that appear to be almost uniformly distributed on large scales.	82
4.11	Histogram showing the median redshifts for all the filaments in the final filament catalogue. The area shaded depicts the region below $z = 0.1$ where the number of filaments drops sharply due to a smaller survey volume.	83
4.12	Group and galaxy distribution in G12 colour coded by environment in 4 declination slices. Black circles represent groups in filaments, blue points are galaxies near filaments, green points are galaxies in tendrils, and red crosses are galaxies in voids.	86
4.13	A section of the G12 field with different galaxy populations shown in each panel. From left to right the populations shown are galaxies in filaments with the filament minimal spanning tree (blue and cyan respectively); galaxies in tendrils (green); galaxies in voids (red); and all three populations in their respective colours.	87
4.14	Tendril galaxies (shown as green points) for two observed GAMA fields and two mock GAMA fields (top and bottom rows respectively). The top row shows tendrils for a slice of the G09 and G15 fields, while the bottom row shows the same slice of sky from two separate volumes of the GAMA mocks. The two populations of tendrils are visually indistinguishable.	89
4.15	Examples of 2D density fields with their corresponding correlation functions $\xi_2(r)$, $\xi_3(r)$ and $l(r)$ shown in green, blue, and red respectively. Shaded regions represent 1σ confidence intervals. The density fields are purposely chosen to have the same two-point correlation function. From top to bottom, the density fields are constructed by superposing 350 randomly oriented Gaussian plane waves; 200 symmetric Gaussian spherical profiles; 30 randomly oriented Gaussian spheres elongated to resemble lines; and a superposition of (a) and (c). Figure reproduced with permission from (Obreschkow et al., 2013).	90

- 4.16 Excess line correlation $\Delta l(r)$ shown for filament (blue), tendril (green) and void galaxies (red), as well as all galaxies combined (black) across all three equatorial GAMA fields. The dashed lines show the median $\Delta l(r)$ for the GAMA mock catalogues, for which we show both the formal $\Delta l(r)$ error as typical error bars, and the spread between mock lightcones as solid lines with no arrowheads. The function $\Delta l(r)$ roughly measures the probability of finding three equidistant points separated by r on a straight line, in a density field that is analogous to GAMA but stripped of all two-point correlation. Shaded regions show lengths where $\Delta l(r)$ is unreliable due to the limited declination range of the GAMA cones ($\sim 40 h^{-1}$ Mpc on average). Galaxies in filaments show more excess line correlation than all galaxies, void galaxies show very little excess line correlation, while tendrils have a distinct, intermediate signature. There is excellent agreement between the data and the mock catalogue structures. 92
- 4.17 Each circled galaxy in this figure is a galaxy in the GAMA structure catalogue that is located inside a void identified in SDSS-DR7 data. The colour of each circle represents the type of structure it is thought to be in in the GAMA structure catalogue: blue circles represent galaxies in filaments, green circles represent galaxies in tendrils, and red circles represent galaxies in voids. Only a quarter of these matched galaxies are identified as voids in the GAMA structure catalogue. We note that we show the full 5° declination for each field in this figure, resulting in increasing projection effects at higher redshifts. 96
- 4.18 Locations of largest filaments in the data (shown in black) and the mocks (shown in red) for all three regions, with each row representing a region. The largest filamentary structures we detect have complex geometries and morphologies, where one might expect long structures that lie along the line of sight. With the exception of G15, all large filaments lie beyond $z \approx 0.14$, further highlighting that the largest networks of galaxies span a large geometric area in two dimensions. 97
- 4.19 The binned distribution of the number density of filaments as a function of length for various components of filaments with Poisson errors, ranging from backbones (solid lines, top panel) through to branches of order $n = 2$ and 3 (dashed and dotted lines, in the middle and lower panels respectively). Black and red lines correspond to filaments from the data and mock regions respectively. The x -axis positions of the points are the median values within that bin. The horizontal error bar on the final point in the red line shows the 1σ spread of the fifteen largest mock filaments across all regions and volumes. Bins with no detections show only an upper limit derived from Poisson statistics. The shaded region marks distances at which the geometry of the GAMA regions means the backbone lengths are poorly constrained. 99
- 4.20 Much like in Figure 4.19, the distribution of distances of tendrils are shown for tendrils in observations (black) and tendrils in simulations (red). Error bars show 1σ deviations about the mean in each bin. Both populations of tendrils are statistically indistinguishable. 100

4.21	<i>Left:</i> In the same manner to Figure 4.19, the number density of filaments with $b = 5.75 \text{ h}^{-1} \text{ Mpc}$ as a function of backbone length is shown. The black, red and blue lines each correspond to filaments in observed groups, groups recovered from simulations using the R11 groupfinder, and the intrinsically known groups from the simulations. <i>Right:</i> The distribution of backbone lengths for filaments with $b = 1 \text{ h}^{-1} \text{ Mpc}$. These filaments are all group-group pairs. Notably, we observe that there are more short filaments of intrinsic groups compared to the observed and FoF groups.	101
4.22	The relationship between backbone length and the number of groups in the backbone in bins containing equal numbers of filaments for observed data and FoF mock groups, shown in black and red respectively. The shaded regions denote 1σ intervals around the mean; points with no shaded region around them are single entries. This data is binned along the x-axis, and there are bins where there is no data; in these cases the point is omitted.	103
4.23	Comparing the ‘complexity’ of filaments in observed data and simulations; whereby complexity we refer to the fraction of branches that are backbones, and higher order branches. The solid lines represent data, and the dashed lines, mocks. Different coloured lines show increasing branch orders, from first to fifth order. As b increases, more complex filaments with more higher order branches are formed. At $b = 5.75 \text{ h}^{-1} \text{ Mpc}$, shown by the shaded region, observed and mock filaments have similar fractions of branches, aside from a slight overabundance of third order branches in observed filaments. The error bars show 1σ uncertainties about the population fraction.	104
4.24	The same region in G09 with different galaxy populations (colour coded as in 4.12), with observed data shown in the bottom right and the other panels consisting of mock data. The similarity between all four fields is apparent, and serves to visually highlight the success of the mock catalogues in reproducing large scale structure.	106
4.25	Contour plot of stellar mass versus colour for a mass complete sample galaxies in filaments, tendrils, and voids. The contours are placed such that they contain the upper 95%, 50%, and 10% of the data respectively.	108
5.1	Postage stamps of randomly selected galaxies in filaments, placed according to their stellar mass and $u - r$ colour. All postage stamps measure $500 \times 500 \text{ h}^{-1} \text{ kpc}$. The postage stamps are three colour images generated from SDSS and UKIDSS <i>Hig</i> bands. The single overly red stamp towards the bottom right is due to bad photometry in the u -band.	113
5.2	Postage stamps of randomly selected galaxies in tendrils, as in Figure 5.1.	114
5.3	Postage stamps of randomly selected galaxies in voids, as in Figure 5.1.	115
5.4	Stellar mass as a function of redshift for all GAMA I galaxies with $z \leq 0.1$ and $M_* \geq 10^{9.5} M_\odot$ coloured in grey. Redshift and mass limits are shown in red lines.	118

5.5	V_{max} adjusted stellar mass functions of galaxies in different environments by 0.1 dex bins in log stellar mass. The region shaded in grey marks the mass range at which the galaxies selected for filaments, tendrils and voids are below the mass completeness limit of the GLSSC. The black shaded region denotes the lower mass limit for all environments.	120
5.6	Kernel density estimates for stellar masses in mass normalised galaxy populations (colour coded in the same was as Figure 5.5).	122
5.7	$u - r$ colour plotted against absolute r -band magnitude for mass normalised galaxies in filaments and voids (black and red respectively). The distributions on the right and top of the figure show distributions in colour and absolute magnitude respectively. Note that colours are extinction and restframe corrected. This data has been taken from StellarMassesv08.	123
5.8	Comparisons of effective radius, ellipticity, absolute magnitude, $u - r$ colour, and r -band Sérsic index for mass normalised galaxies in filaments, tendrils and voids (shown as blue, green and red lines and points respectively). Each frame of this large mosaic plots two of these parameters against each other, while the histograms show the distributions of each one individually. Note that the most distinguishing differences are in these histograms. The data shown in this figure are taken from SersicPhotometryv07.	125
5.9	Metallicities and stellar masses from GasMetallicitiesv01 and StellarMassesv08 shown for mass normalised filament, tendril and void galaxies in blue, green and red respectively.	126
5.10	Effective radius, ellipticity, r -band absolute magnitude, $u - r$ colour, and Sérsic index from SersicPhotometryv07 for galaxies split by three different environment types: large scale structure (top row), halo mass and galaxies not in groups (second row), and paired and unpaired galaxies (bottom row). There is an overall trend for galaxies outside of structure (void galaxies, ungrouped galaxies, and unpaired galaxies) to show similar characteristics.	127
5.11	Elliptical fraction of all galaxies in VisualMorphologyv01 with $M_* \geq 10^{9.5} M_\odot$ as a function of halo mass shown in red. The fraction of non-elliptical galaxies is also shown, in blue (note that this is just the $1 -$ the elliptical fraction). The fraction of ellipticals rises steadily as a function of halo mass until approximately $3 \times 10^{13} M_\odot$, after which it remains relatively stable. Note that elliptical fractions are calculated for halo mass bins of equal size. The right hand panel shows the fraction of elliptical galaxies in a variety of environments. The error bars show 1σ uncertainties about the population fraction.	129
5.12	Total flux densities in 20 bands for mass normalised galaxies in filaments, tendrils, and voids, shown in blue, green and red respectively. These flux values are taken from the MultiBandPhotomv01 catalogue. The CSED shown in grey has been taken from Driver et al. (2012) and represents the mean GAMA energy output, scaled to lie above the points.	130

A.1	Three plots showing the initial distribution of PSF values for all frames in each band before convolution (top panel) and after convolution (middle and lower panels). Note that the colour legend remains unchanged in all three frames. Also note the drastic increase in sharpness of the distribution of PSFs post-convolution, as shown by the scale of the y -axis.	142
A.2	Two examples of bad imaging quality in the raw UKIDSS frames. The left image, in the H -band shows the commonly seen effect of star trails. On the right panel, this time in the K -band is a distinctive ‘donut’-like shape that is present in almost all unfocused images. Note that these images have not been altered past their original state except for having been renormalised to a zeropoint of 30 magnitudes.	143
A.3	Coverage for the v2 mosaics in the $YJHK$ bands, where each band is plotted with a declination offset. Blue areas show regions where there is no flux in the mosaics, indicating a removed frame or missing data. We have not plotted this information for the $ugriz$ data, as the SDSS mosaics are 100% complete saved for masked regions.	143
A.4	An illustration of a one-dimensional representation of how SEXTRACTOR goes through the deblending process. Here the horizontal subdivisions in density represent the different levels determined by DEBLEND_MINCONT, with the greyscale highlighting under the red curve showing the DEBLEND_MINCONT fraction. Here, A is considered to be a branch because B is another branch at a higher level, with a higher fraction of the total flux.	148
A.5	The distribution of pixel separation values for all objects in IOTA. The transparent line represents the smoothed density distribution of the values shown in the histogram, and the dashed blue line lies on the mean value.	150
A.6	Comparison of photometry between v2 and v1 of the matched aperture photometry catalogue in all 9 bands. The rotated density plots shown to the right of the scatter plots show a projected distribution of all the data for that band, and are particularly useful for identifying upwards or downwards shifts in the photometry. The green dashed lines are drawn on the $\Delta \text{mag} = 0$ line and the red lines show means and standard deviations for the density plots.	151
A.7	Colour magnitude plots, with the red text in each panel displaying the bands being compared. The blue and black lines show the distribution for v2 and v1 photometry catalogues respectively. In a small number of cases, most notably for the z -band this distribution sharpens in the new version.	152

List of Tables

2.1	A list of galaxy redshift surveys found in Figure 2.1.	15
3.1	Values for A_c for each group subset in both the $r_{AB} < 19.4$ and $r_{AB} < 19.8$ mock G ³ C mock catalogues using iterative group centres and with all galaxies in group. Including these numbers in Eq. 3.14 gives a median-unbiased estimate for the group mass. Note that without including these factors, the mass estimate can vary by up to a factor of 3, most likely caused by the smoothing FFT caustic fitting method used.	45
3.2	Fraction of groups that have a caustic mass estimate that is within a factor of 2 and 5 from the dynamical mass estimate. The error $\epsilon(N_2)$ is defined as $\sqrt{1/N_2 - 1/N_{\text{tot}}}$ where N_2 is the number of groups within a factor 2 and N_{tot} is the total number of groups for a given bin. The number of groups present in each multiplicity bin drops sharply (from thousands of groups to 10 or less) after the first multiplicity bin, drastically lowering the relevance of these statistics.	57
4.1	Summary statistics of some basic properties of filaments in GAMA. Besides the number of filaments, for each region the following averages are given, the backbone length, sum of the length of all links, number of groups, branches, and galaxies per filament are given. All lengths are given in units of h^{-1} Mpc. The final row contains these values across all three equatorial fields.	85
4.2	Percentage of galaxies, grouped by large scale environment, matched with galaxies in voids located in GAMA fields, for two different absolute magnitude cuts. It is interesting to note that when moving from the brighter to the dimmer magnitude cut, many tendrils are reclassified as filaments, but the fraction of voids remains largely the same.	95
5.1	Parameters for the double Schechter function GSMF fits shown in Figure 5.5. The columns are the shared knee in the Schechter function (M^*), the primary slope at the faint end of the function (α_1), the primary normalisation for the fit (ϕ_1 ; in units of $\text{dex}^{-1} \text{Mpc}^{-3}$), the secondary slope at the faint end (α_2), and the secondary normalisation (ϕ_2), and the goodness of fit χ^2 . The fractional integrated stellar mass is also shown, for subdivisions of environment (marked by the horizontal lines in the table). Errors are estimated from jackknifed re-sampling.	121

Abstract

I have investigated the properties of the large scale structure of the nearby Universe using data from the Galaxy and Mass Assembly survey (GAMA).

I generated complementary halo mass estimates for all groups in the GAMA Galaxy Group Catalogue (G^3C) using a modified caustic mass estimation algorithm. On average, the caustic mass estimates agree with dynamical mass estimates within a factor of 2 in 90% of groups. A volume limited sample of these groups and galaxies are used to generate the large scale structure catalogue. An adapted minimal spanning tree algorithm is used to identify and classify structures, detecting 643 filaments that measure up to $200 h^{-1}$ Mpc, each containing 8 groups on average. A secondary population of smaller coherent structures, dubbed ‘tendrils,’ that link filaments together or penetrate into voids are also detected. On average, tendrils measure around $10 h^{-1}$ Mpc and contain 6 galaxies. The so-called line correlation function is used to prove that tendrils are real structures rather than accidental alignments. A population of isolated void galaxies are also identified. The properties of filaments and tendrils in observed and mock GAMA galaxy catalogues agree well.

I go on to show that voids from other surveys that overlap with GAMA regions contain a large number of galaxies, primarily belonging to tendrils. This implies that void sizes are strongly dependent on the number density and sensitivity limits of the galaxies observed by a survey.

Finally, I examine the properties of galaxies in different environments, finding that galaxies in filaments tend to be early-type, bright, spheroidal, and red whilst those in voids are typically the opposite: blue, late-type, and more faint. I show that group mass does not correlate with the brightness and morphologies of galaxies and that the primary driver of galaxy evolution is stellar mass.

1

Filaments, voids, and the growth of large scale structure

Matter, both visible and dark, has a remarkably nonuniform distribution in the present Universe. The distribution of galaxies and galaxy clusters is not random, and this clustered distribution is largely a consequence of, and caused by the evolution of, initial perturbations in the gravity field of the very early, pre-inflationary Universe (Bond et al., 1996). These initial perturbations were the first seeds of anisotropy in the distribution of matter in the Universe, and expanded rapidly during inflation (Blumenthal et al., 1984). Given this generally ‘lumpy’ distribution of matter in the Universe, it follows that overdense regions would have exerted a greater gravitational force on their surroundings. Since gravitational potential scales with density as $\nabla^2\Phi = 4\pi G\rho$, these overdense regions will continue to accrue matter from nearby underdense regions – this carries the implication that overdense regions are gravitationally unstable. This vicious cycle will eventually result, along with the inclusion of other astronomical processes such as star formation, supernovae and other forces such as radiation pressure,

in galaxies and the Universe as we know it today with cluster diameters in the tens of megaparsecs, and velocity dispersions up to a few hundred kilometres per second.

Baryons tumbling and accumulating in density perturbations, however, are not sufficient to make a universe that resembles the one we exist in. Most modern, conventional theories of the origin of galaxies, groups, and the large scale structure of the Universe require exotic dark matter particles. Over the years, a cold dark matter model coupled with an accelerating expansion model for the Universe has emerged, supported further by analysis of data from recent cosmological probes such as the Wilkinson Microwave Anisotropy Probe (WMAP; Komatsu et al., 2011). One of the predictions of the cold dark matter paradigm is that galaxy groups serve as tracers for halos of underlying dark matter. Prior to the decoupling of matter and radiation, it is thought that dark matter particles, with their smaller interaction cross-section, were at liberty to collapse into the primordial density anisotropies before ‘conventional’ matter. As the Universe cooled and expanded, baryons were left free to fall into these overdense regions already populated by dark matter and form the building blocks of the large scale structure of the Universe (Springel et al., 2005).

The study of the large scale structure is a developing field in astronomy and astrophysics that primarily focuses on observing, quantifying and understanding the distribution of matter (both visible and dark) in the Universe, as well as attempting to determine the mechanisms that drive the formation of such structures (Eke et al., 1996; Springel et al., 2005; Thompson & Gregory, 2011). It serves to create a set of well-understood observational constraints for the clustering of galaxies and the distribution of filamentary structure, all of which act as rigorous benchmarks for computational cosmological models. Since the early 80s (Oort, 1983), astronomers have been mapping out these structures with ever more powerful galaxy surveys; and in much the similar way, the literature and theory surrounding the formation of structures has been evolving with observations (e.g., Einasto et al., 1980; Arnold et al., 1982; Blumenthal et al., 1984). These studies rely on observed or simulated data of the positions of galaxies in the Universe. Observing the large scale structure of the Universe provides astronomers with an insight into the end product of the processes that shaped the Universe, and is vital for furthering not only our understanding of cosmology, but of galaxy formation and evolution.

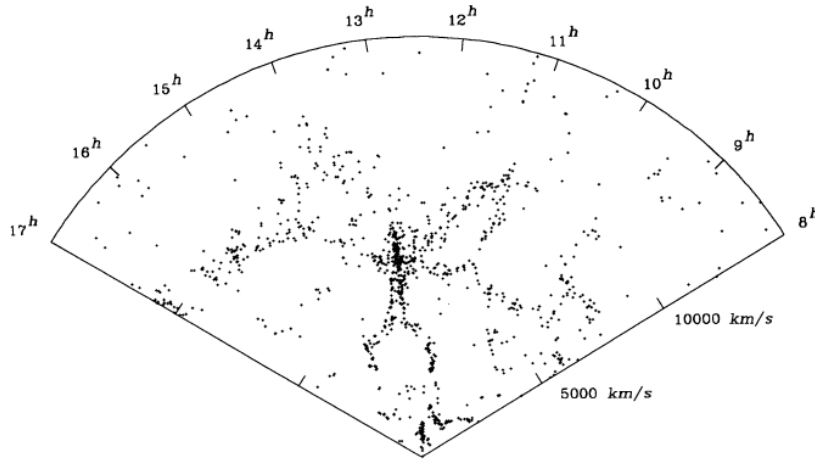


Figure 1.1: One of the earliest (and most puzzling images) of galaxy distribution not being completely uniform, shown in de Lapparent et al. (1986). The ‘CfA Stickman’ has been an iconic image of the startling nature of the distribution of galaxies for many years.

1.1 Large scale structure

When viewed at scales of the order of tens to hundreds of megaparsecs, the Universe looks like a complicated filigree of matter that is often described as a ‘foam,’ or the ‘Cosmic Web.’ The primary components of this web consist of large complexes of galaxies (referred to in the literature as groups or clusters, depending on size) which are themselves linked via linear, one dimensional structures of galaxies commonly called filaments. Filaments can come together to form ‘sheets’ and ‘knots’, which are two and three dimensional structures respectively. Straddling between filaments lie voids: vast, empty regions of the Universe populated only by a handful of isolated galaxies.

There is ample observational evidence of the Cosmic Web. Large scale structures exist on scales around $70 - 100h^{-1}$ Mpc (Sousbie et al., 2009); though smaller filaments that connect close cluster pairs at distances $\leq 5 h^{-1}$ Mpc also exist (Colberg et al., 2005). The Las Campanas Redshift Survey (LCRS; Schectman et al., 1995) was one of the first large scale surveys to provide observational evidence that galaxies are distributed in a non-uniform manner; although evidence for this had been present for a few years by then (Figure 1.1). More recent galaxy surveys such as 2dF (Colless et al., 2001) and the Sloan Digital Sky Survey (SDSS; Abazajian et al., 2009) have demonstrated the existence of filaments, sheets, and knots. Similarly, observations have shown the existence of gigantic voids measuring several tens of Mpc (Joeveer et al., 1978; de Lapparent et al., 1986; Schectman et al., 1996), where there are remarkably low densities of galaxies: down to 10% of the cosmic mean galaxy density (Pan et al., 2012).

It is worthwhile questioning whether or not these results are caused by chance statistical alignments of galaxies. Bharadwaj et al. (2004) have addressed this issue by examining filaments found in the LCRS. By shuffling slices of data from the LCRS randomly, the group were able to demonstrate that any filament longer than the length of the shuffled slices will exist purely out of chance alignment; this variable is found to be roughly $70 - 80 \text{ h}^{-1} \text{ Mpc}$, implying that these are the longest real filaments found in the LCRS.

1.1.1 Constraining cosmology

As discussed above, the large scale distribution of galaxies in the present Universe acts as an excellent benchmark for cosmological theories, but also as a data set for the measurement of fundamental cosmological parameters. Perhaps the most notable example of this was the observational detection of the $100 \text{ h}^{-1} \text{ Mpc}$ baryon acoustic peak in the large-scale correlation function of galaxies (Eisenstein et al., 2005), which served as a verification of the linear growth of structure via gravitational instability and acted as a confirmation of the predicted outcomes of cold dark matter cosmological theory (Peebles & Yu, 1970; Bond & Efstathiou, 1984). These early works predicted characteristic acoustic peaks in the power spectrum of the Universe, imprinted there as a result of perturbations in the relativistic plasma of the early Universe. Once frozen out (following the reduction of the sound speed due to the decoupling of matter and radiation at $z \sim 1000$), these perturbations formed the earliest ‘scaffolding’ into which baryonic matter could gradually accumulate and begin forming the first galaxies; and eventually, large scale structure. The detection of the acoustic peak imprinted onto the correlation function of galaxies in the SDSS Luminous Red Galaxy Sample by Eisenstein et al. (2005) provided direct evidence for the growth of structures via linear perturbation theory following $z \sim 1000$, as well as providing tangible proof of the existence of dark matter in the early Universe (acoustic oscillations would be much greater in a Universe containing only baryonic matter, as the absence of dark matter allows for the further propagation of the oscillations). Given that Λ CDM models predict narrow peaks in the clustering of galaxies, grown out of the initial perturbations imprinted into the distribution of matter at $z \sim 1000$, the detection of this acoustic peak serves as a ‘smoking gun’ confirmation of CDM theory for structure growth. More importantly, alternative non-linear gravitational growth theories do not predict such peaks (e.g. Fry, 1984). This result is a fundamental demonstration of the application of large scale surveys on constraining cosmological parameters. Filaments also serve as rigorous

benchmarks of the accuracy of cosmological models at the largest possible scales, with current studies focusing on comparing the observed properties of filaments to those found in various N-body cosmological simulations (Forero-Romero et al., 2009; Bond et al., 2010; Murphy et al., 2011).

Another probe of cosmology from large scale structure was originally proposed by Alcock & Paczynski (1979), which stipulates that the ratio between an object's observed angular size and its size in radial/redshift size varies according to cosmological parameters. Specifically, the test proposes that a hypothetical sphere in space would have its line of sight scale as a function of the inverse Hubble parameter $H^{-1}(z)$ and its transverse size scale with the angular diameter distance $D_A(z)$; therefore the ratio of the two measures gives $H(z)D_A(z)$. The primary advantage of such an approach is that it is fully independent of galaxy evolution and depends entirely on the geometry of the Universe; however, in practice this measurement is made difficult due to redshift space distortions caused by the peculiar velocities of galaxies in dense regions. Cosmic voids, therefore, make for attractive alternatives for the Alcock-Paczynski test, as they are almost entirely free of galaxies and occupy a large fraction of the Universe and have shapes that can be easily measured, as proposed by Ryden (1995). If we consider the idealised case of a spherical void surrounded by a thin wall of galaxies, the ratio of its depth to its width is given by

$$e_v(z) = \frac{u_v}{H_0 r_v} y(z) z \quad (1.1)$$

where r_v is the radius of the void and u_v is the outward moving velocity of the galaxies that surround it. $y(z)$ is the angular size distance $D_A(z) = H_0 \int_0^z dz/H(z)$. Lavaux & Wandelt (2012) propose a method to measure the expansion history of the Universe by measuring the ellipticity of stacked voids in different redshift bins and this method is applied by Sutter et al. (2012) to SDSS DR-7 data and yields statistically inconclusive results. Despite this, this study provides a robust proof-of-concept of applying the AP test to cosmic voids in future larger redshift surveys.

1.1.2 How does large scale structure impact galaxies?

The impact of environment on galaxy evolution has been studied for many decades. Butcher & Oemler (1984) found spiral galaxies in rich clusters were more red than spirals in the field.

Dressler (1980); Postman & Geller (1984); Dressler et al. (1997) established that there is a relationship between a galaxy's morphology and the density of its local cluster environment. Similar studies have shown that a galaxy's colour (Kreckel et al., 2012), stellar mass (Chabrier, 2003), gas content (Beygu et al., 2013; Benítez-Llambay et al., 2013) and luminosity function (Croton et al., 2005) are all affected by its local density (within 1-2 Mpc). At kiloparsec scales, wet mergers between gas-rich galaxies lead to major changes in morphology, orbital kinematics and star formation (Toomre & Toomre, 1972; Quinn et al., 1993; Balogh, 2007), which in turn lead to a different stellar population compared to a galaxy that is more isolated. Galaxies in clusters tend to be red early-type galaxies, while galaxies in the field and within filaments appear to be blue late-type galaxies (Tanaka et al., 2007), and there exists a correlation between the spin and shape of haloes and the filaments and sheets they reside in (Hahn et al., 2007b,a; Paz et al., 2011). More recently, *Spitzer* observations in the 24 μm wavelength have been used to identify starburst and AGN galaxies in filaments in the outskirts of Abell 1763 (Fadda et al., 2008). While the impact of 'extreme' environmental factors is well understood, the role of large scale structure, or indeed, a general understanding of the impact of environment on all scales on galaxy evolution has not yet been achieved. This is a vast, extant problem in extragalactic astronomy, which, with the advent of modern galaxy surveys and powerful simulations, can be tackled.

Understanding large scale structure also plays an important role in advancing the field of galaxy formation. Through a combination of the agglomeration of matter into pancake-like discs (Zel'dovich, 1970) under the influence of gravity, and models such as hierarchical merging (White & Rees, 1978) or secular evolution (Kormendy & Kennicutt, 2004), matter (baryonic and otherwise) appears to have come together to form galaxies, which themselves have become bound in clusters or groups. In the early Universe, galaxy formation was driven largely by collapsing material within the density fluctuations of cold dark matter; whereas interactions driven by environment are far more important in the Universe we currently exist in. In particular, mergers drive the formation of classical bulges and elliptical galaxies by disrupting disks - this is one of the fundamental impacts of environment on galaxy formation and evolution (Lukash et al., 2011; Gurbatov et al., 2012). These interactions also influence various physical processes within galaxies, as discussed in the preceding paragraph. The properties of galaxies in different large scale structure environments (filaments, knots/clusters, and voids) can be directly compared to the results of semi-analytic models run in conjunction

with high resolution dark matter simulations, where the mass accretion and merging history of each halo is known (Primack, 2009).

As well as impacting on galaxy-galaxy dynamics, filaments are also currently thought to serve as pathways by which matter can accrete onto galaxies: a general study of galaxies in filaments in the 2dFGRS has shown that galaxies infalling onto clusters along filaments tend to have an increased rate of star formation (Porter et al., 2008). Filamentary structures are also important for constraining the mass budget of the Universe, as a large fraction of baryonic matter is thought to reside within inter-cluster space in the form of hot gas (Fukugita et al., 1998; Pimbblet, 2005).

An alternative approach is to study voids, which contain an uncharacteristically few number of galaxies. As well as being tremendously useful for the AP test, studies of voids have also provided a lower constraint for Ω (Dekel & Rees, 1994). Of particular interest is the void probability function (White, 1979) that provides an expression for the probability of there being no galaxies in a volume V based on correlation functions; it was modified by Conroy et al. (2005) to give the probability of finding no galaxies in a sphere of radius R randomly placed in a spherical volume element. Voids and the VPF provide yet another benchmark for the accuracy of theoretical simulations, but one must bear in mind that the VPF is also susceptible to the selection bias caused by magnitude limits of redshift surveys. Early studies of voids (Joeveer et al., 1978; Gregory & Thompson, 1978; Kirshner et al., 1981; de Lapparent et al., 1986) have been supplemented by more recent large galaxy redshift surveys (Colless et al., 2001; Abazajian et al., 2009) that provide comprehensive and complete pictures of voids. The Void Galaxy Survey (VGS, Kreckel et al., 2012) is one such effort that has identified galaxies in voids from the SDSS-DR7 data. Beygu et al. (2013) have recently identified the system VGS_31, which is comprised of three galaxies in a void that appear to be aligned linearly. The system is linked together by a single massive cloud of HI gas. Rieder et al. (2013) have recently shown that such structures can be accurately reproduced in simulations, however; the existence of substructures in voids has been previously predicted, including ‘low-density ridges’ of galaxies within voids (van de Weygaert & van Kampen, 1993; Gottlöber et al., 2003).

The discussion presented so far in this chapter explains the necessity for developing a comprehensive observational catalogue of large scale structure. Being able to rigorously and objectively classify galaxies as belonging to the constituent parts of the Cosmic Web (filaments

and voids) adds an additional parameter space against which other properties of galaxies can be tested. Beyond asserting and confirming known trends such as the tendency of galaxies in dense environments to be brighter and more clustered (Zehavi et al., 2005), a comprehensive catalogue of large scale structure allows us to directly explore the potential impact of some of the earliest density perturbations on the galaxies we observe today.

1.2 Filament finders

While being remarkably easy to detect by eye given a large enough volume of space to look at, an accurate and consensus-driven classification of the Cosmic Web still eludes us. There is much debate on what exactly constitutes a filament, or even a definitive algorithm for detecting them and a number of different methods exist (e.g., Doroshkevich et al., 2004; Pimbblet, 2005; Colberg, 2007; Forero-Romero et al., 2009; Murphy et al., 2011; Smith et al., 2012; Cautun et al., 2013; Alpaslan et al., 2014). Filament finders consist of a set of algorithms that seek out filamentary structure in data sets obtained by galaxy surveys, or from simulated distributions of galaxies or dark matter. Most filament finders can be categorised as finders that approximate the placement of galaxies in the field as a density distribution over which they seek maxima, minima and the lines that connect these. Other methods exist, including the detection of filaments using the warm-hot intergalactic medium (WHIM) which is the result of gravitational heating caused by the formation of filaments. The WHIM can also be detected through x-ray absorptions (Davé et al., 2001; Zappacosta et al., 2010; Williams et al., 2010) and delineates filamentary structure. More general filament finders include an algorithm by González & Padilla (2010) that combines a tessellation based density estimator with a variable that determines the likelihood of two halos being bound based on their dynamical energy. Aragón-Calvo et al. (2010) take a different approach and classify structures by looking at their morphologies, finding that filaments are the most prominent structures as opposed to walls and sheets; and largely mark the structure of the cluster-void dichotomy. More recently, Murphy et al. (2011) have recovered filamentary structure in the 2dFGRS data by running a friends-of-friends finder on galaxies, then rerunning it a second time to detect filaments. Tempel et al. (2013) have used a modified marked point process method to search for filaments within a $0.009 \leq z \leq 0.155$ slice of the SDSS, modelling the filamentary network as a series of connected cylinders. Using narrow cylinders (of radius $0.5 h^{-1}$ Mpc) they identify filaments as having a characteristic length of $60 h^{-1}$ Mpc, and that galaxies in filaments contribute to

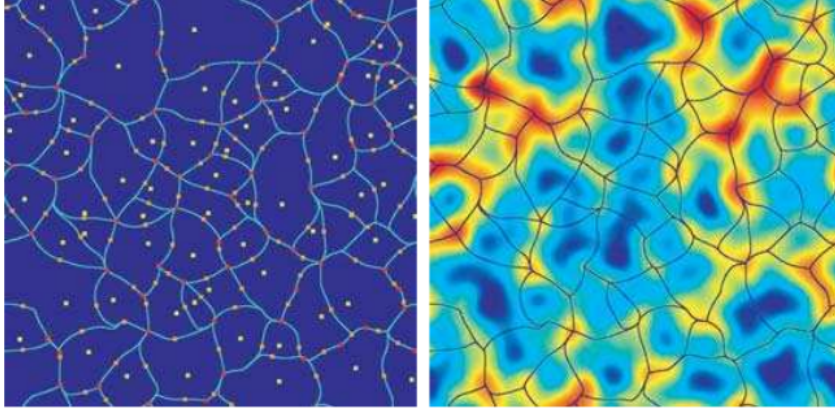


Figure 1.2: *Left:* The skeleton of a random Gaussian field. The yellow points show local minima; orange points are saddle points; and red points are local maxima. *Right:* The skeleton on the left panel overlaid onto a smoothed version of the Gaussian density field it was derived from, showing how the skeleton lines trace out filaments that connect overdensities (Novikov et al., 2006).

35-40% the total galaxy luminosity function.

1.2.1 Density based filament finders

One example of an algorithm that examines the density distribution is the so-called ‘3D skeleton’ (Novikov et al., 2006; Sousbie et al., 2008, 2009) (now referred to as DisPerSE Sousbie, 2011). This method identifies regions of overdensities in the distribution of galaxies in the Cosmic Web, and uses these overdensities to trace out lengths of filaments that connect clusters and nodes (i.e. density maxima and minima respectively). For a density field derived from a distribution of galaxies, it is possible to define a void as being a region of space where all points traveling along the direction of the density gradient $\nabla\rho = \delta\rho/\delta r_i$ converge onto a local minimum. The skeleton is then defined as being the overdense regions that surround these voids, with the following two properties: i) the nodes of the skeleton are the local maxima of the density field, where multiple skeleton lines converge (analogous to nodes in the large scale structure where filaments converge); and ii) any two local maxima must be separated by a saddle point and aren’t directly connected to each other. A two-dimensional schematic of this is shown in Figure 1.2.

The Skeleton method doesn’t have a well-defined way of detecting the ends of filaments. This is taken into account with the Smoothed Hessian Major Axis Filament Finder (SHMAFF) algorithm (Choi et al., 2010; Bond et al., 2010), where the filament end is defined as the point at which the angular rate of change of the primary axis of a filament exceeds a certain curvature parameter C , motivated by the argument that in the Cosmic Web, filaments end

where they intersect each other, or nodes (Bond et al., 1996). Applying DiSPERSE to SDSS data yields filaments that are, on average, 52.5 ± 6.5 Mpc in length (Sousbie et al., 2008).

1.2.2 Friends-of-friends filament finders

An alternative approach to filament finding that has proven successful is to pass a second friends-of-friends algorithm over the results of a first pass; i.e. look at which groups are bound to other groups. This has been successfully applied to the 2dF survey by Murphy et al. (2011) and has resulted in the discovery of filamentary structure. In total, 7603 systems are discovered containing at least two members, and at redshift $z \leq 0.12$ 87 % of all galaxies belong to some form of interconnected structure. The primary challenge of this method is to generate an algorithm that is able to select structures without a preferential direction or orientation, a process made rather difficult by redshift space distortion effects, commonly referred to as the Fingers of God. The FoG effect is characterised by the elongation of groups and clusters of galaxies in redshift space along the line of sight of the observer; it is caused by the random peculiar velocities of galaxies within these dense regions adding an additional Doppler shift to the redshift of the galaxy which deviates from Hubble expansion. A similar, but distinct redshift space distortion effect is referred to as the Kaiser effect (Kaiser, 1987), whereby the coherent peculiar velocities of galaxies infalling into clusters leading to line of sight distortions in redshift space. These effects are circumvented in Robotham et al. (2011) by taking an existing catalogue of galaxy groups, 2PIGG, constructed by Eke et al. (2004) and collapsing each group to a single point on their central positions, effectively scrubbing away the FoG effect. A second FoF algorithm is then run on this distribution and successfully recovers filamentary structure without being affected by the fingers of god (see Figure 2.5).

Comparisons to simulated galaxy catalogues reveal that the properties of structures detected in the 2dFGRS and simulations broadly agree with some discrepancies. On the whole, there are more large luminous structures in the 2dFGRS data than there are in the mocks, and this can be attributed to poorly understood galaxy formation models being used to construct the mocks. Given the existence of the GAMA Group Catalogue (Robotham et al., 2011), discussed in the following chapter, I have chosen to pursue this approach for my own filament finding algorithm.

The second chapter of this thesis serves as a general overview of the Galaxy and Mass Assembly survey, whose data has been used for all of the work done in this thesis, and collects

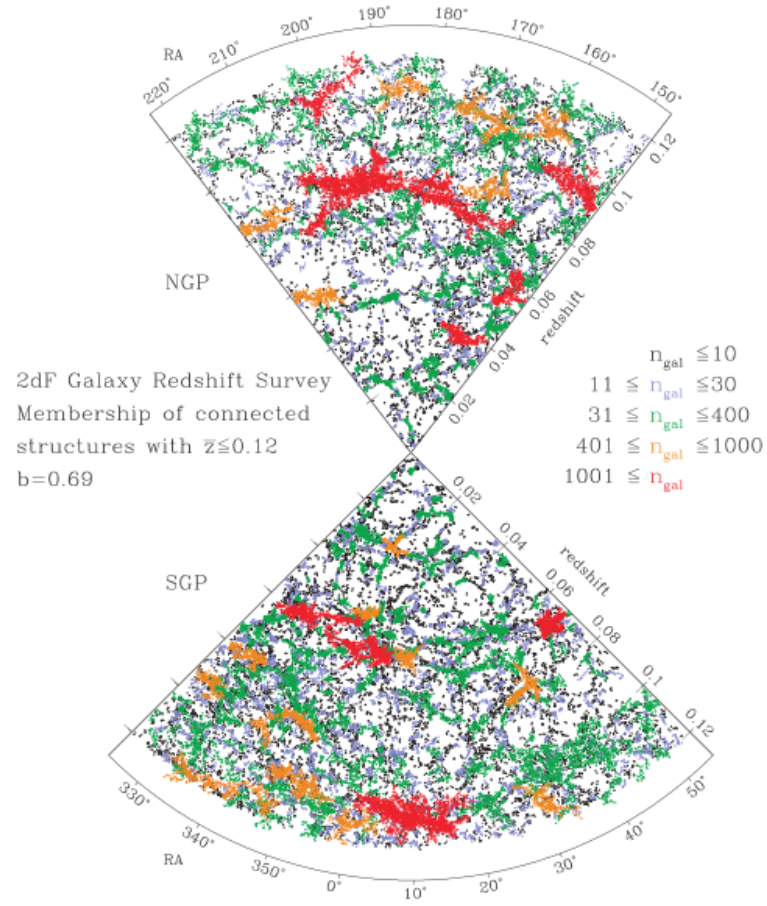


Figure 1.3: The RA- z plane of the 2dFGRS survey showing both the northern and southern galactic poles and the distribution of galaxies in connected systems colour-coded according to the multiplicity of the structures they belong to (Murphy et al., 2011).

most of the relevant information about the survey. In the third chapter of this thesis, I will describe the process by which I have been able to verify the masses of the GAMA groups using a complementary mass estimation method. This is an important first step to filament finding, as it verifies the accuracy of the GAMA groups. As these form the initial particles for the filament finding method I have developed, this verification ensures that any calibration of the filament finding algorithm parameters that depend on group parameters will be reliable. In chapter 4, I discuss my filament finding algorithm, the GAMA Large Scale Structure Catalogue, and make comparisons to filaments found in the GAMA mock catalogues. Chapter 5 presents a preliminary study on the impact of large scale structure on observable galaxy properties. Finally, Chapter 6 is a summary of work and conclusions. Throughout this thesis, a cosmology of $H_0 = 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_M = 0.25$, and $\Omega_\Lambda = 0.75$ is used.

2

The Galaxy and Mass Assembly survey

Since the discovery of the ‘Great Wall’ in the CfA survey, (Geller & Huchra, 1989), galaxy surveys have come a long way in providing larger and larger data sets that are used in detecting and quantifying the statistics of large scale structures, as well as addressing other questions in extragalactic astronomy (see Table 2.1 for a detailed list of recent extragalactic surveys and their selection criteria and a summary of the properties of current galaxy surveys can be found in Figure 2.1). Galaxy surveys have now expanded to cover over 45000 deg^2 and recover 357 million unique objects (Abazajian et al., 2009) at their largest extent at close redshifts, with other surveys choosing to focus on smaller regions but collecting a much greater density of galaxies – up to 1,000 galaxies per square degree (Driver et al., 2011). The importance of increasing object density is to be able to recover to a high fidelity the distribution of galaxies in halos; this information can then be used to detect galaxy clusters and filaments.

Every survey has a different goal, but generally improvements come in the form of deeper and more precise observations of more galaxies, though not necessarily in bigger regions. The huge volume of data generated by these surveys addresses many problems that exist in

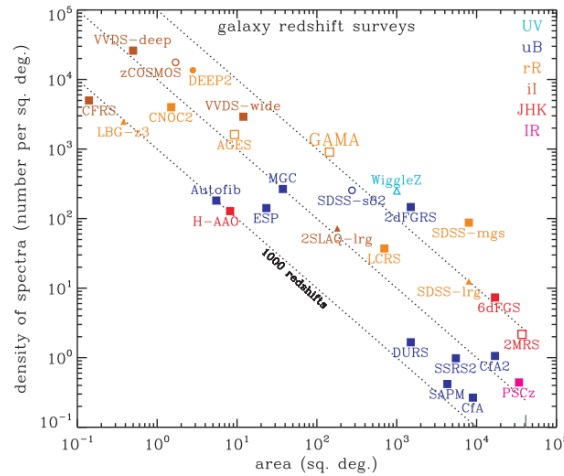


Figure 2.1: A summary of the properties of modern galaxy surveys taken from Baldry et al. (2010) that shows spectral density against survey area, with a colour legend showing the primary spectrum range for each survey. Squares represent surveys that are limited by magnitude cut-offs; circles are surveys that use colour-cuts while triangles represent surveys that predominantly observe targeted objects. Table 2.1 summarises the selection criteria of these surveys.

modern extragalactic astronomy. Examples of this using data from recent surveys include work on the galaxy luminosity function (Loveday et al., 2012), the galaxy stellar mass function (Baldry et al., 2012), the bimodality of the distribution of red and blue galaxies (Baldry et al., 2006), star formation history of galaxies in halos (Weinmann et al., 2006), and the cosmic spectral energy distributions (Driver et al., 2012), as well as providing observational tests for cosmology (Blake et al., 2011). Moving from single to multiple wavelengths is increasing the understanding we have of the different populations of stars present in galaxies as well as the role of dust, and challenging old paradigms such as the classification of galaxies into ‘red and dead’ and ‘blue and starbursting’ (Kelvin et al., 2012).

Table 2.1: A list of galaxy redshift surveys found in Figure 2.1.

Abbreviation	Survey name	Magnitude limits	Area (deg ²)	Reference
CFRS	Canada-France Redshift Survey	$I_{AB} < 22.5$	0.14	(Lilly et al., 1995)
LBG-z3	Lyman Break Galaxies at $z \sim 3$ Survey	$R_{AB} < 25.5$	0.38	(Steidel et al., 2003)
VVDS-deep	VIMOS VLT Deep Survey deep sample	$I_{AB} < 24.0$	0.5	(Fèvre et al., 2005)
CNOC2	Canadian Network for Obs. Cosmology 2	$R < 21.5$	1.5	(Yee et al., 2000a)
zCOSMOS	Redshifts for the Cosmic Evolution Survey	$I_{AB} < 22.5$	1.7	(Lilly et al., 2007)
DEEP2	Deep Evolutionary Exploratory Probe 2	$R_{AB} < 24.1$	2.8	(Davis et al., 2003)
Autofib	Autofib Redshift Survey	$b_J < 22.0$	5.5	(Ellis et al., 1996)
H-AAO	Hawaii+AAO K-band Redshift Survey	$K < 15.0$	8.2	(Huang et al., 2003)
AGES	AGN and Galaxy Evolution Survey	$R < 20, B_W > 20.5$	9.3	(Watson et al., 2009)
VVDS-wide	VIMOS-VLT Deep Survey wide sample	$I_{AB} < 22.5$	12.0	(Garilli et al., 2008)
ESP	ESO Slice Project	$b_J < 19.4$	23.3	(Vettolani et al., 1997)
MGC	Millennium Galaxy Catalogue	$B < 20$	37.5	(Liske et al., 2003)
GAMA	Galaxy and Mass Assembly Survey	$r < 19.8, z < 18.2, K_{AB} < 17.6$	144	(Driver et al., 2009)
2SLAQ-lrg	2SLAQ Luminous Red Galaxy Survey	$i < 19.8$	180	(Cannon et al., 2006)
SDSS-s82	SDSS Stripe 82 surveys	$u \leq 20, r < 19.5$	275	(Adelman-McCarthy et al., 2006)
LCRS	Las Campanas Redshift Survey	$R < 17.5$	700	(Shectman et al., 1996)
WiggleZ	WiggleZ Dark Energy Survey	$NUV < 22.8$	1000	(Drinkwater et al., 2010)
2dFGRS	2dF Galaxy Redshift Survey	$b_J < 19.4$	1500	(Colless et al., 2001)
DURS	Durham-UKST Redshift Survey	$b_J < 17$	1500	(Ratcliffe et al., 1996)
SAPM	Stromlo-APM Redshift Survey	$b_J < 17.1$	4300	(Loveday et al., 1992)
SSRS2	Southern Sky Redshift Survey 2	$B < 15.5$	5500	(da Costa et al., 1998)
SDSS-mgs	SDSS Main Galaxy Sample	$r < 17.8$	8000	(Strauss et al., 2002)
SDSS-lrg	SDSS Luminous Red Galaxy Survey	$r < 19.5$	8000	(Eisenstein et al., 2001)
6dFGS	6dF Galaxy Survey	$K < 12.7$	17000	(Jones et al., 2009)
CfA2	Center for Astrophysics 2 Redshift Survey	$B < 15.5$	17000	(Falco et al., 1999)
PSCz	IRAS Point Source Catalog Redshift Survey	$60\mu\text{m}_{AB} < 9.5$	34000	(Saunders et al., 2000)
2MRS	2MASS Redshift Survey	$K < 12.2$	37000	(Erdogdu et al., 2006)

2.1 Galaxy and Mass Assembly

One such survey, the Galaxy and Mass Assembly (GAMA)¹ survey (Driver et al., 2009, 2011) is an ongoing joint European-Australasian galaxy survey based around a spectroscopic campaign using the Anglo-Australian Telescope, with complementary imaging covering 21 bands of the electromagnetic spectrum from the ultraviolet through to far infrared using a multitude of ground and space based telescopes: GALEX (*FUV*, *NUV*), SDSS (*ugriz*), UKIDSS (*YJHK*), VIKING (*ZYJHK*), WISE (*W1-4*; *NIR*), and H-ATLAS (*PACS*, *SPIRE*; *FIR*) and GMRT and ASKAP (radio). The GAMA input catalogue is based on data taken from SDSS and UKIDSS. The aims of the survey are to cover a region of $\sim 360 \text{ deg}^2$ and to obtain $\sim 400\text{k}$ redshifts for galaxies out to a magnitude of $m_r = 19.8 \text{ mag}$. Phase one of GAMA (referred to internally as GAMA I) has observed three fields of 4×12 degrees centred at $\alpha = 9\text{h}$, $\delta = 1 \text{ deg}$ (G09), $\alpha = 12\text{h}$, $\delta = 0 \text{ deg}$ (G12) and $\alpha = 15\text{h}$, $\delta = 0 \text{ deg}$ (G15). Phase two (GAMA II) has expanded these to 5×12 degrees as well as gathering data in two additional southern fields: $\alpha = 34 \text{ deg}$, $\delta = -7 \text{ deg}$ (G02) and $\alpha = 345 \text{ deg}$ and $\delta = -32.5 \text{ deg}$ (G23). GAMA is funded by the STFC (UK), the ARC (Australia), the AAO, and the participating institutions. Figure 2.2 shows an all-sky projection with regions mapped out by different galaxy surveys, as well as GAMA.

All three equatorial regions are more than 98% spectroscopically complete (Driver et al., 2011), and the two southern fields are in the process of being completed. GAMA has been designed to have over 1000 objects per square degree. This is achieved through a greedy tiling algorithm that performs the target selection for the survey (Robotham et al., 2010). By ensuring that a particular section of sky is revisited up to 10 times, GAMA is able to detect not just the most bright sources in every square degree, but fainter background objects. The survey's coverage of such a large variety of wavelengths ensures maximum sensitivity to a large range of physical processes that take place within galaxies, while at the same time providing information on all the different constituents of a galaxy (dust in the IR, star formation in the UV, stellar mass in the optical, and so on). All this complexity originates from the many different processes that go through a galaxy in its lifetime, including mergers, infall of gas once it cools, star formation, and tidal interactions. It is therefore necessary for a comprehensive galaxy survey to sample as much of the spectrum as possible. The GAMA database contains, amongst others, catalogues of galaxy morphology (Kelvin et al., 2012), stellar masses (Taylor

¹<http://www.gama-survey.org/>

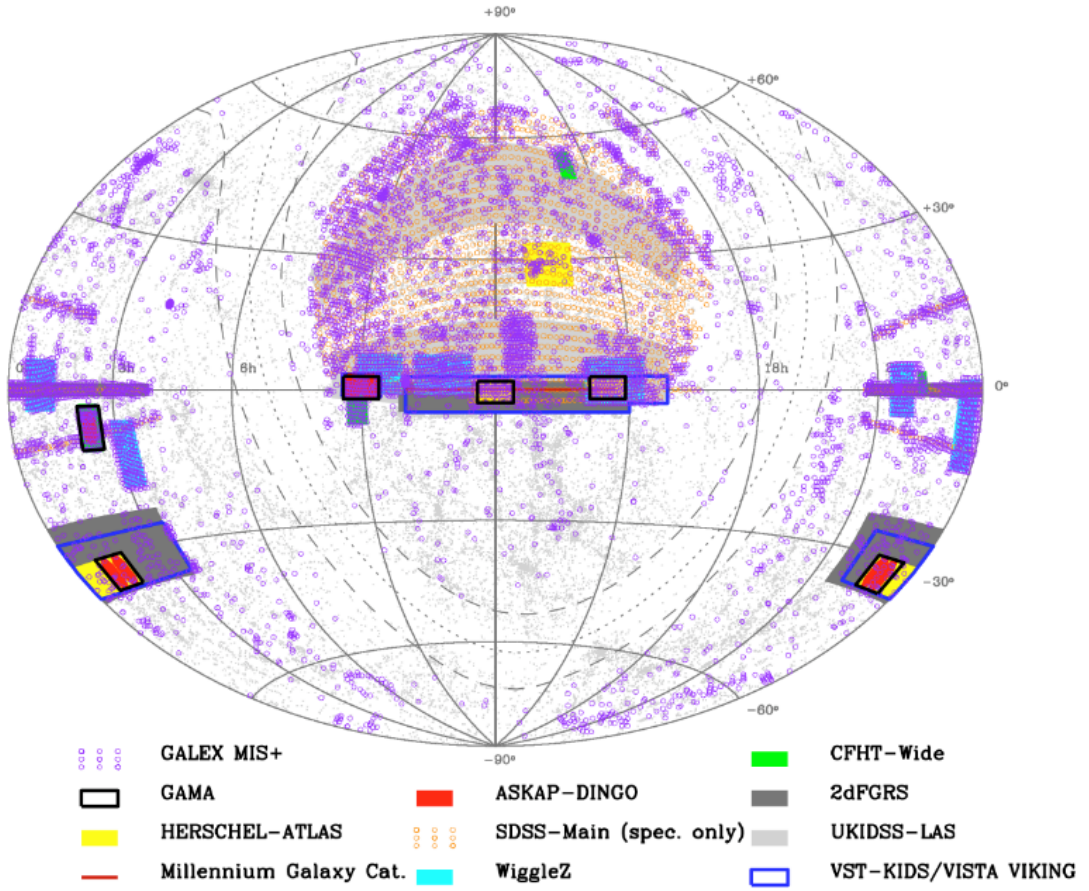


Figure 2.2: An all-sky projection map of pointings and regions for a variety of galaxy surveys conducted with different instruments and telescopes, and GAMA, marked by the thick black lines. The three equatorial regions are the most complete, with the two southern fields currently being completed. Figure taken and reproduced with permission from Simon Driver (priv. comm.)

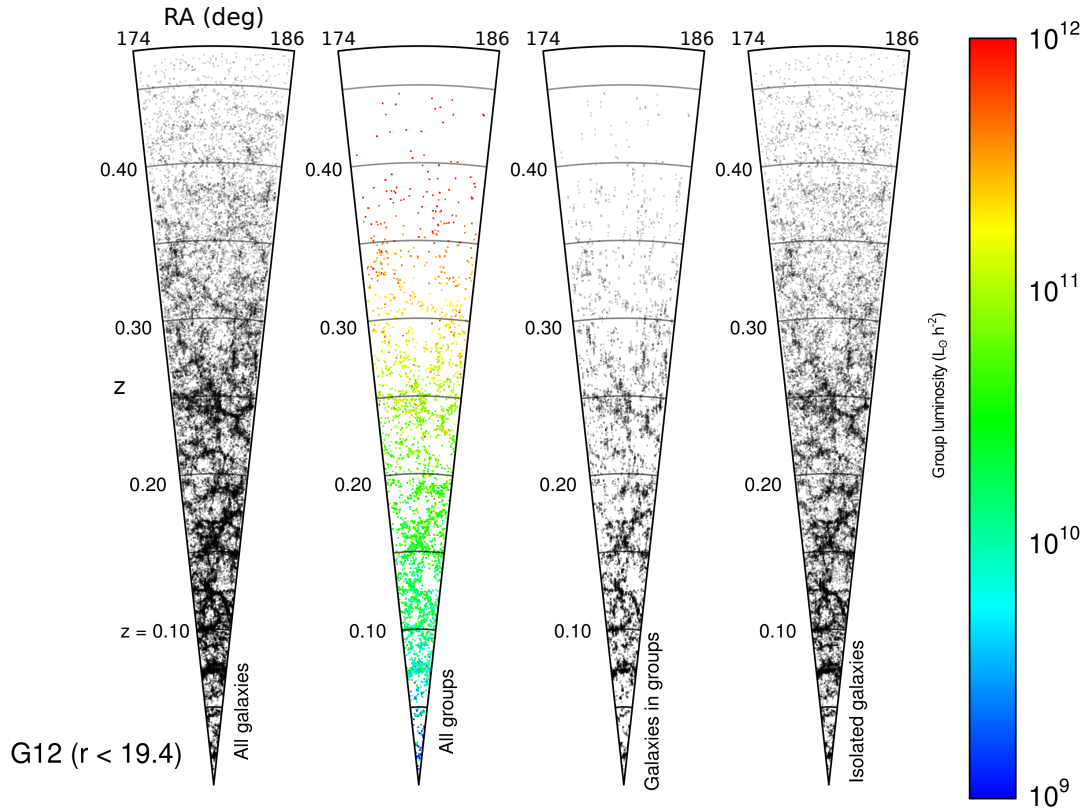


Figure 2.3: The GAMA G12 field shown for four different distributions, with $m_r > 19.4$ mag. The two leftmost panels show all galaxies and groups in this field (groups coloured according to total luminosity in L_\odot); and the rightmost panels show all galaxies within groups and all galaxies outside of groups, in the field. It is very telling to see that galaxies present within groups show clear examples of redshift space distortion, whereas this is removed in the plot showing all groups. Note also how galaxies outside of groups also trace out large scale structure.

et al., 2011), metallicities (Lara-Lopez et al., in prep), star formation rates (Gunawardhana et al., 2013), matched aperture photometry (Hill et al., 2011, Liske et al., in prep) and the GAMA Group Catalogue (Robotham et al., 2011) which is described below and referenced in subsequent chapters.

2.2 The GAMA Group Catalogue

One of GAMA's primary scientific goals is to better understand the relationship between dark matter halos and the galaxies that reside in them. The largest concentrations of galaxies, in clusters and groups, act as beacons that help us pinpoint dark matter halos; so a qualitative understanding of the distribution of galaxy clusters and their associated halo masses is a rigorous benchmark for dark matter simulations and models (Driver et al., 2009) (see Figure 2.4). Cosmological parameters, particularly Ω_0 have a profound effect on the presence, abundance and evolution of dark matter halos (Murray et al., 2013). By identifying the most massive

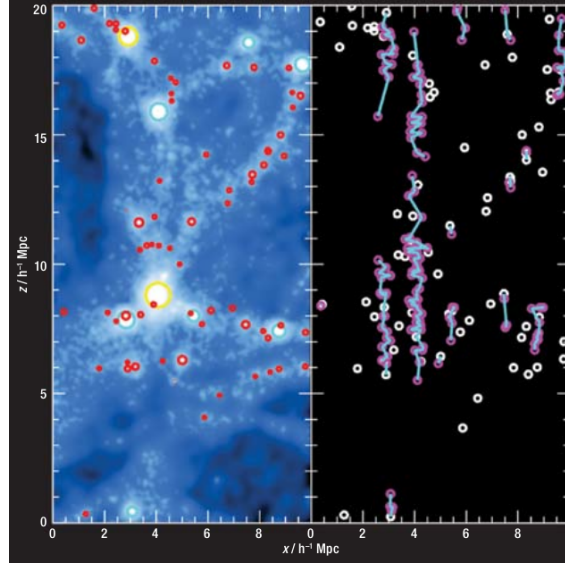


Figure 2.4: Galaxy groups are unique tracers of dark matter halos. The panel on the left shows dark matter halos generated with the Millennium catalogue with coloured circles indicating masses of groups within halos (yellow: $10^{13} - 10^{14} M_{\odot}$; cyan $10^{12} - 10^{13} M_{\odot}$; red $< 10^{12} M_{\odot}$). The right panel is the same field with only group members shown in redshift space, with prominent group member galaxies shown in white and others in magenta, with links drawn in cyan lines (Driver et al., 2009).

clusters, we can make good estimates of the amplitude of density fluctuations in the early Universe (Jenkins et al., 2001).

With GAMA, the large scale structure of the Universe and the nature of dark matter have so far been investigated by looking at the relationship between dark matter haloes and their galactic contents, focusing on analysing the low end of the halo mass function. The halo mass function (Peacock & Smith, 2000; Jenkins et al., 2001; Berlind & Weinberg, 2002) measures the number of galaxies hosted by a halo as a function of its mass. While the masses, numbers and locations of galaxies within halos should be driven more by the phenomena such as gas dynamics, star formation, galactic interactions and radiative cooling, the masses and spatial distributions of halos is expected to depend on gravitational dynamics and the nature of dark matter, as well as different cosmological models. All of these can be robustly tested by applying a groupfinding algorithm to GAMA data and mocks, and comparing these results.

The groups presented in the GAMA Galaxy Group Catalogue (G^3C) (Robotham et al., 2011) have been identified using a slightly modified friends-of-friends (FoF) algorithm (Press & Davis, 1982). FoF algorithms work by associating groups that are within certain distances of each other into common groups. To correctly account for the distortion effects of peculiar velocities, the algorithm used by Robotham et al. (2011) uses a two step approach considers projected separations separately to the radial redshift positions; in other words, the algorithm groups

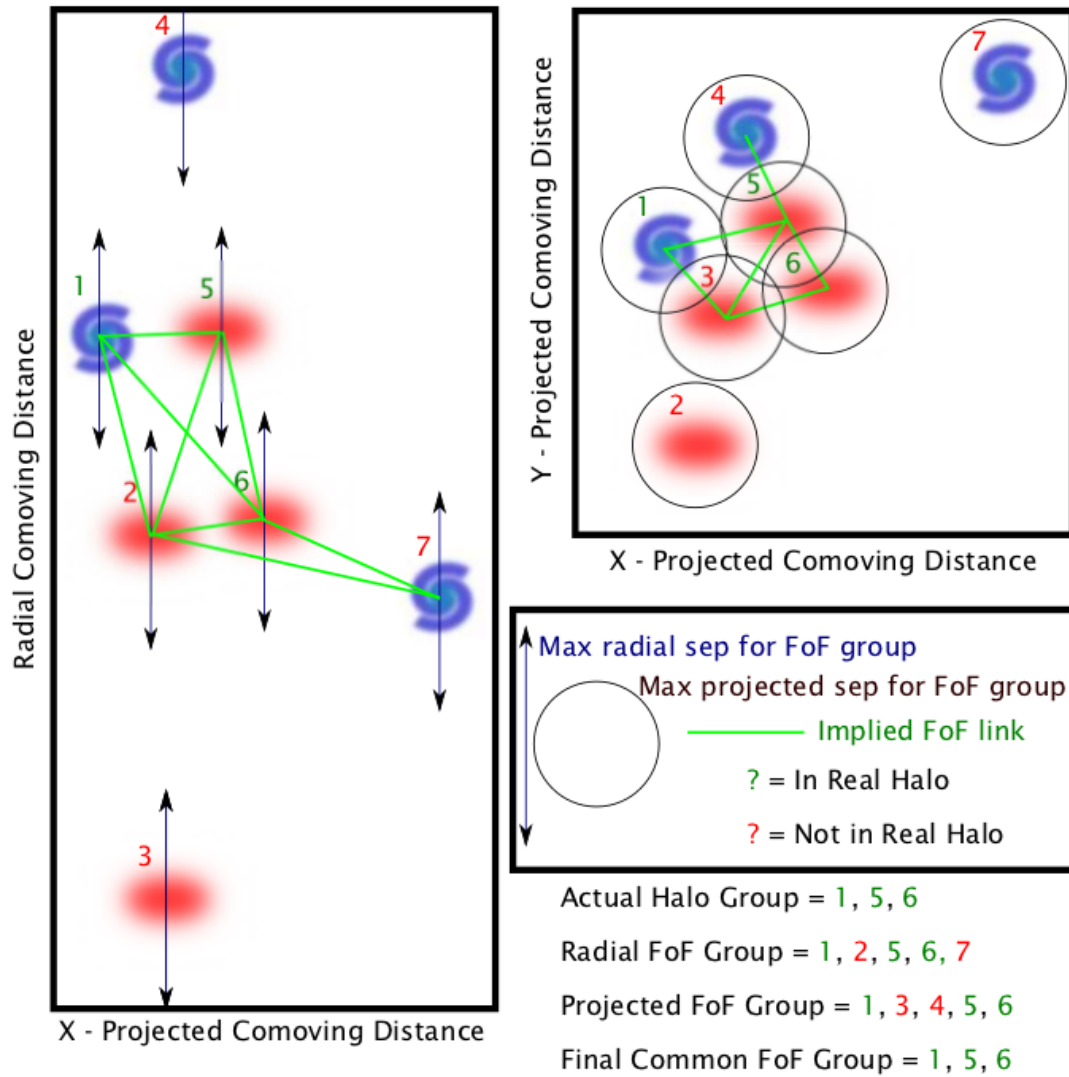


Figure 2.5: A schematic demonstration of the modified FoF algorithm of Robotham et al. (2011) applied to an idealised set of galaxies. Namely, this figure demonstrates the two step approach whereby galaxies are grouped both along the line of sight and projected on the sky. This uses both projected and radial separations and recovers the underlying group (galaxies 1, 5, and 6). This figure has been reproduced with permission from Robotham et al. (2011).

galaxies together both along the line of sight, as well as projected onto the sky. Galaxies considered to be grouped in both projections are then assigned to the same group. A schematic of this is shown in Figure 2.5. This approach therefore requires two linking lengths, with the radial linking length designed to be considerably greater than the projected projected one to properly account for the effects of peculiar velocities.

The two key parameters for the FoF algorithm are the projected linking length b , and the radial expansion factor applied to it to account for the peculiar motions of galaxies, R . Assuming static versions of these parameters falls short of accurately accounting for the effects

of peculiar velocities in clusters, so b and R are allowed to scale as a function of local density (similar to the approach of Eke et al. (2004)). b and R now therefore depend on position (\mathbf{r}) and faint magnitude limit m_{lim} :

$$b(\mathbf{r}, m_{\text{lim}}) = b_0 \left(\frac{1}{\Delta} \frac{\rho_{\text{emp}}(\mathbf{r}, m_{\text{lim}})}{\bar{\rho}(\mathbf{r}, m_{\text{lim}})} \right)^{E_b} \quad (2.1)$$

$$R(\mathbf{r}, m_{\text{lim}}) = R_0 \left(\frac{1}{\Delta} \frac{\rho_{\text{emp}}(\mathbf{r}, m_{\text{lim}})}{\bar{\rho}(\mathbf{r}, m_{\text{lim}})} \right)^{E_R} \quad (2.2)$$

where $\bar{\rho}$ is the average local density of GAMA at \mathbf{r} as a function of survey depth and is calculated from the selection function, ρ_{emp} is calculated directly from the number density of galaxies in a comoving cylinder centred on \mathbf{r} with projected radius r_{Δ} and radial extent l_{Δ} . Δ is calculated such that a galaxy within a local volume that is Δ times overdense as $\bar{\rho}$ will not have its links altered; Δ , E_r , and E_R are calculated by calibrating the FoF algorithm on a series of mock galaxy catalogues, designed specifically to replicate the GAMA luminosity function and geometry (the details of these mocks and the simulations used to obtain them are described in more detail in Section 2.2.1). To clarify, I refer to groups taken directly from the mocks as *mock groups* and groups that the FoF algorithm recovers from the mocks as *FoF groups*.

Calculating these parameters therefore becomes an optimisation problem: with the mock catalogues the true grouping of galaxies is known; one can therefore iteratively run the FoF finder and vary the parameters mentioned in the previous paragraphs such that the grouping quality of the algorithm is maximised. This grouping quality is defined as the group cost function. In Robotham et al. (2011), grouping quality is determined with two global measures: how accurately the algorithm recovers groups and the galaxies within them, and the significance of the grouping.

Accurate group retrieval means that at least 50% of the recovered FoF group galaxies belong to their actual mock group; in Robotham et al. (2011) this is defined as a bijective match. This can be converted into an efficiency statistic, E_{Tot} , which measures global halo finding efficiency and will be 1 if all groups are bijectively matched, and 0 if none are so. The significance of group retrieval is given by the total quality parameter, Q_{Tot} . This is defined as the product of the relative fractions of members that belong to the recovered FoF group and the intrinsic group and should equal 1 for the perfectly recovered group. These two parameters can be

combined into a final summary statistic:

$$S_{\text{Tot}} = E_{\text{Tot}} Q_{\text{Tot}} \quad (2.3)$$

that will span the range 0-1. The optimisation is performed such that this value is maximised. Those parameters that maximise S_{Tot} are then used to generate the group catalogue on observed data. Refer to Tables 1 and 2 in Robotham et al. (2011) for the values of each parameter and their errors.

The final group catalogue contains 23,838 galaxy groups with ≥ 2 members out to $r_{\text{AB}} = 19.8$ mag which corresponds to about 40% of the galaxies in the GAMA catalogue. Notably, most groups (63.2%) found in the catalogue are galaxy-galaxy pairs that span across the entire redshift range. The groupfinding algorithm has been calibrated by being run on a series of GAMA mock catalogues (Merson et al., 2012), discussed in the following section. A broad visual overview of the outputs of the group catalogue is shown in Figure 2.6 where the three equatorial GAMA fields are shown, with all groups in the catalogue shown.

GAMA is a highly complete spectroscopic survey (98% as of the creation of the group catalogue, with measured redshifts having an uncertainty $\sigma_v \approx 50 \text{ km s}^{-1}$). The average target density is 1050 galaxies per square degree, out to $m_r < 19.8$ mag. This means that some galaxies that may previously have been considered to be in the field and isolated are now seen to be part of an underlying group of faint galaxies (Robotham et al., 2011). Similarly, regions thought to contain few galaxies are now seen to contain not just more galaxies, but a considerable amount of structure. This is one of the principal strengths of GAMA, and is fundamental to why it is so well suited for studies of structure.

The group catalogue provides a number of useful properties for each group, including mass, radius, velocity dispersion, and luminosity. Masses are estimated by using the relation $M \propto \sigma^2 R$ where σ and R correspond to the velocity dispersion and the projected group radius. Group radius estimates were calculated for radii that contain 50%, 68% and 100% of the galaxies within each group (Rad_{50} , Rad_{68} and Rad_{100} respectively), and the projected group centre was located by three different methods, discussed below. The group velocity dispersion is calculated using the GAPPER algorithm introduced in Beers et al. (1990) where all the recession velocities for N galaxies within a group are ordered, and the velocity dispersion is estimated

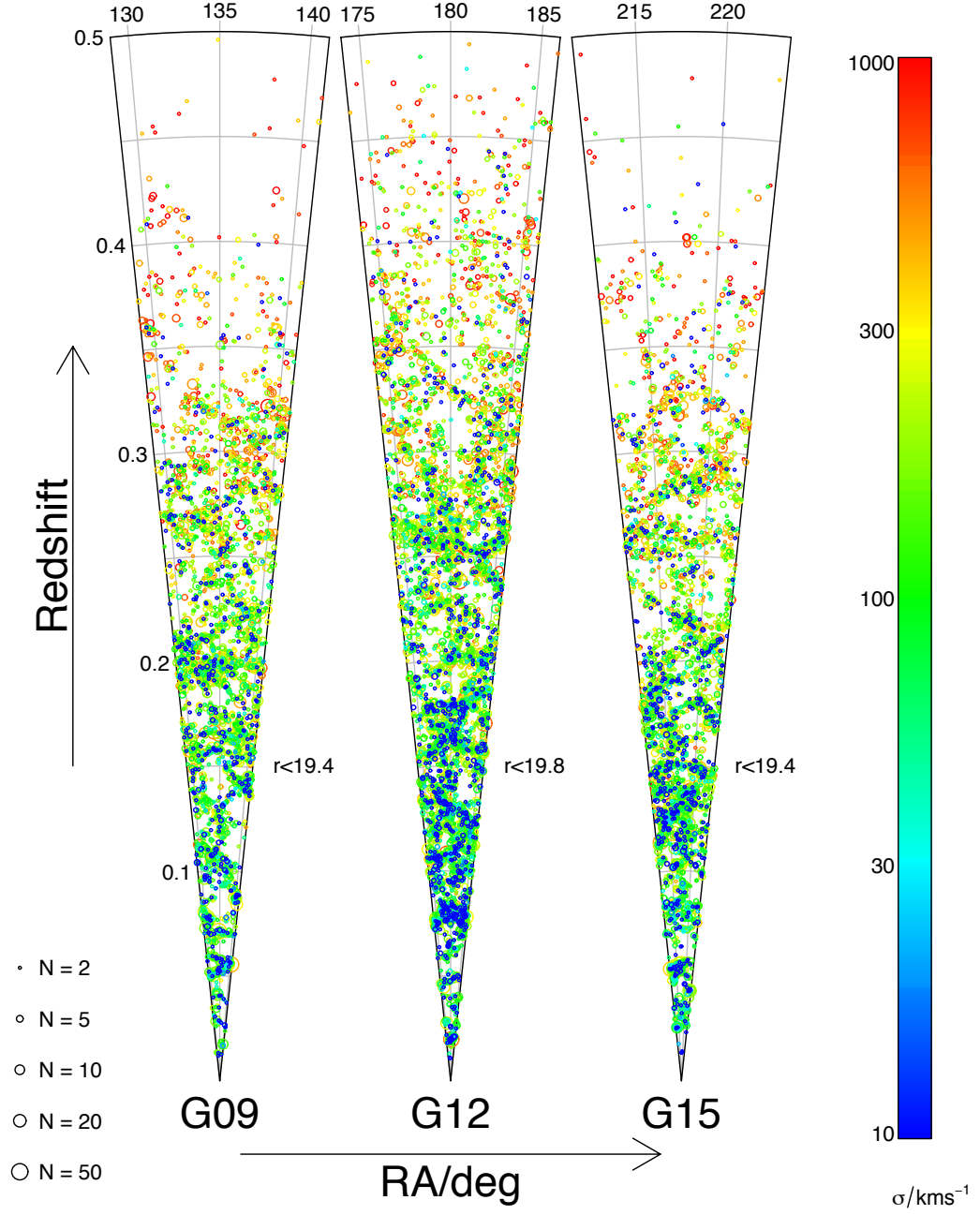


Figure 2.6: Top-down view (looking down in Declination) of the three equatorial GAMA fields and the groups that they contain, as per the GAMA group catalogue. Each group is denoted by a circle placed on its centre, the size of which denotes the number of galaxies in the group. The circles are coloured according to the velocity dispersion of the group. This image has been reproduced with permission from Robotham et al. (2011).

using calculated values for the gaps between velocity pairs $v_{i+1} - v_i$ for $i = 1, 2, \dots, N - 1$. The dynamical mass for each group is given by

$$\frac{M_{\text{FoF}}}{h^{-1}M_{\odot}} = \frac{A}{G/(M_{\odot}\text{km}^2\text{s}^{-2}\text{Mpc})} \left(\frac{\sigma_{\text{FoF}}^2}{\text{km}^2\text{s}^{-2}} \right) \frac{\text{Rad}_{\text{FoF}}}{h^{-1}\text{Mpc}} \quad (2.4)$$

where G is the gravitational constant ($6.673 \times 10^{-20} \text{ km}^2 \text{ s}^{-2} \text{ Mpc}$), Rad_{FoF} and σ_{FoF} are the radius and velocity dispersion of the group obtained by the methods described above. A is a scaling factor that is required to obtain a median-unbiased estimate of the friends-of-friends mass with respect to the real halo mass. It varies depending on multiplicity and median redshift of a group.

The group luminosity isn't defined as simply the total observed r -band luminosity of all galaxies considered to be in a group, as this does not account for any residual selection effects. Instead, the total observed luminosity is corrected for each group by integrating the GAMA luminosity function to the r -band absolute magnitude limit of each group. In other words, the group luminosity L_{FoF} is given by:

$$L_{\text{FoF}} = BL_{\text{obs}} \frac{\int_{-30}^{-14} 10^{-0.4M_r} \phi_{\text{GAMA}}(M_r) dM_r}{\int_{-30}^{M_{r-\text{lim}}} 10^{-0.4M_r} \phi_{\text{GAMA}}(M_r) dM_r} \quad (2.5)$$

where L_{obs} is the observed total r -band luminosity of the group, B is the scaling factor required to produce a median unbiased luminosity estimate (similar to A for masses), and $M_{r-\text{lim}}$ is the effective r -band absolute magnitude limit for the group and depends on the redshift of the group and the apparent magnitude limit of GAMA ($m_r = 19.8 \text{ mag}$). The limits of -30 and -14 for the numerator are effectively infinite limits, as a typical density of a luminosity function is covered by a couple of magnitudes of M_* .

It is constructive here, as an aside, to discuss how the absolute magnitudes of galaxies in the group catalogue (and in subsequent parts of this thesis) are calculated. To begin with, the distance modulus of an object at z is calculated, using the cosmological luminosity distance D_L to that object at that redshift. In other words,

$$DM = 5 \log D_L + 25 \quad (2.6)$$

with $D_L = (1+z)R_0S_k(r)$, where $R_0S_k(r)$ refers to the radial comoving distance, D_M , all given in h^{-1} Mpc. The radial comoving distance is defined as the distance between the observer and the galaxy measured along a radial path defined at the present cosmological time. There is another detail that one must be mindful of: given that the absolute magnitude calculation is being done using flux from one filter (the r -band in this case), one must bear in mind that that this filter will only be sensitive to a certain fraction of the galaxy's full spectrum, redshifted into the rest frame. One must therefore apply a so-called k -correction to all fluxes in order to be able to compare them. The absolute magnitude of an object with $m_r = 19.8$ mag at that redshift z can be calculated using the k -correction values taken from R11:

$$(k + e)(z) = \sum_{i=0}^N a_i(z_{\text{ref}}, z_p)(z - z_p)^i - Q_{z_{\text{ref}}}(z - z_{\text{ref}}) \quad (2.7)$$

where z_{ref} is the redshift to which all galaxies are corrected to; $Q_{z_{\text{ref}}}$ is a single luminosity evolution parameter (Yee et al., 2000b); z_p is the redshift used in the polynomial fit to the median KCORRECT-v4.2 (Blanton & Roweis, 2007) k -correction of GAMA-I galaxies and $a_i(z_{\text{ref}}, z_p)$ are the coefficients of that polynomial fit. We replicate the values used in R11: $z_{\text{ref}} = 0$, $Q_{\text{ref}} = 1.75$, $z_p = 0.2$, $N = 4$, $a_{1,4} = 0.2085, 1.0226, 0.5237, 3.5902, 2.3843$. This is not as effective as calculating individual k -correction values for each galaxy based on stellar population synthesis models as done in Taylor et al. (2011) but is computationally far less expensive and provides results that are statistically similar enough to warrant the use of this faster method.

Returning to the properties of groups, the projected group centre is defined in three different ways in the G³C: the first centre corresponds to the r_{AB} luminosity centre of light (CoL) of all the galaxies associated with the group. The second definition, the iterative group centre, is estimated by calculating the r_{AB} CoL of the group, and then rejecting the most distant galaxy from this centre. This process is then repeated until only two galaxies remain, at which point the brightest one of these is selected as the group centre. In the final method, the brightest group galaxy in the r_{AB} -band is selected as the group centre (BGG).

The G³C provides a very good starting point for a project that involves the detection of filament finding using a second-pass friends-of-friends algorithm as described in the previous chapter and in Murphy et al. (2011). Using groups as the foundations of filaments successfully

removes redshift space distortions from any detected structures (see Figure 2.3). A second FoF pass on the G³C should therefore extract filamentary structure free from the bias of redshift space distortions. Furthermore, the existence of mock catalogues allows for comparisons between real and simulated filaments.

2.2.1 The mock GAMA group catalogue

The GAMA mock catalogues (R11; Merson et al., 2012) are 9 mock light cones that match the geometry of the three equatorial GAMA fields. The primary purpose of mock galaxy catalogues in GAMA, at the time of writing of this thesis, was to provide a benchmark for the grouping algorithm used to generate the galaxy group catalogue. More broadly, mock galaxy catalogues are an essential tool in understanding how the construction of an observational survey and its selection effects impact on the estimation of astronomical statistics and measurements, such as luminosity functions and correlation functions. The obvious advantage of mock galaxy catalogues is that the ‘true’ values of these statistics are known by design, and so it is possible to compare statistics extracted from the mock catalogue with conventional methods to the ‘actual’ value.

The construction of the GAMA mocks is detailed in Merson et al. (2012). Broadly speaking, the procedure by which the mock catalogues are constructed from numerical simulations is as follows: (i) a population of galaxies is generated using dark matter haloes or the distribution of dark matter, (ii) place these galaxies into a cosmological volume, and (iii) apply the angular and radial selection functions of the survey that the mock catalogue is mimicking. In the case of the GAMA mock galaxy catalogue, the dark matter haloes are taken from the Millennium Simulation (Springel et al., 2005). This is an N-body dark matter simulation that provides the full spatial information of each halo and provides merger trees, which allows for clustering information to be extracted. The cosmological parameters adopted in the Millennium Simulation (and therefore present in the GAMA mock catalogues) match those measured from the WMAP and are as follows: the total density of baryons, $\Omega_b = 0.045$; the total matter density, $\Omega_m = 0.25$; the dark energy density, $\Omega_\Lambda = 0.75$; the Hubble constant, $H_0 = h100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ with $h = 0.73$; the primordial scalar spectral index, $n_s = 1$; and a fluctuation amplitude of $\sigma_8 = 0.9$.

The haloes provided by the Millennium Simulation are populated with galaxies by using semi-analytic models, or SAMs (as opposed to hydrodynamical simulations which are better

suited to simulate smaller scales). The SAM used for the GAMA mock galaxy catalogues is the Bower et al. (2006) GALFORM semi-analytic galaxy formation model. GALFORM functions by modeling, within each halo, numerous physical processes that play a part in the formation and evolution of galaxies: collapse and mergers of dark matter halos; heating and cooling of gas; star formation in disks; feedback from supernovae, active galactic nuclei and the ionisation of the interstellar medium; chemical enrichment of stars and gas; and dynamical friction from mergers. The galaxies are adjusted to match the GAMA survey luminosity function (Loveday et al., 2012) in the r -band. The mock galaxy catalogues are considered to be complete to approximately $M_r - 5 \log h \approx -17$, and the GALFORM model used is complete to around $10^8 M_\odot h^{-2}$ (Peder Norberg, priv. comm.). This completeness is estimated by testing different galaxy formation models and running them through the snapshots of the Millennium Simulation in order to measure the robustness of the predicted stellar mass function. Having generated galaxies within haloes, 9 lightcones matching the geometry of each of the 3 GAMA fields are placed within the simulation box. The result is a series of 27 mock lightcones (9 for each GAMA field) with a galaxy population whose luminosity function matches that of GAMA, and whose full clustering information is known.

Having a series of bespoke mock galaxy and group catalogues is extremely useful as it allows for calibration work (such as the calculation of scaling parameters such as A and B used in Eqns 2.4 and 2.5). The GAMA mock galaxy catalogue is utilised extensively in this thesis. In chapter 3, the mocks are used as a benchmark to calibrate the parameters used in the caustic mass measurement algorithm, and in chapter 4 the properties of filaments and tendrils found in the mocks are compared to ones found in observational data.

3

Verification of the GAMA group catalogue masses and groupfinder effectiveness

In this chapter we look to verify the GAMA group masses derived by Robotham et al. (2011). In Robotham et al. group masses were estimated from the velocity dispersion combined with the group radius. Measuring the mass of a galaxy group is both one of the most challenging and most fundamental ways to understand the distribution of the dark matter it sits in. Traditionally, mass estimates of groups are calculated virially, via measurements of the velocity dispersions of their members, e.g., Hughes (1989); Carlberg et al. (1996); Girardi et al. (1998); Tucker et al. (2000). By assuming that the group is in virial equilibrium, the dynamical mass of the group follows the relation $M \propto \sigma^2 R$. The obvious limitation of this method is that the relation will hold only out to the virial radius of the galaxy group, so mass estimates made using galaxies beyond this radius become less reliable. The dynamical mass will, however, measure the mass of whatever is contained within the virial radius of the group: visible and dark matter alike. Comparing the dynamical mass measurement to the estimate of the mass

of the visible matter of a group is a very reliable estimator of dark matter mass. More accurate mass estimates at large radii are possible using weak lensing (Kaiser et al., 1995), however the obvious drawback of this approach is the observational challenge involved in measuring the lensing signal for a large galaxy group sample.

An alternative and complementary approach to group mass estimation is to look at the distribution of the galaxies within a group in redshift space (defined as the projected distance r from the group centre and the line-of-sight velocity v with respect to the median group redshift for every member of the group) and estimate the group escape velocity by analysing the distinct shape of this distribution, described by some to resemble the horn of a trumpet. This method of analysing galaxy group members in redshift space was first introduced as a mass estimator by Diaferio & Geller (1997) and Diaferio (1999) (hereafter DG97 and D99) for large clusters with 200+ members and is referred as the caustic mass estimation technique by its developers. An early version of the method was first used in an attempt to constrain cosmological parameters, specifically Ω_M (Regős & Geller, 1989) with little success. By subsequently identifying that velocity measurements of galaxy groups were heavily affected by random motions, as well as by comparing observed caustics to those predicted by cosmological models, DG97 determined that this method is unreliable for determining Ω_M . Navarro et al. (1997) later went on to demonstrate that the density profiles of dark matter halos do not vary with cosmological parameters, putting attempts to use the caustic method to measure cosmological parameters to rest for good. However, the amplitude $A(r)$ of these caustics, which is defined as the area within caustics fit to the minimum of the upper and the lower line-of-sight velocities v_u and v_l , does provide a measure of the gravitational potential $\phi(r)$ of the group. This can then be used to estimate the mass of the group. The method has been subsequently used by Diaferio et al. (2005); Diaferio (2009); Serra et al. (2011) as a robust way of calculating masses of galaxy groups using all of the positional information from a group (out to R_{200}) and as a test for cluster membership (Rines & Diaferio, 2006). Additional research has shown that for a set of rich X-ray luminous clusters, caustic mass estimates agree to within a ratio of 1.03 ± 0.11 with mass estimates obtained via lensing analysis (Rines et al., 2003).

This chapter details the application of the caustic method to the GAMA Galaxy Group catalogue and its accompanying mocks (described in the previous chapter) and attempts to provide complementary caustic mass estimates to the dynamical mass estimates of the group halos within the catalogue. Verifying group masses is crucially important to being able to

use groups as the foundations for large scale structure in the next chapter: the caustic mass method fails for poorly defined groups (as discussed later in this chapter), so if the caustic masses agree well with the dynamical masses, this implies that the groups are defined well. In the filament finding method discussed in Chapter 4, groups are used to trace filaments, so it is vital that they are defined well. Accurate mass measurements are therefore the first step to developing a large scale structure catalogue. By also applying the method to mock light cones that mimic the GAMA data I am able to carefully calibrate my algorithm to produce median-unbiased mass estimators for each galaxy group using only redshift and positional information out to radii that are well beyond the virial radius.

3.1 The caustic mass method - background

The caustic method is based on analysing the distribution of group members within redshift space, which is defined in D99 as the plane (r, v) of the galaxies, where r is the projected radial separation from the group centre, and v is the line-of-sight velocity relative to the group centre of mass. The spherical infall model of Regős & Geller (1989) predicts the existence of two ‘lines’ that form a trumpet shape on this plane where the phase-space density in redshift space is infinite; and in practice these trumpets are observed when looking at both simulated and real groups. In the spherical model, plotting galaxies in a group in redshift space causes them to collapse to the peak redshift of the group along the line-of-sight, placing them somewhere between the centre of the group and the turnaround radius. In redshift space this translates to the trumpet-like lines that describe the escape velocity of the group. By definition, galaxies outside of these caustics are beyond the turnaround radius of the group – in other words, their peculiar velocity is so great that they are escaping the group. If we assume that galaxies lying outside the caustics are considered to be escaping the group, it follows that the caustic describes the escape velocity $v_{\text{esc}}^2(r)$ of a cluster as a function of distance r from its centre. An idealised caustic fit to a simulated cluster, taken from Serra et al. (2011) is shown in Figure 3.1.

Here follows a brief overview of the physical justification behind the caustic method. Full details of the model can be found in DG97, D99 and Serra et al. (2011).

Assuming a spherically symmetric model for a group, the escape velocity within a shell of radius r for a group is given by $v_e^2(r) = -2\phi(r)$. Given our position as observers, it is the line-

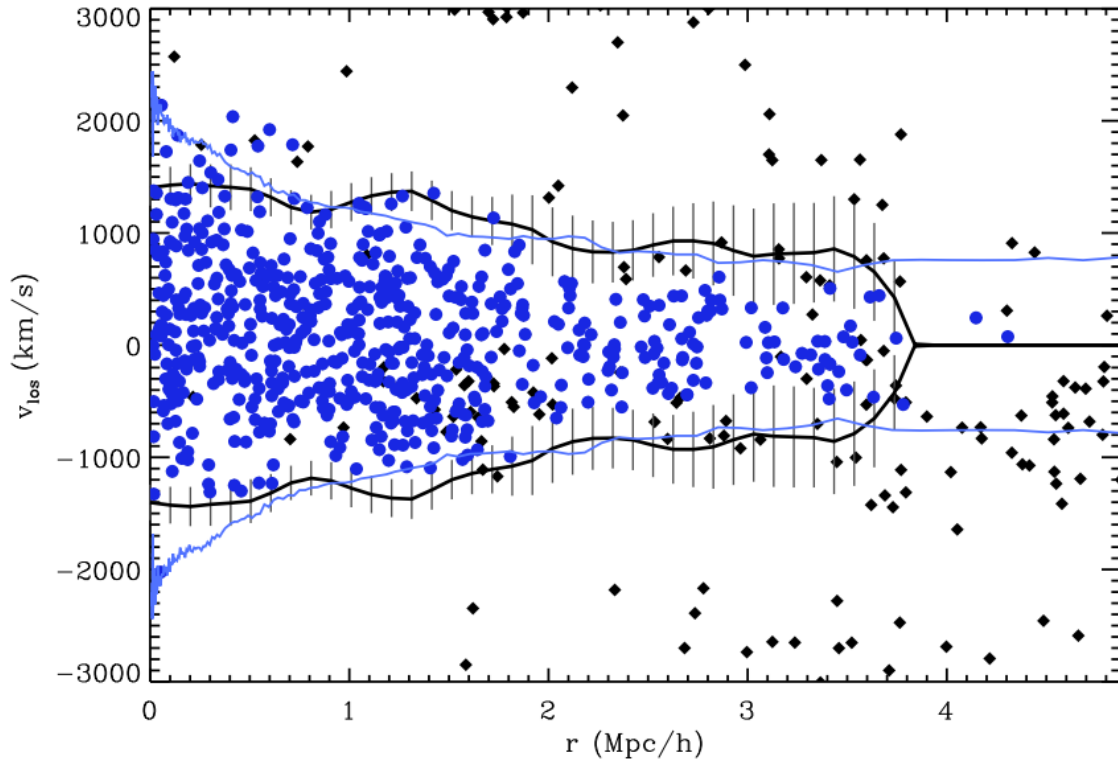


Figure 3.1: In this figure taken from Serra et al. (2011), a simulated galaxy cluster is plotted in redshift space, with blue points denoting galaxies within the cluster, and black points galaxies outside the cluster. The cyan lines are true caustics for this cluster, and the black lines represent the fitted caustic, with 1σ errors shown.

of-sight component (v_{los}) of this escape velocity that determines the location of the caustics, but this value depends upon the escape velocity profile of the cluster which we may not always know. Instead, we require an expression for the caustic amplitude that is independent of the escape velocity profile. Serra et al. (2011) determine such an expression, summarised below:

We begin by looking at the velocity anisotropy parameter $\beta(r) = 1 - (\langle v_\theta^2 \rangle + \langle v_\phi^2 \rangle) / 2\langle v_r^2 \rangle$ where v_θ , v_ϕ and v_r correspond to the longitudinal, azimuthal and radial components of an individual galaxy's velocity respectively. $\beta(r)$ therefore measures the velocity anisotropy of the group by taking the averages of these velocity measurements for each group member. If we assume that group rotation is negligible, $\langle v_\theta^2 \rangle = \langle v_\phi^2 \rangle = \langle v_{\text{los}}^2 \rangle$ and $\langle v_r^2 \rangle = \langle v^2 \rangle - 2\langle v_{\text{los}}^2 \rangle$ where v_{los} is the line-of-sight component of the velocity. Rearranging this for $\langle v^2 \rangle$ and incorporating the equality of all velocity components into the expression for $\beta(r)$ gives:

$$\langle v^2 \rangle = \langle v_{\text{los}}^2 \rangle \left(\frac{3 - 2\beta(r)}{1 - \beta(r)} \right) \equiv \langle v_{\text{los}}^2 \rangle g(\beta) \quad (3.1)$$

where

$$g(\beta) = \frac{3 - 2\beta(r)}{1 - \beta(r)} = \frac{2\langle v_{\text{los}}^2 \rangle + \langle v_r^2 \rangle}{\langle v_{\text{los}}^2 \rangle} \quad (3.2)$$

The potential of a system is related to its escape velocity in the form of $-2\phi = \langle v_{\text{esc}}^2(r) \rangle$, so it is possible to link the potential to the caustic amplitude if we make the assumption that $\mathcal{A}^2(r) = \langle v_{\text{esc,los}}^2 \rangle$ in the form

$$-2\phi(r) = \langle v_{\text{esc,los}}^2 \rangle g(\beta) = \mathcal{A}^2(r) g(\beta) \quad (3.3)$$

which combines all unknowns into a single parameter, β . The final link in the chain is to consider the mass of an infinitesimal shell:

$$G \, dm = -2\phi(r) \mathcal{F}(r) \, dr = \mathcal{A}^2(r) g(\beta) F(r) \, dr \quad (3.4)$$

with $\mathcal{F}(r) = \frac{-2\pi G \rho(r) r^2}{\phi(r)}$. Combining the expression for $\mathcal{F}(r)$ and $G \, dm$ brings us back to the familiar result for the mass of a shell; $G \, dm = 4\pi G \rho(r) r^2 \, dr$. Integrating Eq. 3.4 gives

$$GM(< r) = \int_0^r \mathcal{A}^2(r)g(r)F(r)dr \quad (3.5)$$

This expression is close to what we need, but is limited by the fact that the density profile of the system needs to be known in order to get a value for the mass. To overcome this, D99 assume that $\mathcal{F}_\beta(r) = \mathcal{F}(r)g(\beta)$ is a slowly varying function with respect to the radius of the system in hierarchical clustering scenarios, and this result is confirmed by Serra et al. (2011). We therefore set $\mathcal{F}_\beta(r)$ to be constant and adopt the value of $\mathcal{F}_\beta = 0.7$ of Serra et al. (2011) (this is not critical, as we later adjust our mass estimates by a scaling factor, discussed below), giving the final expression

$$GM(< r) = \mathcal{F}_\beta \int_0^r \mathcal{A}^2(r)dr \quad (3.6)$$

which can be used to provide mass estimates for our data sample. The Serra et al. (2011) $\mathcal{F}_\beta = 0.7$ value is measured from simulated galaxy groups.

3.1.1 The caustic mass method - algorithm

A successful caustic mass estimation algorithm must be able to accurately infer the continuous dark matter distribution in the halo that the group resides in, from a discrete set of points – the galaxies within the group. The most important goal of this algorithm is to correctly determine the location of the caustic for a group. Given that caustics are lines where densities tend to a certain value, locating them involves calculating a smooth 2D density distribution for the group members in redshift space (this is typically done by convolving the 2D discrete data with a kernel; e.g. Pisani, 1993). Based on this density distribution the algorithm then determines a threshold at which the caustic is placed.

For a given galaxy group for which member positions are known (α , δ , and z), the projection into redshift space takes place via the following transformations:

$$r = \frac{cD_A(z_c)}{H_0} \tan \psi \quad (3.7)$$

and

$$v = c \frac{z - z_c}{1 + z_c} \quad (3.8)$$

where D_A is the comoving distance to the galaxy, z_c is the redshift of the group centre, and ψ is the angular separation of a given member galaxy from the group centre at redshift z along the line of sight.

Having placed each group member galaxy into redshift space, we can now calculate the 2D density distribution of these galaxies within redshift space. Consider N galaxies in a cluster distributed in a redshift diagram with coordinates $\mathbf{x} = (r, v)$. Using an adaptive kernel method (Silverman, 1986), the density distribution of these galaxies is described as

$$f_q(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N \frac{1}{h_i^2} K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h_i}\right) \quad (3.9)$$

where K is the adaptive kernel

$$K(\mathbf{t}) = \begin{cases} 4\pi^{-1} (1 - t^2)^3 & \text{if } t < 1, \\ 0 & \text{otherwise} \end{cases} \quad (3.10)$$

and $h_i = h_c h_{\text{opt}} \lambda_i$ is the local smoothing parameter. $\lambda_i = [\gamma / f_1(\mathbf{x}_i)]^{1/2}$ where f_1 is equation (3.9) where $h_c = \lambda_i = 1$ for any i and $\log \gamma = \sum_i \log [f_1(\mathbf{x}_i)] / N$. The motivation for using an adaptive kernel estimator is to have a density estimator that can adapt to density distributions where the true probability density changes quickly (Pisani, 1993); this is generally a caveat of fixed kernel estimators that risk to oversmooth or undersmooth the probability distribution. The true distribution of galaxy members within a group in redshift space is not always known (the grouping algorithm is imperfect, nor are we able to observe all galaxies in the Universe). Finally, the optimal smoothing parameter h_{opt} is

$$h_{\text{opt}} = \frac{3.12}{N^{1/6}} \left(\frac{\sigma_r^2 + \sigma_v^2}{2} \right)^{1/2} \quad (3.11)$$

where σ_r and σ_v are respectively the uncertainties in the galaxy coordinates. The positional uncertainty σ_r is calculated from the astrometric uncertainties in GAMA for each galaxy and is negligible, while for σ_v we use uncertainty in the redshift measurement of the galaxy. This

was typically around 80 km s^{-1} when this work was first done; however GAMA has since automated its redshift measurement process and the typical uncertainty has decreased to 50 km s^{-1} .

Performing this calculation can take a great deal of time, particularly when it comes to calculating f_q and optimizing for the best value of h_c ; an initial attempt to make the calculation on a personal computer was abandoned after two days. A faster, less computationally intensive and time saving way of obtaining the density estimator is to use Fast Fourier Transforms (FFT) to convolve a two-dimensional histogram of the data in redshift space with the adaptive kernel. The process is described in Silverman (1986) for one dimension, but can easily be extended to two.

This process begins by creating a two-dimensional normalised histogram of the galaxies in redshift space (projected radius from the group centre versus line-of-sight velocity), and over the same parameter space, a histogram of calculated values for the kernel given in Eq. (3.10). The smoothing parameters λ_i , h_{opt} and h_c are used to adjust the values of the data histogram, but ultimately have little impact on the final calculated value of the density estimate. The density estimate $f_q(r, v)$ is defined as the inverse FFT of the product of the forward FFT of the data and kernel histograms, or in other words:

$$f_q(r, v) = \mathcal{F}^{-1}(\mathcal{F}(\text{data}) \times \mathcal{F}(\text{kernel})) \quad (3.12)$$

where \mathcal{F} and \mathcal{F}^{-1} denote forward and inverse fast Fourier transforms respectively. The end result of this calculation is a two-dimensional matrix that describes the density distribution of the group galaxies in redshift space; note that the modulus of this matrix is used in order to discard any phase shifts caused by the Fourier transforms. The matrix has dimensions of $2^8 \times 2^8$; any size lower than $2^7 \times 2^7$ does not provide the algorithm enough resolution to give reliable results and using values larger than $2^8 \times 2^8$ produces negligible improvements in resolution at the cost of computer time. The caustics are drawn on the density distribution on locations where $f_q(r, v) = \kappa$, and κ is obtained by minimising the function $S(\kappa, R)$ taken from D99:

$$S(\kappa, R) = |\langle v_{\text{esc}}^2 \rangle_{\kappa, R} - 4\langle v^2 \rangle_R|^2 \quad (3.13)$$

where the term $\langle v_{\text{esc}}^2 \rangle_{\kappa, R}$ is the square of the mean escape velocity at R for a given value of κ . This corresponds to the average size of the caustic amplitude from the group centre to the maximum projected radial distance R for a given value of κ and $\langle v^2 \rangle_R$ is the group velocity dispersion taken from the G³C. Using the group velocity dispersion from the G³C means that this method is not fully independent from the dynamical masses used in the group catalogue (as they both depend on velocity dispersion). The value of κ that minimises $S(\kappa, R)$ is the parameter that most affects the final calculation of where the caustics lie and the resulting mass estimate, as opposed to the parameters that go into calculating $f_q(r, v)$.

To minimise the function $S(\kappa, R)$, the value of κ for which the average caustic size is equal to $4\langle v^2 \rangle_R$ is used. The R (R Development Core Team, 2011) function `OPTIM` is used to do this; `OPTIM` is a general-purpose optimisation function based around the Nelder-Mead algorithm (Nelder & Mead, 1965), which lends itself particularly well to this task as $S(\kappa, R)$ is a parabolic function with only one well-defined minimum – more complicated optimisation algorithms are therefore not needed. Once the location of the contour is determined (black line in Figure 3.2), the algorithm draws the caustics (green lines in Figure 3.2 along $\min\{|v_u|, |v_l|\}$ where v_u and v_l correspond to the upper and lower values of the line-of-sight velocity of the group along the contour. The algorithm scans through the density distribution in bins of r , and for each bin selects the minimum of these two velocities, $v_u(r)$ and $v_l(r)$, and reflects it along the line-of-sight velocity axis (i.e. $v = 0 \text{ km s}^{-1}$). The caustic amplitude beyond the maximum extent of the group is artificially set to 0, even though often the caustic closes before the maximum radial extent of the group is reached (see Figure 3.2).

Based on mock catalogues of galaxies built with N-body simulations in DG97, there is a constraint on the logarithmic derivative of the caustic amplitude: $d \ln \mathcal{A} / d \ln R \leq 2$. Any values of $\mathcal{A}(r)$ for which this derivative does not hold are considered to be the result of the caustic algorithm coming up with the wrong location for the caustic at that particular radius, often due to excessive foreground/background galaxies. In other words, DG97 find that for a vast range of simulated groups, galaxies never have velocities such that the logarithmic derivative of the caustic amplitude is greater than 2. Instead, in these cases a caustic amplitude such that $d \ln \mathcal{A} / d \ln R = 1/4$ is used, as in Serra et al. (2011).

To summarise, the algorithm works as follows:

1. Convert the galaxy positions in redshift space into a two-dimensional histogram of galaxy

number densities,

2. create another histogram of the same dimensions containing values for the kernel as per Eq. (3.9),
3. calculate $f_q(r, v)$ using Eq. (3.12),
4. calculate the best value for κ with Eq. (3.13),
5. fit the caustics by reading off the minimum value of $|v_u|$ and $|v_l|$ along r , whilst ensuring that the derivative inequality holds,
6. integrate between the caustics to estimate the mass of the group using Eq. (3.14) and scale it accordingly using caustic mass estimates of mock galaxy groups to obtain a median-unbiased estimator.

The final mass expression is

$$\frac{M_c}{h^{-1}M_\odot} = \frac{0.7A_c}{G/(M_\odot\text{km}^2\text{s}^{-2}\text{Mpc})} \int_0^r \mathcal{A}(r)^2 dr \quad (3.14)$$

where r is given in units of h^{-1} Mpc and \mathcal{A} in s^{-1} km. $\int_0^r \mathcal{A}^2(r) dr$ is calculated by discretising $\mathcal{A}(r)$ over a set of equally spaced steps and A_c is the caustic mass scaling factor. Figure 3.2 shows example caustic fits for four friends-of-friends mock galaxy groups of descending total luminosity (from $10^{12} h^{-2} L_\odot$ to $10^9 h^{-2} L_\odot$).

Despite the computational efficiency of the FFT method to calculate the density estimate, there are a number of drawbacks that need to be addressed. To begin with, the area over which the 2D histograms for the data and the kernel are created need to be larger than the area the data spans. This is to avoid the kernel (and the resulting density estimate) wrapping around the borders due to the periodic nature of Fourier transforms and results in a very smooth density distribution compared to those shown in DG97 and D99. This will artificially increase the sizes of the caustics and cause a systematic overestimation of the caustic mass (though the scaling factor A_c corrects for this); the smoothing is independent of group size as shown in Figure 3.2. The presence of background galaxies is not considered in this analysis, and it is assumed that the friends-of-friends algorithm used in Robotham et al. (2011) has recovered group members as accurately as possible. This is tested later in this chapter.

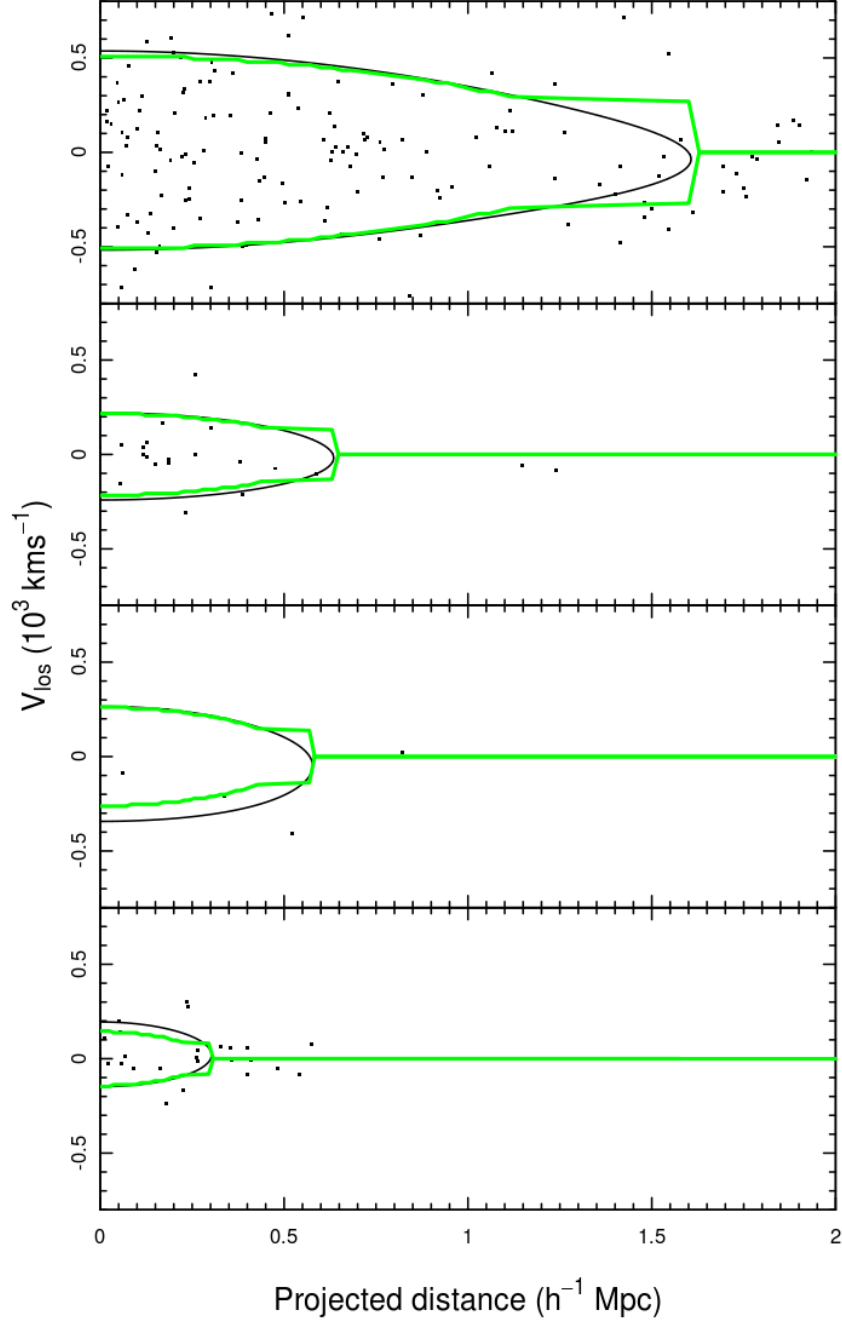


Figure 3.2: Four examples of the placement of the contour κ (black) and the caustics fitted to it (green) for four FoF mock galaxy groups (whose galaxies are shown as the black points) from the G³C in descending order of luminosity. It is evident here that the FFT method used to estimate the density distribution causes the final caustics to be very smooth with respect to the caustics drawn in DG97 and D99. This introduces a source of error in our caustic mass estimates. Notice how the gradient requirement $d \ln \mathcal{A} / d \ln R = 1/4$ is enforced towards the ends of the caustics in all panels, where the green lines no longer follow the black caustics.

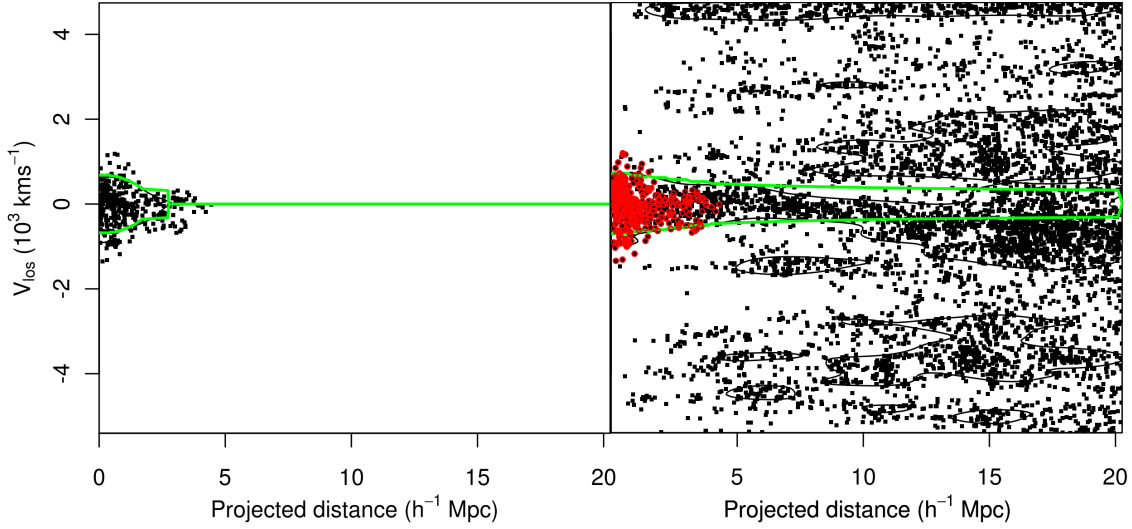


Figure 3.3: Side-by-side comparison showing how the caustic mass estimation performs for a large mock galaxy group. In both panels, the black points represent the locations of galaxies in redshift space, the black lines show the contour $f_q(r, v) = \kappa$ and the green lines are the caustics drawn along $\min\{|v_u|, |v_l|\}$. The left panel is a caustic fit only for galaxies present within this mock group and shows a clear example of the ‘trumpet’ distribution. The right panel includes nearby galaxies in redshift space out to $\pm 4R_{100}$ and $|z - z_{\text{med}}| \leq 4 \times \max|z - z_{\text{med}}|$ whilst circling the original group members in red. In this case the trumpet distribution is lost and the caustics are artificially closed at the end of the sample.

3.2 Caustic mass estimates of groups

Before applying this algorithm to the actual group catalogue, it is important to understand how well it performs when estimating masses for a set of mock catalogues that have been prepared alongside the G³C, where true masses are known. We therefore calibrate our caustic mass estimation algorithm using a set of 9 GAMA mock galaxy catalogues (described above).

In these mock catalogues, the true grouping of galaxies is known, so a well-informed calculation of their halo mass is possible. This acts as a benchmark for our caustic mass estimation algorithm and allow us to experiment with different implementations of the method. Serra et al. (2011) show that fine-tuning the parameters of the caustic mass algorithm (h_c , q , κ) does not provide a considerable improvement of its results; as described above, parameters such as h_{opt} do not change the final mass estimate by much when changed. Instead, to test for the impact of group centre definition, the algorithm is run using different values for the group centre given in the G³C: the centre of light, the iterative group centre and the brightest cluster galaxy. This provides a useful way of testing the stability of the caustic algorithm to different definitions of the group centre, as well as confirming the conclusions from Robotham et al. (2011) of which definition is the most appropriate.

3.2.1 Sensitivity to definition of group centroid

In Robotham et al. (2011), it is shown using comparisons to mock catalogues that the iterative method seems to be particularly robust at picking out the true group centre even in the presence of outliers. These tend to throw off the estimated group centre when using the centre of light method, as one would expect any mean to be susceptible to outliers; the centre of light method depends most explicitly on the group being recovered correctly. The brightest cluster galaxy approach is also robust to outliers, but analysis reveals that the iterative method recovers the most group centres that match the mock groups: the iterative centre of light method picks out the correct group centre 90% of the time. For the caustic mass algorithm, the most robust definition of the group centre should produce the most stable results; in other words, the variance in the scale factors A_c for each multiplicity and redshift bin should be minimal for the most stable group centre. This is confirmed by running the caustic algorithm on the mock group catalogue and changing only the location of the group centre from the iterative centre-of-light group centre to the brightest group galaxy and then to the centre-of-light: $\sigma^2(A_{\text{IterCen}}) = 0.192$ whereas $\sigma^2(A_{\text{BGG}}) = 0.201$ and $\sigma^2(A_{\text{CoL}}) = 0.361$. It is expected that the tendency of the centre-of-light method to incorrectly define the group centre to be near an outlying bright galaxy to throw off the placement of the caustics by deforming the density distribution of galaxies in redshift space. However, the caustic algorithm is expected to be robust to minor perturbations in the placement of the galaxies in redshift space, as the difference in the variances given above is minor.

3.2.2 Sensitivity to definition of group radius

A second test involves changing the number of galaxies that I consider when calculating the mass for each group, i.e. artificially increasing group membership. This will test the caustic mass algorithm's robustness to incorrectly defined groups. I do this by extending the boundary of the group in redshift space to include some nearby galaxies. In redshift space, this extension is defined as:

$$|z - z_{\text{med}}| \leq \Delta z; \Delta z = 4 \times \max |z - z_{\text{med}}| \quad (3.15)$$

Spatially, I increase the maximum distance a given galaxy can be from the group centre while it is still considered to be in the group. This effectively checks whether the caustic

algorithm is sensitive to other interloping groups or poorly defined groups. Including extra galaxies that in some cases belong to other groups not only systematically increases the mass estimates made by the caustic algorithm as one might expect, it also increases the mean spread of the results (defined as the ratio of the logarithm of the true and estimated mass) from $\langle\sigma^2\rangle = 8.33 \times 10^{-3}$ for caustic mass estimates made with the groups as they are, to $\langle\sigma^2\rangle = 0.0126$. One expects the overall mass estimate to increase as a result of including more galaxies, but the increase in spread is unexpected: the caustic algorithm ought to work better with a greater number of galaxies. Instead, the algorithm is unable to correctly place the caustics in the redshift space diagram because the extended search cut includes galaxies that are most likely inside other independent groups (Figure 3.5).

Visual examination of the redshift space plots in Figure 3.3 of these extended group cuts shows that the inclusion of nearby galaxies disrupts the distinctive trumpet shaped distribution seen in ideal spherical groups – galaxies that belong in other groups sit off at high values of line of sight velocity. Figure 3.4 shows a mass comparison between the group catalogue data and the extended group cut. Shown are a subset of groups; those that with redshift between 0 and 0.1, and with 5 to 9 members (before including nearby galaxies). The caustic mass estimates made with the extended group cut (in dashed black lines) show a much greater spread, due to the reasons described above. This result highlights the importance of carefully determining group membership when using the caustic algorithm, as the presence of galaxies that are not associated with the group being considered can have a catastrophic effect on the locations of the caustics, ultimately resulting in an incorrect mass estimate. For this reason, much of the work that has been done with caustics (DG97, DG99 and Serra et al., 2011) also incorporates a groupfinding method. In the case of this work, combining the caustic method with the friends-of-friends algorithm used to determine group membership for the G³C and mocks allows me to significantly reduce the probability of the caustics being placed incorrectly due to the presence of interloper galaxies within a group.

3.2.3 Caustic mass estimation in mock groups

Just as the mock catalogue was used to calibrate the groupfinding algorithm by Robotham et al. (2011), so can the catalogue be used to calibrate the caustic mass algorithm. Using the original grouping from the G³C with the iterative group centre, I calculate caustic mass estimates for every group in the mock catalogues, as well as A_c for a set of redshift and mul-

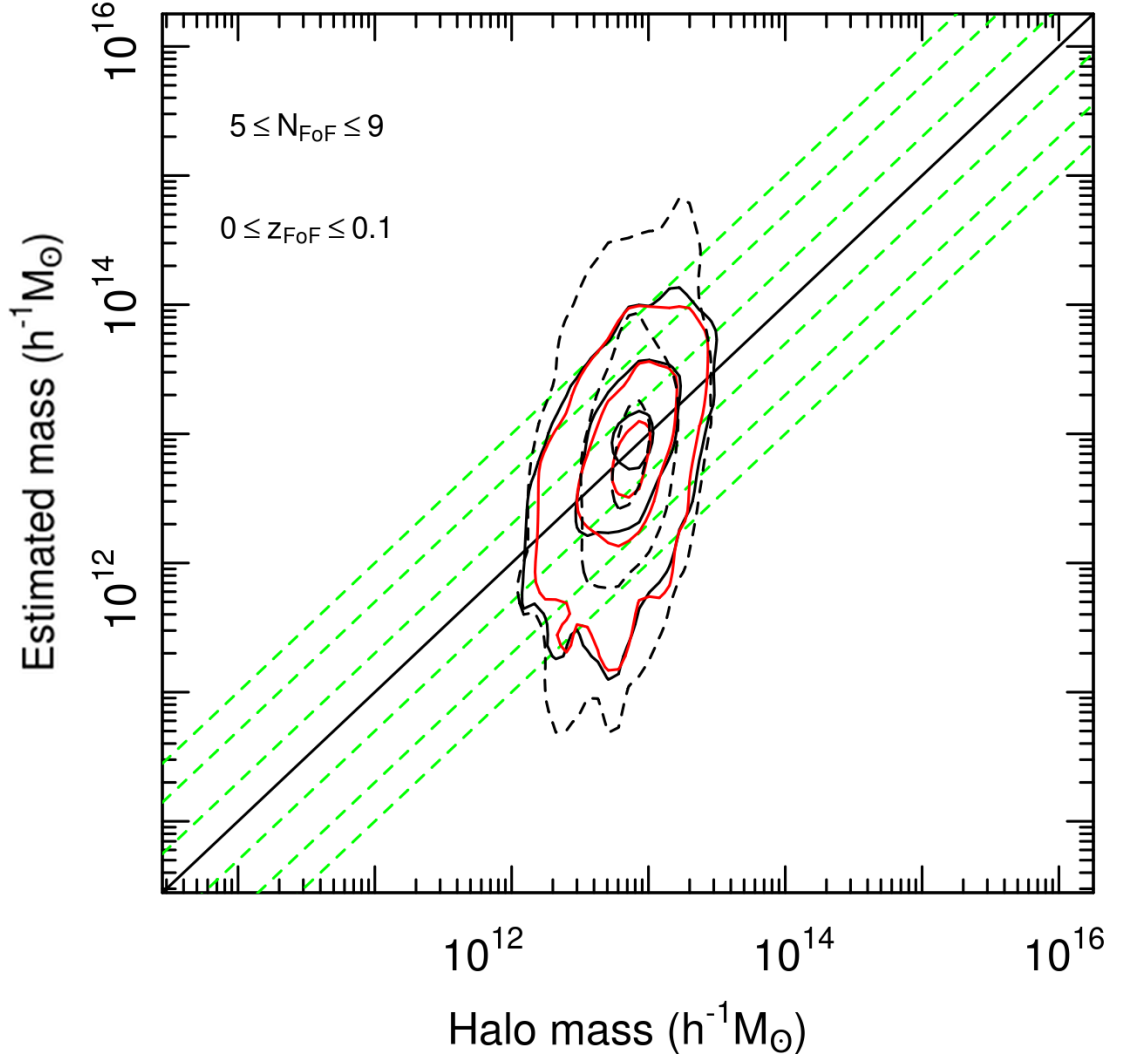


Figure 3.4: The black dashed contours show the results of running the caustic algorithm on the mock catalogues and including galaxies nearby in the field, or in other groups. By comparison, the solid black contours represent the mass estimates made when considering only galaxies known to be in each group from the simulation. Finally, the red contours represent the distribution of the dynamical mass estimates from the G³C. All three distributions have been adjusted to the same median as the dynamical mass estimates, shown in the red contours. Each contour line contains 10, 50 and 90% of the points, and the dashed green lines are 2/5/10 times away from the median.

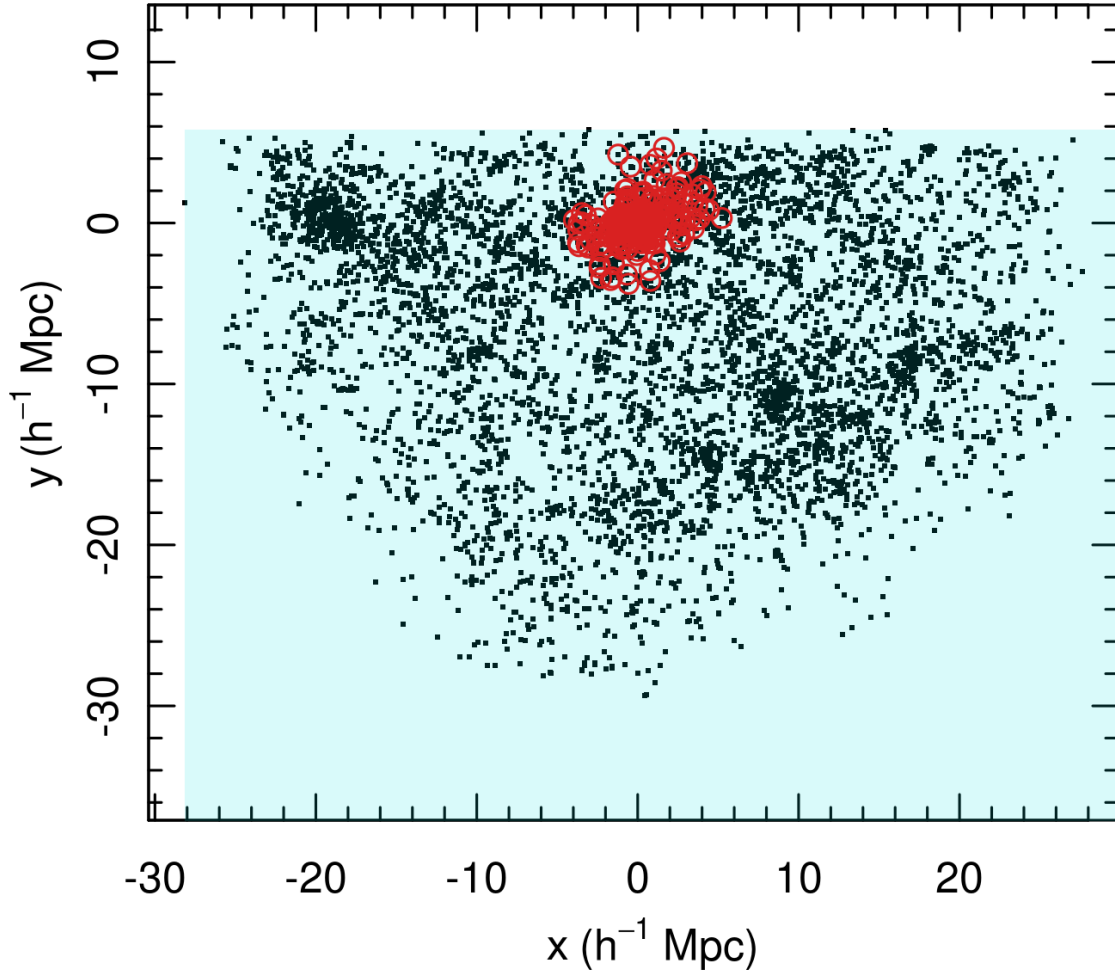


Figure 3.5: A projection of the mock group in Figure 3.3 showing all galaxies within the mock group circled in red, and all galaxies that are detected when the algorithm includes nearby galaxies out to $\pm 4R_{100}$ and $|z - z_{\text{med}}| \leq 4 \times \max|z - z_{\text{med}}|$. The area shaded in blue represents the physical size of the simulation region at the median redshift of the group. At this range it is evident that the algorithm is including galaxies that are likely to belong to other separate groups.

3.2. Caustic mass estimates of groups

	$2 \leq N_{\text{FoF}} \leq 4$		$5 \leq N_{\text{FoF}} \leq 9$		$10 \leq N_{\text{FoF}} \leq 19$		$20 \leq N_{\text{FoF}} \leq 1000$	
	19.4	19.8	19.4	19.8	19.4	19.8	19.4	19.8
$0 \leq z_{\text{FoF}} \leq 0.1$	1.63	1.63	0.43	0.43	0.45	0.46	0.41	0.41
$0.1 \leq z_{\text{FoF}} \leq 0.2$	1.58	1.59	0.43	0.43	0.42	0.42	0.38	0.39
$0.2 \leq z_{\text{FoF}} \leq 0.3$	1.52	1.53	0.42	0.44	0.36	0.38	0.35	0.35
$0.3 \leq z_{\text{FoF}} \leq 0.5$	1.18	1.21	0.29	0.31	0.29	0.29	0.24	0.26

Table 3.1: Values for A_c for each group subset in both the $r_{\text{AB}} < 19.4$ and $r_{\text{AB}} < 19.8$ mock G³C mock catalogues using iterative group centres and with all galaxies in group. Including these numbers in Eq. 3.14 gives a median-unbiased estimate for the group mass. Note that without including these factors, the mass estimate can vary by up to a factor of 3, most likely caused by the smoothing FFT caustic fitting method used.

tiplicity bins in order to make caustic mass estimates for observed groups median unbiased (Table 1). Figure 3.6 shows the results of these calculations for each multiplicity and redshift bin, comparing the distribution of caustic mass estimates (in black) to the dynamical mass estimates (red). As it is designed to do, the caustic method performs best in very populated groups ($N \geq 200$; of which there is only one in the G³C). This is shown by the fact that the scatter of the distribution of mass estimates shown in Figure 3.6 decreases as a function of increasing multiplicity. The scaling factor A_c should be unity for an ideal galaxy; instead the caustic mass is systematically greater than the masses of the mock groups, i.e. $A_c < 1$. The only exception to this is for groups with two to four members, where the caustic algorithm is much more likely to fail to find appropriate contours, and thus defaults the mass to a specific value of $10^4 h^{-1} M_\odot$. This artificially adds a tail to the distribution of masses for that group subset. There may be a few reasons for this: (a) the simulated galaxy groups in the mock catalogues not being perfectly spherical, (b) the groups not being virialised (one of the basic assumptions of the spherical infall theory, on which the caustic algorithm is based, is that the group is virialised), (c) the extra smoothing in our caustic introduced by using FFT to calculate the density distribution $f_q(r, v)$, and (d) the fact that the caustic method is designed to work on groups with more than 200 members. Note also that the caustic mass estimate is very susceptible to the velocity dispersion measured for the group, which is always less of a robust measurement when considering groups with few members. The overall variation in the scale factor A_c is roughly a factor of 4.5, which is of the order of the range of scaling factors used to calibrate the G³C dynamical mass estimates (see Table 3 in Robotham et al. (2011)). In contrast to the dynamical mass estimate scaling factors, the caustic mass scaling factors vary less as a function of redshift, but are far more sensitive to variations in group multiplicity. However, both mass estimates perform equally badly when dealing with low multiplicity groups.

As expected, the least satisfactory results are for groups that are galaxy pairs, where there is little velocity dispersion information for the caustic algorithm to use. This explains the presence of large tails in the $2 \leq N_{\text{FoF}} \leq 4$ panels of Figure 3.6. However, note that both methods tend to fail at these low multiplicities, which is to be expected as dynamical mass measurements rely on accurate velocity measurements as well. For higher multiplicity and redshift cuts the caustic mass estimate PDFs shown in this figure agree extremely well with those of the dynamical mass estimates; indicating again similarities between both algorithms when it comes to performing badly for certain galaxy groups. This can be seen in the scatter plots as both contours tend to follow roughly the same profile, meaning that it is likely that a group that performs badly in one algorithm is likely to do so with the other. This is particularly visible for the $10 \leq N_{\text{FoF}} \leq 19$, $0.3 \leq z_{\text{FoF}} \leq 0.5$ bin where there is a secondary concentration of high mass groups that is present in both distributions.

Figure 3.7 converts the information shown in Figure 3.6 into a set of probability density functions where the ratio between the caustic and dynamical mass estimates and the known halo mass (shown in black and red respectively) is displayed alongside the ratio between both mass estimates (in blue). The PDFs are all generated by convolving the data with a rectangular kernel with a bandwidth of $0.1/\sqrt{12}$ (this value is chosen to be small so that the data isn't oversmoothed). As shown on Figure 3.7, for groups with a mid-range multiplicity ($5 \leq N_{\text{FoF}} \leq 19$) the caustic mass estimates have a greater spread than the dynamical mass estimates; this is true across all redshift bins. For groups with $N_{\text{FoF}} \leq 4$ the scatter is comparable across all redshifts. Both methods produce estimates that are within a factor of 2 in agreement with each other, which is very good. The large tails seen in Figure 3.6 for low multiplicity groups are again visible here: the small 'bumps' forming on the right hand of the PDF are seen, once again demonstrating that both methods tend to fail in similar ways. Despite the caustic algorithm being designed for high multiplicity groups, it is still possible to make reasonable estimates of the group mass for groups with only a handful of galaxies.

The caustic mass estimate should recover the intrinsic halo mass of each group as accurately as possible, as the grouping for these is known. However, when applying the algorithm to the real data in the G³C, I run the algorithm not on intrinsic groups, but on groups put together by the friends-of-friends algorithm used in Robotham et al. (2011). An important test therefore is to see how the caustic mass estimation algorithm performs on bijectively matched groups drawn out from the mock catalogue (instead of intrinsic mock halo groups) using the

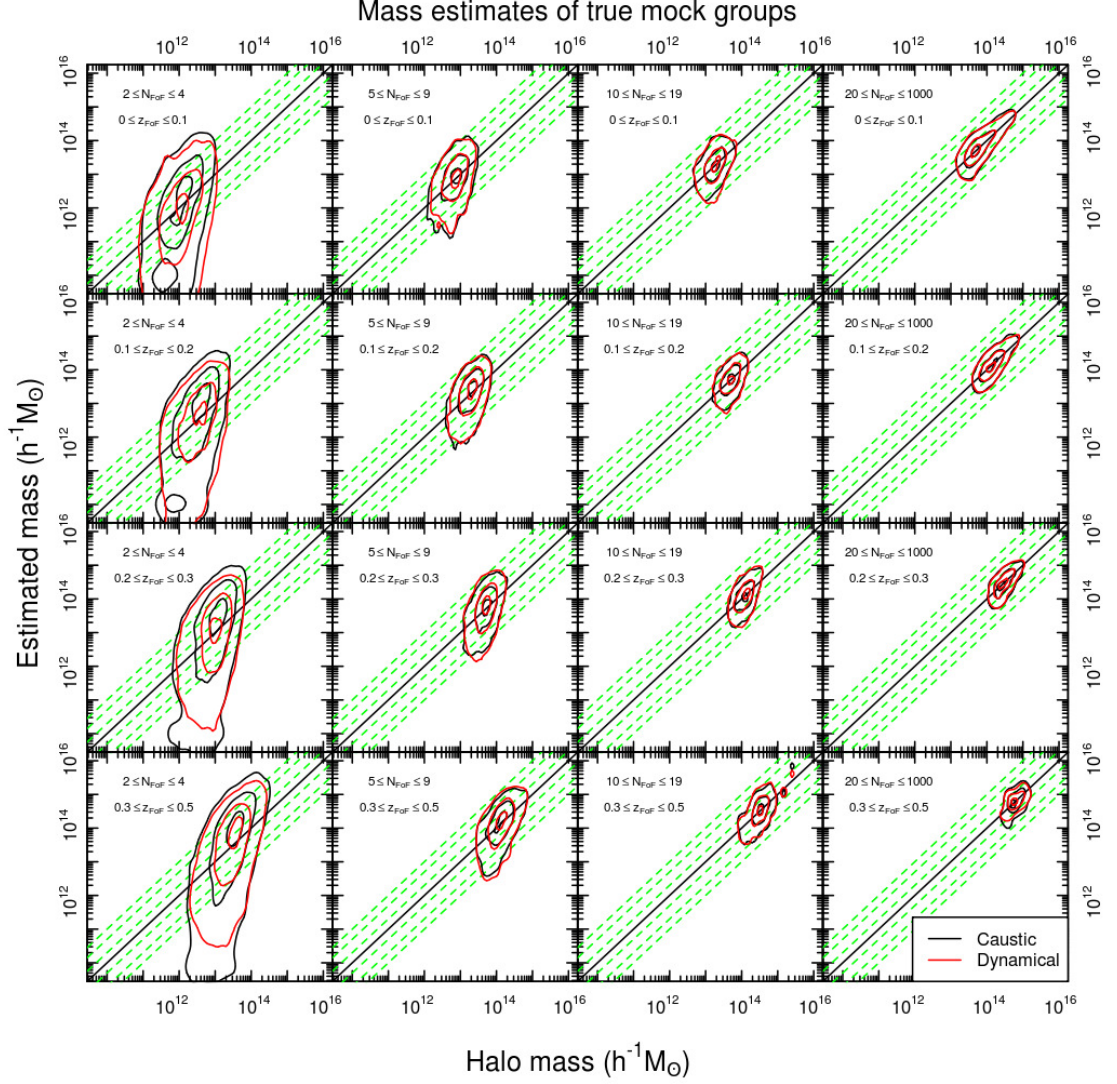


Figure 3.6: Distribution of caustic (dynamical) masses for the intrinsic mock groups as a function of halo mass, drawn in black (red) for $r_{AB} \leq 19.4$. For each panel, both mass estimates have been corrected to be median-unbiased. The contours represent areas containing 10, 50 and 90% of the groups and the green lines are regions where the mass estimate is 2/5/10 times off the true mass. Of particular interest is the tendency for both distributions to follow each other very closely, particularly when over or under estimating the true halo mass.

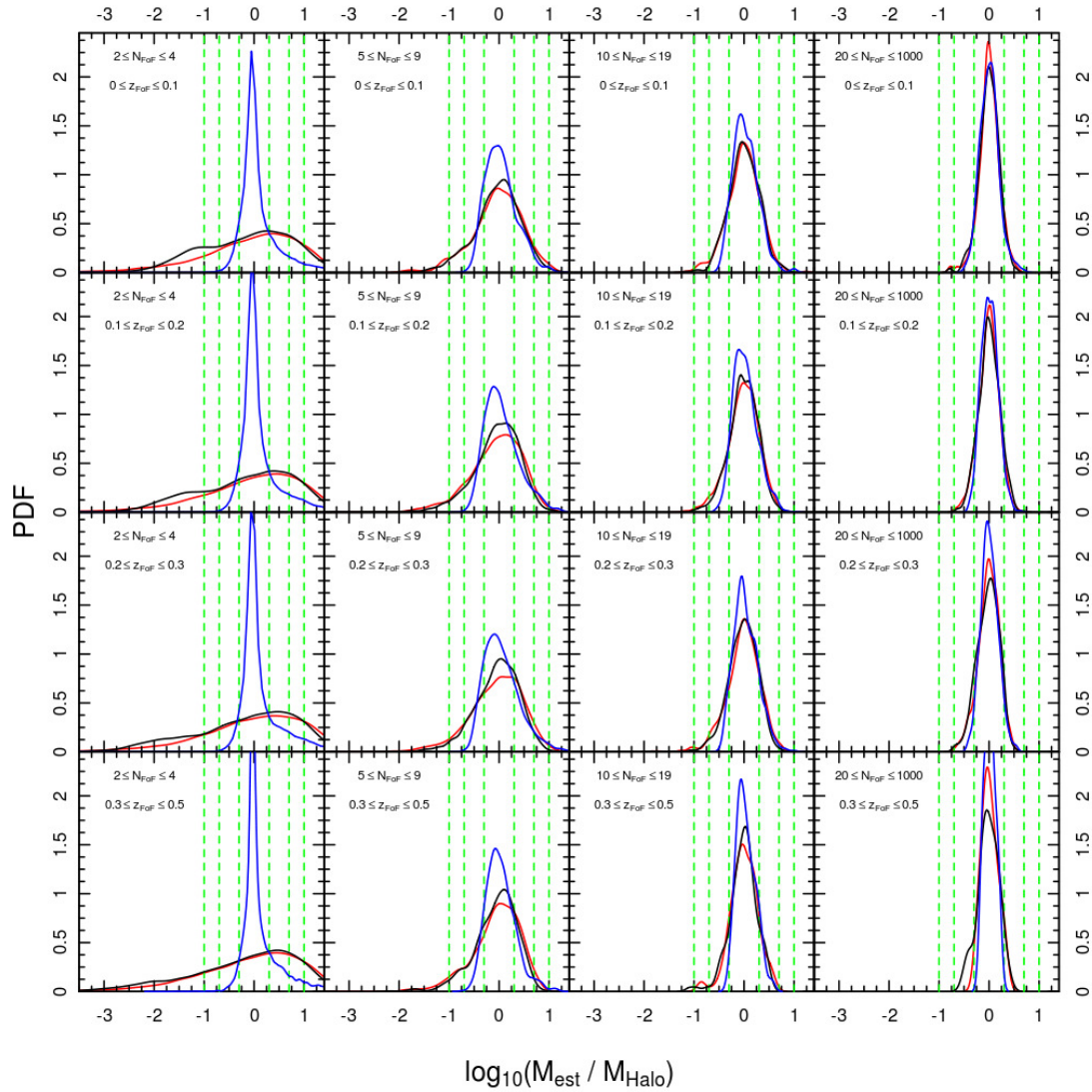


Figure 3.7: Probability distribution functions of median-unbiased values of $\log \frac{M_{\text{est}}}{M_{\text{Halo}}}$ for the $r_{\text{AB}} < 19.4$ sample, with caustic masses drawn in black, and dynamical masses in red. The blue line is the PDF of $\log \frac{M_c}{M_{\text{dyn}}}$ and highlights the agreement between the two methods. The green dashed lines indicate regions that are factors of 2/5/10 away from the ‘true’ mass. The difference between scatter in the caustic and dynamical mass estimates is often minimal, with the caustic method showing less scatter for groups with $5 \leq N_{\text{FoF}} \leq 19$.

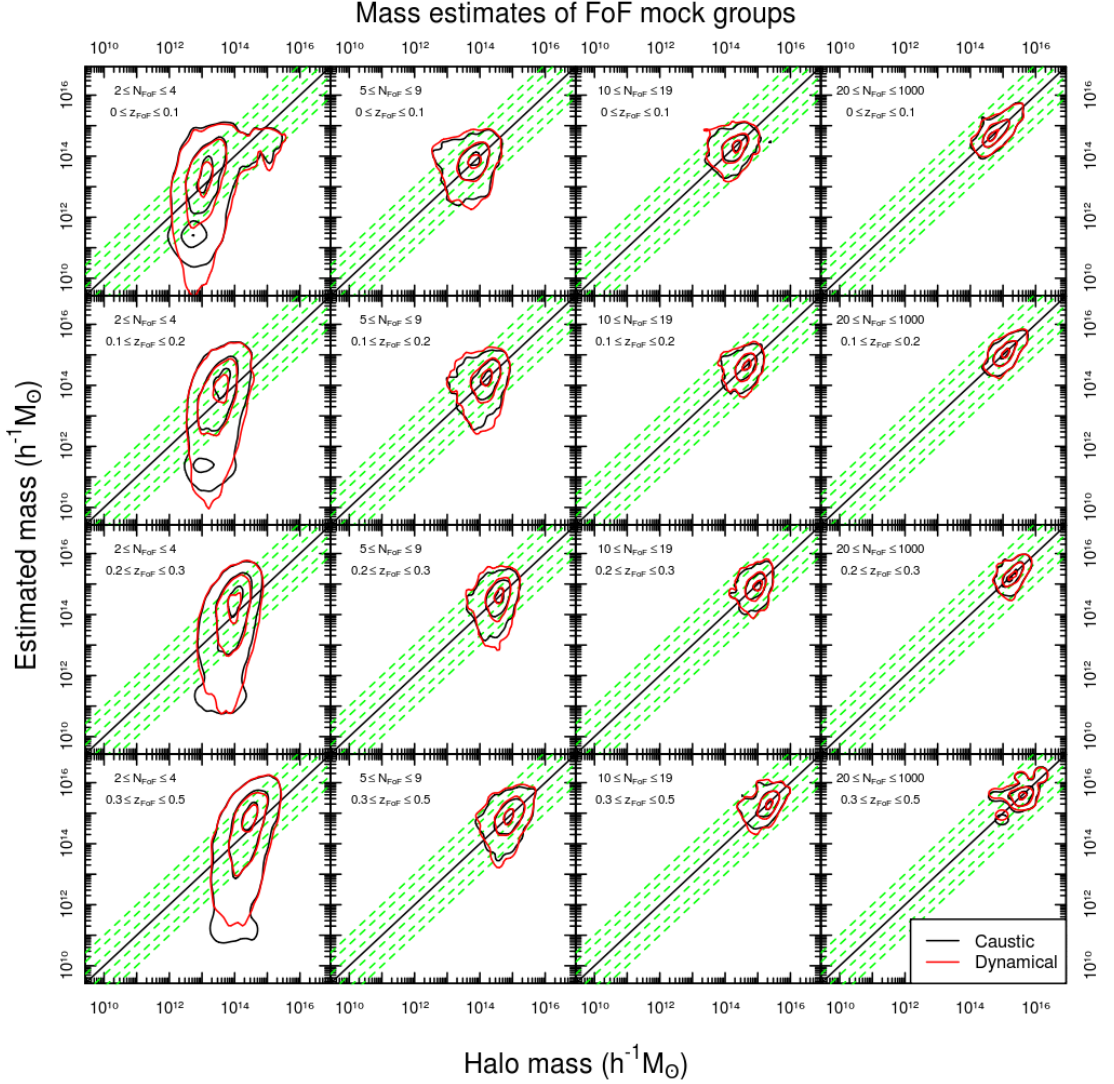


Figure 3.8: As Fig. 3.6 but for bijectively matched groups identified by the friends-of-friends algorithm instead of the true known intrinsic grouping. The coloured lines represent regions containing 10, 50 and 90 % of groups. The black contours compare the caustic mass to the FoF halo mass, and FoF dynamical mass estimate to the FoF halo mass estimate are shown by the red contours.

same friends-of-friends algorithm, and the results of this can be seen in Figures 3.8 and 3.9. In both figures the distribution of data mimics that of the trends for the caustic masses on the intrinsic mock groups, with these showing less scatter compared to the caustic masses of the FoF mock groups. Given that the design of the friends-of-friends algorithm used in Robotham et al. (2011) to construct the groups in both the mocks and the final group catalogue is to reject groups that have significant outliers, it is expected that the caustic method does better for intrinsic groups, where the grouping is perfect.

One final check is to examine how the caustic mass estimates perform as a function of

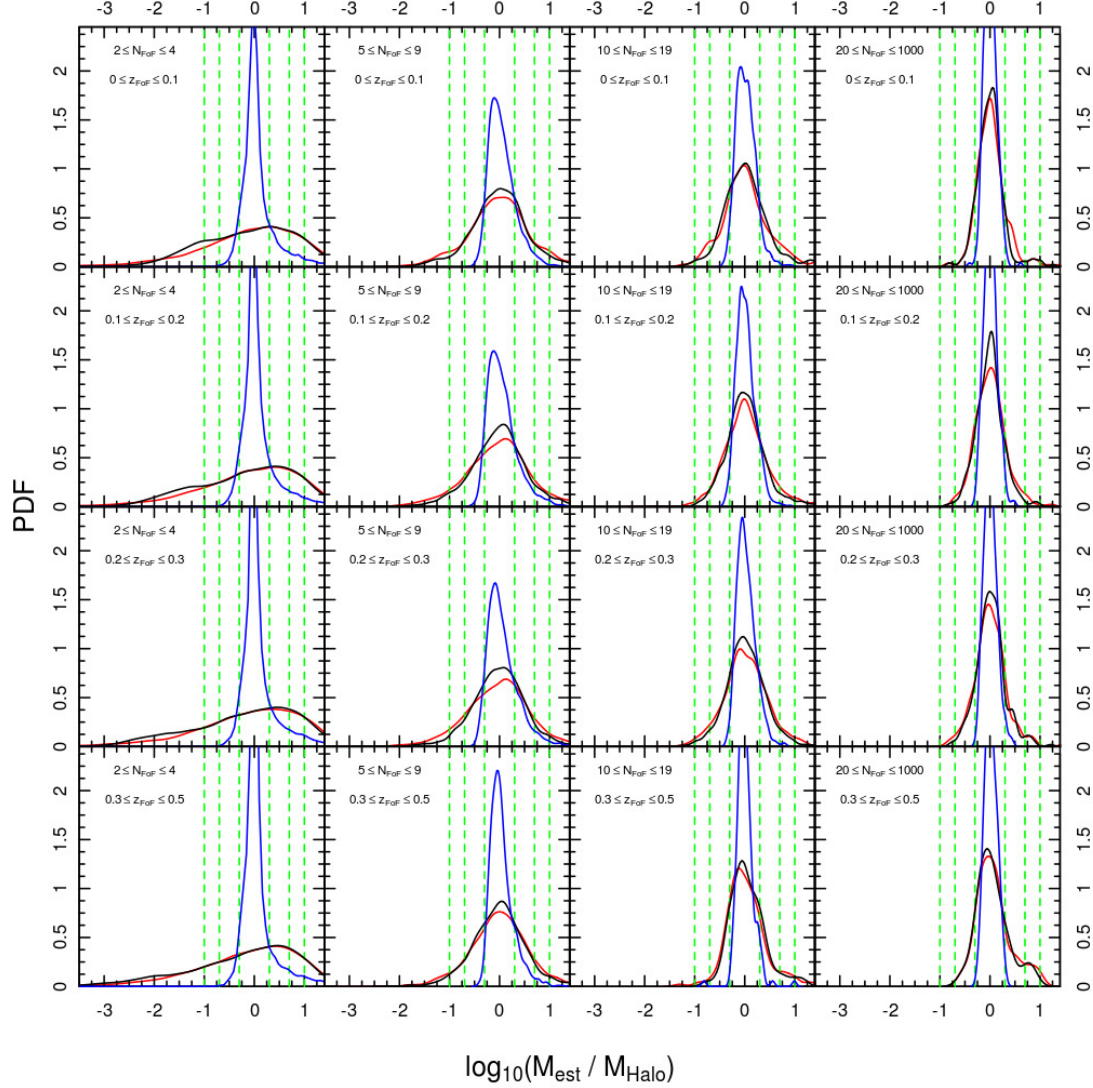


Figure 3.9: As Fig. 3.7 but for the bijectively matched FoF groups from the mock galaxy catalogues. The different lines show median-unbiased values of $\log \frac{M_{\text{est}}}{M_{\text{Halo}}}$ for the $r_{\text{AB}} < 19.4$ sample. The black line shows the ratio between the caustic mass and the FoF halo mass, while the red line shows the ratio between the FoF dynamical and halo masses. As before, the blue line shows the distribution of the ratio between the caustic mass and the FoF dynamical mass.

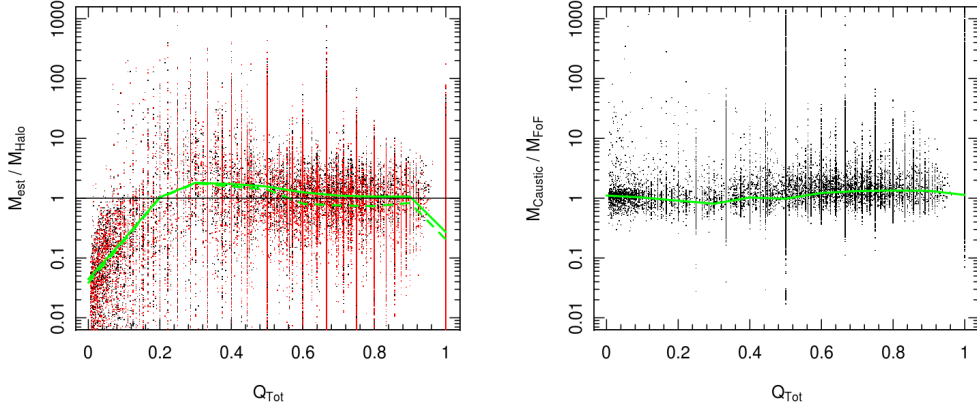


Figure 3.10: *Left:* A comparison of the performance of the two mass estimation methods as a function of the grouping quality parameter, Q_{Tot} . Black points correspond to caustic masses, and red ones to dynamical masses. Both methods perform worst when Q_{Tot} is close to 0, but quickly recover at approximately 0.2. The green lines show a rolling median in each 0.1 bin for the caustic mass (solid line) and for the dynamical mass (dashed line). *Right:* The ratio between the caustic and the dynamical mass estimates as a function of Q_{Tot} , with the rolling median in each 0.1 bin overplotted in green.

the quality of the grouping Q_{Tot} , which is defined in Robotham et al. (2011) and described in the previous chapter. Figure 3.10 shows how the two mass estimates behave as a function of Q_{Tot} . The left panel shows that the caustic mass estimates behave in just the same way as the FoF mock dynamical mass estimates, performing well after a total quality factor of about 0.2. This is further demonstrated in the right panel of this figure, where the ratio between the caustic and dynamical mass estimates for the FoF mock groups remains close to unity as a function of Q_{Tot} , with a small tendency for the caustic mass to be systematically greater than the dynamical mass as Q_{Tot} approaches 1. This may be caused by the fact that for $Q_{\text{Tot}} = 1$ (i.e. a perfectly recovered group), the scale factors that account for the imperfections in grouping and ‘correct’ the caustic mass estimates are no longer needed.

Based on these results, the caustic mass estimates for the mock FoF groups in the catalogue are made using group centres obtained with the iterative CoL rejection method and only include galaxies considered to belong to each group. The same redshift and multiplicity cuts used in Robotham et al. (2011) are used. I calculate the scaling factors A_c for each bin as the necessary value to ensure that the median of the ratio of the mass estimate to the true mass is unity. Values for A_c are listed in Table 3.1.

3.2.4 Caustic mass estimation in observed groups

With the appropriate scale factors in Table 3.1, I run the algorithm on the actual group catalogue itself, using the same redshift and multiplicity cuts, and based on the results from the previous section, use the grouping from the catalogue and the iterated CoL. This provides me with a full set of halo mass measurements for every group in the catalogue that are complementary of the dynamical masses already existing in the G³C. Figure 3.11 shows the ratio between the caustic and dynamical masses as a function of dynamical mass without scaling the caustic masses to be median unbiased with respect to the mocks. The figure shows the importance of the scale factor A_c and shows that the caustic masses tend to improve as a function of rising dynamical mass, except for a handful of extreme groups with $M_{\text{dyn}} \sim 10^{16}$ where the caustic mass seems to wildly overestimate the mass of the group.

Both the caustic and the dynamical mass estimates are adjusted by their appropriate scale factors, and then compared. The results of this comparison can be seen in Figure 3.12 for the $r_{\text{AB}} < 19.4$ mag limited sample. By calculating the ratio of the caustic to the dynamical mass, I show that $90.8\% \pm 6.1\%$ groups have a caustic mass estimate that is within a factor of two of the dynamical mass estimate, which is a remarkably good agreement. In a vast majority of cases the two mass estimates agree very well with each other, particularly for groups at high redshift (bottom row). This is largely due to the fact that at high redshift, only the most massive groups are recovered, and so have very good velocity dispersion measurements. In all cases the scatter of the data is minimal, with the scatter decreasing with increasing group membership through a combination of better grouping and velocity dispersion measurements.

In order to make a comparison between the groups and mocks as fair as possible, it is necessary to account for the fact that velocity uncertainties have not been included in the mocks as this directly affects the kernel, through Eqn. (3.11). This issue is addressed in Robotham et al. (2011), whereby all groups with $\sigma^2 \leq 130 \text{ km}^2 \text{ s}^{-1}$ are removed from any comparisons between the groups and mocks. It is argued that below this cut the velocity dispersion of a group would be significantly affected by this uncertainty; since the dynamical mass estimate is directly proportional to σ^2 , a poor velocity dispersion calculation will throw off the final mass estimate. I apply the same velocity cut to the group and mock catalogues. Figures 3.12 and 3.13 display the resulting comparison between the ratio M_c/M_{dyn} of the caustic and dynamical masses for the groups (black) and mocks (black) for this subset of groups. There is still some

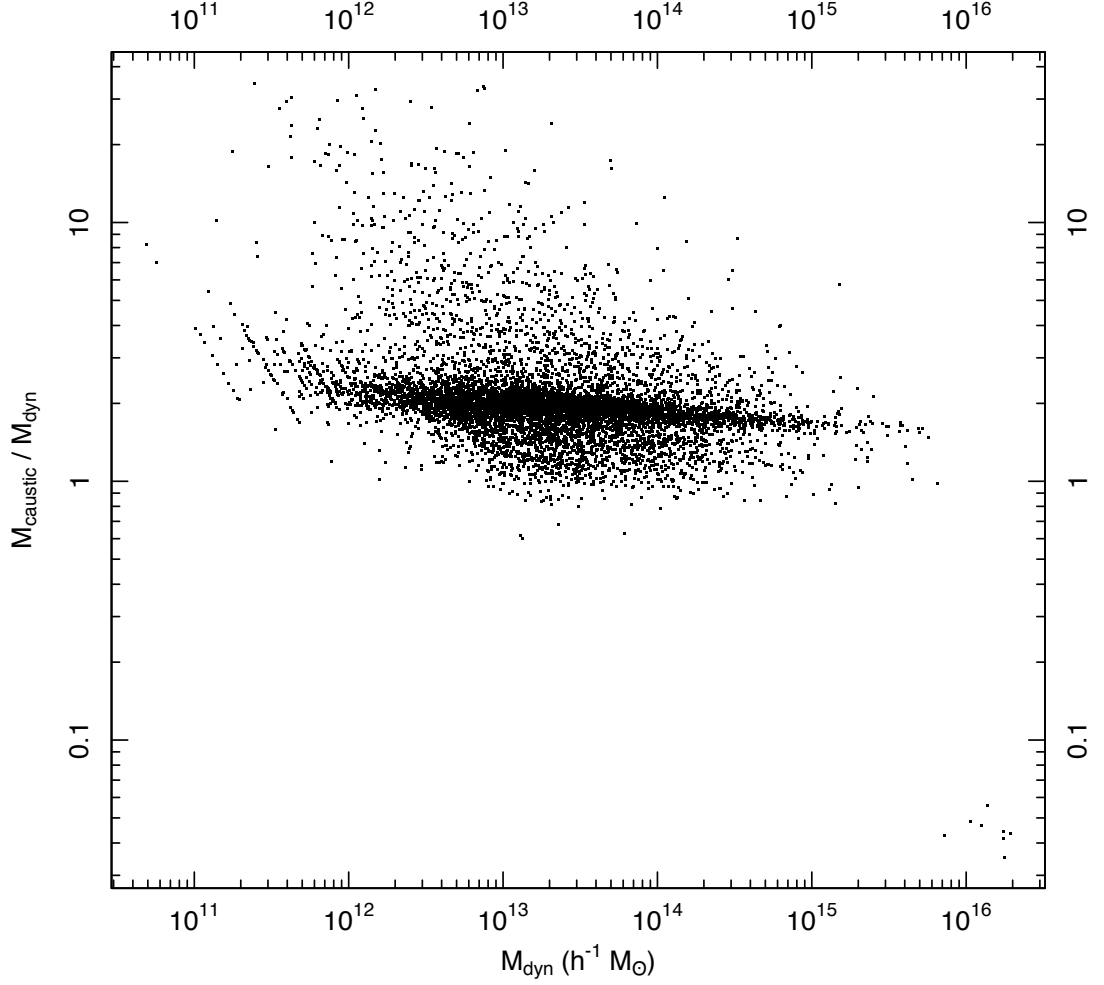


Figure 3.11: The mass ratio of caustic and dynamical masses as a function of dynamical mass. Note that the caustic masses in this figure have not been median unbiased with respect to the mocks using A_c , to show the importance of doing so. See Figure 3.12 for mass comparisons with the proper scaling factors applied.

difference between the scatter of the mass estimates for the groups and the mocks and the disagreement is most noticeable in cases where the caustic mass is greater than the dynamical mass. However, if this velocity cut is not included, and instead include the mock groups with poorly defined masses, the discrepancy between the two distributions increases.

Taking these considerations into account, these new mass estimates appear to agree with the existing dynamical mass estimates for the G^3C . While both methods utilise the same velocity and positional information in different forms, the dynamical mass method is not as sensitive to the full 2D velocity profile of each group and does not extend radially as far as the caustic mass method. This is caused by the dependence of v and r to each other, which is another effect that is considered in the caustic method, whereas the velocity dispersion and group radius used to calculate M_{dyn} are treated independently to each other. The good agreement between the mass estimates is also a good indicator that the GAMA groups have been well defined, since the caustic mass method is quite sensitive to a well defined group.

3.3 Summary and discussion

Using the caustic mass estimation algorithm introduced by DG97 and D99 complementary caustic mass estimates are made for observed and mock groups in the G^3C . This implementation of the algorithm is calibrated by running it on mock GAMA group catalogues where the intrinsic grouping is known; and for bijectively matched friends-of-friends mock groups. A scaling factor A_c is calculated for different bins of redshift and multiplicity and is used to calculate median-unbiased mock group mass estimates. In all cases there is very good agreement between the caustic mass and dynamical mass estimates. Of great note is the tendency for both PDFs shown in Figure 3.6 to match each other extremely well, implying that both algorithms perform equally well and equally badly for mock groups: that both estimate methods fail for low multiplicity ($2 \leq N \leq 4$) groups highlights an interesting systematic bias present in both methods that warrants further study.

Having been calibrated the algorithm on the mock catalogues, it is applied to the G^3C to obtain caustic mass estimates for real groups, and the results shown in Figure 3.12. As with the mocks, both mass estimates are generally consistent with each other: on average, 90.8 ± 6.1 % of groups have caustic and dynamical mass estimates that agree to within a factor of two, which is an excellent result. This is strong evidence for the reliability of the caustic mass estimation

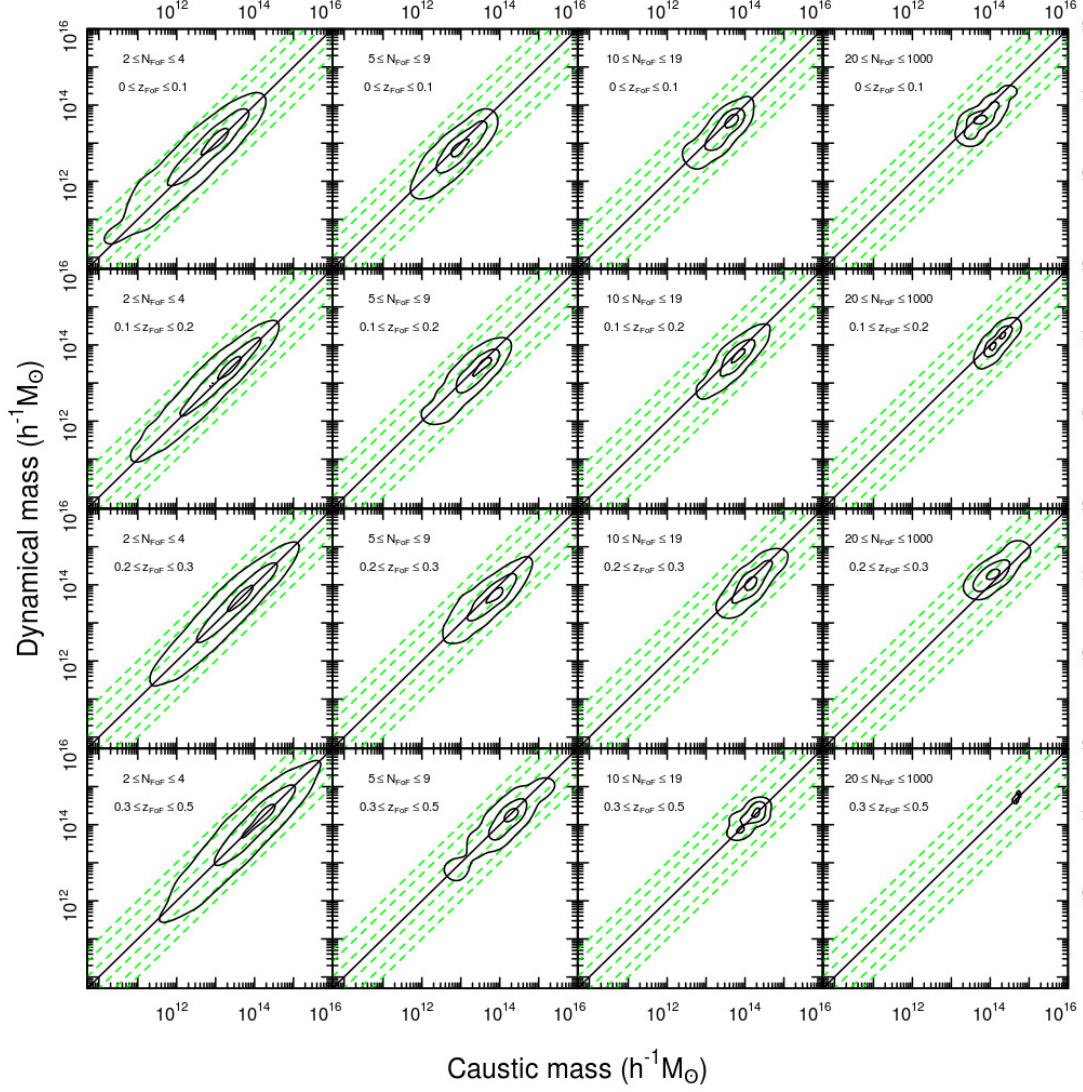


Figure 3.12: Distribution of the caustic masses compared to dynamical masses for real galaxy groups from G³C with $\sigma^2 \geq 130 \text{ km}^2 \text{ s}^{-1}$. The dashed green lines are as Figure 3.6, and the solid black lines represent regions containing 10, 50 and 90% of the data. There is a good correlation between the caustic and dynamical mass estimates as already shown in Fig 3.8 and 3.9 for the mocks.

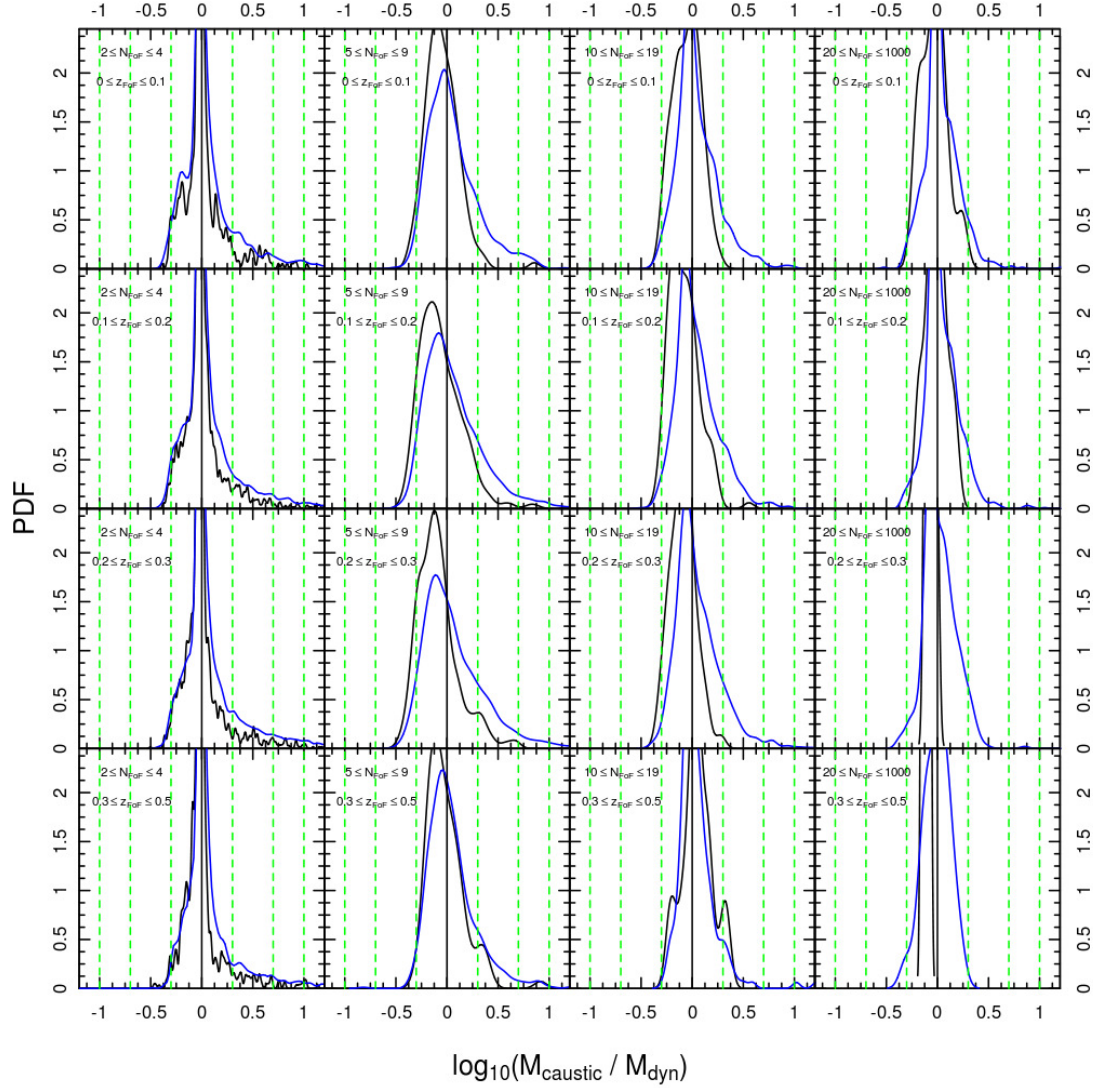


Figure 3.13: As in Figure 3.7 this figure shows the PDF of the logarithm of the mass ratio between the caustic and dynamical mass estimates for all groups with $\sigma^2 \geq 130 \text{ km}^2 \text{ s}^{-2}$. The blue line shows the same distribution but for the bijectively matched FoF mock catalogues with the same velocity cut to illustrate the smaller scatter between the two methods for the real data.

	$2 \leq N_{\text{FoF}} \leq 4$			$5 \leq N_{\text{FoF}} \leq 9$			$10 \leq N_{\text{FoF}} \leq 19$			$20 \leq N_{\text{FoF}} \leq 1000$		
	2	5	$\epsilon(N_2)$	2	5	$\epsilon(N_2)$	2	5	$\epsilon(N_2)$	2	5	$\epsilon(N_2)$
$0 \leq z_{\text{FoF}} \leq 0.1$	0.878	0.970	0.010	0.862	0.983	0.030	0.958	1.0	0.030	0.769	1.0	0.152
$0.1 \leq z_{\text{FoF}} \leq 0.2$	0.898	0.977	0.005	0.836	0.904	0.020	0.949	1.0	0.020	0.978	1.0	0.022
$0.2 \leq z_{\text{FoF}} \leq 0.3$	0.904	0.979	0.006	0.794	0.997	0.028	0.952	1.0	0.028	1.0	1.0	0
$0.3 \leq z_{\text{FoF}} \leq 0.5$	0.909	0.978	0.009	0.904	0.979	0.034	0.889	1.0	0.118	1.0	1.0	0

Table 3.2: Fraction of groups that have a caustic mass estimate that is within a factor of 2 and 5 from the dynamical mass estimate. The error $\epsilon(N_2)$ is defined as $\sqrt{1/N_2 - 1/N_{\text{tot}}}$ where N_2 is the number of groups within a factor 2 and N_{tot} is the total number of groups for a given bin. The number of groups present in each multiplicity bin drops sharply (from thousands of groups to 10 or less) after the first multiplicity bin, drastically lowering the relevance of these statistics.

method, particularly when considering the less than ideal conditions the algorithm faces when working with so many groups with few members. Despite being designed to work with groups with over 200 members, the caustic method is able to successfully determine accurate mass estimates for groups with as few as two members, highlighting the adaptability and strength of this method when combined with a friends-of-friends algorithm. Previous work using the caustic mass estimation method has highlighted the importance of running it on well-defined groups, so its success in this work is a credit to the robustness of the G^3C groups. The tightness of the contours shown in the first column of Figure 3.12 and the large fraction of accurately determined masses given in Table 3.2 are very convincing demonstrations that the caustic method is able to perform well even for very low multiplicity groups, and that these mass estimates are indeed correct.

In Figure 3.13, the ratio between the caustic and dynamical mass estimates in the groups is very sensitive to the application of a velocity cut at $\sigma^2 \geq 130 \text{ km}^2 \text{ s}^{-2}$, which is the minimum velocity for groups in the mock catalogue to have realistic velocity dispersions. The mock catalogues have not been analysed with realistic velocity errors, and so by discarding groups whose velocity dispersions (i.e. groups with $\sigma^2 \leq 130 \text{ km}^2 \text{ s}^{-2}$) would be badly affected by the inclusion of velocity errors, a fair comparison to the group catalogue can be made. Once this cut is applied, it is shown that the distribution of ratio between the caustic and dynamical masses is consistent between the mocks and groups, particularly for groups with 2 to 4 members.

These caustic mass estimates serve to add further credibility to the dynamical mass estimates, as the caustic algorithm is sensitive to the escape velocity profile of a given group out to its full extent. That both dynamical mass and caustic mass estimates for the mocks and the groups give such similar results despite their relative independence from each other is a positive result that not only reinforces the dynamical mass estimates, but also demonstrates

the power of the caustic approach when combined with a friends-of-friends algorithm. A non-GAMA specific version of the algorithm is freely available to use and can be downloaded from <http://www.gama-survey.org/pubs/>.

Further than verifying the dynamical masses for the GAMA groups, the work done in this chapter strengthens the assumption that the GAMA groups have been robustly defined. Several times in this chapter I have referred to the fact that the caustic mass method drastically underperforms and miscalculates mass estimates when run on groups that are poorly defined. Therefore, the fact that these caustic masses are matching well to the dynamical masses goes beyond the verification of the dynamical masses to say that the groups are indeed very well defined. This is particularly true for low multiplicity groups, where the caustic mass method still gives reliable results. This verification of the quality of the grouping in the GAMA group catalogue is a strong foundation for the work done in the following chapter, where groups serve as the most fundamental building block for filamentary network of the Universe.

4

The GAMA Large Scale Structure Catalogue

Both visible, and dark matter in the Universe is distributed in a non-uniform way. The earliest galaxy surveys, such as the CfA Redshift Survey (de Lapparent et al., 1986), paved the way in recognising that there is structure and order to the distribution of galaxies in the Universe. The dominant model of structure formation today is that galaxies tend to cluster into groups, which themselves form the building blocks of large scale structure we observe (Press & Schechter, 1974; Bahcall, 1988; Bond et al., 1996; Eke et al., 2004). The so-called ‘cosmic web’ is composed of clusters and superclusters of galaxies that are connected to each other by groups of galaxies (e.g. Bharadwaj et al., 2004; Colberg et al., 2005; Pimbblet, 2005; Novikov et al., 2006; Soubie et al., 2008, 2009). These structures themselves surround voids, which are extremely underdense regions containing a small number of isolated galaxies: voids typically have a galaxy density that is 10% less than that surrounding a typical field galaxy (Pan et al., 2012). The few galaxies that do exist in voids are subject to dynamics that are unique to these underdense regions (Blumenthal et al., 1992; Sheth & van de Weygaert, 2004). Galaxies can be classified as belonging to different types of density regions: filaments, clusters, or voids,

with each classification presenting a unique environment to that galaxy.

Our understanding of large scale structure has developed over recent years. Structure formation has developed through advanced simulations such as Angulo et al. (2012); Habib et al. (2012) and large galaxy surveys like the 2dFGRS (Colless et al., 2001), the MGC (Liske et al., 2003), the SDSS-DR7 (Abazajian et al., 2009), the 6dFGS (Jones et al., 2009) and GAMA (Driver et al., 2011; Liske et al. in prep) have helped to characterise large scale structure in the Universe. There is still some work to be done, however, on bridging the gap between observations and simulations, in order to establish whether or not the larger scale environment (i.e. $\geq 1h^{-1}$ Mpc) of galaxies influences their evolution. In other words, is a galaxy in a filament discernibly different from a galaxy in a void? If so, how can we use direct observations and simulations to find out?

Answering these questions requires a robust, reliable, and reproducible definition of the constituent parts of large scale structure: filaments, voids, and everything in between. The field of filament finding and classification has been expanding in recent years, with numerous algorithms currently being used to detect and define large scale structure (Sahni et al., 1998; Pimblet, 2005; Forero-Romero et al., 2009; Aragón-Calvo et al., 2010; Stoica et al., 2010; Murphy et al., 2011; Sousbie, 2011; Hoffman et al., 2012; Smith et al., 2012). In complement to this, there is a large volume of work that is currently being done to identify voids (El-Ad & Piran, 1997; Peebles, 2001; Hoyle & Vogeley, 2002, 2004; Aragon-Calvo et al., 2010; Thompson & Gregory, 2011; Pan et al., 2012).

In this chapter I discuss the heart of my thesis: an algorithm to identify and classify large scale structures in the three equatorial GAMA fields, and a series of catalogues it produces that identify different populations of galaxies as belonging to distinct types of large scale environments. Filaments of groups and galaxies are detected, as well as smaller coherent structures formed by individual galaxies on the peripheries of filaments, dubbed ‘tendrils’, as well as galaxies that lie in very underdense regions of space, referred to void galaxies.

4.1 Data

4.1.1 GAMA Group Catalogue

As discussed earlier in this thesis, the GAMA Group Catalogue, or G³C (hereafter I refer to the paper that describes its creation as R11), provides a comprehensive catalogue of 23838

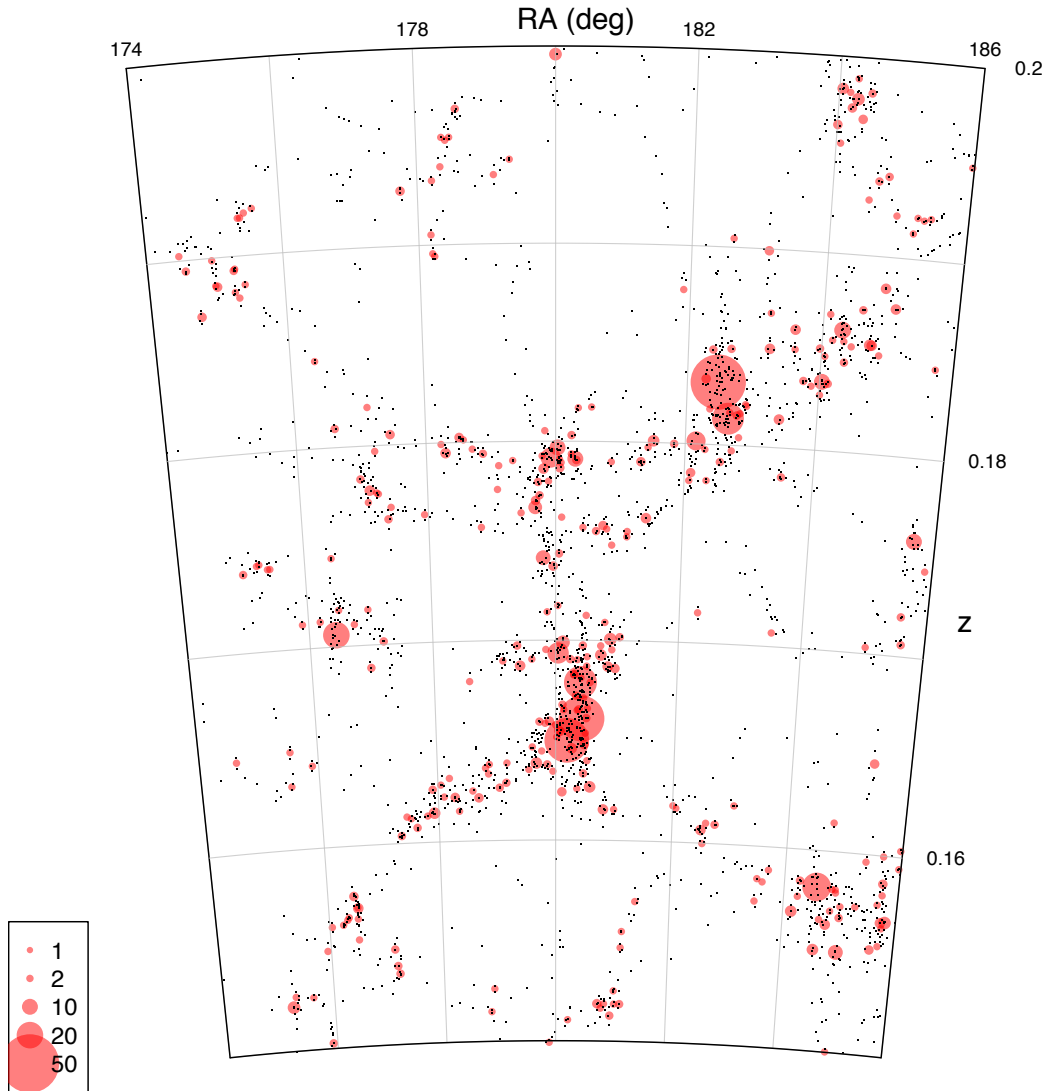


Figure 4.1: A cutout of the G12 field, with the $0.15 \leq z \leq 0.2$ and $-1^\circ \leq \delta \leq 0^\circ$ limits. The red circles represent group centres, sized according to the number of galaxies in each group. Black points represent galaxies inside and outside of groups.

galaxy groups across the three equatorial GAMA fields out to $m_r < 19.8$ mag. The catalogue groups 73298 galaxies out of a possible 180979, roughly 40% of all galaxies. A cutout of the G12 field is shown in Figure 4.1, with groups and galaxies shown together. Group centres are marked by red circles sized according to each group’s multiplicity. The black points represent all galaxies within and without groups.

As discussed in Chapter 2, the G^3C is generated with a friends-of-friends algorithm that operates on projected and radial separations independently (see Figure 1 in R11). This is a very important step, as it allows the algorithm to take redshift space distortions into account and effectively removes them. Figure ?? displays four panels with different populations of

galaxies and groups for the G12 region. The two panels to the left show, respectively, all galaxies that are within this region, and all groups recovered by the FoF group finder (coloured by their group luminosity in $L_{\odot}h^{-2}$). The third panel shows all galaxies in groups. The final panel shows all galaxies not in groups, which are defined as being isolated galaxies. This final population is very important, as it still traces large scale structure despite not being dense enough to be detected as structure by metrics that rely on local overdensities of galaxies. This ‘isolated’ population emphasises that large-scale structure exists on all scales, persisting even down to rather low values of local galaxy density. A complete quantification of large-scale structure must, therefore, not rely solely on a threshold in density, but must take into consideration the spatial distribution of galaxies themselves.

The G³C provides estimates of group centres and a number of size estimates (whose accuracy is tested in Section 3.2.1 of the previous chapter), integrated magnitude and luminosity measurements, and other properties for each group. Of greatest importance to the large scale structure work are the position estimates for each group, as the groups provide the first step in generating filamentary structure in the algorithm, as discussed in the following sections. Collapsing groups to their central point ensures that the filaments are free from redshift space distortions. The group centre used for this work is iterated centre of light, which in the previous chapter and in R11, is shown to be the most robust definition of the group centre when compared to mock groups.

4.1.2 Sample selection

As with any other body of observed data, it is important to ensure that the sample of galaxies and groups used in this work are as free as possible from any intrinsic bias, most often caused by observational effects and the inescapable limitations found in any galaxy survey. Absolute magnitude limited samples are particularly important for a study on large scale structure like this, where different populations of galaxies in varying density environments span many magnitude ranges (Driver et al., 2011); it also allows any linear structure finder to use a constant search length instead of varying it as a function of redshift, which is a great shortcut. To maximise the number of galaxies kept after a luminosity cut in the r -band, I seek out the absolute magnitude limit for a given redshift z such that I retain the largest possible number of galaxies and groups. Within the G³C, the proxy for absolute magnitude for a group is given by the TotFluxProxy parameter. This is defined as the total luminosity for the group, and

is corrected to account for selection effects and missing flux, and is given in units of L_* . The absolute magnitude of each galaxy can be calculated from its apparent magnitude.

For a given redshift, the absolute magnitude of the faintest possible galaxy that can be seen in GAMA is given by $M_r^h(z) = 19.8 - DM(z) - k(z)$. All galaxies whose magnitude $M_{\text{gal}} > M_r^h(z)$ for a given redshift z are discarded. After this any groups that have fewer than 2 member galaxies remaining are discarded, retaining only groups that would still have been detected with this absolute magnitude cut. The redshift that retains the largest number of galaxies and groups is $z = 0.213$, where $M_r^h = -19.77$ mag.

The sample selected by $z = 0.213$ and $M_r^h = -19.77$ mag can be seen in Figure 4.2. The numbers on the top left of this panel refer to the number of groups that are kept after the absolute magnitude cut is applied, and those that are discarded. Only groups that are kept are plotted. The numbers in the bottom left of this plot show the number of galaxies below the redshift limit, and those above (shown in the region shaded in red). The notable feature of this plot is that the group luminosity distribution is effectively flat below the redshift cut. The final sample contains a total of 45542 galaxies and 6000 groups across the three equatorial GAMA regions. It would be possible to include groups from the G³C with only one remaining galaxy after making a volume limited sample, but this would lead to more noisy large scale structures with many links to small, isolated groups.

Figure 4.3 highlights my sample within the context of the entire G³C, with the group luminosity plotted as a function of redshift for all groups. The points in red show all the groups in our sample, which have at least two members left. For any group left in the sample, I use its full group properties as listed in the G³C. I also apply this same sample selection to the mock galaxy and group catalogues.

The three panels in Figure 4.4 show, for the three equatorial fields from G09 to G15, all of the galaxies (grouped and ungrouped) in the selected sample. In all three regions the number density of galaxies increases sharply after $z \approx 0.1$. Lowering the absolute magnitude limit for the sample, thereby selecting more faint galaxies, would reveal more faint galaxies at low redshifts; however this would result in a much smaller sample size.

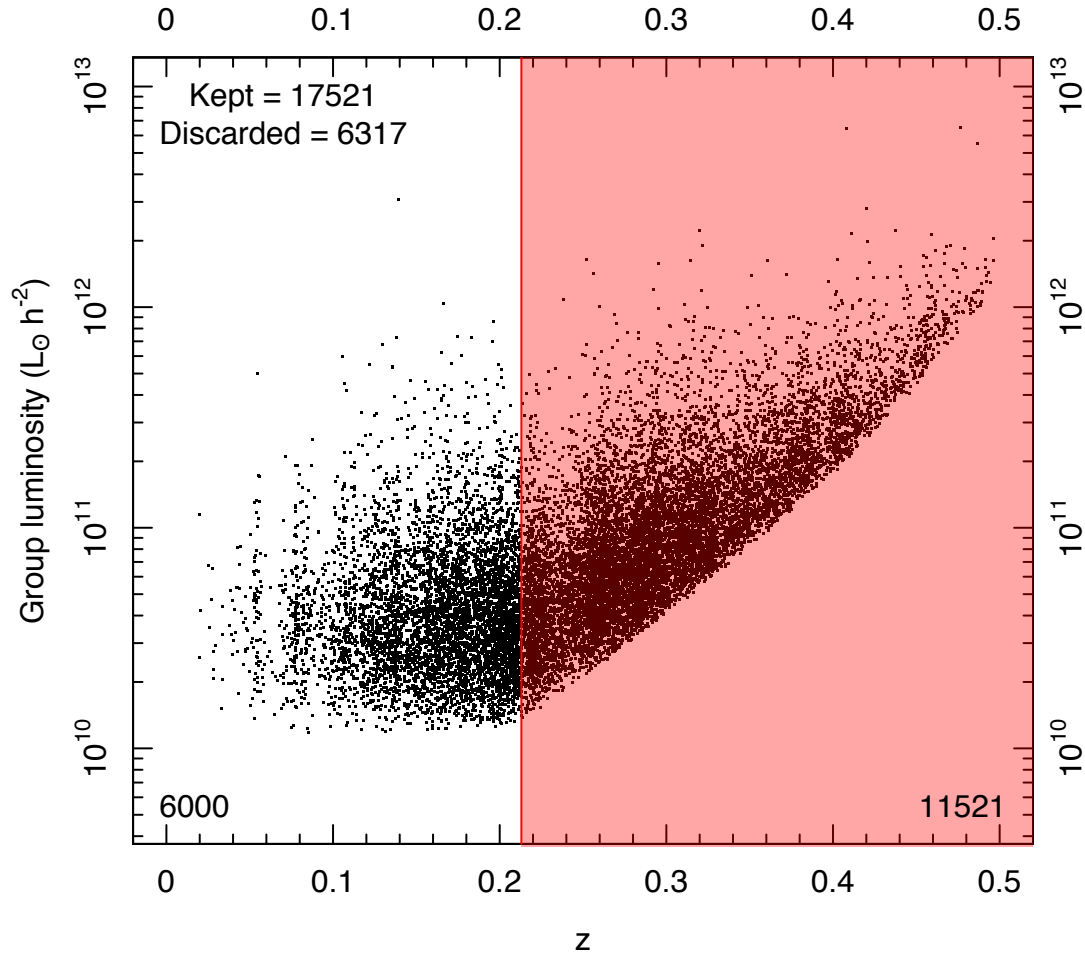


Figure 4.2: Distribution of the total group r-band luminosity (see R11 for more details) as a function of redshift, after the sample selection process has been applied. The numbers in the top left display the number of groups kept and discarded after removing galaxies - the kept galaxies are plotted in the figure. The region shaded in highlights the region with $z > 0.213$ and is no longer volume limited, and the numbers in the bottom corners show how many groups are above and below the redshift cut. We are therefore left with 3,589 groups across all three GAMA regions, with $z \leq 0.213$ and with at least two or more galaxies with $M_r \leq -19.77$. This sample selection ensures the structures we detect are volume limited.

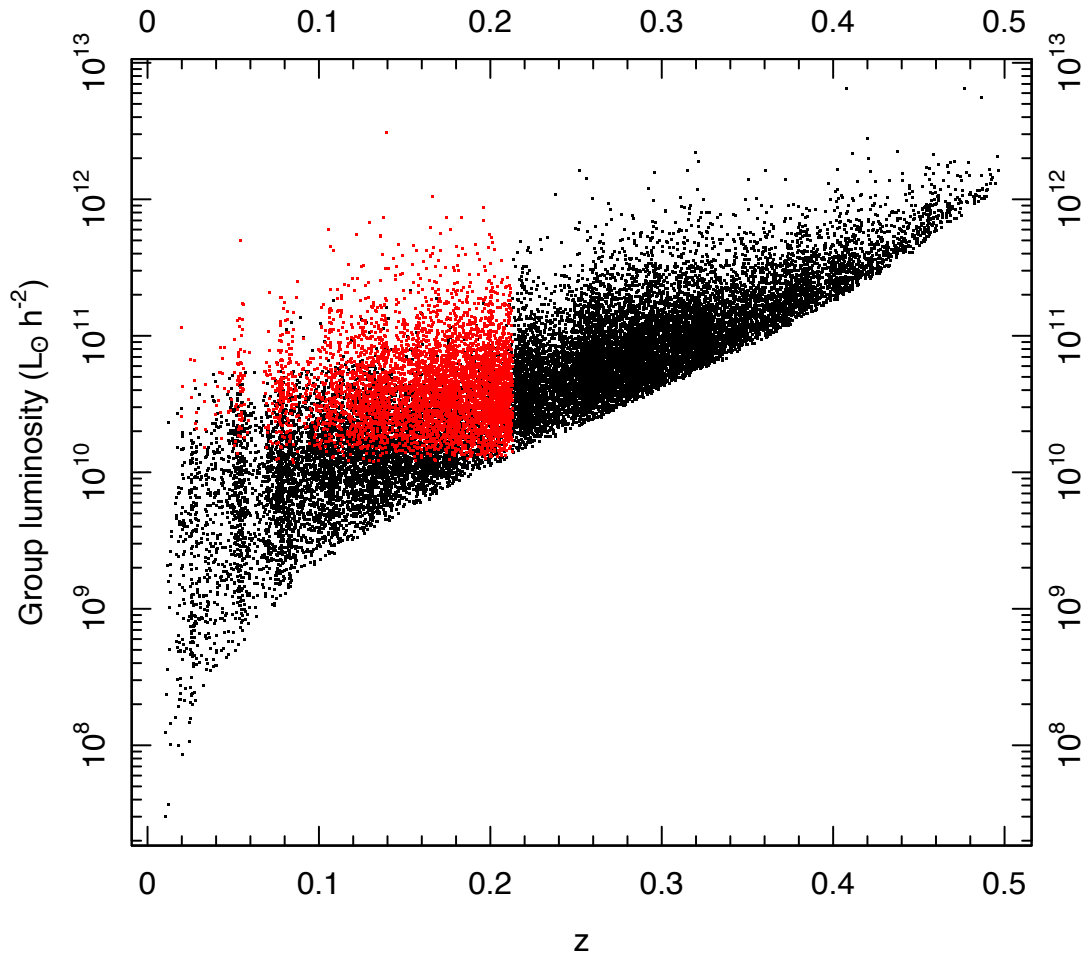


Figure 4.3: All groups in the G³C are plotted here, with our final sample shown in red. The redshift limit of $z = 0.213$ is easily seen here. The red sample corresponds to all groups in the unshaded region in the left.

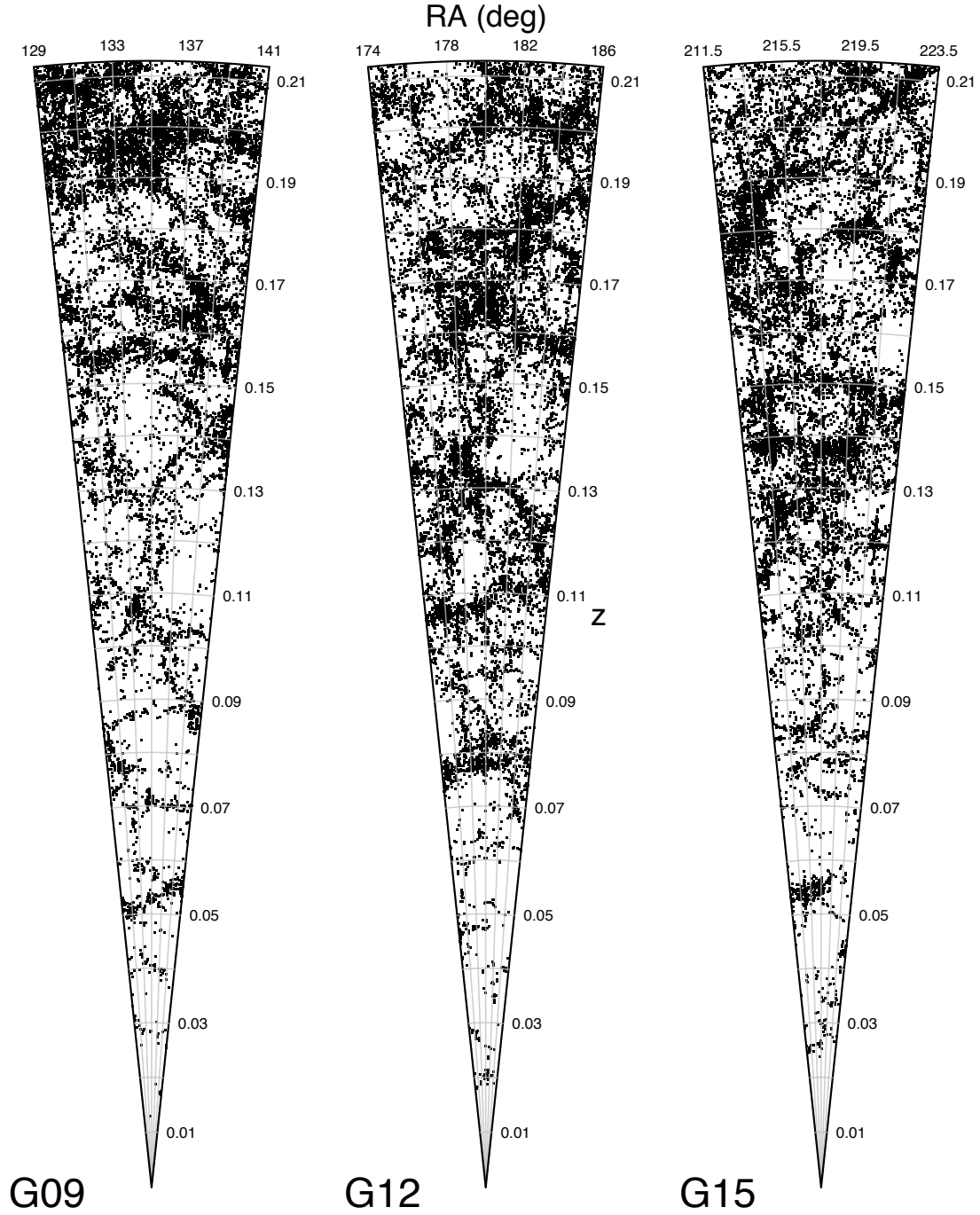


Figure 4.4: Three side by side cones showing the remaining galaxy sample after the selection described in Figure 4.2 for the G09, G12 and G15 regions respectively out to $z = 0.213$. All three cones span the full 5° declination range, which results in increasing projection effects at higher redshifts.

4.2 Filaments and large scale structure

4.2.1 Minimal spanning trees and Scooper

After selecting an appropriate sample, it is possible to move on to running the data through the structure finding algorithm. This classification method is designed to be easily repeatable and be as objective as possible with regards to classifying large scale structure. The algorithm functions on the basis of two assumptions: (1) that all bright, high luminosity groups tend to ‘live’ in knots and filaments and that (2) void galaxies are only clustered at extremely small scales. Both of these assumptions are robustly backed up by observational evidence.

The structure finding algorithm is based primarily on minimal spanning trees (MST, Iyanaga, 1980), which have been used previously by others (Barrow et al., 1985; Graham et al., 1995; Doroshkevich et al., 2004; Colberg, 2007) on a combination of observed and simulated data to examine the large scale structure of galaxies and haloes. Here, the minimal spanning tree approach is applied to groups of galaxies, instead of individual galaxies, as has been done in the aforementioned works. This approach of using groups instead of galaxies as tracers of filaments is very similar to the method used in Murphy et al. (2011), which is discussed in the introduction to this thesis. MSTs are a product of graph theory most famously associated with Paul Erdős, and are commonly used in a number of scientific fields, including computer science, sociology, scientometrics and epidemiology. They are particularly useful for picking out ‘skeletal’ patterns and linear associations within point data sets as well as for distinguishing clustering and structure in a systematic and quantitative way. This makes them ideal tools to objectively detect large scale structure in the Universe.

Within the context of graph theory, a *graph* is a collection of *nodes* (in this case, groups) and *edges* (straight lines connecting nodes). A *path* is defined as a sequence of edges that joins nodes, and a graph where a path is possible between any pairs of nodes is a *connected* graph. A *spanning tree* is defined as a graph where a single path connects all nodes and has no loops. If this path is the shortest possible path that connects all nodes, then it is a *minimal spanning tree* (MST). MST-based algorithms are analogous to FoF-based ones, as an MST is simply one specific solution of a FoF algorithm. In the case of this work, each node is a galaxy or group, and edges are the distances between them.

For a selected sample of groups and galaxies, the structure finding algorithm is composed

of 5 main steps:

1. Generate an MST on group centres, and remove excessively lengthy edges (see Section 4.2.1). The structures that are left over are defined as *filaments*; in other words all groups that are in the same set of unbroken edges, or path are considered to be part of the same structure.
2. Examine the morphology of each filament by subdividing it into a series of branches, including the *backbone*, which is the longest link that travels from one end of the filament to the other through its most central node. This is discussed in Section 4.2.1.
3. Travel along each filament, scooping up galaxies that lie within a certain orthogonal distance r from each filament. These are referred to as *galaxies near filaments*. Galaxies not in and around filaments are referred to as *unassociated galaxies*.
4. Having removed galaxies near filaments from the sample, I generate and trim another MST on these unassociated galaxies. These structures are defined as *tendrils*, containing *tendrill galaxies*; as with filaments, all galaxies that belong to the same unbroken chain are considered to be part of the same tendril.
5. Any galaxies not in tendrils or near filaments are finally classified as being *void galaxies*. This means that the definition of a void galaxy used in this work depends on linear proximity to another galaxy, as opposed to the commonly used method of finding underdensities and then naming any galaxy within them as a void galaxy.

A visual representation of this algorithm is shown in Figure 4.5, where the data and output of R11 is shown in the region enclosed by the dashed black lines. The groups are then put through an MST and placed into filaments; which are then combined with galaxies using the Scooper algorithm to identify galaxies near filaments (shown in blue) and isolated galaxies. A second MST is generated using unassociated galaxies, and these are classified into tendrils (shown in green) and voids (shown in red). The algorithm then outputs a series of interlinked catalogues that give summary statistics for each filament and tendril, and the associated groups and galaxies within filaments, tendrils and voids through a series of unique identifiers. We now describe the steps given above in greater detail.

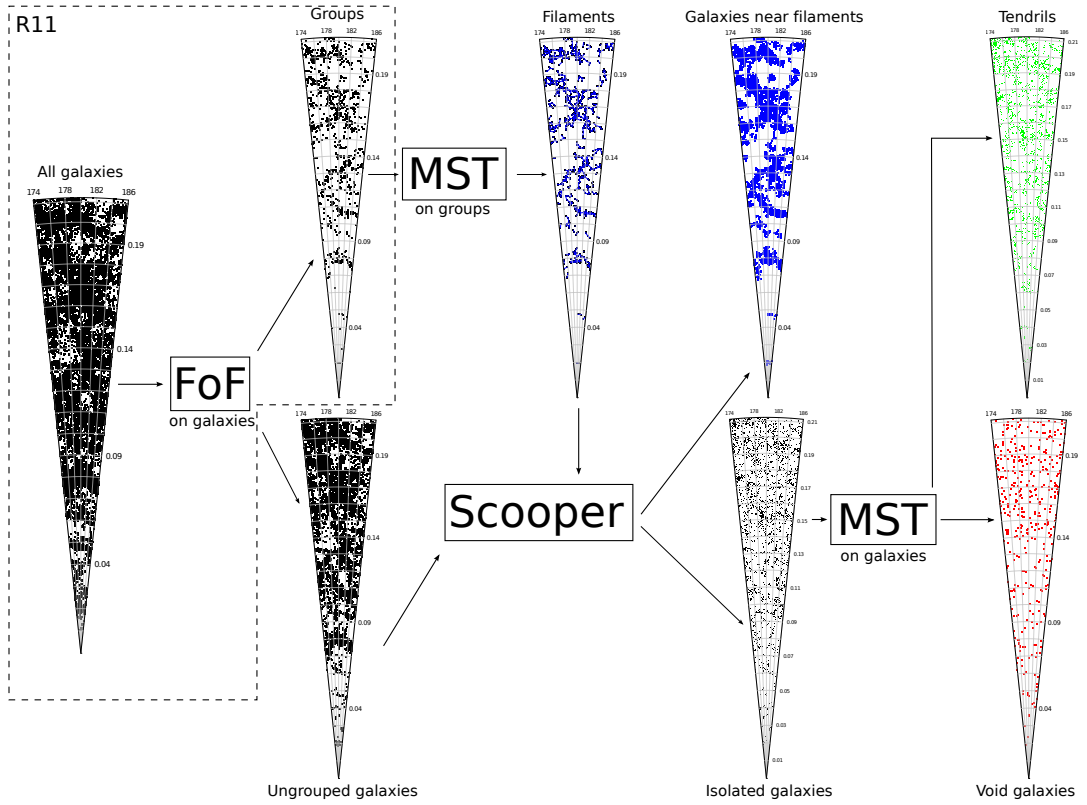


Figure 4.5: Flowchart schematically describing, for one example region, all the steps taken to go from a distribution of galaxies to a network of filaments, tendrils and voids. All groups are shown on the top panel, and all ungrouped galaxies on the lower panel. The groups are then put into a minimal spanning tree and the longest edges are trimmed. Ungrouped galaxies are then scooped up around each filament, giving the large network of galaxies near filaments (shown in blue). All ungrouped galaxies are then classified as being tendril galaxies (in green) or void galaxies (shown in red).

Minimal spanning tree on groups, and filaments

The construction of the MST on the groups (and, subsequently, the galaxies) is done using the `nnclust` package within the R programming language (R Development Core Team, 2011). The function `mst` within `nnclust` constructs a minimal spanning tree for a set of points on a 2D or 3D Cartesian plane using Prim’s algorithm (Prim, 1957). Prim’s algorithm functions on the basis of knowing the distance between all nodes in a graph. Starting from a random node, the algorithm travels along an edge to the nearest node. It then travels to the node nearest to either of the nodes it has already visited, and continues this process iteratively until all nodes have been visited. The path it has taken to do this is the minimal spanning tree. This is the most computationally efficient way of generating MSTs.

Comoving Euclidean coordinates of group centres are used to define the locations of the groups (nodes) and are fed into `mst`, whose output is a set of links between nodes, and their distances. The links are given by unique ID names between the start and end of an edge. I remove any edges whose length is beyond a certain threshold value b (the value of this parameter and how it is determined is explained below). This allows the identification of distinct sub-structures and removes unrealistically long links between objects in low density regions. This is important, because `Scooper` will go on to collect galaxies along edges and associate them with filaments. For an untrimmed MST, some edges will be between distinct filaments and travel through underdense regions, meaning that `Scooper` will associate possible void galaxies as being in a filament. Objects that remain in unbroken chains are then grouped together as an individual filament.

The choice of the maximum edge length b , is an important one. Examples of different maximum linking lengths are shown in Figure 4.6 where b is given in units of h^{-1} Mpc. Within each cone, each point is a group and all points of the same colour belong to the same filament. As b tends to higher and higher values, all groups will be clustered into one massive super-filament, which is unrealistic given that we expect structural collapse in large scale structure to stabilise at scales less than $15 h^{-1}$ Mpc (Chris Power, private communication). Conversely, if the linking length is too small, dominant superstructures are broken up into several short sub-structures. Additionally, as b drops, the total number of groups included in filaments also drops (because a group with no edges leading from it is no longer part of the MST). Therefore at $b = 1 h^{-1}$ Mpc we are effectively sampling the distribution of group-group pairs that lie

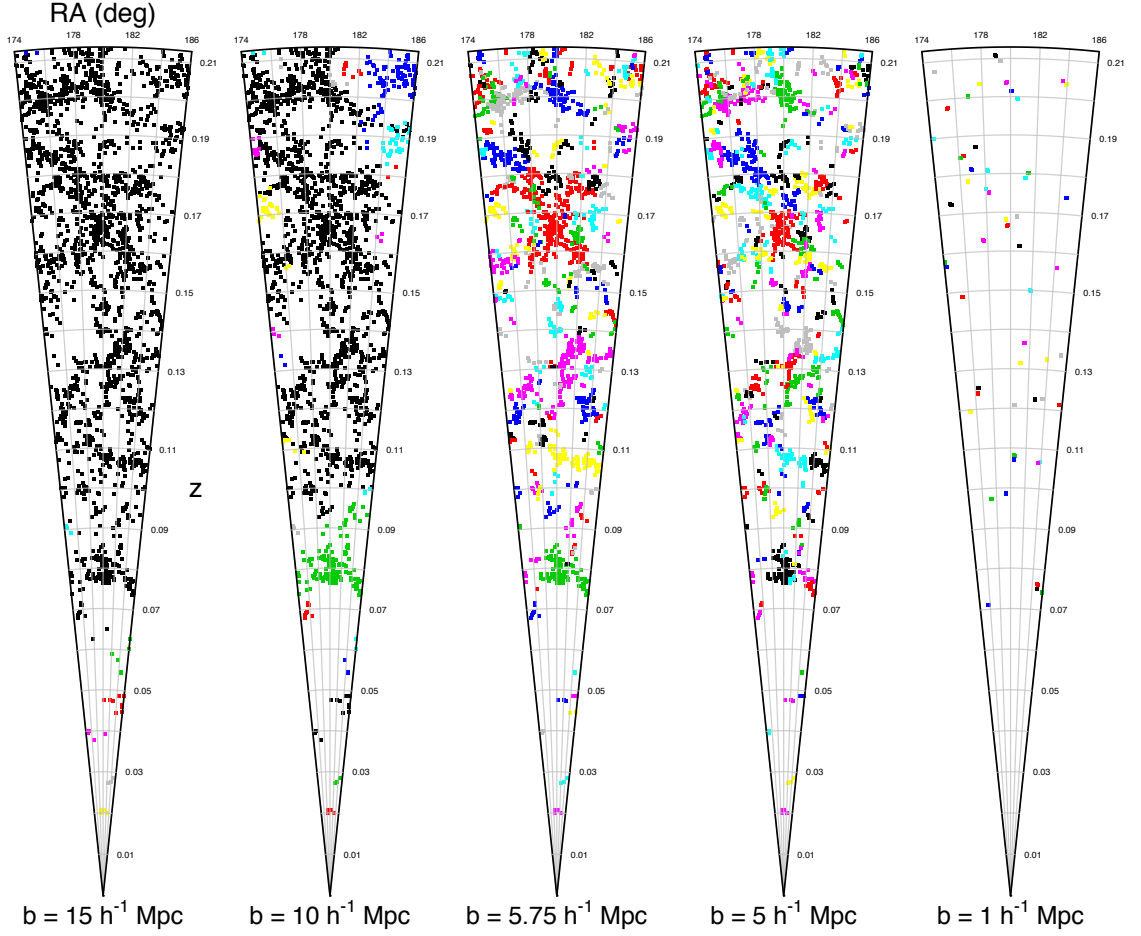


Figure 4.6: Filaments constructed from the same minimal spanning tree, but with different maximum edge lengths, decreasing from left to right. Groups in the same filament are coloured in matching colours. As the linking length increases, all galaxies tend towards being in one single huge filament, while as it decreases, leaving only groups that are in close proximity to each other. The number of groups in filaments also drops as b decreases, as groups with no links to other groups are not considered to be filaments (a filament needs at least 2 groups).

within $1 h^{-1}$ Mpc of each other.

To make the selection of b as objective and unbiased as possible, the largest, brightest groups should belong to filaments; this is assumption (1). A bright group is defined as one with $L_{\text{Group}} \geq 10^{11} L_{\odot} h^{-2}$, where L_{Group} is given by the total group luminosity given in the G³C (the TOTFLUXPROXY parameter). This value is roughly equivalent to the 98.65% quantile in the total range of group luminosities in the sample used. The algorithm requires at least 90% of these brightest groups to be in filaments. Figure 4.7 shows the fraction of groups in filaments as a function of $\log(L_{\text{group}}/L_{\odot} h^{-2})$ for a different set of values for b . This process gives $b = 5.75 h^{-1}$ Mpc as the minimum length at which the condition given above is fulfilled; therefore edges longer than this value are trimmed. The MST and filaments shown in Figure

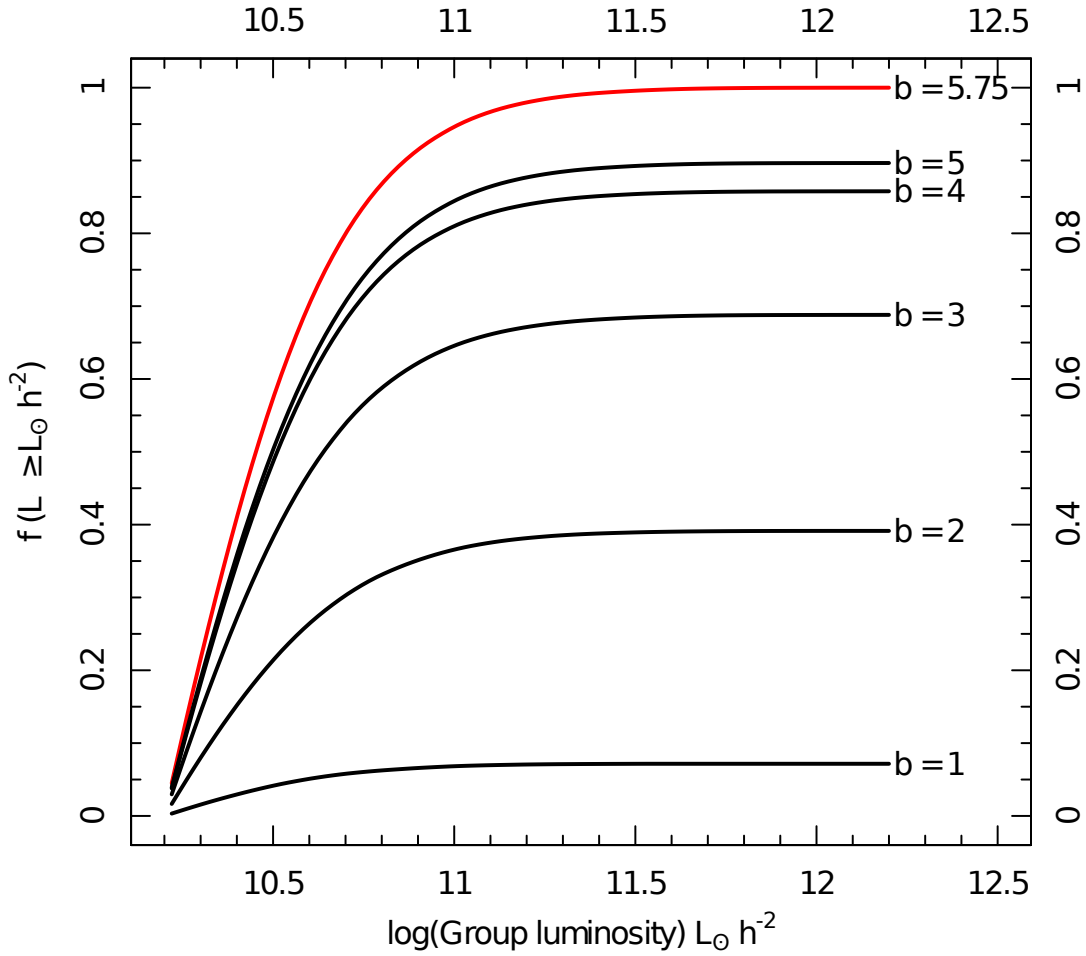


Figure 4.7: Cumulative fraction of groups in filaments as a function of their total r-band group luminosity, shown for different maximum edge lengths in the minimal spanning tree (given by b). As b decreases only filaments between pairs of groups that are in extreme close proximity are constructed, and the fraction of high mass groups in filaments drops to 0. Raising b to a much higher value results in all groups being in a single giant filament. I therefore select the minimum value for b at which 90% or more of galaxies with $L_{\text{Group}} \geq 10^{11} L_{\odot} h^{-2}$ are in filaments; or in other words, $f(L_{\text{Group}} \geq 10^{11} L_{\odot} h^{-2}) \geq 90\%$.

4.5 are constructed with $b = 5.75 h^{-1}$ Mpc. As expected, as b increases, more groups are linked to the same filament, finally leading to a single massive superstructure.

Filament morphology

Going from a series of links that groups together some points into a common structure, to a rigorous description of the shape and morphology of that structure is non-trivial. One must define where the edges of the structure are (it may be possible, for example, for a node to exist

geometrically near other edges and nodes, but be a dead end itself) as well as the most central part of the filament. To this end, we have developed an algorithm to analyse the structure of a filament called `walk`. The purpose of this algorithm is to step (or ‘walk’) through the filament and record, for each node, the number of steps required to exit the filament from the nearest end; this is referred to as the *count* for that node. A second property that is recorded is the so-called ‘branch order’ of each node; this value represents the number of branches, or forks in the path, between the node and the nearest end. Nodes on branches with one end are said to have a branch order of 1, and this value increases with each intersection or fork in the path. The output of `walk` after going through this process is a simple table that contains, for each node, the count value which represents a distance, in terms of nodes, between that node and the nearest end of the filament, and its branch order. This approach of splitting filaments into individual branches has previously been done by Colberg (2007).

The output of `walk` is fed into a secondary function called `makebranch` along with the original list of links for the filament. In knowing the count and branch order of each node as well as all the links in the filament, this function can travel along the branches of the filament from any given starting point, and search ‘upwards’ or ‘downwards.’ This is a setting specified by the user, and the direction refers to searching ‘up’ to find the centre of the filament, or ‘down’ to get to the nearest end; in other words, the algorithm walks along a path in ascending count value, or descending count value. Potentially, therefore, a user could choose to start at the ends of the filament and find the fastest way to the centre, or vice versa.

An important setting of `makebranch` allows the user to instruct the function to avoid reusing nodes that have already been visited by the function. This option is important in determining the primary branches of a filament, which is dubbed its backbone. To do this, I run `makebranch` without avoiding nodes first to determine all possible paths that lead from the ends of the filament to the filament centre. By then rearranging these branches in descending order and rerunning `makebranch`, this time avoiding visited nodes, and starting from the two biggest branches, it is possible to determine the longest primary filament that starts at an end, travels to the centre, and moves to another end. In this case ‘longest’ can be determined either by number of nodes, or physical distance. A detailed example of this process, as well as a step-by-step analysis of how the algorithm ‘walks’ through the filament is now shown.

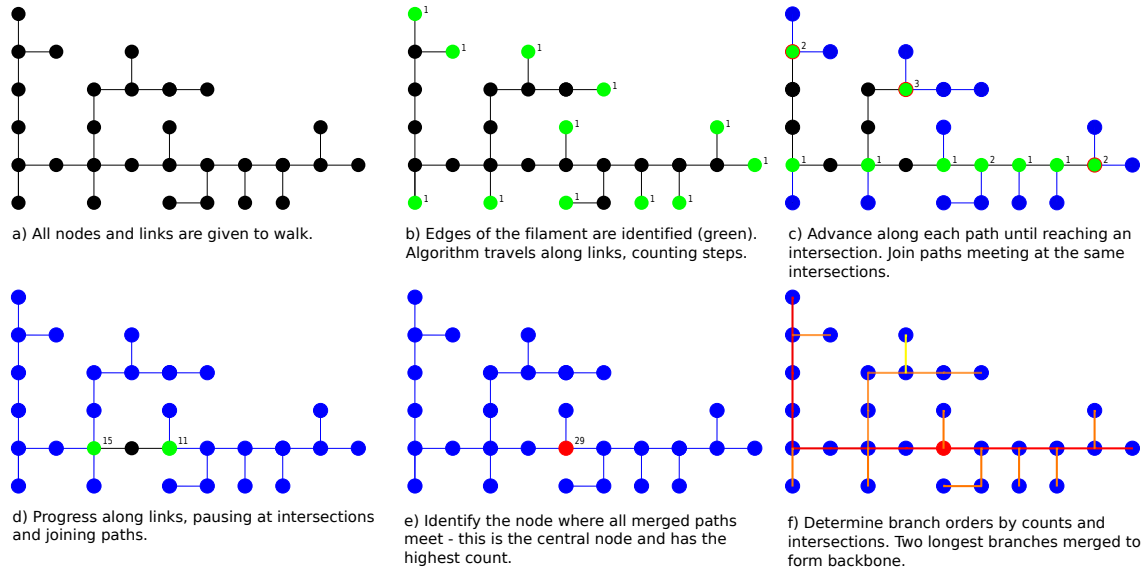


Figure 4.8: An overview of the process by which the topology of a filament is determined. The top left panel shows all nodes and links for an example filament (circles and lines respectively). In each panel, green objects represent where the algorithm is, while blue ones represent visited objects and black ones, unvisited objects. From here, *walk* identifies all the ends of the filament (shown in panel b) and travels along them, stopping at intersections and merging all paths that reach the same intersection (panels c and d). The algorithm associates a count value to each node (shown for the green nodes on the top right in each panel), which is the number of steps required to reach the end of the filament. The count of the centre of the filament at an intersection, the counts along each branch are summed up and assigned as the count value for that node. Therefore, the node at which all branches meet will have the highest count, and be determined to be the centre of the filament, as shown by the red node in panel e. The output of *walk* is fed into *makebranch*, which analyses this output and uses it to construct branches for the filament, and assign orders to them. These are shown in panel f, with first, second and third order filaments shown in red, orange and yellow respectively. The backbone is then defined as the single path that travels along the two first order branches.

Filament walker and finding backbones

The purpose of these functions is to be able to systematically determine the internal structure of a filament; that is to say, where its edges are, where the most dense nodes of the filament are, and how to travel from one region of the filament to another in the most efficient path possible.

To illustrate the function of all of our algorithms, they are applied to an idealised 2D filament, which is shown on the top left panel of Figure 4.8. In this schematic, each circular point marks a galaxy group (or node) and lines represent links (or edges) between nodes.

4.2.2 walk

The most important algorithm in analysing the structure of a filament is *walk*. This algorithm's purpose is to start at the ends of the filament (defined as being nodes that have only one edge)

and travel along all the links in the filament until they have all been visited. As described above, for each node, walk assigns a *count*, which is effectively the number of steps required to get from that node to the nearest filament end. The algorithm also keeps track of the *order* of each node. Nodes at the filament ends are said to be of order 1, and this increases each time the algorithm goes past an intersection; or a node with three or more edges that lead away from it. In the example filament shown in Figure 4.8, the nodes shown in green in panel b have orders of 1, while nodes shown in green in panel c have orders of 2. One important note to make is that the count and branch order given for a particular node are not with respect to the nearest end, but are a sum of all branches leading out from it.

Briefly, the walk algorithm works as follows:

- To begin with, walk identifies all nodes in the tree that have only one edge. This implies that they lie at the ends of branches. In the example filament, these are the green nodes in panel b.
- The algorithm then goes through each end node and progresses along that path until it reaches an intersection. In the case of this example filament, the first intersections are the green nodes in panel c, with the blue nodes and links representing the path the algorithm has taken to reach those nodes.
- If an intersection has been reached by more than one branch, these are then merged, before the algorithm continues to progress along the edges leading from them. Intersections that have only been reached by one edge do not progress. In panel c, walk will only continue walking from the green nodes circled in red, as these have two links leading out of them.
- This pattern of stopping at intersections and merging branches continues until all nodes have been visited. In panel d, the process is shown for the step at the intersection where the final two pathways are about to meet - while these nodes had been reached by walk as early as panel c, they have only now been arrived at from two links, meaning the algorithm can walk along them now. In panel e, all paths have merged at the central node, shown in red.

The output of walk after going through this process is a simple table that contains, for each node, the count value which represents a distance, in terms of nodes, between that node

and the end of the filament, and its branch order.

4.2.3 makebranch

The output of `walk` describes, for each node, how far away it is from the ends of the filament, and how many intersections it takes to get from that node to the end. This information is fed into an algorithm called `makebranch` that is capable of starting at any node specified by the user and travel ‘upwards’ or ‘downwards’ along the links in the filament; that is to say, from a given node it can either travel to the neighbouring node with a higher or lower count respectively. Travelling downwards will lead it to the nearest filament end, while traveling upwards will lead it to the filament centre.

To determine the location of the filament’s backbone, I first run `makebranch` by starting it on nodes with only one edge and tell it to travel upwards; this gives all possible paths to the filament centre. After rearranging these branches in descending order of counts to get to the ends, `makebranch` travels along them again, this time avoiding revisiting nodes it has already been to. By starting at the biggest branch and travelling to the centre, then doing the same for the second biggest branch and so on, this algorithm is able to determine the longest continuous path that goes from one end of the filament to the other while travelling through the central node. This path is referred to as the first order branch or the filament backbone; any paths that branch off from the backbone are second order branches, and so on. In other words, a path that intersects with a branch of order n is assigned an order of $n + 1$. This implies that there will only be two branches of order 1, which are subsequently joined together to form the backbone, but any number of higher order branches. The higher the number of branches in a filament, the more complicated its morphology.

All the branches in the sample filament are shown in panel f of Figure 4.8. First order branches are shown in red; therefore the single unbroken red path is the backbone of the filament. Orange paths are second order branches, and the single yellow path is the third order branch, as it is the only path that intersects with a second order branch.

With the backbone and branches for a given filament, it is possible to objectively look at its morphology. The backbone will always represent the most central route through the filament, and branches will always refer to links emerging from the backbone. The backbone therefore serves as a good measure of the overall extent of the filament, while examining the lengths and

sizes of branches, as well as their relative abundances, and provides a measure for the ‘spread’ of the filament. For example, a filament with one large backbone and few to no branches is topologically the same as, or similar to, a straight line; while one with a short backbone and many branches can be seen as less linear.

Galaxies near filaments

There is now a set of filaments, each of which contains a number of groups. To associate galaxies with filaments, the structure finding algorithm now looks through each filament and travels along all of its branches, identifying all galaxies within an orthogonal distance r from the filament, with a function called Scooper. For each link along the filament, Scooper identifies all galaxies within a locus at a distance r from the straight line, or vector, that describes that link. In three dimensions, this locus is a cylinder with hemispheres at each end. If the distance of the galaxy $d \leq r$ then the galaxy is considered to be associated with that filament. Should a galaxy be within r of multiple branches, it is associated with the branch it is closest to.

Note that in order to diminish the effects of redshift space distortions, in this step of the process any galaxy (within the sample) that is known to be associated with a group within a filament is automatically assigned to that filament and branch. Instead of the distance to the filament, the distance to the iterative group centre is considered.

Tendrils and voids

To identify any underlying structure outside or between filaments, all elements belonging to filaments from the data set are removed. These so called ‘isolated galaxies’ or unassociated galaxies are shown in Figure 4.5 and are themselves used as points for another MST with a maximum edge length q (the choice for q is detailed below). Once again, galaxies that are part of a single uninterrupted chain of links are classified as being in the same tendril, which are structures akin to filaments but on much smaller scales, formed entirely out of galaxies. Characteristically they branch out from filaments and penetrate into voids, or connect two distinct filaments. Tendrils are discussed in greater detail below. Void galaxies are the last remaining galaxies – those that were rejected from the MST used to identify the tendrils after the edge trimming. Tendrils and voids are shown as the green and red distributions of galaxies in Figure 4.5.

Constraining the maximum edge length q requires the second assumption: that void galaxies are only clustered on extremely small lengths. In other words, the spatial two point correlation function of void galaxies should show less signal than the two point correlation function of galaxies in and around filaments.

The two point correlation function is computed using the estimator from Landy & Szalay (1993), namely:

$$\xi(r) = \frac{N_r(N_r - 1)}{N_d(N_d - 1)} \frac{DD(r)}{RR(r)} - \frac{2(N_r - 1)}{N_d} \frac{DR(r)}{RR(r)} + 1 \quad (4.1)$$

where $\xi(r)$ corresponds to the spherically averaged two point correlation function. $DD(r)$, $DR(r)$, and $RR(r)$ refer to the number of pairs separated by a distance $r \pm dr$ for data-data pairs, data-random pairs and random-random pairs, and N_r and N_d refer to the number of random points and data points respectively. The random distribution is in the same geometric space as each GAMA region, and is filled with 100,000 randomly generated points in a spherical pointing; this is to match my volume limited sample. I calculate $\xi(r)$ for each GAMA field separately and normalise results for each bin in distance. The resulting two point correlation functions for galaxies in structures and voids are shown in Figure 4.9. An attempt to estimate uncertainty in the two point correlation functions is made by jackknifing each GAMA region into several sub-regions and recalculating the correlation functions for the whole GAMA field, excluding one region at a time. The shaded areas in Figure 4.9 around each line show the maximum range of different $\xi(r)$ for each jackknifed estimate. Because these different regions are so correlated, strictly interpreting these regions as uncertainties is not advisable. Note that when doing jackknife error analysis, it is important to ensure that the size of the area removed is sufficiently large so as to minimise this correlation; in practice, a selection of a jackknife angle θ such that the geometric area it removes is around $100 \text{ h}^{-1} \text{ Mpc}$ is advisable.

All galaxies near filaments and in tendrils are classified as belonging to large scale structure (filaments or tendrils), and voids are defined as being the remaining population of galaxies. Running the entire algorithm, from generating the filaments to detecting tendrils and voids, takes just over one minute (the most time consuming step is the scooping up of galaxies near filaments). Because of this, it does not take long to generate a multitude of tendril and void galaxy distributions using different values for r and q . The final values for r and q are chosen such that they minimise the integral $\int R^2 \xi(R) dR$, with $\xi(R)$ being the correlation function of

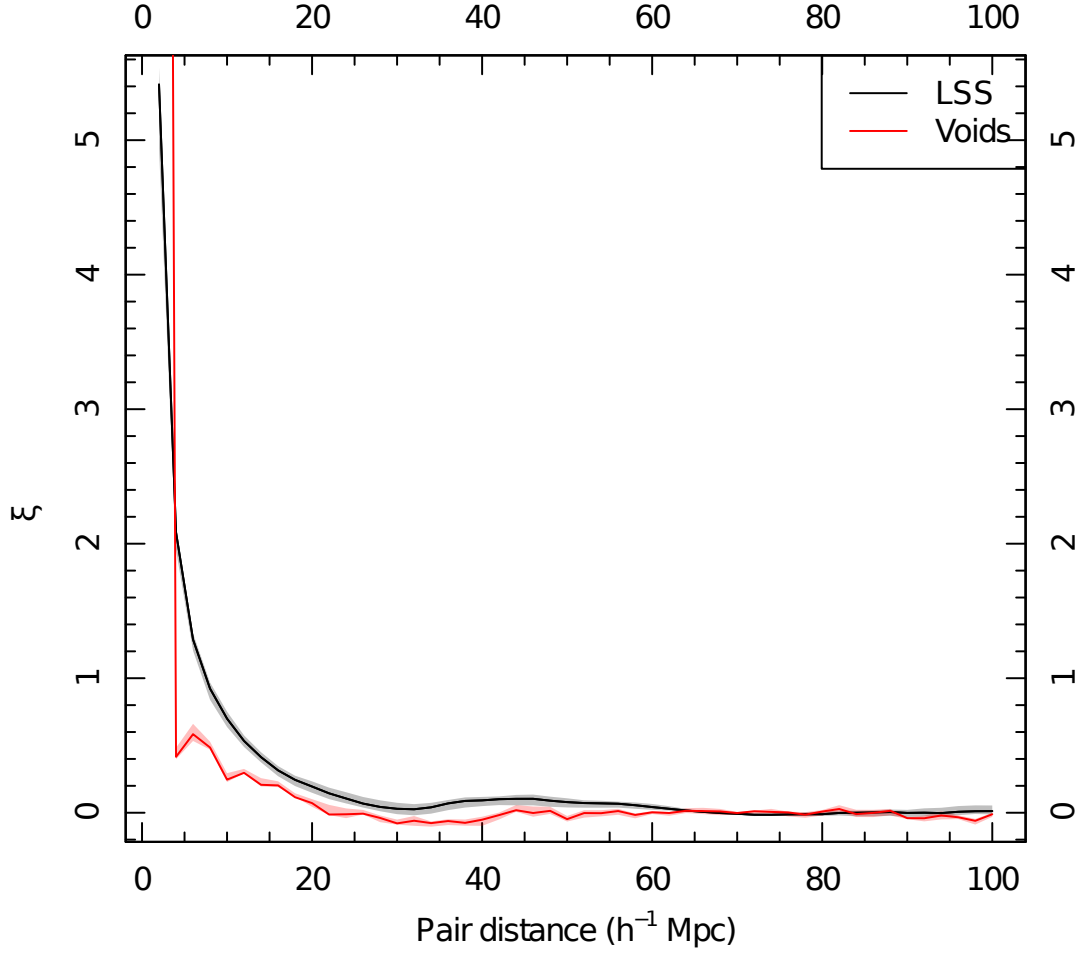


Figure 4.9: Two point correlation functions as a function of comoving distance for two different galaxy populations. The black line represents galaxies in groups that are in filaments, and within $4.56 h^{-1}$ Mpc of filaments; and the red line shows the function for galaxies in voids. By setting the maximum edge length between filaments to be $5.75 h^{-1}$ Mpc and $4.13 h^{-1}$ Mpc between galaxies in tendrils, we ensure that the resulting distribution of void galaxies has no correlation signal. The grey and pink shaded areas show, for each bin, the range of values for two point correlation functions calculated by sub-sampling the GAMA regions, with a jackknife method, and serve as uncertainty estimators.

void galaxies. This is the expression for the volume average correlation function, i.e. $\xi(< R)$. Optimising to minimise this integral gives $r = 4.13 h^{-1}$ Mpc and $q = 4.56 h^{-1}$ Mpc. The final parameters used are $b = 5.75 h^{-1}$ Mpc, the trimming length for the MST that identifies filaments, $r = 4.13 h^{-1}$ Mpc, the maximum distance allowed between a galaxy and a filament, and $q = 4.56 h^{-1}$ Mpc, the trimming length for the tendrils MST.

Note that the parameter selections are, to some extent, arbitrary, just as there are no formal definitions for filaments currently in the literature. The value for b is chosen such that we maximise the number of groups included in our filaments while making sure that most of the brightest galaxies are in filaments, and r and q are chosen such that the void galaxy correlation function is minimised over large distances ($\geq 20 h^{-1}$ Mpc). While it is possible to change and refine the parameter selection process, the overall hierarchy of the large scale structure is very stable with respect to changes in b , r and q . Varying any of these parameters by $\pm 1 h^{-1}$ Mpc results in a shift of approximately 5% of galaxies from filaments into tendrils, and tendrils into voids and vice-versa. There is a negligibly small effect on the comparisons to mock filaments discussed later in the chapter. The classifications are far more susceptible to survey depth; as a preliminary version of the GLSSC made with GAMA I data ($m_r < 19.4$ mag) had various tendrils that were reclassified as filaments with the GAMA II ($m_r < 19.8$ mag) version of the GLSSC.

There is one final caveat to consider with this algorithm, which is that while it tries to classify large scale structure as objectively as possible, it is still not making any assertions about the actual structuring of galaxies and groups. That is to say, it is valid to claim that group A belongs in a filament and group B doesn't, or that galaxy C is in a void and galaxy D is in a tendrils; but I cannot claim that the filaments and tendrils that I detect are necessarily the definitive structures found in space.

During all of these steps, galaxies and groups that are flagged for being within $q h^{-1}$ Mpc from the survey volume edges are kept track of. Any filament, branch, group or tendrils that has a constituent (group or galaxy) that is within this limiting edge is flagged. Therefore this data is still usable, however a user can specify to discard all structures that contain elements that are cut. GAMA contains a sufficiently large amount of data within its volume that flagging and ignoring filaments near its edges in any study would still yield a very large amount of useable data.

4.2.4 Filament catalogue

The algorithm described above is run on all three equatorial GAMA fields, as well as the GAMA mock catalogues. An ‘overhead’ view of all three equatorial GAMA regions, side by side, out to $z = 0.213$ is shown in Figure 4.10. Here, cyan points show groups in filaments, blue points correspond to galaxies near filaments, green points to galaxies in tendrils, and red points to void galaxies. It is strikingly easy to visually discern the skeletal pattern traced out by the filaments and their associated groups in blue, although at higher redshifts there will be projection effects due to the cones spanning the entire declination range. Tendrils of galaxies appear to be wispy, coherent structures that emerge from dense filamentary regions and either bridge across to other filaments, or terminate within voids. They span a range of different morphologies, as filaments do, with some being very linear (such as the tendril on the upper right region of the top right panel in Figure 4.10) while others are more clustered (middle right of the top left panel in Figure 4.10). Void galaxies lie in more isolated regions, and may have undergone fewer dynamical and chemical interactions with other galaxies during their evolution compared to galaxies in filaments.

While it is illuminating to identify large structures such as filaments, it is important to remember that a group at the end of one massive complex-like filament will not necessarily directly affect a group at the other end; however, both structures belong to the same superstructure and environment, therefore it is sensible to expect that they have evolved, dynamically speaking, in similar circumstances.

The full catalogue contains just under 650 filaments, with each filament having on average, 8 groups in it. The average length of the backbone of the filament, or in other words, the distance from the most extreme end of the filament to the other, is $\sim 13 h^{-1}$ Mpc, while the total of all the links in the filaments is close to $21 h^{-1}$ Mpc. Most filaments have 3 branches, and contain, on average, 46 galaxies. Table 4.1 provides average values for the properties of filaments in all three regions, and for the whole volume limited sample. The distribution of filaments is similar between all three regions. On average, I find shorter filaments than those detected in Tempel et al. (2013); however, this discrepancy is entirely to differences in data and methodology (Tempel et al. use data from SDSS and a filament finder whereby cylinders are linked together to form networks). The median redshifts for all filaments in the catalogue are shown as a histogram in Figure 4.11, where I also highlight the region below $z = 0.1$ where

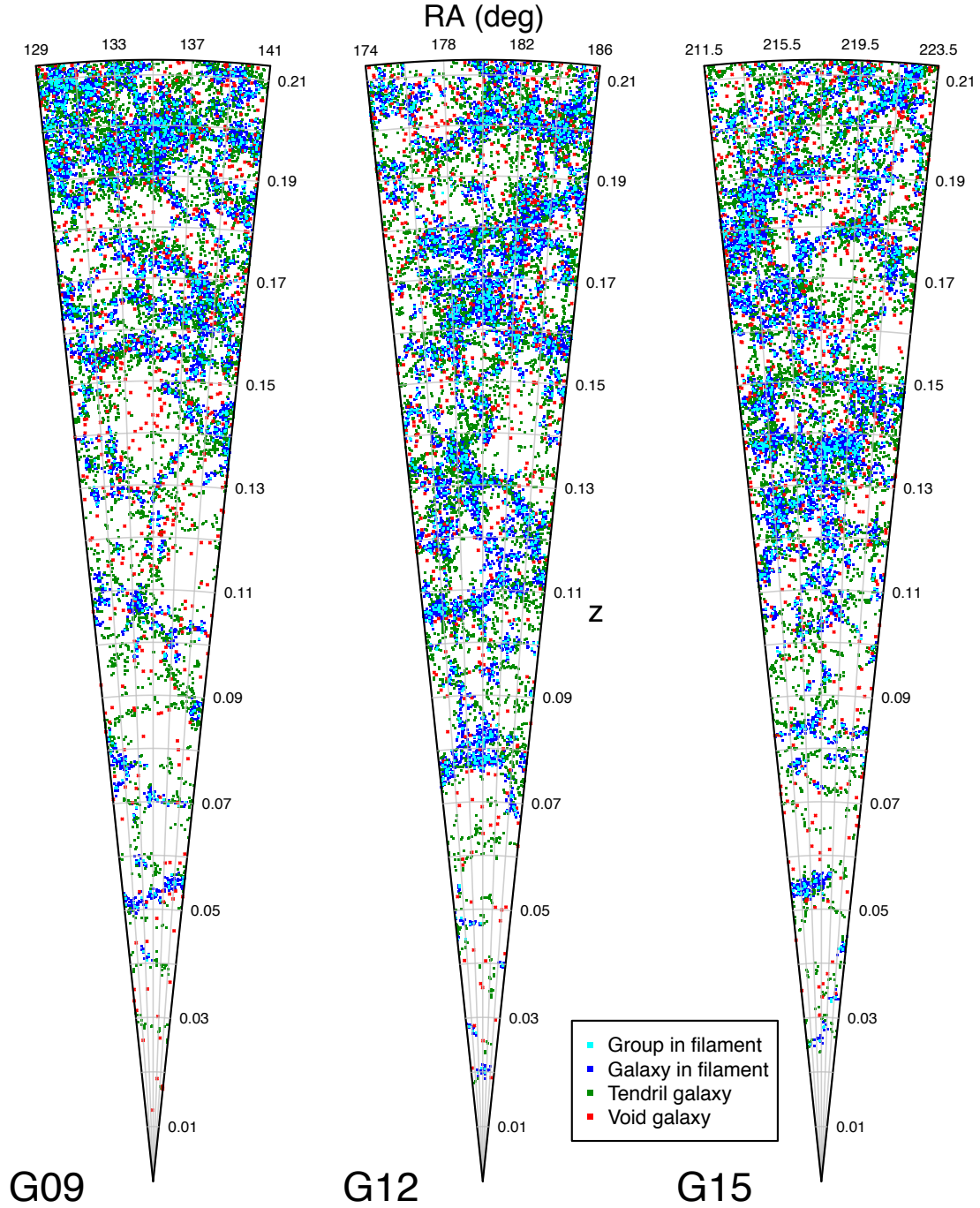


Figure 4.10: Galaxy distribution in the three GAMA regions, colour coded according to their environment, with groups in filaments, galaxies near filaments, galaxies in tendrils, and void galaxies shown in cyan, blue, green and red respectively. Groups, and galaxies near them form the bulky complexes of large scale structure, tendrils spreading from them in filamentary structures into voids, which seem to be populated by galaxies that appear to be almost uniformly distributed on large scales.

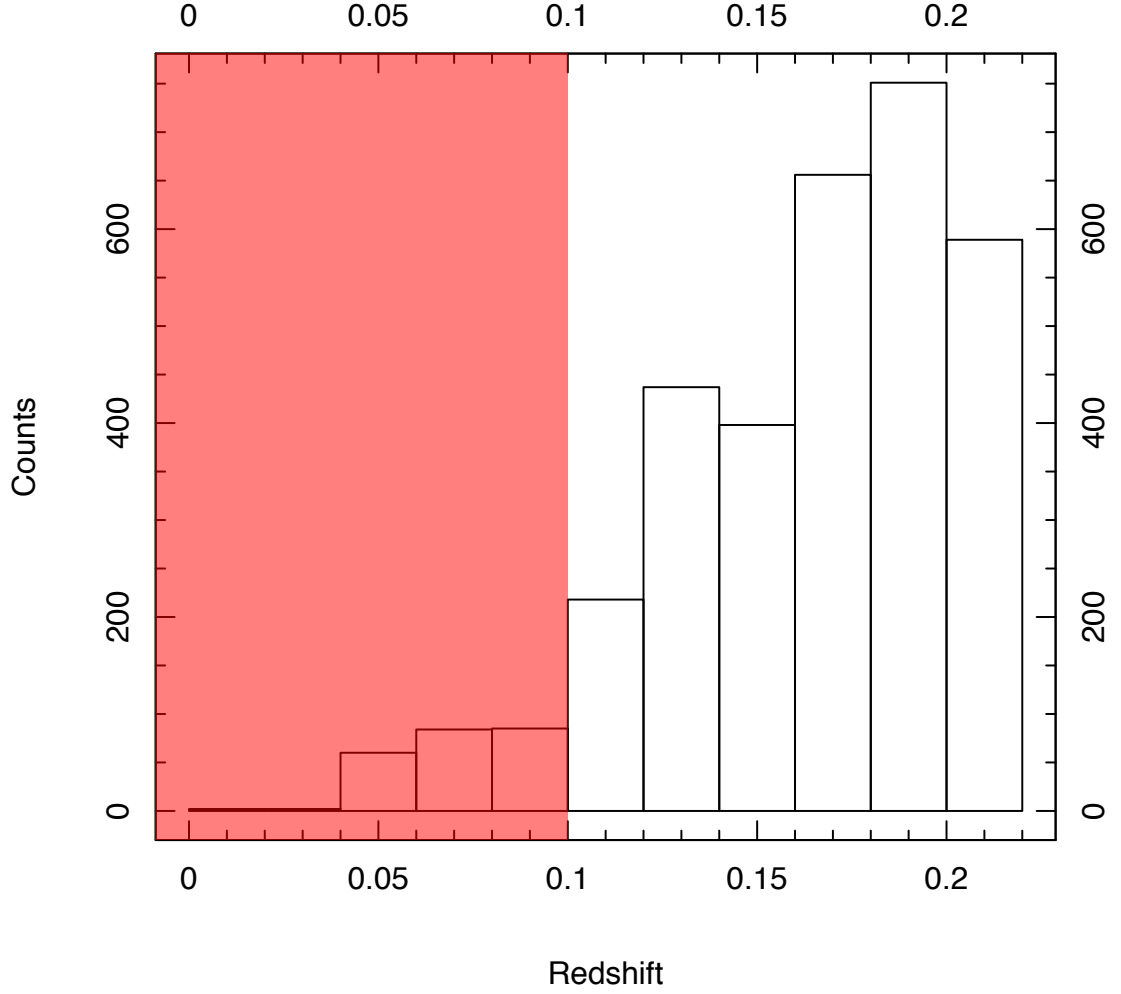


Figure 4.11: Histogram showing the median redshifts for all the filaments in the final filament catalogue. The area shaded depicts the region below $z = 0.1$ where the number of filaments drops sharply due to a smaller survey volume.

there are far fewer filaments due to a combination of the small survey volume at low redshift and the absolute magnitude cut used.

The full catalogue is composed of the following sub-catalogues:

- **Filaments**

This lists all filaments composed of groups of galaxies, giving them a unique identifier in FilID - the first digits of which correspond to the equatorial region the filament is in. There is also information pertaining to the number of branches the filament has, as well as the number of groups. The total length of all links in the filament is given, as well as

the length of the backbone.

- **FilBranches**

This catalogue lists all branches present within filaments in FilCat. Each branch is given a unique identifier and the filament it belongs to is identified as well. The order of the branch is given, as well as the number of groups it has, and its length.

- **FilGroups**

This catalogue contains the groups that are within filaments. They are identified by their GroupID as given in R11's catalogue, as well as their RA, Dec and median redshift. The groups' 3D comoving cartesian coordinates are also provided, as well as the branch they belong to, its order, and the filament they belong to.

- **FilGals**

In this catalogue, all galaxies that are within a certain orthogonal distance of filaments are listed. The GAMA CATAID (an internal unique galaxy identifier) for each galaxy is given, along with 3D comoving cartesian coordinates, as well as the orthogonal distance to the nearest branch of a filament, whose IDs are given.

- **FilLinks**

This simply contains a list of links between groups used to construct the filaments. The groups are identified by their GroupIDs. This catalogue can be used to reconstruct, visually, the links between groups in filaments, but can also be used to identify groups that are 'intersections' - that is to say, groups that have 3 or more links to other groups.

- **Tendrils**

Moving from groups to galaxies, this catalogue is analogous to FilCat in that it contains the top level structures formed by galaxies that are not included in filaments. Each tendril is given a unique ID, and their length and number of galaxies are specified.

- **TendrilGals**

This is the catalogue of all galaxies in tendrils. Their CATAID is given, as well as their 3D comoving cartesian coordinates, and the ID of the tendril they belong to.

- **TendrilLinks**

	N_{fil}	$\bar{L}_{\text{fil,BB}}$	\bar{L}_{fil}	\bar{n}_{group}	\bar{n}_{branch}	\bar{n}_{gal}
G09	213	11.6	18.4	7.5	2.8	42.3
G12	200	14.7	23.6	9.2	3.4	51.6
G15	230	12.5	20.0	8.0	3.0	44.2
All	643	12.90	20.7	8.2	3.0	46.0

Table 4.1: Summary statistics of some basic properties of filaments in GAMA. Besides the number of filaments, for each region the following averages are given, the backbone length, sum of the length of all links, number of groups, branches, and galaxies per filament are given. All lengths are given in units of h^{-1} Mpc. The final row contains these values across all three equatorial fields.

A second list of links, this time for the tendrils. Now, galaxies are identified by their CATAIDs.

- **VoidGals**

Finally, this catalogue lists all galaxies that are not associated with any filaments or tendrils.

Each catalogue links with each other using a series of unique identifiers for each type of structure: filament, branch, group, and galaxy. Separate catalogues describe tendrils and the galaxies in them in a similar way, and voids are isolated from them all. The links within structures are also given (i.e. the links of the minimal spanning tree after edges are cut), both for filaments and tendrils. All of this allows a user to fully reconstruct the large scale structure of the GAMA regions easily. For example, a user may wish to identify all galaxies associated with the longest filament in G09; this is easily done first by using **Filaments** to search for the longest filament whose identifier begins with 9, then going to **FilGals** and selecting all galaxies with a filament ID that matches the filament found.

A more detailed view of a region in the G12 ($174^\circ \leq \alpha \leq 186^\circ$) field is shown in Figure 4.12, with each panel representing a declination slice of 1° , for a redshift range of $0.15 \leq z \leq 0.2$. In this zoomed view it is possible to see the detailed interplay between filaments (blue) and tendrils (green); with the latter branching out from the former and penetrating into voids. It is also possible to see with detail the coherence of structure formed by individual galaxies in tendrils; a notable example is in the top right area of the top right panel of Figure 4.12, where a delicate string of galaxies is seen to be curving out of a filament. This is the same region as plotted in Figure 4.1; notice how many galaxies that are outside of that group in that figure are now shown to be tendril galaxies.

While there is no attempt made to identify actual voids in this catalogue, it is possible

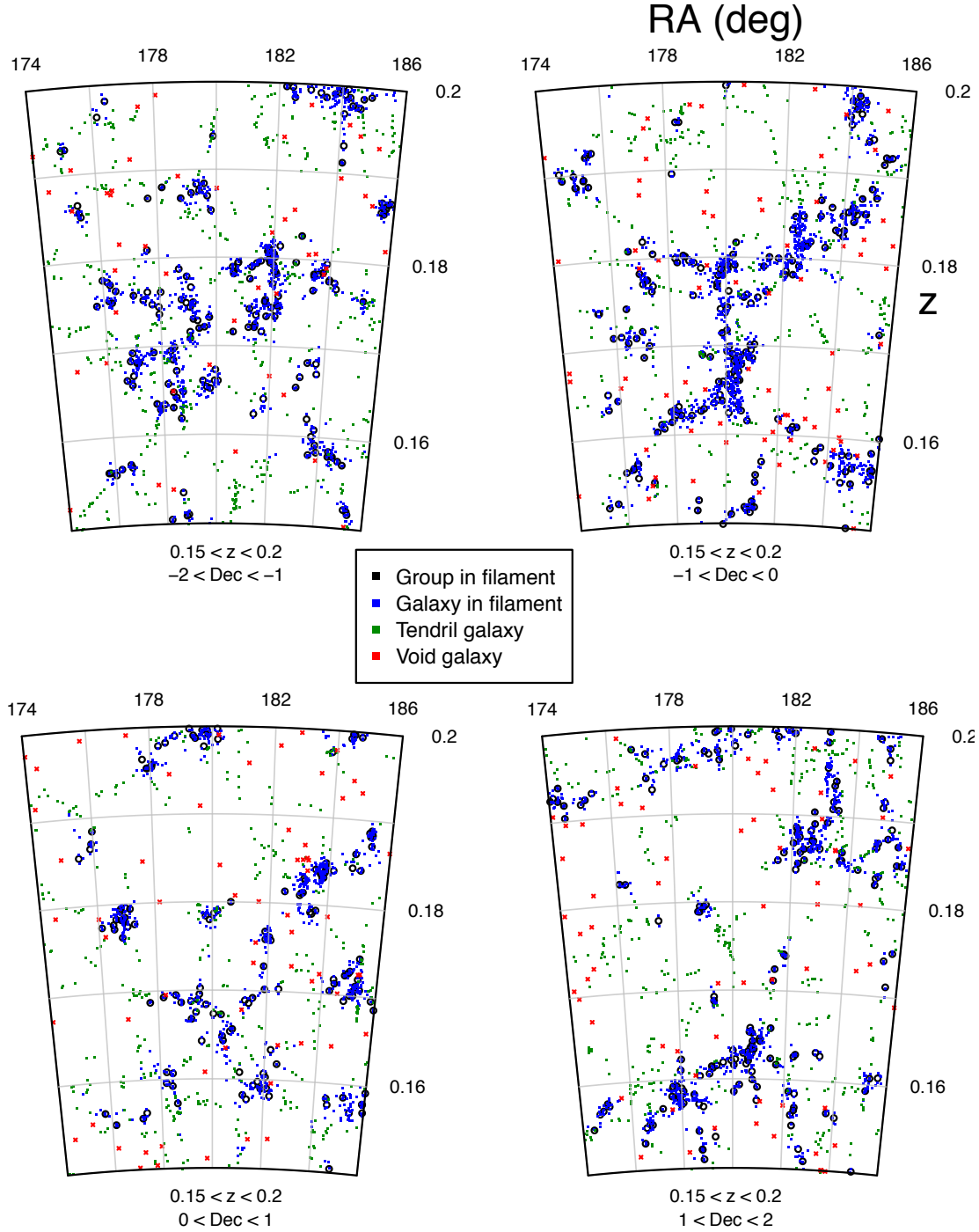


Figure 4.12: Group and galaxy distribution in G12 colour coded by environment in 4 declination slices. Black circles represent groups in filaments, blue points are galaxies near filaments, green points are galaxies in tendrils, and red crosses are galaxies in voids.

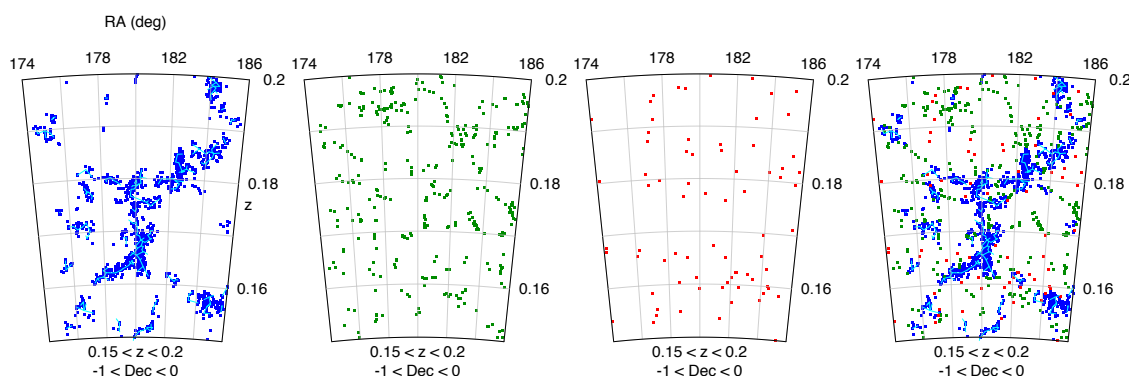


Figure 4.13: A section of the G12 field with different galaxy populations shown in each panel. From left to right the populations shown are galaxies in filaments with the filament minimal spanning tree (blue and cyan respectively); galaxies in tendrils (green); galaxies in voids (red); and all three populations in their respective colours.

to accurately recover galaxies within voids; these objects can be considered to be extremely isolated with regards to their environment, and can be considered as a separate population of galaxies. The distribution of void galaxies exhibits no inherent structure, although it must be noted that this is in part due to design, as we have selected parameters for our filament finder that produce such a result.

4.3 Tendrils

Figure 4.13, shows the different galaxy populations in a 1×12 degree slice of the G12 field, with $0.15 \leq z \leq 0.2$ – this is the same region as shown in Figure 4.12 in the previous chapter. The left panel displays all filament galaxies in blue, with the MST links shown on top in cyan. Tendril galaxies are shown in the next panel, voids in the third, and all four populations of galaxies are shown together in the rightmost panel. This final panel shows once again that tendrils generally branch off from filaments and penetrate into voids, often bridging separate filaments together. Alternatively, they enter voids and end there. Most importantly, tendrils arc across ranges of right ascension and declination as well as redshift, meaning that they are not statistical alignments by chance or caused by redshift space distortions, but actual coherent structures.

Tendrils are detected both in the observed GAMA data, as well as the GAMA mock catalogues. Figure 4.14 displays slices of the GAMA G09 and G15 fields from observed and mock data (top and bottom row respectively), showing only tendril galaxies. In both the observed and simulated cases, tendrils appear to be structures that are morphologically similar to filaments, but at smaller scales, spanning shorter distances and containing fewer galaxies. They

are also morphologically less complex. Tendrils in the observed and mock catalogues are not only visually similar, but are also statistically indistinguishable from one another.

4.3.1 The excess line correlation of LSS

This subsection has been primarily written by Danail Obreschkow.

It is possible to quantify the filamentarity, defined as the linearity of structure, of the filament, tendril and void galaxies using their line correlation $l(r)$. This correlation is an estimator of spatial statistics introduced by Obreschkow et al. (2013) to characterize the cosmic density field $\delta(\vec{r})$ or its Fourier transform (FT) $\hat{\delta}(\vec{k})$. Since the isotropic power-spectrum $p(k) \propto k^{-2} \int d\phi d\theta |\hat{\delta}(\vec{k})|^2$ exhaustively exploits the information contained in the amplitudes $|\hat{\delta}(\vec{k})|$ of a homogeneous and isotropic field, Obreschkow et al. chose to investigate the residual information in the phase factors defined as:

$$\hat{\epsilon}(\vec{k})|\hat{\delta}(\vec{k}) = e^{i \arg[\hat{\delta}(\vec{k})]} \quad (4.2)$$

The inverse FT $\epsilon(\vec{r})$ has a vanishing two-point correlation $\xi_2(r)$; thus the lowest non-trivial correlations of $\epsilon(\vec{r})$ are three-point correlations. The line-correlation $l(r)$ is defined as a suitably normalised version of the isotropic three-point correlation of $\epsilon(r)$ for three points on a straight line, separated by r . Geometrically, it turns out that $l(r)$ measures the degree of straight filamentarity on length scales r in about the same way that $\xi_2(r)$ measures the clustering on scales r . Figure 3 in Obreschkow et al. (2013) displays $l(r)$ for a series of density fields and illustrates how $l(r)$ can distinguish linear over-densities from spherical ones, unlike $\xi_2(r)$ and more robustly than the traditional three-point correlation. A demonstration of how the line correlation function measures structure in distinct density fields is shown in Figure 4.15. The isotropic line correlation function is defined as

$$l(r) = \frac{V^{1/2} r^{3D/2}}{(2\pi)^{2D}} \iint_{|\mathbf{k}||\mathbf{q}| \leq \pi/r} d^D \mathbf{k} d^D \mathbf{q} w_D(|\mathbf{k} - \mathbf{q}|r) \frac{B(\mathbf{k}, \mathbf{q})}{|B(\mathbf{k}, \mathbf{q})|} \quad (4.3)$$

where $D = 3$ and $w_D = \sin(x)/x$, $B(\mathbf{k}, \mathbf{q})$ is the bispectrum (i.e. the power spectrum of the three-point correlation function $\xi_3(r)$) with \mathbf{k} and \mathbf{q} representing the wavevectors in Fourier space for $\xi_2(r)$ and $\xi_3(r)$.

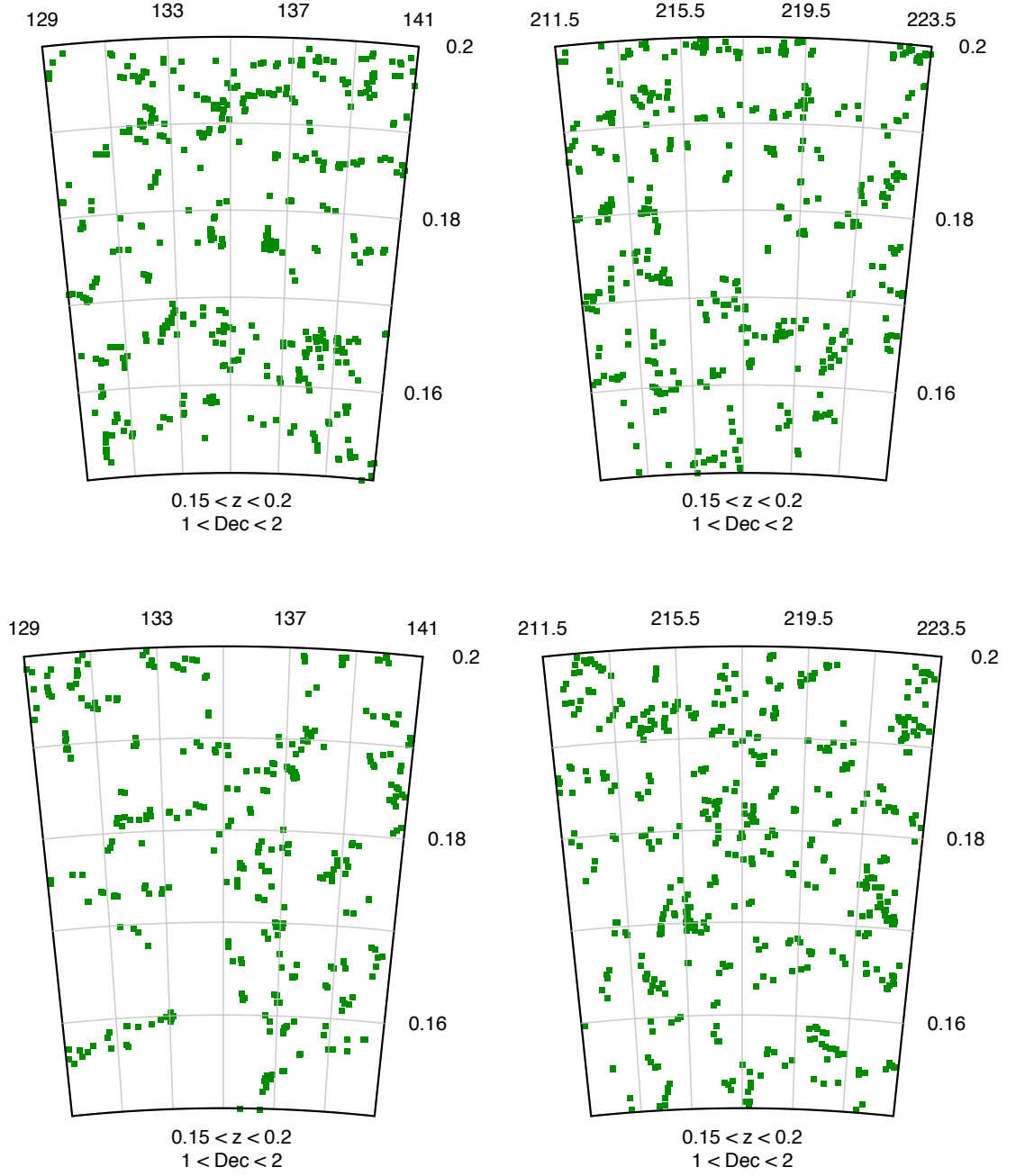


Figure 4.14: Tendril galaxies (shown as green points) for two observed GAMA fields and two mock GAMA fields (top and bottom rows respectively). The top row shows tendrils for a slice of the G09 and G15 fields, while the bottom row shows the same slice of sky from two separate volumes of the GAMA mocks. The two populations of tendrils are visually indistinguishable.

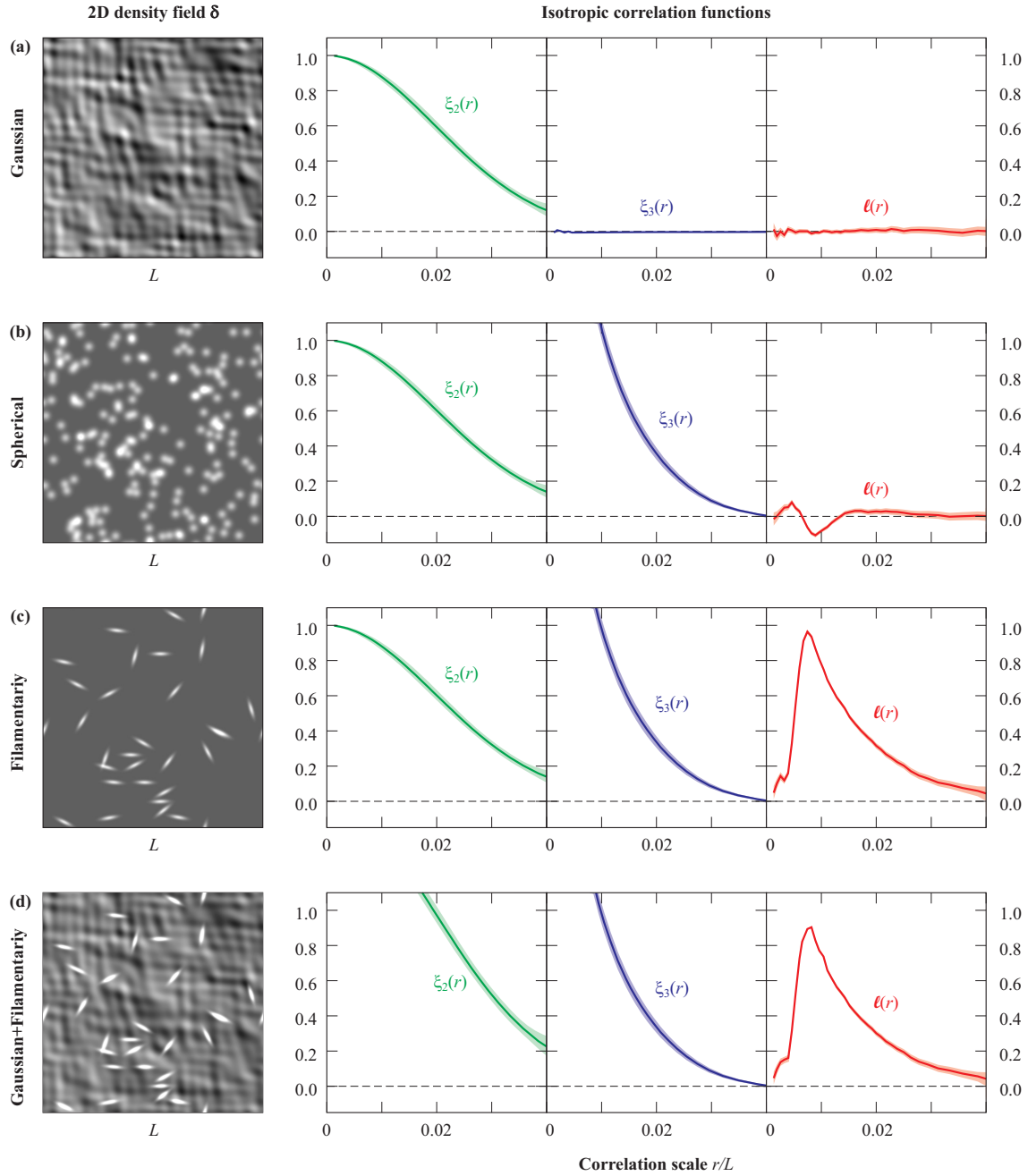


Figure 4.15: Examples of 2D density fields with their corresponding correlation functions $\xi_2(r)$, $\xi_3(r)$ and $\ell(r)$ shown in green, blue, and red respectively. Shaded regions represent 1σ confidence intervals. The density fields are purposely chosen to have the same two-point correlation function. From top to bottom, the density fields are constructed by superposing 350 randomly oriented Gaussian plane waves; 200 symmetric Gaussian spherical profiles; 30 randomly oriented Gaussian spheres elongated to resemble lines; and a superposition of (a) and (c). Figure reproduced with permission from (Obreschkow et al., 2013).

Given the non-cartesian volume of the GAMA cones and the varying number of galaxies in each sample, the filamentarity is measured via the excess line correlation $\Delta l(r) = [l(r) - l_0(r)]\sqrt{f}$. Here, $l(r)$ is the line correlation of the galaxies, taken as points of equal mass, truncated to distances larger than $400h^{-1}$ Mpc and fitted in a cubic box of $215 h^{-1}$ Mpc side length. The reference function $l_0(r)$ is the line correlation of an equal number of points, randomly distributed in an equivalent survey volume. The factor f is the volume fraction of the GAMA cones within the cubic box. The term \sqrt{f} ensures that $\Delta l(r)$ is approximately independent of the volume of the cubic box (according to Section 3.4 of Obreschkow et al., 2013).

$\Delta l(r)$ is calculated separately for filament, tendril and void galaxies, as well as all galaxies combined. This calculation is performed individually for each of the three equatorial GAMA fields as well as for the GAMA mock catalogues. Figure 4.16 shows the functions $\Delta l(r)$ of the observed data (solid lines) with measurement uncertainties (1σ error bars). These uncertainties arise from the limited number of independent Fourier modes in the finite survey volume. They do *not* include cosmic variance due to correlated large scale structure. In turn, dashed lines and their error bars represent the median and standard deviation of $\Delta l(r)$ of the nine GAMA mock catalogs. These standard deviations naturally include both, uncertainties in the measurements of $\Delta l(r)$ and cosmic variance in the survey volume. Since these error bars are much larger than those of the observed data, cosmic variance is the dominant uncertainty in $\Delta l(r)$ across all considered scales. There is remarkable agreement between data and mocks; however the variations in mock data appear to be greater than the variation of the observed $l(r)$ between fields. This suggests a possible correlation of structure between fields that the mock fields cannot capture, as they are randomly pointed lightcones within a simulation box.

Observed and simulated filaments, tendrils and voids show very similar values for $\Delta l(r)$ on all scales across the three GAMA fields. In all cases, filament galaxies alone exhibit a higher correlation $\Delta l(r)$ than all galaxies together. Void galaxies show very little excess line correlation at all scales, indicating that for the magnitude limited sample in the GLSSC, void galaxies are free from structure. Tendril galaxies show a clear intermediate line correlation, leaving no doubt that those galaxies contain real filamentarity well beyond that of a random point set. This example effectively demonstrates the usefulness of higher-order statistical estimators such as $l(r)$ to characterize large scale structure.

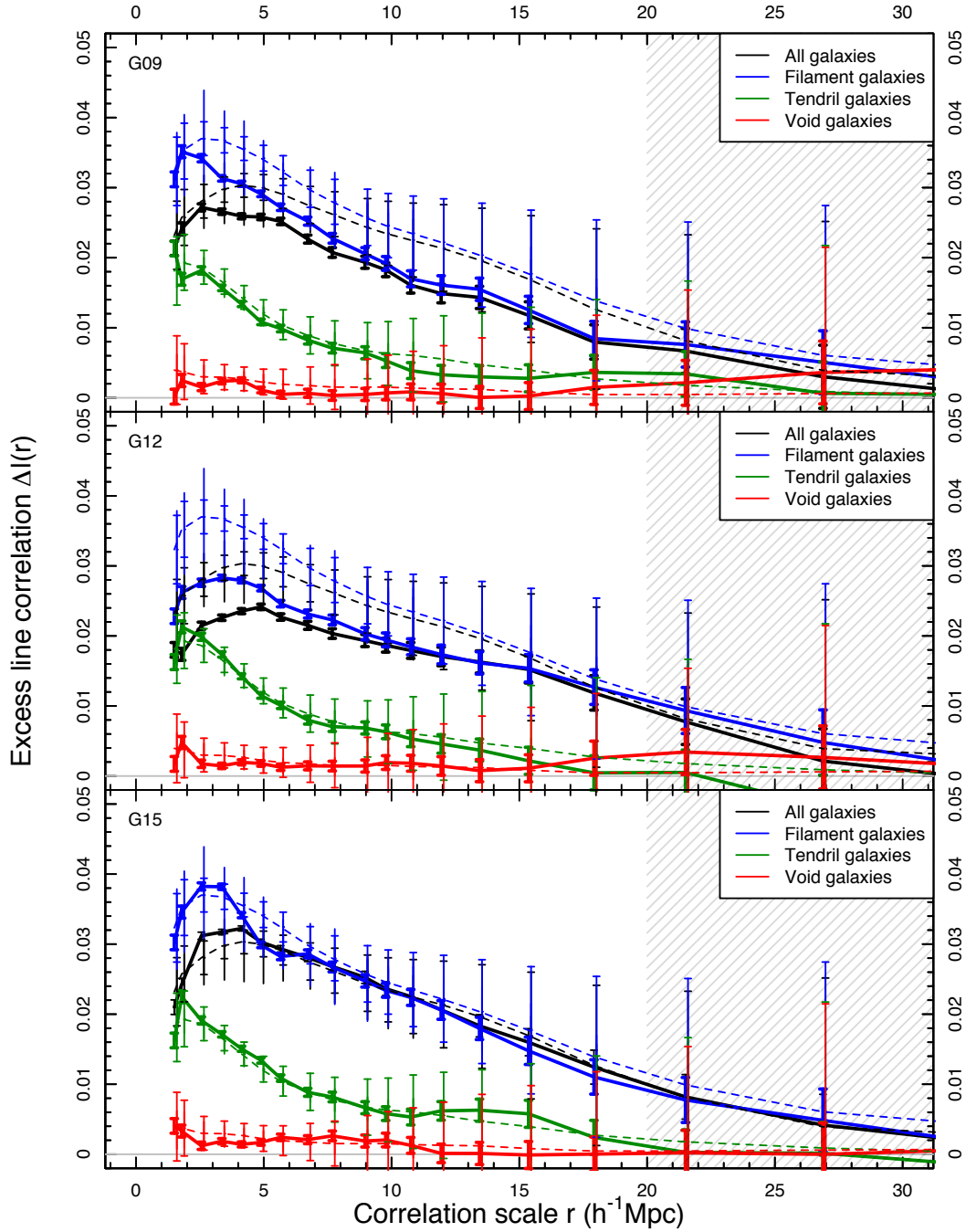


Figure 4.16: Excess line correlation $\Delta l(r)$ shown for filament (blue), tendril (green) and void galaxies (red), as well as all galaxies combined (black) across all three equatorial GAMA fields. The dashed lines show the median $\Delta l(r)$ for the GAMA mock catalogues, for which we show both the formal $\Delta l(r)$ error as typical error bars, and the spread between mock lightcones as solid lines with no arrowheads. The function $\Delta l(r)$ roughly measures the probability of finding three equidistant points separated by r on a straight line, in a density field that is analogous to GAMA but stripped of all two-point correlation. Shaded regions show lengths where $\Delta l(r)$ is unreliable due to the limited declination range of the GAMA cones ($\sim 40 h^{-1}$ Mpc on average). Galaxies in filaments show more excess line correlation than all galaxies, void galaxies show very little excess line correlation, while tendrils have a distinct, intermediate signature. There is excellent agreement between the data and the mock catalogue structures.

The lack of signal in the excess line correlation for voids in Figure 4.16 is a strong indicator that this work has reliably identified galaxies within voids in GAMA. It is also a reflection of the parameter selection that requires r and q in the filament finding algorithm to be selected such that void galaxies have a minimised signal in their two point correlation function. However, there is ample room for future surveys to discover deeper structures embedded in and around the void galaxies presented in the GLSSC. This is a somewhat plausible scenario, as an earlier version of the GLSSC made with GAMA I (with a magnitude limit of $m_r < 19.4$ mag instead of $m_r < 19.8$) data had several tendrils which were reclassified as filaments in the GAMA II version of the catalogue. Furthermore, the excess line correlation function of voids in GAMA I data shows a similar lack of signal across all distances.

4.3.2 Shrinking voids

Early observational and theoretical studies of voids (Joeveer et al., 1978; Gregory & Thompson, 1978; Kirshner et al., 1981) have been supplemented through more recent large galaxy redshift surveys (Colless et al., 2001; Abazajian et al., 2009) that provide provide comprehensive and complete pictures of voids, complemented by numerous void-finding techniques (e.g. El-Ad & Piran (1997); Hoyle & Vogeley (2002); Aragon-Calvo et al. (2010)). Voids are an easily recognised (and, initially quite unexpected) feature of the Cosmic Web, spanning between 20 and $50 h^{-1}$ Mpc. We now know that the $> 1 h^{-1}$ Mpc environment of a galaxy in a void is very empty relative to that of a galaxy in a dense cluster, making such objects ideal for studying galaxy evolution independent of environmental processes. The few galaxies that do exist in voids are subject to dynamics that are unique to these underdense regions due to the lack of neighbouring galaxies (Blumenthal et al., 1992; Sheth & van de Weygaert, 2004). Voids can serve as tools for constraining cosmological parameters, or for testing the accuracy of large cosmological simulations, as shown in Dekel & Rees (1994); Lavaux & Wandelt (2010); Park et al. (2012) and many other works.

In the GLSSC, a void galaxy is defined as a galaxy that is at least $4.13 h^{-1}$ Mpc away from the nearest galaxy that belongs to a tendril (and tendril galaxies themselves must be at least $4.56 h^{-1}$ Mpc away from the nearest filament), so they exist in the most underdense regions of the GAMA fields. We do not determine where voids are, or their sizes, only the galaxies that exist in very underdense regions. This is a significantly different way of defining a galaxy in a void compared to the traditional approach of detecting voids using voidfinding

algorithms (examples of which are given at the start of this chapter) that depend on locating underdensities and selecting the galaxies within them as void galaxies.

Recently, Pan et al. (2012) have released a catalogue of void galaxies obtained from the SDSS-DR7 data using a voidfinder similar to the one introduced by El-Ad & Piran (1997). In Pan et al. (2012), voids are identified within a volume limited subsample of 120606 SDSS galaxies with $M_r < -20.09$ mag and $z < 0.107$: galaxies are first determined to be within the field or not, using the third nearest neighbour distance d_3 and the standard deviation of this distance σ_{d_3} . Any galaxy with $d = d_3 + 1.5\sigma_{d_3} > 6.3h^{-1}$ Mpc is considered to belong to a part of large scale structure: a filament or a cluster. These galaxies are afterwards referred to as ‘wall’ galaxies. All other galaxies are classified as ‘field galaxies’ and are removed from the sample. The remaining wall galaxies are then gridded into cells of side $5 h^{-1}$ Mpc and all empty cells are considered to be possible centres of voids. A sphere is grown from each cell until it is bounded by four wall galaxies, and any two spheres with more than 10% overlap are considered to belong to the same void. This process is continued iteratively until no more spheres can be grown or merged. A sphere must have a radius of at least $10 h^{-1}$ Mpc to be considered a void.

As the GAMA regions overlap with the SDSS, there are some voids in the Pan et al. (2012) catalogue that lie in GAMA regions. In order to be as conservative as possible, a sample of all GAMA galaxies that lie within the inner two thirds of SDSS voids in GAMA regions is taken. This sample contains 1130 galaxies with $0.001 \leq z \leq 0.1$ and $M_r < -19.77$. Of these galaxies, 132 are part of the GLSSC – that is to say that they were part of the sample that the algorithm described on the preceding chapter was run. This small number is due to the fact that SDSS is a more faint survey with a lower target density. Figure 4.17 shows all galaxies in the GAMA large scale catalogue for the three equatorial fields in grey and circles the 132 galaxies that are matched to the 1130 GAMA galaxies in the SDSS voids. The circles are coloured according to the structures those galaxies are classified as being in the GLSSC: blue circles are galaxies in filaments, green circles are galaxies in tendrils, and red circles are galaxies in voids. Only 25% of the GLSSC galaxies found in SDSS voids are also galaxies that are identified to be in voids in the GLSSC, with the vast majority (65%) considered to be in tendrils, and a further 11% associated with filaments. For a subsample of 79 galaxies with $M_r < -20.06$ mag, 15% are filament galaxies, 61% in tendrils and 24% in voids. This highlights the importance of using a deep survey with a high target density when investigating the properties of void galaxies,

	Filaments	Tendrils	Voids
$M_r < -19.77$	11%	64%	25%
$M_r < -20.06$	15%	61%	24%

Table 4.2: Percentage of galaxies, grouped by large scale environment, matched with galaxies in voids located in GAMA fields, for two different absolute magnitude cuts. It is interesting to note that when moving from the brighter to the dimmer magnitude cut, many tendrils are reclassified as filaments, but the fraction of voids remains largely the same.

as voids defined in a shallow survey may contain additional undetected galaxies. Despite the differences in void galaxy definition, it is clear that regions considered to be empty in the SDSS do in fact contain galaxies.

The implication of these results is that voids are largely susceptible to the diagnostics of observational surveys used to detect them, and that it is not always advisable to use surveys with a shallow magnitude limit for searching for these structures – and less so for characterising the properties of galaxies that reside in them. Having said that, there is no evidence that there will not be a future survey that revisits the GAMA regions with a much deeper magnitude limit and finds that the voids presented in this work are just as full of more faint, less massive objects. The fact that the excess line correlation function for voids in Figure 4.16 shows no signal does seem to indicate that the void galaxies in the GLSSC are truly free from structures, but this is not evidence that there are no more undected sources in the background. However, the fact that the fraction of void galaxies in SDSS voids remains virtually unchanged when moving from a $M_r > -19.77$ cut to a $M_r > -20.06$ cut might imply that there are not many deeper void galaxies to find.

4.4 Comparison of LSS to mocks

I now concentrate on comparing the overall properties of filaments in the observed GAMA data, to filaments generated from the mock galaxy and group catalogues. For both data sets, the exact same algorithm is used with identical parameters $b = 5.75h^{-1}$ Mpc, $r = 4.13h^{-1}$ Mpc, and $q = 4.56h^{-1}$ Mpc. The same sample selection process as described in Section 4.1.2 is used, and the same hierarchy of catalogues is generated, producing a total of 10 sets of large scale structure catalogues for each region. Filaments in both data sets are largely the same, aside from some very subtle morphological differences. Figure 4.18 shows the locations and populations of the largest filaments in each field, for the observed data (shown in black) and all 9 mock simulations for that region (shown in red). Most of the large filaments are at higher redshift. A combination of effects causes this: the selected sample is such that at

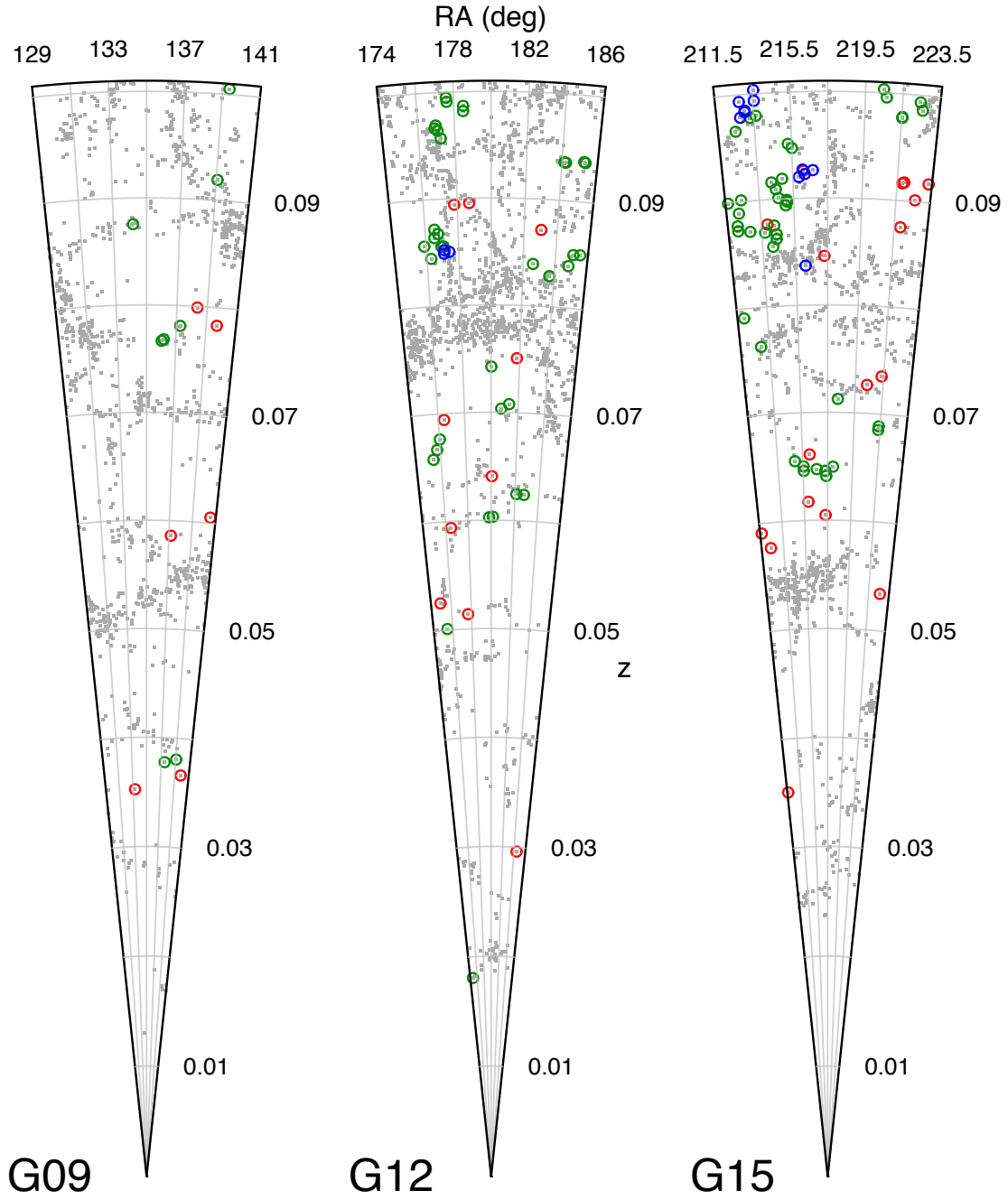


Figure 4.17: Each circled galaxy in this figure is a galaxy in the GAMA structure catalogue that is located inside a void identified in SDSS-DR7 data. The colour of each circle represents the type of structure it is thought to be in in the GAMA structure catalogue: blue circles represent galaxies in filaments, green circles represent galaxies in tendrils, and red circles represent galaxies in voids. Only a quarter of these matched galaxies are identified as voids in the GAMA structure catalogue. We note that we show the full 5° declination for each field in this figure, resulting in increasing projection effects at higher redshifts.

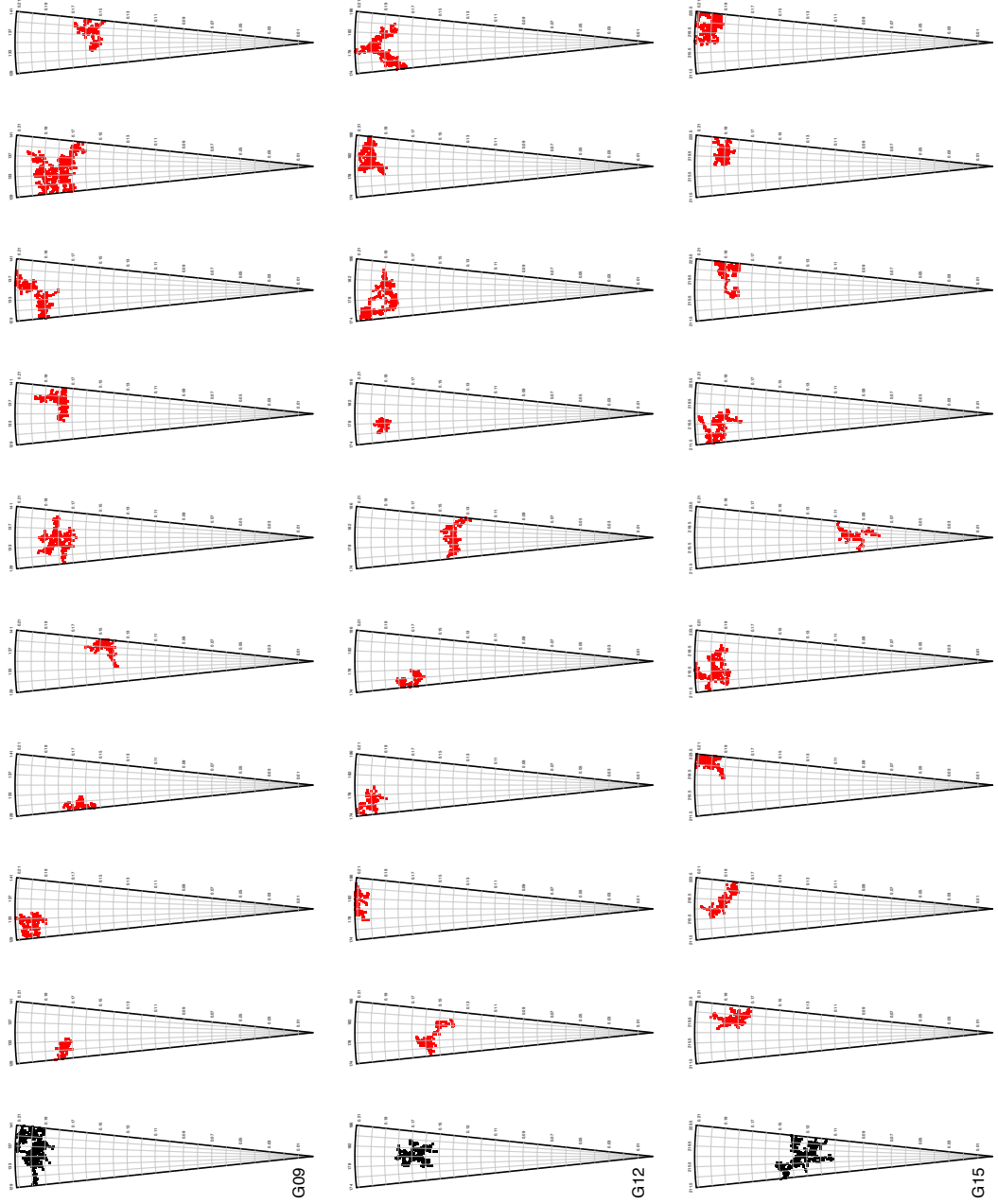


Figure 4.18: Locations of largest filaments in the data (shown in black) and the mocks (shown in red) for all three regions, with each row representing a region. The largest filamentary structures we detect have complex geometries and morphologies, where one might expect long structures that lie along the line of sight. With the exception of G15, all large filaments lie beyond $z \approx 0.14$, further highlighting that the largest networks of galaxies span a large geometric area in two dimensions.

low redshift there are fewer galaxies, and at higher redshift the survey benefits from a wider volume, which allows the MST to expand along three axes (RA, Dec and z). The largest structures found with this method appear to be confluences of many filaments, or perhaps even massive superclusters, or galaxy complexes. The largest filaments are also the most morphologically complex filaments that exist.

It is important to note that while I am making comparisons between real and mock filaments, these comparisons can only apply to filaments found in the GAMA mocks; moreover, these comparisons are applied on the basis that the algorithm is run on both data sets using the same set of parameters to ensure consistency and that the GAMA mocks successfully replicate the number density and luminosity function of the observed GAMA data. Since b , r , and q depend most strongly on galaxy number density and luminosity, it is acceptable to use the same values for these parameters across both data sets. To ensure that these assumptions are valid, values of b , r , and q are calculated for each individual mock catalogue, and arrive at the following values: $b = 5.75 \pm 0.2$, $r = 4.3 \pm 0.8$, $q = 3.9 \pm 0.7 h^{-1} \text{ Mpc}$ with errors of 1σ about the mean. The parameters all agree within their errors with values derived for the observed data set.

A comparison of filament lengths is shown in Figure 4.19. This figure displays, for observed and mock data (the black and red lines respectively), the binned abundance distribution of the lengths of filament components as a function of the number density of filaments. The backbone length is a good indicator of the overall span of a filament across its dominant axis as it traces the longest possible path from one end of the filament to the other through its central node. These are shown by the solid lines in Figure 4.19, with vertical error bars denoting 1σ uncertainty ranges from the variance between the mocks. The horizontal error bar on the final point for the mock filaments marks the 1σ distribution in maximum filament lengths across mock filaments. This error highlights the uncertainty of these results at high values for length, as one runs into sample variance (only one or two filaments are that long). For bins with only one object, I calculate upper error limits based on Poisson statistics. To first order there is a remarkably good agreement between the mock and observed filaments across 2 orders of magnitude of scale; however, future comparisons of this nature may benefit from calculating the variance in filament backbone lengths between fields.

An analogous analysis shows that this holds true for tendrils as well (see Figure 4.20). Both statistically and visually, tendrils in simulations are identical to those in observations, which is a surprising result as the GAMA mocks were not designed or created with tendrils in mind – these appear to be just as much a consequence of the simulations as a naturally occurring population of galaxies. Future incarnations of the GAMA mocks will have simulated photometry and comparisons between the two populations will become far more interesting and in-depth.

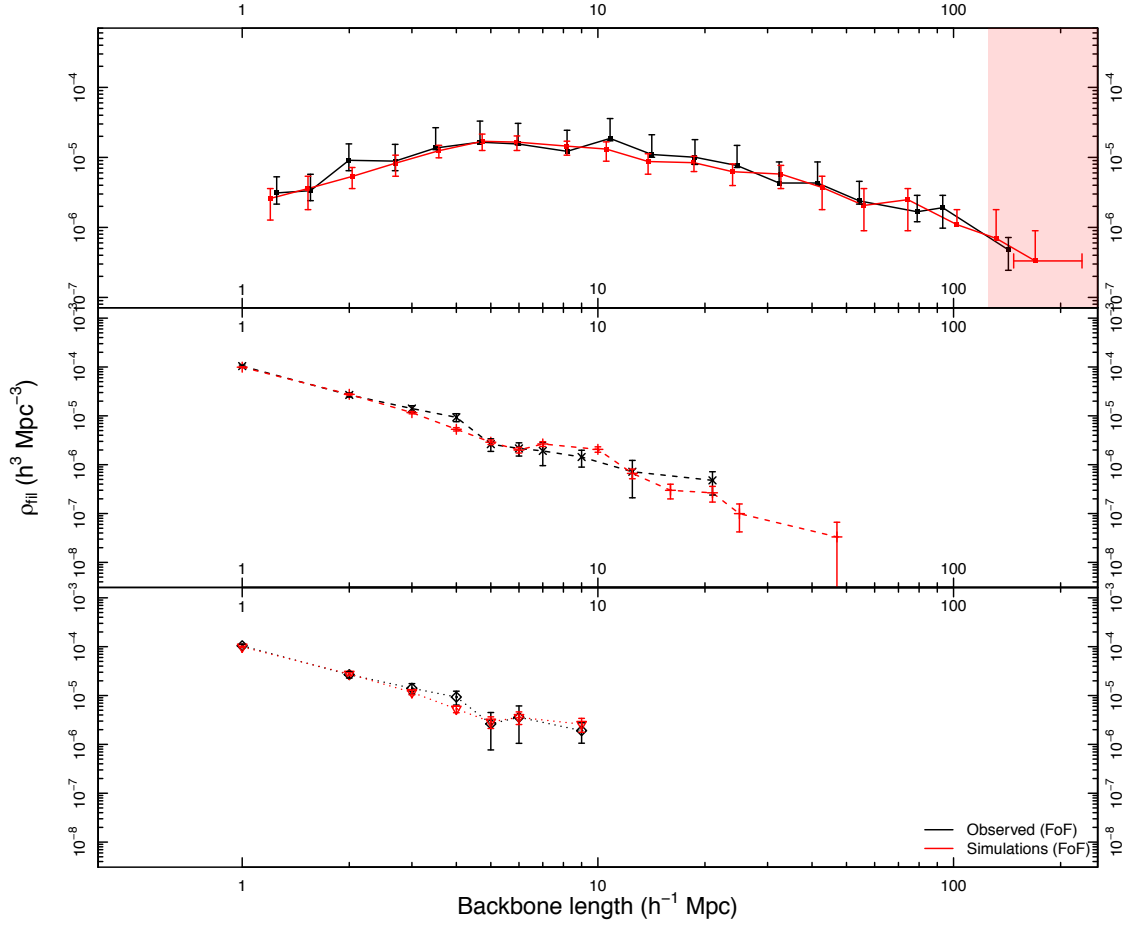


Figure 4.19: The binned distribution of the number density of filaments as a function of length for various components of filaments with Poisson errors, ranging from backbones (solid lines, top panel) through to branches of order $n = 2$ and 3 (dashed and dotted lines, in the middle and lower panels respectively). Black and red lines correspond to filaments from the data and mock regions respectively. The x-axis positions of the points are the median values within that bin. The horizontal error bar on the final point in the red line shows the 1σ spread of the fifteen largest mock filaments across all regions and volumes. Bins with no detections show only an upper limit derived from Poisson statistics. The shaded region marks distances at which the geometry of the GAMA regions means the backbone lengths are poorly constrained.

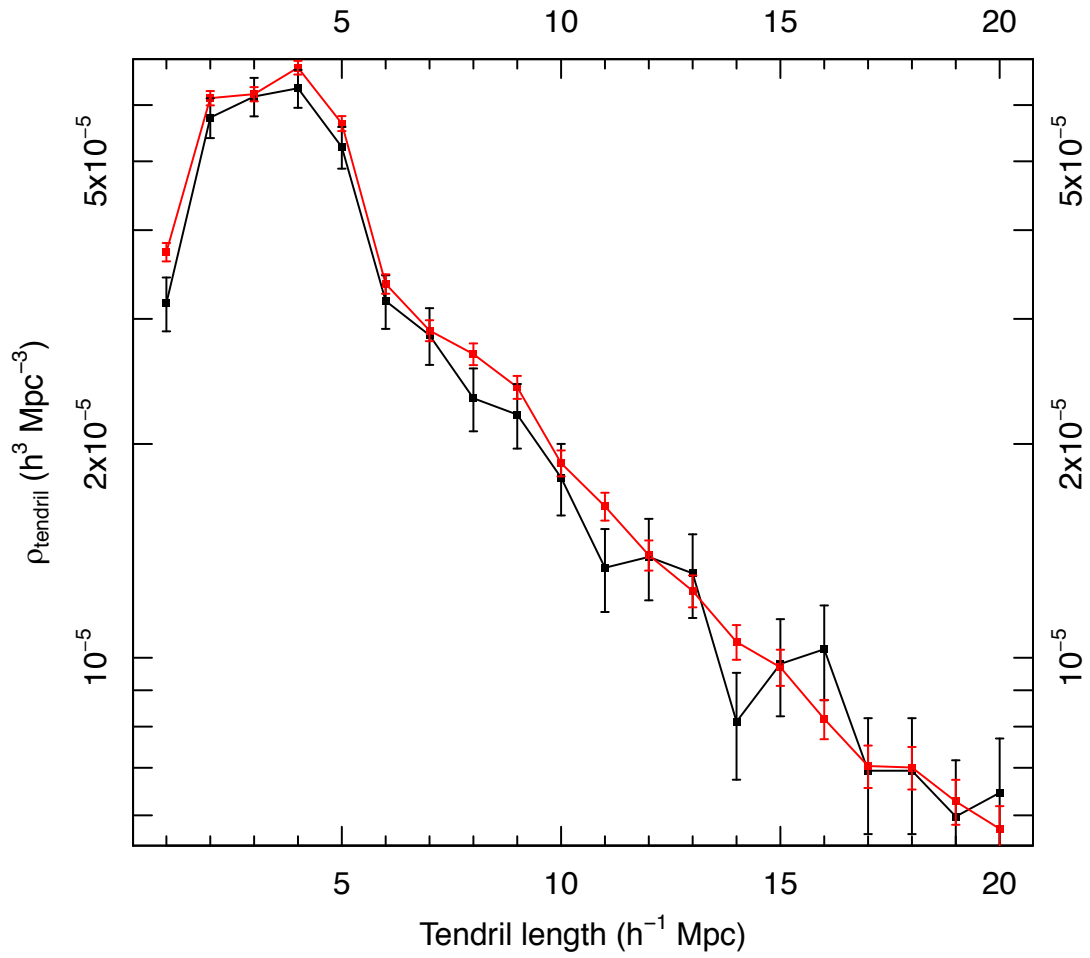


Figure 4.20: Much like in Figure 4.19, the distribution of distances of tendrils are shown for tendrils in observations (black) and tendrils in simulations (red). Error bars show 1σ deviations about the mean in each bin. Both populations of tendrils are statistically indistinguishable.

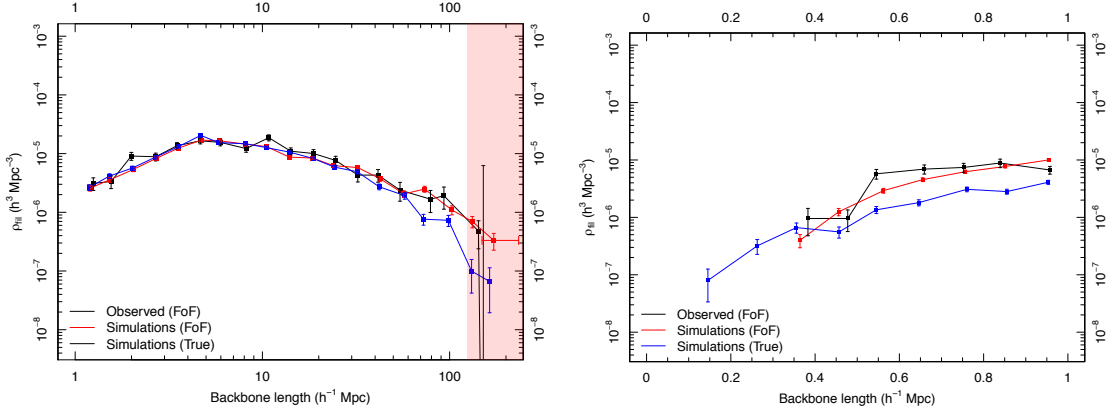


Figure 4.21: *Left:* In the same manner to Figure 4.19, the number density of filaments with $b = 5.75 h^{-1} \text{ Mpc}$ as a function of backbone length is shown. The black, red and blue lines each correspond to filaments in observed groups, groups recovered from simulations using the R11 groupfinder, and the intrinsically known groups from the simulations. *Right:* The distribution of backbone lengths for filaments with $b = 1 h^{-1} \text{ Mpc}$. These filaments are all group-group pairs. Notably, we observe that there are more short filaments of intrinsic groups compared to the observed and FoF groups.

Given that the G^3C has been generated by calibrating the FoF algorithm against mock galaxies whose intrinsic grouping is known, it is an interesting exercise to generate filaments using haloes (whose grouping is known in the simulations) instead of groups. I apply the same algorithm, with the same sample selection (I select haloes instead of groups of galaxies) and generate filaments of haloes, whose backbone length is shown on the left panel in Figure 4.21 in blue. Using the same values for b , r , and q in this case ensures that any difference in results for these filaments will be due to how the groupfinder in R11 breaks haloes apart into groups; just like the caustic mass analysis done for halos and groups in the previous chapter. These can be considered the ‘true’ mock filaments, as they are not subject to biases in the FoF algorithm. Halo filaments are remarkably similar to mock and observed filaments, as shown in the left panel of Figure 4.21. It is expected that FoF filaments are longer than halo filaments, however, as the FoF algorithm will occasionally break a halo into multiple groups; this effectively means that the MST has an extra stepping stone between two haloes and is therefore able to form structures with shorter links that are less likely to be trimmed for a given value of b . There is an equal chance for the groupfinding algorithm to merge multiple haloes into a single group, depending on the halo mass range being considered.

Similarly by reducing the maximum edge length to $1 h^{-1} \text{ Mpc}$, one can begin to examine the group-group pairs that are within $1 h^{-1} \text{ Mpc}$ of each other; all the FoF filaments in this sample consist of two neighbouring groups (most halo filaments are composed of 2 or 3 haloes). We

can see in the right panel of Figure 4.21 that for intrinsic group filaments there exist shorter group-pair filaments compared to the observed data. The GAMA mocks are known to have a high value of σ_8 , which may explain why halo-halo pairs are much closer to each other than observed galaxy pairs

Both observed and simulated (FoF) data have filaments that grow larger in a similar way as a function of number of groups. Figure 4.22 shows the length of the backbone of the filament as a function of the number of groups in the backbone of the filament, for observed and mock data in black and red respectively. The shaded regions about each point show 1σ spreads about the mean for the filaments in that length bin, and binning is made so that each contains 20 filaments. The growth of filaments is very similar between observed and mock data, with no statistically significant differences.

In a similar way, it is possible to examine the complexity of filaments as a function of the maximum linking length allowed between groups during the MST process. Here, complexity refers to the overall intricacy of the topology of each filament; or the number of branches it has. A filament with a single backbone is referred to as a ‘simple’ filament, while one with numerous branch orders (i.e. 4 or more) is thought of as being ‘complex.’ As the maximum linking length in the MST tends towards smaller values, the complexity of filaments decreases, as trees are only allowed to exist between very close group neighbours, and these tend to be simple group-group pairs, as with the population of filaments shown in Figure 4.21. This does not allow for many branches to be formed. In Figure 4.23, fractional complexities for filaments are shown as a function of maximum linking length. By fractional complexity, I am referring to the fraction of filaments with 1, 2, ... , n branches. As in Figure 4.6, these filaments are all constructed from the sample of galaxies and groups, with only the MST parameter b varying. The solid and dashed lines show the fractions for observed and mock data respectively; and the colour of the line represents the branch order. Blue, purple, green, orange and red show the fraction of backbones, second, third, fourth and fifth order branches. The errors for each point confidence estimates on population proportions derived from a beta distribution¹, as described in Section 3 of Cameron (2011).

The points in the shaded region of Figure 4.23 show the relative branch fractions for the

¹The beta distribution is derived from the measured success fraction, or the observed population proportion, defined as k/n ; or observing k successes in n observations. Cameron (2011) show that the beta distribution estimates confidence intervals on population proportions with small sample sizes better than other estimation methods, which is why it is used in this analysis.

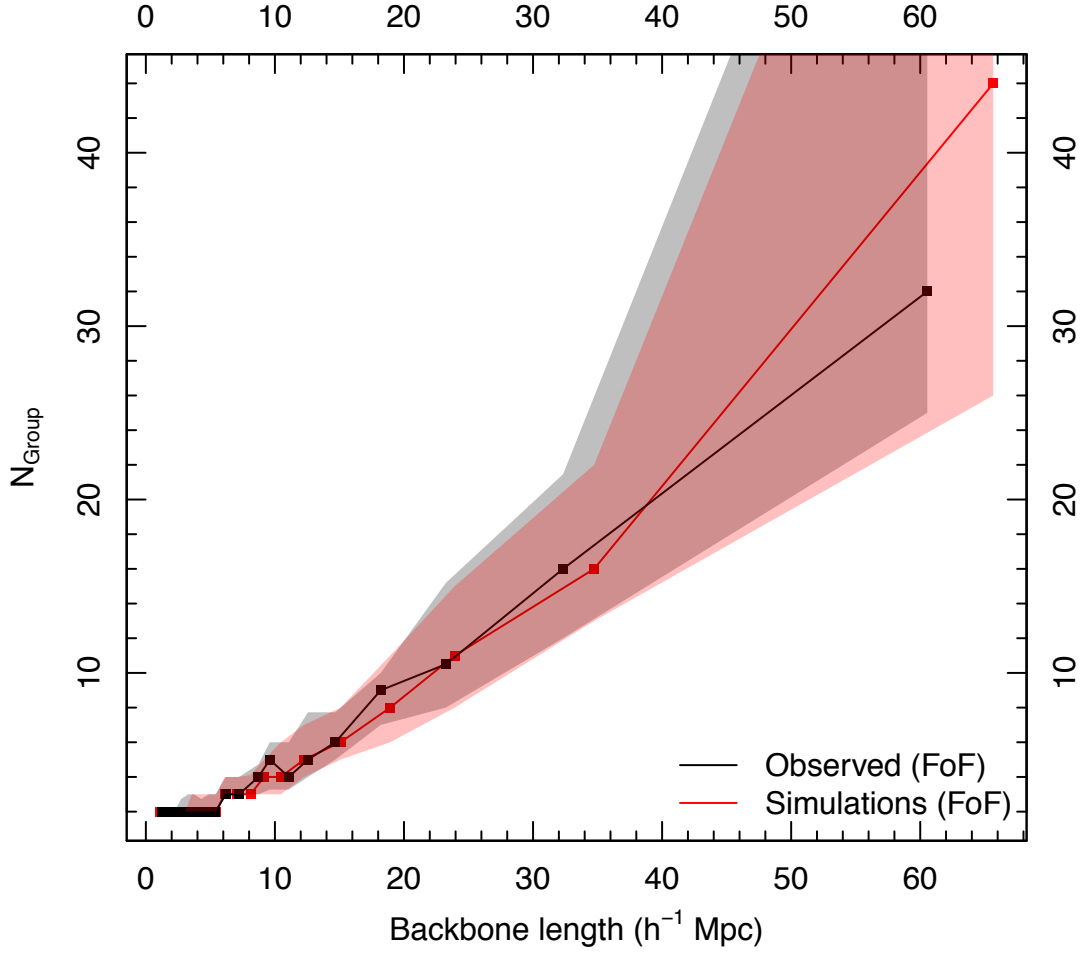


Figure 4.22: The relationship between backbone length and the number of groups in the backbone in bins containing equal numbers of filaments for observed data and FoF mock groups, shown in black and red respectively. The shaded regions denote 1σ intervals around the mean; points with no shaded region around them are single entries. This data is binned along the x-axis, and there are bins where there is no data; in these cases the point is omitted.

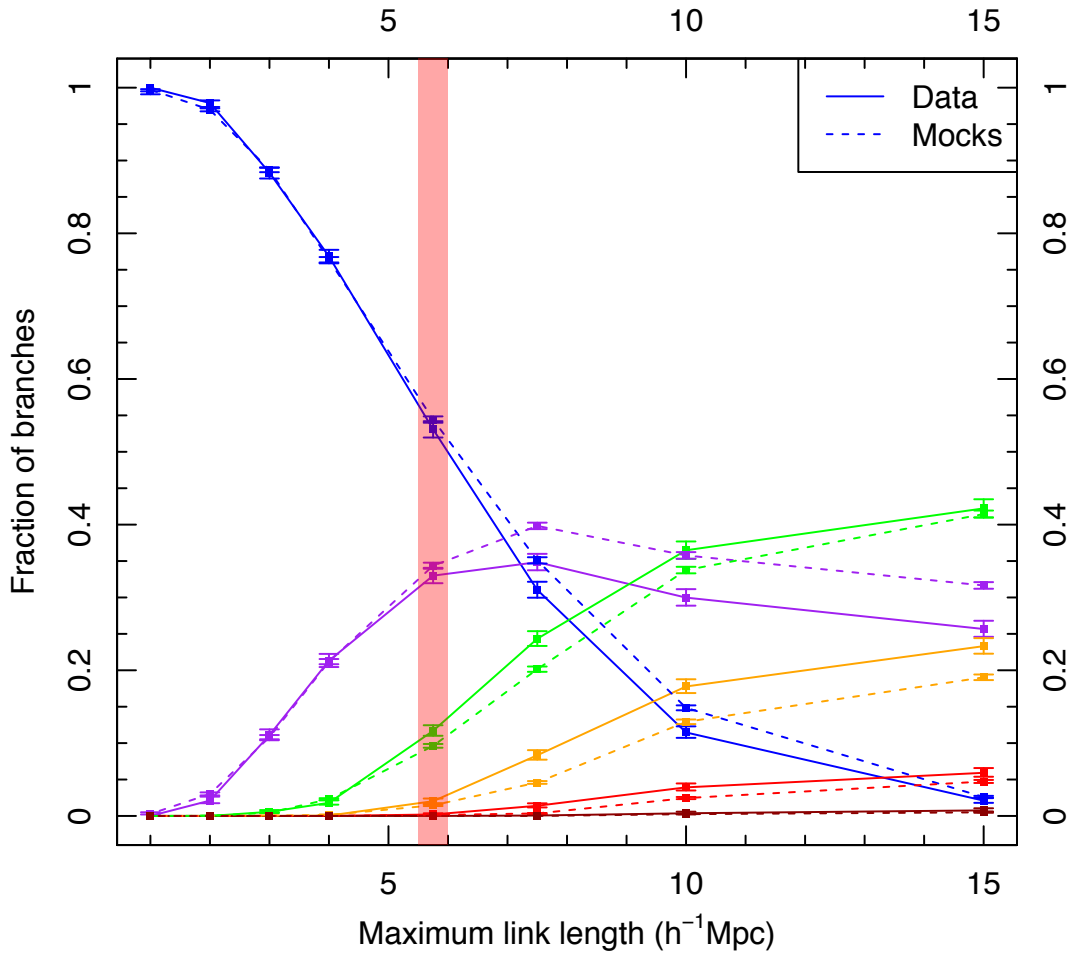


Figure 4.23: Comparing the ‘complexity’ of filaments in observed data and simulations; whereby complexity we refer to the fraction of branches that are backbones, and higher order branches. The solid lines represent data, and the dashed lines, mocks. Different coloured lines show increasing branch orders, from first to fifth order. As b increases, more complex filaments with more higher order branches are formed. At $b = 5.75h^{-1}$ Mpc, shown by the shaded region, observed and mock filaments have similar fractions of branches, aside from a slight overabundance of third order branches in observed filaments. The error bars show 1σ uncertainties about the population fraction.

same filaments displayed in Figures 4.19 and 4.21. The relative fraction of third order branches is slightly higher for observed filaments, otherwise both sets of filaments are very similar in their morphology. The difference between the two populations decreases sharply at lower values of b and for $b \leq 5h^{-1}$ Mpc there is no difference between the data and the mocks, however at $b = 15h^{-1}$ Mpc there is a more notable difference between the fractions of some components, most notably second and third order branches.

Beyond the power of the two-point correlation function, this analysis confirms that the GAMA mocks successfully reproduced the observed distribution of galaxies on large scales. It is very difficult to visually distinguish between real and mock data when looking at large scale structure maps, as shown in Figure 4.24. Once again, figures 4.21 and 4.22 show the overall similarity between observed and mock filaments.

4.5 Discussion and summary

This chapter has presented a method to systematically identify and categorise large scale structures in the Universe, as well as identify different populations of galaxies in different structural environments. This algorithm is based on using minimal spanning trees to identify filaments composed of groups, around which nearby galaxies are identified and these are associated with each filament. The remaining population of galaxies is then classified as tendrill galaxies or void galaxies using a second minimal spanning tree. The parameters used for this approach are selected by optimising for a large scale structure that obeys two assumptions: that the brightest groups be in filaments, and that the distribution of void galaxies show much less structure than for filaments and clusters, particularly at large scales. Large scale structure catalogues are generated for the three equatorial GAMA fields, as well as 9 mock galaxy volumes for each field, adding to a total of 3 observed LSS catalogues and 27 mock LSS catalogues.

With GAMA we benefit from an extremely complete survey that has a very high target density, revealing that there is far more underlying structure behind the brightest galaxies and groups that form the skeletal signatures of filaments and large scale structure. This is the largest, most complete, and deepest observational catalogue of large scale structure to date. As we revisit the same patch of sky and conduct deeper, more complete observations, we find more complex substructures, such as tendrils. Tendrils are a potentially new class of galaxy large scale structure that connect filaments to each other, or emerge from filaments and extend into underdense regions, terminating within voids. Tendril galaxies contain 23.8% of all stellar mass (compared to around 70% for filaments) for a magnitude and mass limited sample of the structure catalogue; however they are tenuous enough to be difficult to detect using structure finders that rely on density smoothing (as most structure finders do), or only consider galaxies in groups. These methods will not fully take into account less massive galaxies outside of the most high density regions, and therefore miss out on these structures that contain a significant amount of stellar mass.

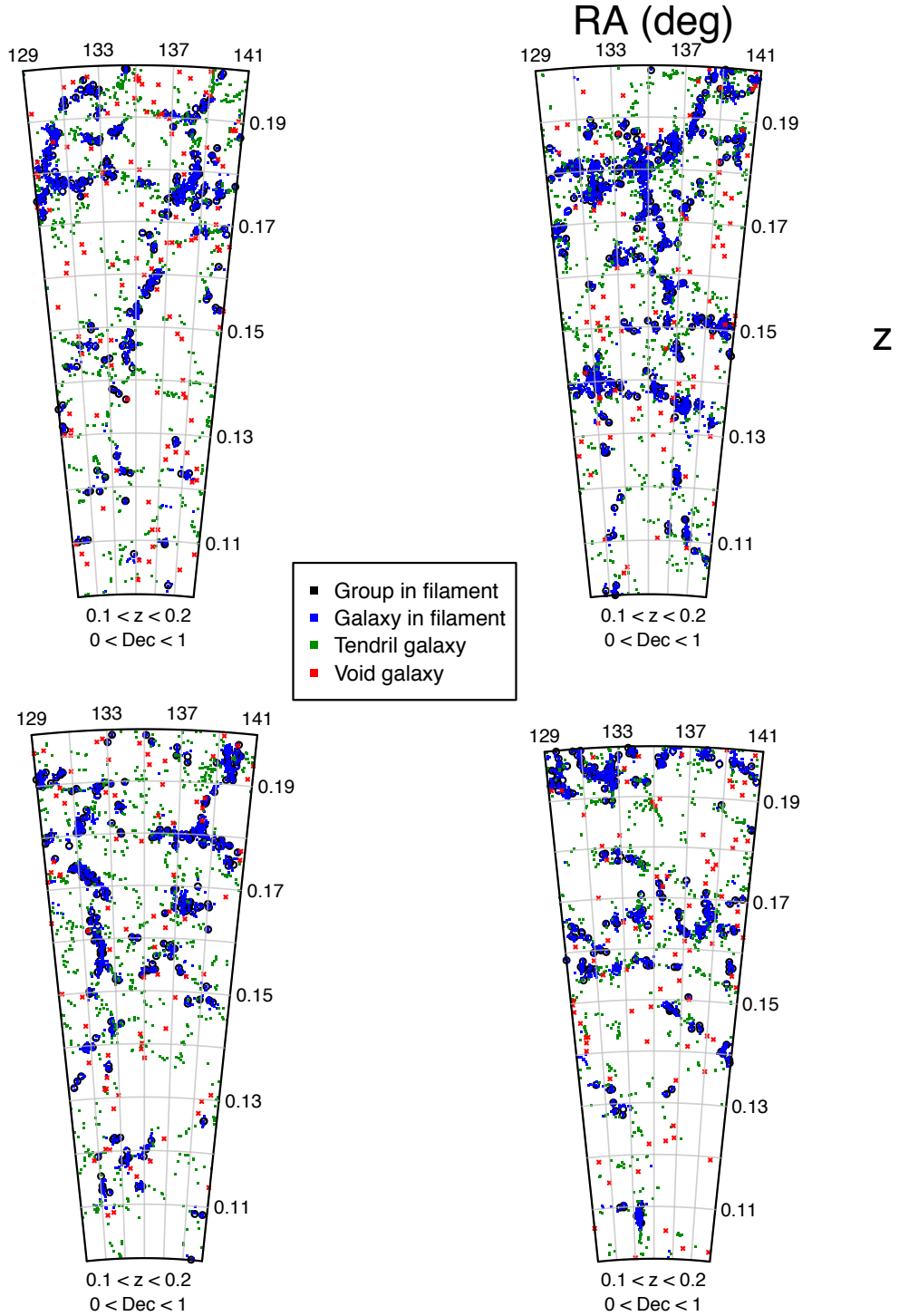


Figure 4.24: The same region in G09 with different galaxy populations (colour coded as in 4.12), with observed data shown in the bottom right and the other panels consisting of mock data. The similarity between all four fields is apparent, and serves to visually highlight the success of the mock catalogues in reproducing large scale structure.

As structures, tendrils appear to be morphologically distinct from filaments in that they are less dense and span shorter distances. On average, a tendril will contain just under 6 galaxies and measure roughly $10 h^{-1}$ Mpc. As with filaments, Figure 4.20 shows that the lengths of tendrils in simulations match well with those in observations. The similarity of simulated and observed filaments and tendrils is remarkable. When examining the filamentarity of tendrils using the line correlation function $l(r)$ we find that they are distinguished from filaments by having an intermediate excess line correlation and are certainly distinguishable from voids. A visual inspection of tendrils, shown as the green points in the second and fourth panels of Figure 4.13, confirms that they are distinct structures to filaments. Galaxies in tendrils are most likely in the intermediate stages of evolution from a low density galaxy to one in a high density region. Figure 4.25 shows the distribution of stellar masses and colours for a mass complete subset of galaxies in filaments, tendrils and voids, with contours enclosing the upper 95%, 50%, and 10% of galaxies respectively. From this Figure it is possible to see that voids have a more distinct distribution of masses and colours compared to filaments and tendrils; however, a far more detailed analysis is required to accurately measure these differences. This analysis, as well studies of how other galaxy properties change as a function of environment will be discussed in far greater detail in the following chapter.

To understand the impact of tendril galaxies on traditional void classifications, I identify a subset of galaxies in the GLSSC that lie within voids found with a modified voidfinder algorithm applied to the SDSS-DR7 data set. As the SDSS overlaps in space with GAMA, some of these voids lie in GAMA fields. There are 1130 GAMA galaxies in these overlaps, 132 of which are in the GLSSC. Only a quarter of these galaxies are determined to be void galaxies in the GLSSC, and that almost 65% of them are in fact tendril galaxies. Given that tendrils show some linear structure, it is important to ensure that any void galaxy selection only includes galaxies that are as isolated as possible, in order to ensure that studies of isolated growth are conducted on truly isolated galaxies. This result may go some ways to explaining the abundance of red galaxies in voids in studies such as Kreckel et al. (2012). I stress the importance of selecting isolated galaxies and voids from the same survey, and note that voids appear to be less empty when observed at fainter magnitudes. Future void galaxy surveys will benefit from a deeper target selection and a high target density in order to identify the most isolated galaxies in the local Universe.

Overall, mock large scale structure strongly resembles observed large scale structure and is

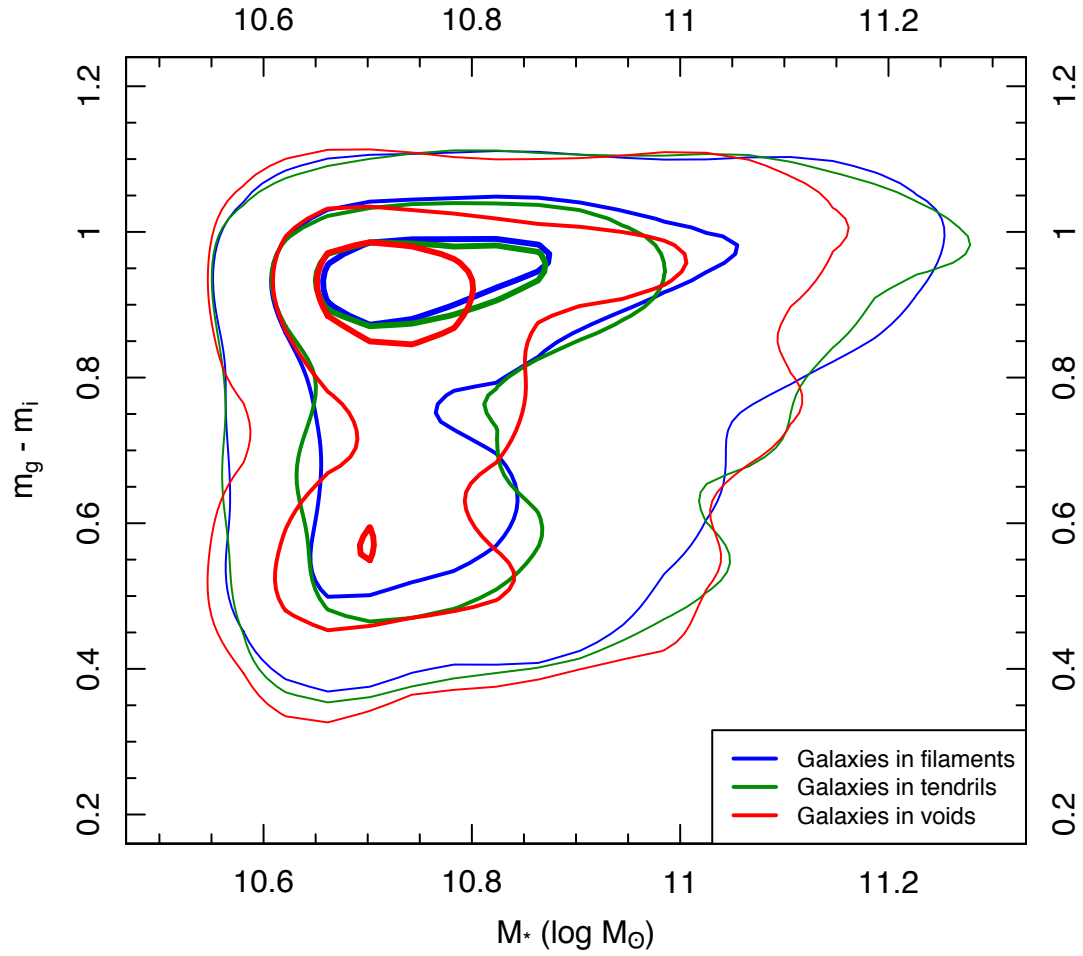


Figure 4.25: Contour plot of stellar mass versus colour for a mass complete sample galaxies in filaments, tendrils, and voids. The contours are placed such that they contain the upper 95%, 50%, and 10% of the data respectively.

virtually indistinguishable by eye (as shown in Figures 4.23 and 4.24). The filament analysis also decodes the topology of filaments into a primary backbone of links that travels from one end of the filament to the other across its centre, and various tributary branches that connect up to this central spine. Figure 4.23 shows that filaments in simulations have complexities that match very closely with observed filaments, for the value of b that we have used. If there are any differences, they are subtle: the morphology of simulated filaments is marginally closer to being less complex than observed filaments, hinting at some discrepancies in the simulations used to generate the GAMA mocks. However, these differences are minor and difficult to prove.

The GAMA Large Scale Structure catalogue is an important tool for understanding the role of environment on galaxy evolution. In the following chapter I will combine this catalogue with several existing catalogues of observational properties available in GAMA, and investigate the role of filaments and voids in the lifetime of a galaxy.

5

The properties of galaxies as a function of environment

The GLSSC provides an unprecedented catalogue of galaxies classified by environmental type. The previous chapter examined voids in greater detail, as well as verifying the filamentarity of tendrils using the excess line correlation function. While a detailed and thorough study of the impact of environment on galaxy evolution would require a whole thesis itself, a very preliminary quick-look exploration of the impact of environment on the properties of galaxies is presented in this chapter.

The evolutionary fate of a galaxy is closely tied to its immediate surroundings, within 1 megaparsec (e.g. Hahn et al., 2007b). Many observable properties of a galaxy are greatly influenced by the presence of other galaxies nearby; stellar populations in particular are very susceptible to environment. The proximity of galaxies has been shown to trigger dormant regions of gas within a galaxy into infall, leading to an increased rate of star formation (Porter et al., 2008). The local environment of a galaxy has impact on many other properties, in-

cluding colour (Kreckel et al., 2012), stellar mass (Chabrier, 2003), gas content (Beygu et al., 2013; Benítez-Llambay et al., 2013), luminosity function (Croton et al., 2005; and morphology (Butcher & Oemler, 1984; Dressler et al., 1997).

As described in the introduction to this thesis, the variations of some galaxy properties with local galaxy density are well understood. A few key facts are known about the impact of large scale structure on galaxies: namely the apparent suppression of star formation as galaxies fall into clusters (Lewis et al., 2002; Balogh, 2007). Norberg et al. (2002); Zehavi et al. (2005, 2011); Masaki et al. (2013) and others show links between galaxy clustering and luminosity and colour. Robotham et al. (2013) show that star formation is suppressed in pairs. However, there has been no broad look at galaxy evolution from filaments down to pairs in a single survey. Following the mantra of always inspecting the data, I start this chapter by showing postage stamps of randomly selected galaxies in filaments, tendrils, and voids in Figures 5.1, 5.2, and 5.3, with each galaxy placed according to its stellar mass and $u-r$ colour. Each stamp measures $500 \times 500 \text{ h}^{-1} \text{ kpc}$. The postage stamps are false colour images generated using *Hig* band images from SDSS and UKIDSS imaging data. The galaxy images shown in the figures suggest that the influence of environment will be subtle as all galaxy types and colours visible in all panels. One possible exception however is the dearth of higher stellar mass systems visible in the void environment, though this may simply be due to a smaller sample size.

In order to broaden our investigations beyond the simple filament, tendril and void subsamples and beyond stellar mass and colour we also look to now incorporate some of the additional information available within the GAMA database, i.e., pairings and groups. The environmental types examined in this chapter are therefore: filaments, tendrils, voids (as defined in the GLSSC), galaxies in high mass halos, mid mass halos and low mass halos (defined as $M_H \geq 10^{14} M_\odot$, $10^{13} \leq M_H \leq 10^{14} M_\odot$ and $M_H \leq 10^{13} M_\odot$ respectively), galaxies not in groups, galaxies in pairs, and galaxies not in pairs. Some later figures also examine the central members of galaxy groups, as well as a small sample of morphologically classified galaxies. However, this chapter is not intended to be an exhaustive study of this topic (which is beyond the scope of this thesis) but to provide some early indication of trends worth exploring. Ultimately a detailed study of the gas content will also be required and this will commence in mid-2014 via the ASKAP DINGO (Meyer, 2009) survey.

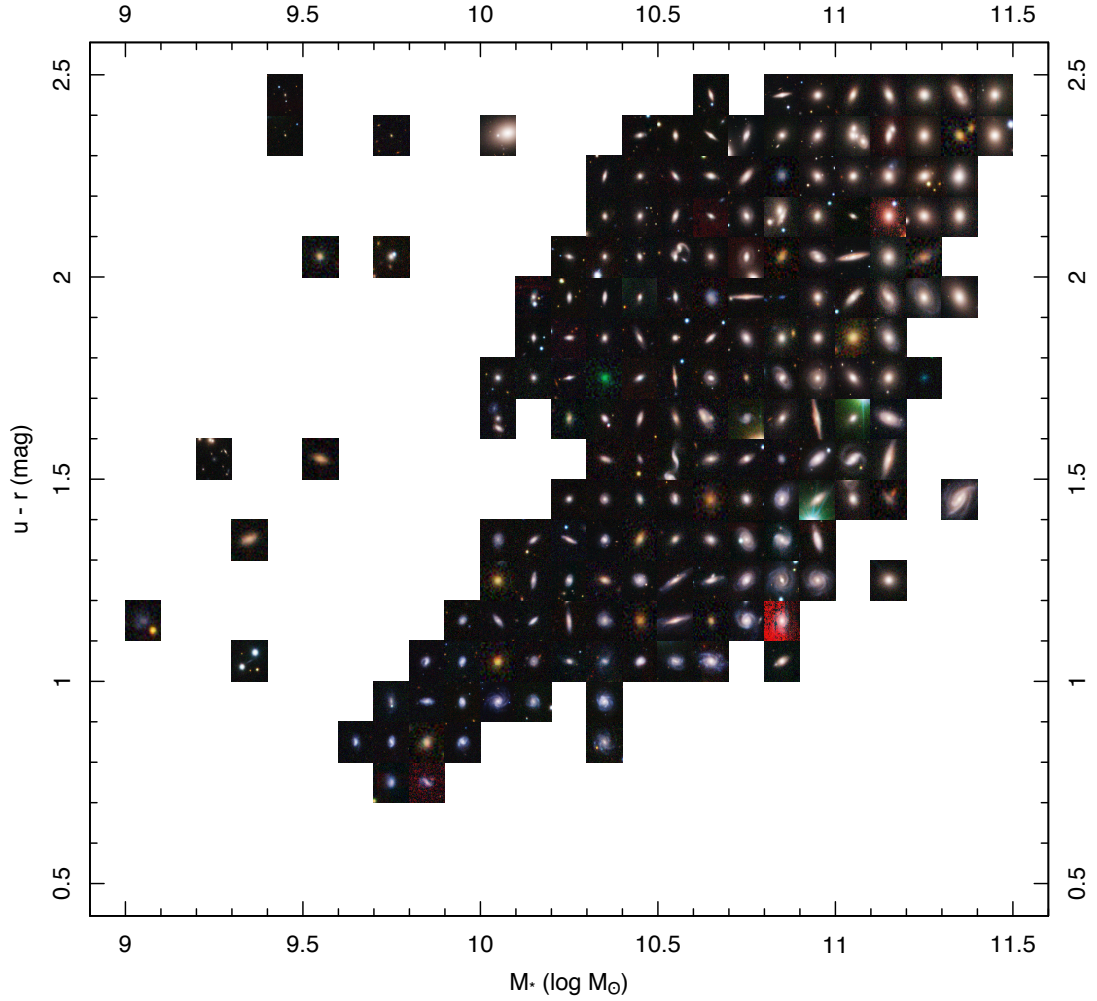


Figure 5.1: Postage stamps of randomly selected galaxies in filaments, placed according to their stellar mass and $u - r$ colour. All postage stamps measure $500 \times 500 \text{ h}^{-1} \text{ kpc}$. The postage stamps are three colour images generated from SDSS and UKIDSS *Hig* bands. The single overly red stamp towards the bottom right is due to bad photometry in the u -band.

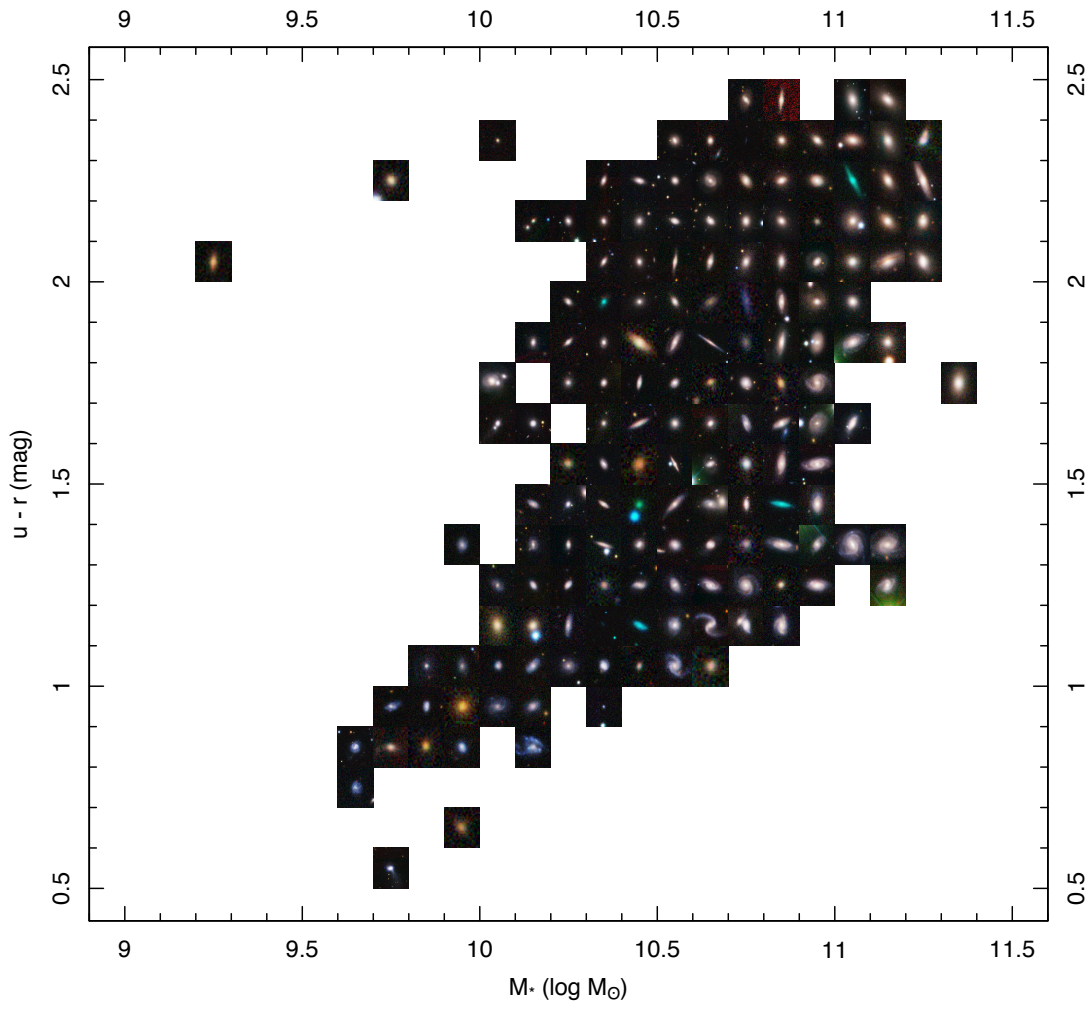


Figure 5.2: Postage stamps of randomly selected galaxies in tendrils, as in Figure 5.1.

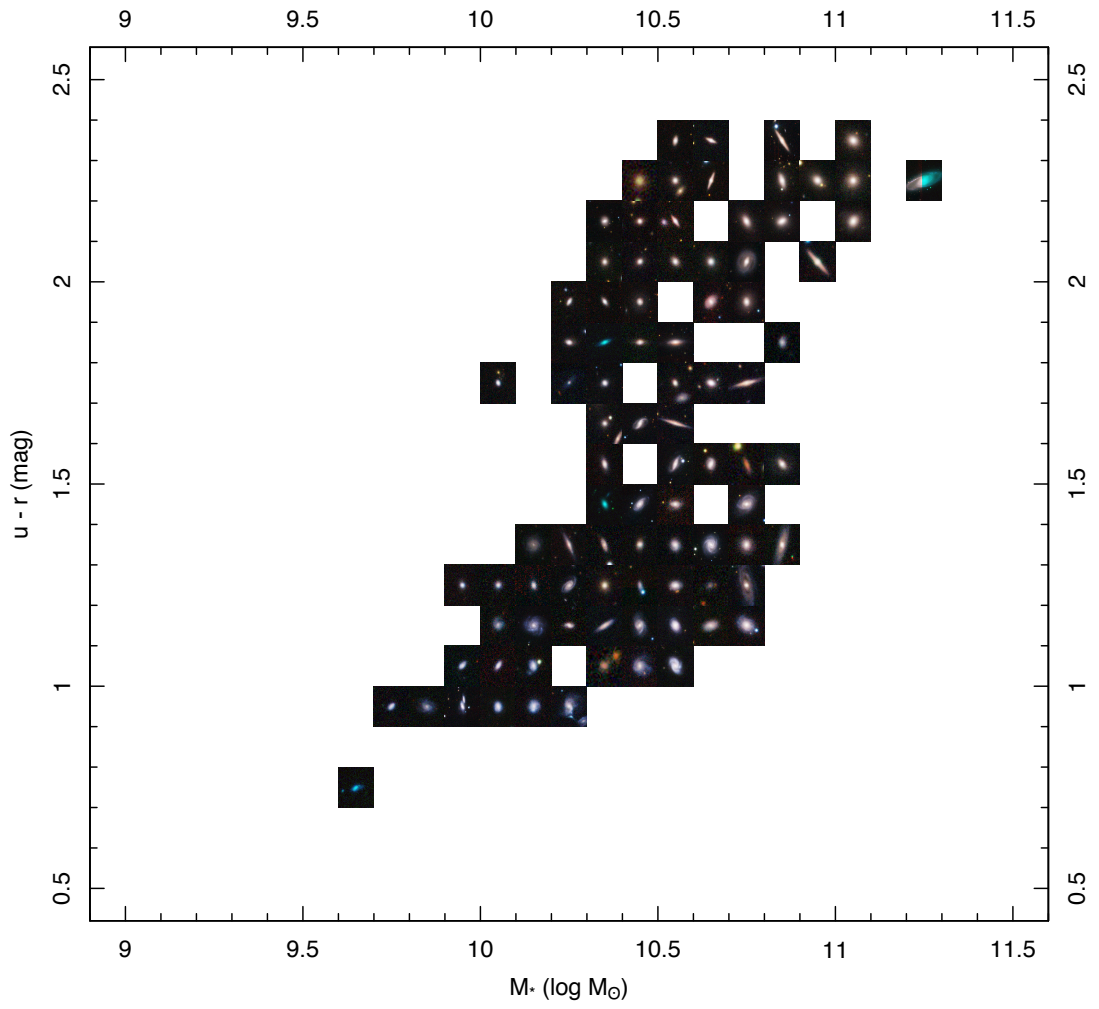


Figure 5.3: Postage stamps of randomly selected galaxies in voids, as in Figure 5.1.

5.0.1 GAMA catalogues

The work described in this chapter relies on a vast number of GAMA catalogues, available in the GAMA database or through individual GAMA members that have shared these catalogues (referred to internally as data management units, or DMUs). They are as follows:

- **FilamentFindingv01** (Alpaslan et al., 2014): The GAMA Large Scale Structure Catalogue.
- **ApMatchedCatv03** (Liske et al., in prep): Aperture matched r -band photometry derived using SEXTRACTOR in $ugrizYHJK$ bands for GAMA II galaxies. A summary of how this catalogue was prepared is provided in the appendix.
- **StellarMassv08** (Taylor et al., 2011): This catalogue provides stellar masses, rest-frame and extinction corrected photometry and colours, and other related stellar population parameters for all $z < 0.65$ galaxies in the GAMA I sample.
- **GroupFindingv02** (Robotham et al., 2011): This is the GAMA Galaxy Group Catalogue and provides grouping information for GAMA II galaxies in the equatorial regions.
- **SpecLineSFRv04** (Gunawardhana et al., 2013): This catalogue provides emission line measurements from AAT and derived star formation rates for GAMA I galaxies.
- **SersicPhotometryv07**: (Kelvin et al., 2012): This catalogue provides single-component Sérsic fits to GAMA I galaxies.
- **GasMetallicitiesv01**: (Maritza Lara-Lopez, priv. comm.): This catalogue contains estimates of gas metallicities derived from line ratios for GAMA I galaxies.
- **MultiBandPhotomv01**: (Simon Driver, priv. comm.): This catalogue collects photometry for GAMA II galaxies in 20 bands (FUV , NUV , $ugriZYJHK$, $W1-4$, $SPIRE$ 100, 160, 250, 350, 500).
- **VisualMorphologyv01**: (Rebecca Lange & Simon Driver, priv. comm.): This catalogue contains eyeballed classifications of GAMA I galaxies with $0.013 > z > 0.1$ as being elliptical or non-elliptical from colour images.

5.1 Sub-sample selection and mass normalisation

To ensure that any trends of galaxy properties are free from underlying mass selection effects, it is important to define a mass and (ideally) volume limited sample of galaxies. Taking the GAMA stellar masses from Taylor et al. (2011), Figure 5.4 shows the redshift-stellar mass distribution for all GAMA galaxies. From this, a sub-sample of galaxies with $z \leq 0.1$ and $M_* \geq 10^{9.5} M_\odot$ is chosen for the analysis shown in this chapter. These limits in redshift and stellar mass are chosen as many GAMA DMUs I wish to use are limited to $z \leq 0.1$ (such as VisualMorphologyv01). Samples with higher upper limits for redshift and stellar mass were also tested, but yielded less statistically significant results; using a higher redshift cut means that one moves into a narrow and high mass limited sample where galaxies become very unimodal (i.e. all bright systems are mostly early-type E/S0a regardless of environment). The high-mass end-points of galaxy evolution tend to be the same regardless of environment.

Note that the GLSSC is only mass complete to $M_* \geq 10^{10.61} M_\odot$. This means that for the low mass sample, it is necessary to include galaxies that are not originally analysed by the filament finding algorithm. This is done by associating each galaxy with $z \leq 0.1$ and $M_* \geq 10^{9.5} M_\odot$ not in the GLSSC as being in the same environment type as their nearest neighbour that is in the GLSSC. For filaments and tendrils, 40% of the low mass sample galaxies are already in the GLSSC. Only 4% of low mass void galaxies are also in the GLSSC galaxy sample. However, as explained in the previous chapter, the definition of void galaxies in the GLSSC is very robust, so it is likely that any galaxy near a pre-defined void galaxy is also within a void. The two point correlation function of this sub-population of void galaxies shows a similar lack of structure as in Figure 4.9.

Having defined a mass limited sample, I now look at the Galaxy Stellar Masss Function (GSMF, Baldry et al., 2012), this is the mass equivalent of the galaxy luminosity function. It gives the effective number of galaxies per unit volume in the stellar mass interval $M + dM$ in the form of a Schechter function (Schechter, 1976), and is typically shown in logarithm space. The GSMF is given by the following expression:

$$\Phi(M) dM = \ln(10) \phi^* 10^{(M-M^*)(\alpha+1)} \exp(-10^{(M-M^*)}) dM \quad (5.1)$$

Figure 5.5 shows stellar mass functions for galaxies selected from distinct environments

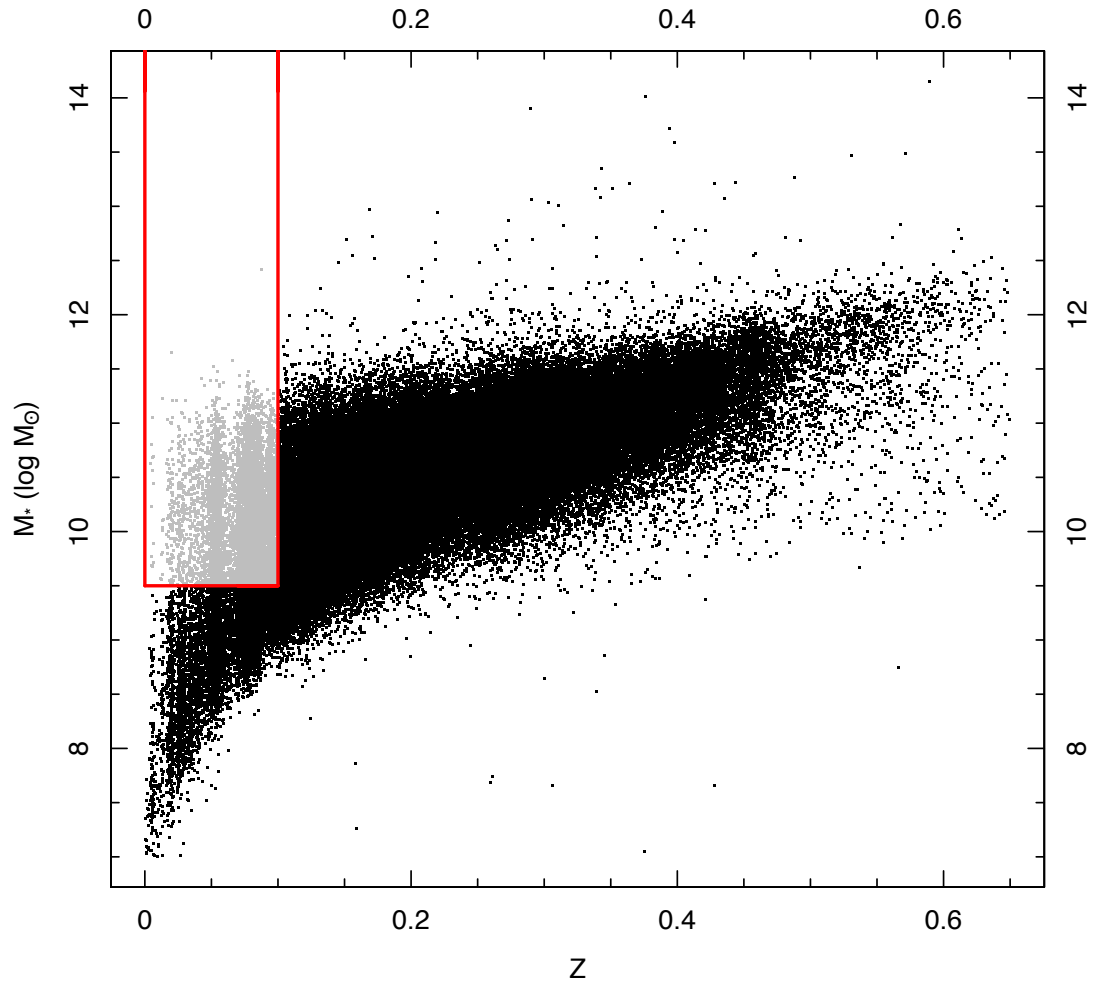


Figure 5.4: Stellar mass as a function of redshift for all GAMA I galaxies with $z \leq 0.1$ and $M_* \geq 10^{9.5} M_\odot$ coloured in grey. Redshift and mass limits are shown in red lines.

(as indicated in the inset panel). The area shaded in grey denotes the mass range where filament, tendrill and void galaxy samples have been assigned by association as described at the start of this chapter rather than via the formal filament, tendrill, void analysis presented in earlier chapters; all other populations are unaffected at this low mass range. To ensure that the number densities plotted are completeness-corrected, *StellarMassesv08* provides limiting redshifts for each object. This redshift z_{max} is defined as the maximum redshift at which each galaxy would be detected given the magnitude limits of GAMA I: $m_r < 19.4$ mag for the G09 and G15 fields and $m_r < 19.8$ mag. These values are derived by Taylor et al. (2011) by taking the single best fitting model spectrum for each galaxy and studying how the k-corrected r -band magnitude varies with redshift. From z_{max} , one can calculate the effective comoving survey volume V_{max} in cubic megaparsecs for that particular galaxy and weight its number count appropriately. It is only necessary to apply V_{max} corrections to the filament, tendrill and void galaxy populations, as all other samples are complete. The number densities shown in Figure 5.5 have all been V_{max} corrected and show the different mass regimes that different environments span. Note that for the vast majority of galaxies the V_{max} correction is identical as the sample, except for some systems with extreme colours (mass-to-light ratios), is volume selected: the V_{max} uncorrected stellar mass function for all galaxies is statistically indistinguishable from the one shown in Figure 5.5. The errors shown are $\sqrt{(N)}$ Poisson noise errors.

From Figure 5.5 it is easily visible that different environments contain galaxies with very different stellar masses. Void galaxies are the most strikingly different population, containing far fewer massive galaxies than all other distributions. As expected, high mass halos are dominated by high mass galaxies. This variation in stellar mass as a function of environment is likely driven by the expected higher-merger rates in the higher-density environments. A clear progression is seen from voids to tendrills to filaments and a smaller but significant progression from ungrouped to low to mid mass groups. This potentially suggests as well as local density (Elbaz et al., 2007) and halo mass (Reddick et al., 2013; Cen, 2013), large scale structure is also an important driver of stellar mass evolution.

The fit parameters for the stellar mass functions are given in Table 5.1. The errors in the parameters are estimated via a jackknife resampling:

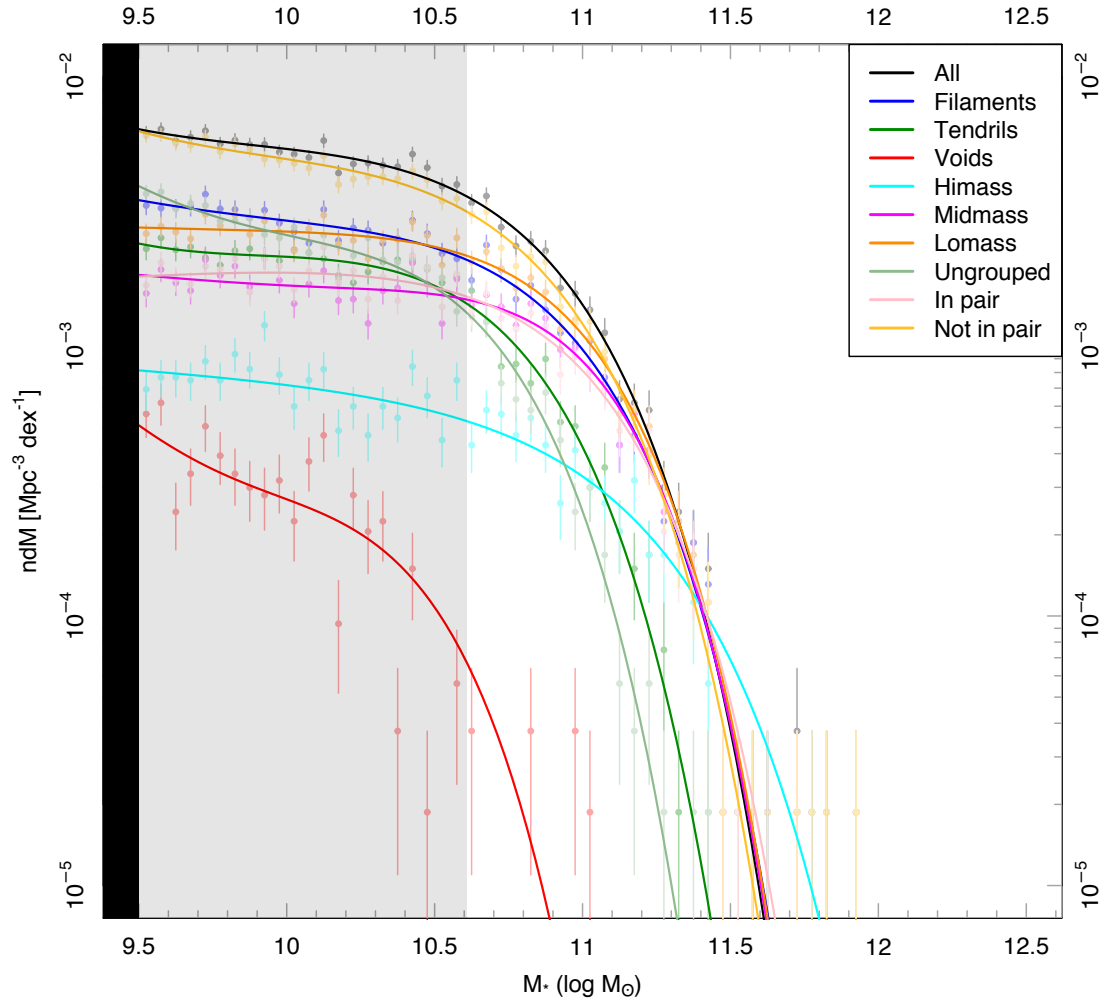


Figure 5.5: V_{\max} adjusted stellar mass functions of galaxies in different environments by 0.1 dex bins in log stellar mass. The region shaded in grey marks the mass range at which the galaxies selected for filaments, tendrils and voids are below the mass completeness limit of the GLSSC. The black shaded region denotes the lower mass limit for all environments.

5.1. Sub-sample selection and mass normalisation

	M^*	α_1	$\phi_1^*/10^{-3}$	α_2	$\phi_2^*/10^{-3}$	χ^2/ν	$\int \Phi(M) dM$
All	10.75 ± 0.04	-0.85 ± 0.08	3.30 ± 0.27	-2.00 ± 0.01	0.04 ± 0.09	1.12	100%
Filaments	10.79 ± 0.04	-1.48 ± 1.40	0.22 ± 2.41	-0.69 ± 1.25	0.002 ± 0.002	1.17	57%
Tendrils	10.59 ± 0.06	-0.68 ± 1.81	1.72 ± 2.26	-2 ± 1.85	0.03 ± 2.30	1.07	39%
Voids	10.13 ± 16.94	-2 ± 4.78	0.04 ± 6790	-0.19 ± -3.57	0.23 ± 0.56	2.52	4%
HiMass	11.16 ± 0.10	-1.06 ± 1.51	0 ± 0.05	-1.07 ± -0.69	0.28 ± 0.47	1.33	10%
MidMass	10.72 ± 0.017	0	1.08 ± 0.09	-1.13 ± 0.04	0.55 ± 0.05	1.42	24%
LoMass	10.74 ± 0.05	-0.19 ± 0.85	1.27 ± 0.44	-1.06 ± 0.97	0.96 ± 0.52	1.13	36%
Ungrp	10.45 ± 0.09	-0.49 ± 1.47	1.92 ± 1.72	-1.89 ± 1.31	0.18 ± 1.71	1.32	30%
inPair	10.79 ± 0.09	-0.92 ± 2.36	1.03 ± 1.64	0 ± 2.91	0.43 ± 1.59	1.35	30%
notinPair	10.74 ± 0.04	-0.85 ± 1.09	2.91 ± 2.72	-1.99 ± 1.33	0.05 ± 2.75	1.01	70%

Table 5.1: Parameters for the double Schechter function GSMF fits shown in Figure 5.5. The columns are the shared knee in the Schechter function (M^*), the primary slope at the faint end of the function (α_1), the primary normalisation for the fit (ϕ_1 ; in units of $\text{dex}^{-1} \text{Mpc}^{-3}$), the secondary slope at the faint end (α_2), and the secondary normalisation (ϕ_2), and the goodness of fit χ^2 . The fractional integrated stellar mass is also shown, for subdivisions of environment (marked by the horizontal lines in the table). Errors are estimated from jackknifed resampling.

$$\sigma^2 = \frac{N-1}{N} \sum_{i=1}^N (x_j - x)^2 \quad (5.2)$$

where x is the best fit parameter and x_j is the best fit parameter from a jackknife resampled variant of the data set, and N is the number of jackknife volumes (set to 10). While most of the fits are satisfactory, the fit for the void GSMF has notably large uncertainties. These functions also give integrated mass densities $\int \Phi(M) dM$, which can be used to see which environment contains the most stellar mass. These are calculated such that they add up to 100% for distinct subdivisions: filaments, tendrils, and voids; high mass, mid mass, low mass halos and ungrouped galaxies; and pairs and unpaired galaxies.

Given the extreme variation in the stellar mass distributions it is obvious that comparison of other properties may be driven more by stellar mass rather than the environmental marker selected. In order to remove the role of stellar mass it is possible to mass normalise our samples to tease out the specific influence of environment alone. By resampling the various galaxy populations so that for a given stellar mass bin they match the void galaxy stellar mass function, one is left with populations of galaxies that are not only mass complete, but have observed mass distributions that are matched. This means that for this sub-sample, all comparisons between the galaxy populations in different environments will be free from mass bias and will not simply be recasting the differences in the GSMF. Note however that matching to the void mass function means that the results shown in this chapter are confined to a small range of stellar masses.

For this chapter, unless otherwise specified, all figures and results are for the sample shown

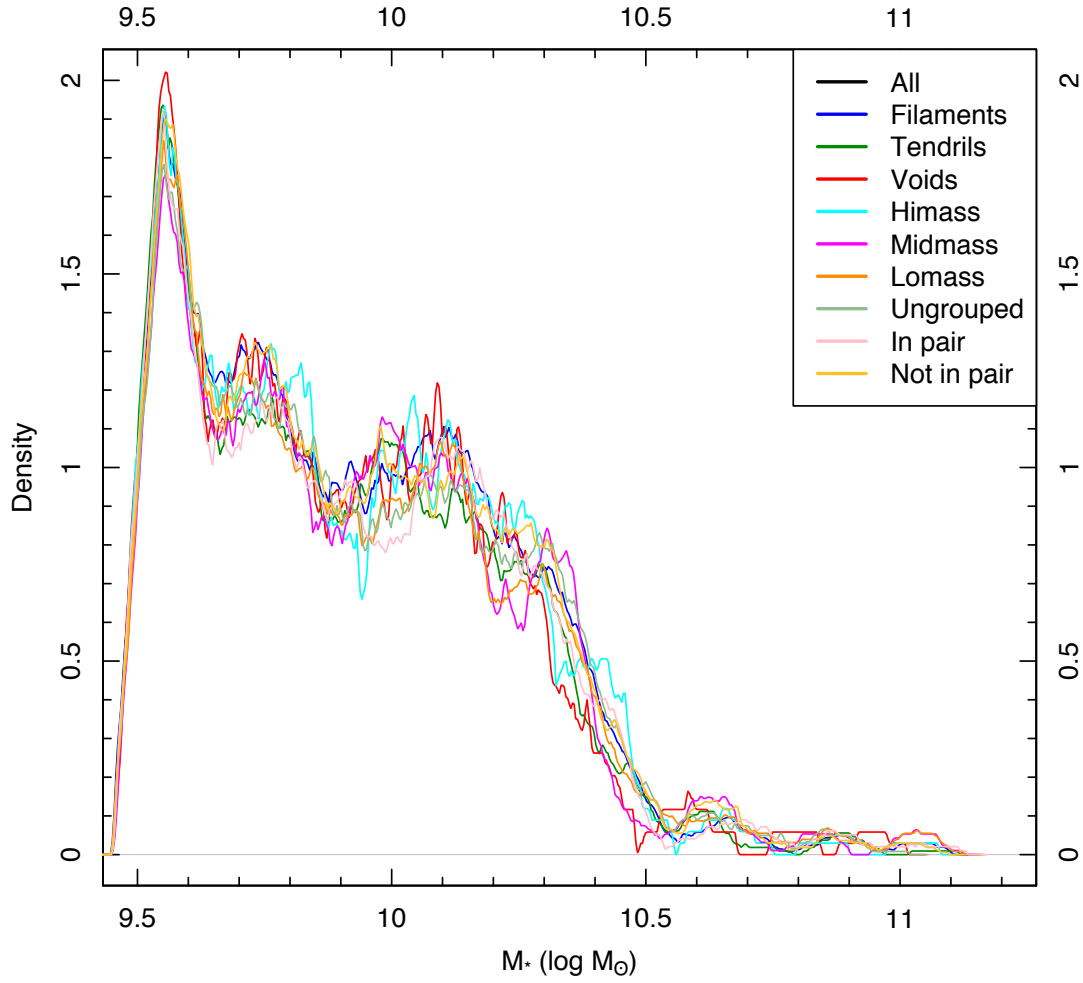


Figure 5.6: Kernel density estimates for stellar masses in mass normalised galaxy populations (colour coded in the same way as Figure 5.5).

in grey in Figure 5.4; i.e. $z \leq 0.1$ and $M_* \geq 10^{9.5} M_\odot$. If a sample is described to have been ‘mass normalised,’ galaxies in different populations will have been mass normalised to have the same mass distribution as void galaxies, as described above. The results of the mass normalisation process are shown in Figure 5.6, where the kernel density estimates for stellar masses in mass normalised populations are shown.

5.2 The impact of large scale structure on galaxy properties

Kreckel et al. (2012) show that for a sample of galaxies in voids, taken from the Void Galaxy Survey (VGS), there is a tendency for galaxies in voids to be bluer and fainter than their

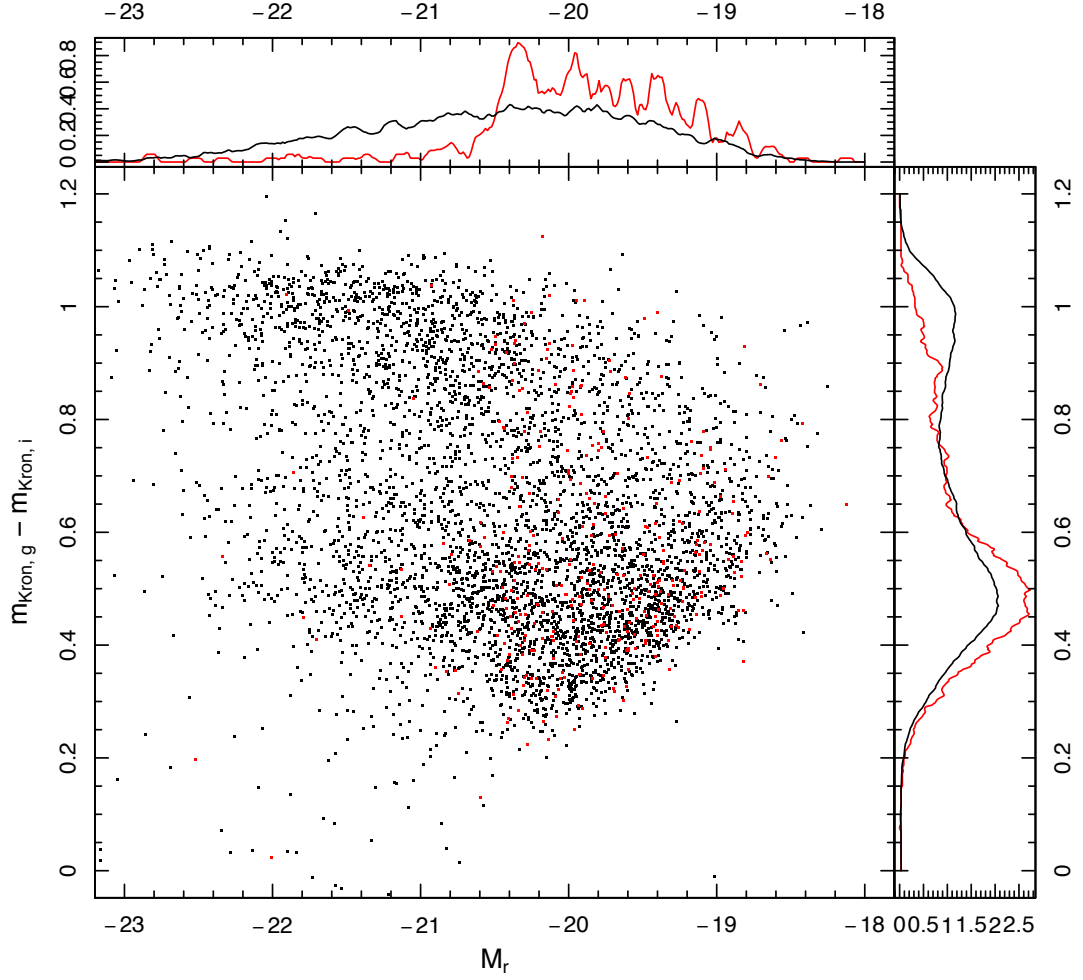


Figure 5.7: $u - r$ colour plotted against absolute r -band magnitude for mass normalised galaxies in filaments and voids (black and red respectively). The distributions on the right and top of the figure show distributions in colour and absolute magnitude respectively. Note that colours are extinction and restframe corrected. This data has been taken from *StellarMassesv08*.

counterparts in overdense regions. This result is shown to hold for the mass normalised subsample of galaxies in the GLSSC in Figure 5.7, where it is evident that galaxies in voids are significantly less bright and bluer than their counterparts in filaments. The colour bimodality is no longer present in void galaxies, which show an overwhelming tendency to be blue.

5.2.1 Broad 2D distributions

A wider comparison of various parameters can be seen in Figure 5.8 where effective radius, ellipticity, absolute magnitude, $u - r$ colour, and r -band Sérsic index are all compared against

each other for galaxies in filaments, tendrils, and voids, which have been mass normalised. Sections of this figure where observables are plotted against each other are not as informative as the one dimensional distributions for individual parameters. Most two dimensional trends and correlations between observables are reflected across all three populations of galaxies. That differences exist implies that large scale structure does have some effect on the properties of a galaxy, however there are no cases where a particular population of galaxies is the only population taking up a certain area of the parameter space. This figure strengthens the case for void galaxies being marginally later, fainter, and bluer. The properties of tendril galaxies are more similar to those of filament galaxies, which is an indication that our tendrils are indeed distinct structures more closely related to filaments and that large scale structure trends only become evident at the extreme void environment. Coupled with the stellar mass distributions seen in Figure 5.5 the suggestion might be that while the mechanisms and processes occurring in filaments and tendrils drive towards the same properties, the process is more accelerated in filaments (because of the advanced mass function).

This figure provides some insight into the ‘typical’ galaxy that resides in a filament or a void. Galaxy properties most affected by large scale structure appear to be absolute r -band magnitude, $u - r$ colour, Sérsic index and ellipticity. A galaxy in a void tends to be bluer, fainter, more ellipsoidal (flattened) and with a lower Sérsic index than a galaxy within a filament, or tendril. The difference between filaments and tendrils is a lot more subtle; this may be an indication that these parameters are sensitive to galaxy growth in extremely isolated environments.

5.2.2 Metallicity

Figure 5.9 displays the mass-metallicity relation for galaxies as a function of large scale structure, showing that all three populations have similar metallicities, with voids perhaps showing slightly less chemical enrichment, but this is difficult to establish given the noise in the data. The metallicity is taken from GasMetallicitiesv01 where it is calculated using the O3N2 index, which is defined as:

$$\text{O3N2} = \log \left(\frac{[\text{OIII}]\lambda 5007 / \text{H}\beta}{[\text{NII}]\lambda 6583 / \text{H}\alpha} \right) \quad (5.3)$$

from which it is possible to calculate the O/H metallicity indicator as $[12 + \log(\text{O}/\text{H}) = 8.73 -$

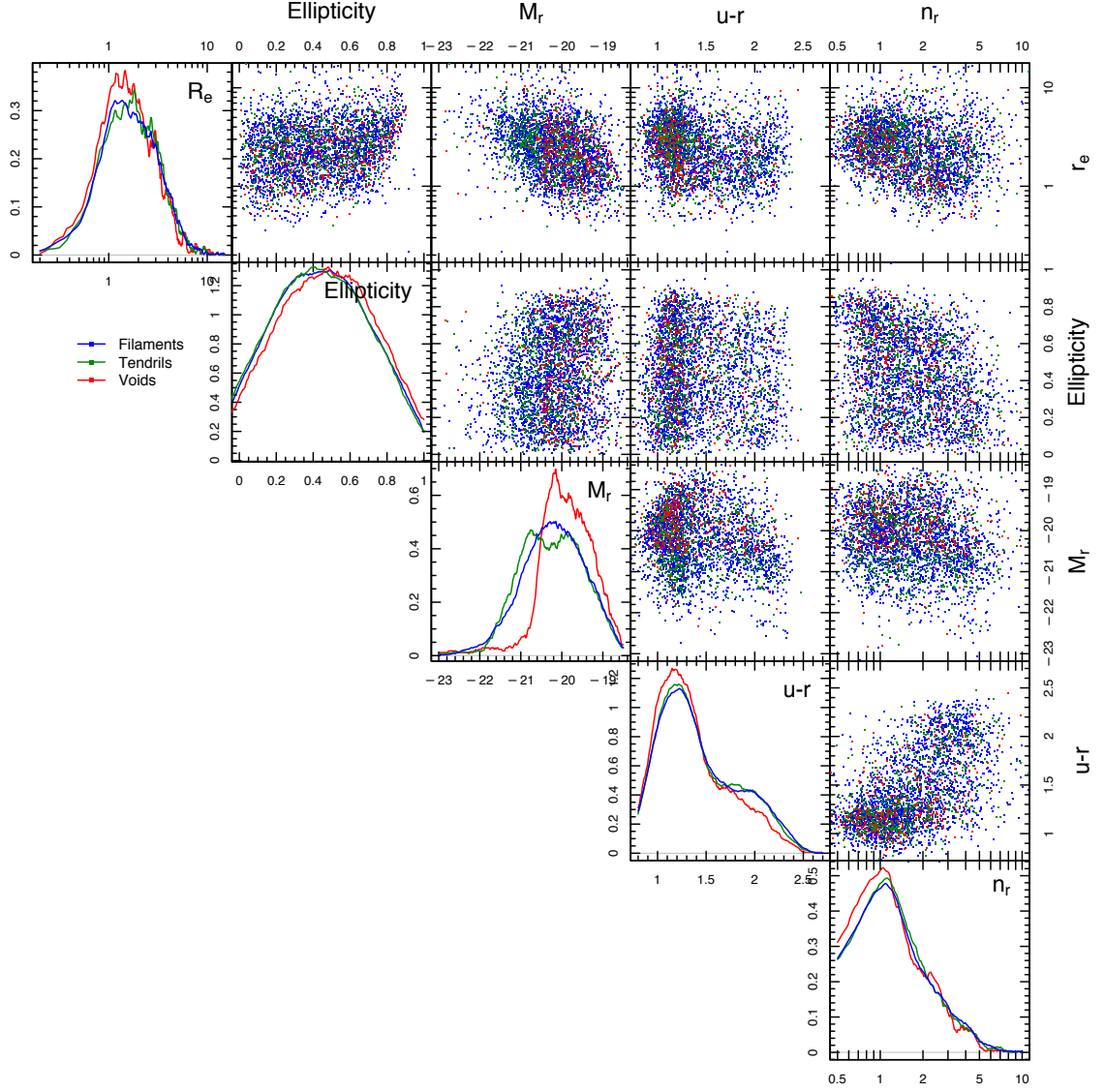


Figure 5.8: Comparisons of effective radius, ellipticity, absolute magnitude, $u - r$ colour, and r -band Sérsic index for mass normalised galaxies in filaments, tendrils and voids (shown as blue, green and red lines and points respectively). Each frame of this large mosaic plots two of these parameters against each other, while the histograms show the distributions of each one individually. Note that the most distinguishing differences are in these histograms. The data shown in this figure are taken from *SersicPhotometryv07*.

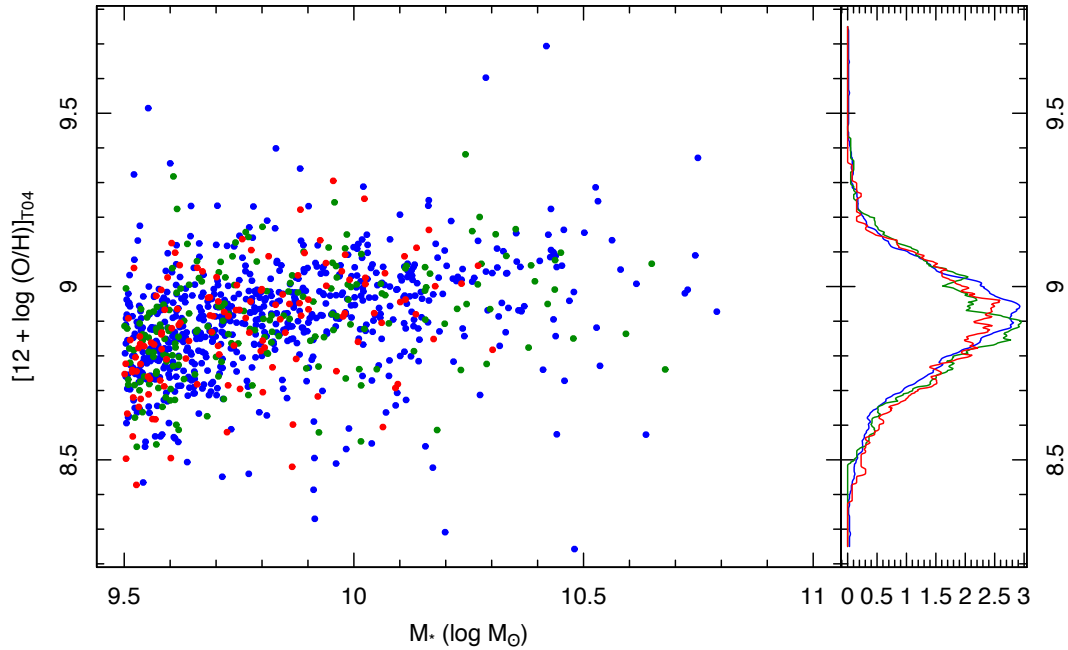


Figure 5.9: Metallicities and stellar masses from GasMetallicitiesv01 and StellarMassesv08 shown for mass normalised filament, tendril and void galaxies in blue, green and red respectively.

$0.32 \times \text{O3N2}$ as per the prescription of Pettini & Pagel (2004). Line strengths are emission measurements and have been corrected for dust attenuation; and this estimator does require that there be emission in This estimator is selected in order to be able to compare GAMA metallicities to provide the best comparison to SDSS metallicities (Lara-Lopez, priv. comm.).

5.2.3 1D distributions

Figure 5.10 shows the distribution of effective radius, ellipticity, r -band absolute magnitude, $u - r$ colour, and Sérsic index for galaxies split by three different types of environment: large scale structure, group mass (and group membership), and presence in a pair. There is a tendency for all galaxies outside of dynamically bound structures (i.e. in voids, or not in groups, or not in pairs) to exhibit similar characteristics: lower luminosity, bluer colours, lower Sérsic index, and higher ellipticity. This figure also shows that halo mass is not a very good predictor of morphology (as both ellipticity and Sérsic index do not vary much with halo mass) and that a galaxy's presence within or without a group is more important in determining its shape. Halo mass impacts colour, but presence within or without a halo has a much more prominent

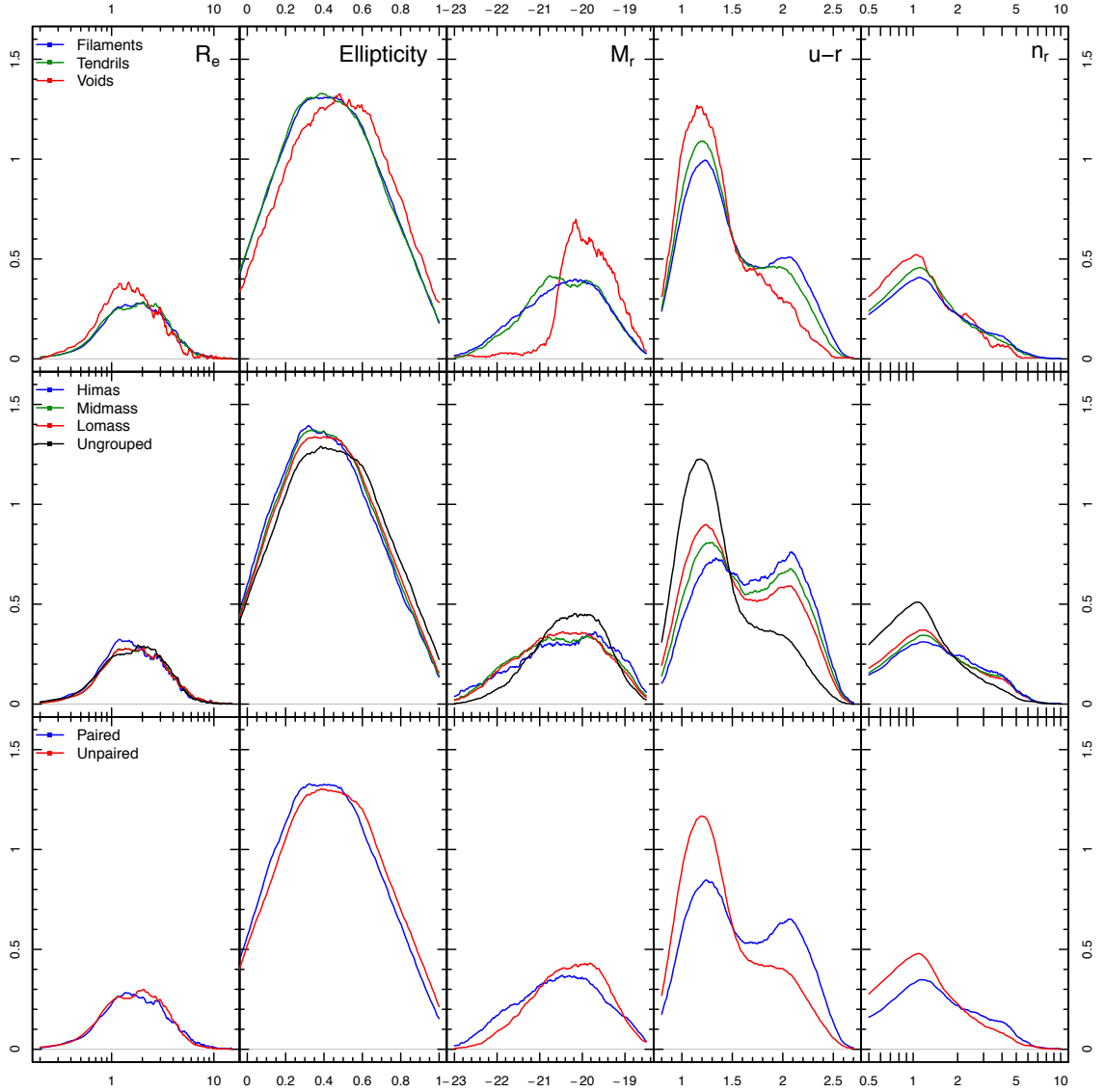


Figure 5.10: Effective radius, ellipticity, r -band absolute magnitude, $u-r$ colour, and Sérsic index from *SersicPhotometryv07* for galaxies split by three different environment types: large scale structure (top row), halo mass and galaxies not in groups (second row), and paired and unpaired galaxies (bottom row). There is an overall trend for galaxies outside of structure (void galaxies, ungrouped galaxies, and unpaired galaxies) to show similar characteristics.

impact on colour. Colour is also the only property where a difference between filament and tendrill galaxies is significant; for all other properties both populations of galaxies show the same distribution, with only voids differing significantly.

An interesting feature of Figure 5.10 is that the variation of galaxy properties isn't very susceptible to halo mass, as shown by the coloured lines in the second row of the figure. This is in agreement with recent work, e.g. Lopes et al. (2013), who find that halo mass (for haloes $10^{13} \leq M_H \leq 10^{15} M_\odot$) has no impact on the colour, spectral type, and concentration of a galaxy, and that the dominant process that impacts these parameters is local density; however,

their analysis does not contain a similar mass normalisation approach as the one used in this thesis. The authors argue that most of the changes to a galaxy interacting with a halo occur as it falls into the halo. Valentinuzzi et al. (2011) show that the red sequence of galaxies within a cluster do not depend on halo mass, but instead on local density. These results are at odds with more recent work by Velandier et al. (2014) who measure the relation between the halo mass and baryonic mass of weak lensing galaxies and show that for red galaxies, the halo mass increases with baryonic mass at a higher rate than for blue galaxies.

5.2.4 Elliptical fraction

The left panel of Figure 5.11 shows elliptical fractions as a function of halo mass for a set of eyeballed GAMA galaxies with $z \leq 0.1$, taken from `VisualMorphologyv01`. The trend is for the fraction of elliptical galaxies within bins of halo mass to rise as halo mass rises, until roughly $3 \times 10^{13} M_{\odot}$ where the trend stops and the elliptical fraction stays constant, implying a limit to the impact of halo mass on the morphology of galaxies. The right hand panel of this figure shows the elliptical fraction for mass normalised galaxies in variety of environments and shows that filaments have higher elliptical fractions compared to voids, which have a lower fraction of elliptical galaxies compared to any other environment type. Galaxies are much more likely to be ellipticals if they are in a pair, and the fraction of ellipticals rises as a function of halo mass; this echoes the result shown in the left hand panel. Filaments have a higher fraction of elliptical galaxies than the entirety of the $z \leq 0.1$, $M_* \geq 10^{9.5} M_{\odot}$ sample. The errors for each point on both panels show confidence estimates on population proportions derived from the Beta function, as described in Cameron (2011).

5.2.5 Total energy (CSED)

Figure 5.12 displays the total flux densities for a sub-sample of mass normalised galaxies in filaments, tendrils and voids across 20 bands, taken from `MultiBandPhotomv01`. For each filter, galaxies with flux detections in all bands (in other words, only galaxies in areas of the sky that have coverage in all 20 bands) have their fluxes summed, and their errors summed in quadrature. The resulting data shows the cosmic spectral energy distribution (CSED; Driver et al., 2012) of each environment type. The CSED describes the energy of photons present within a specific volume of space and provides a broad overview of the properties of galaxies that contribute to this energy at a given epoch. All points are normalised by the K -band and

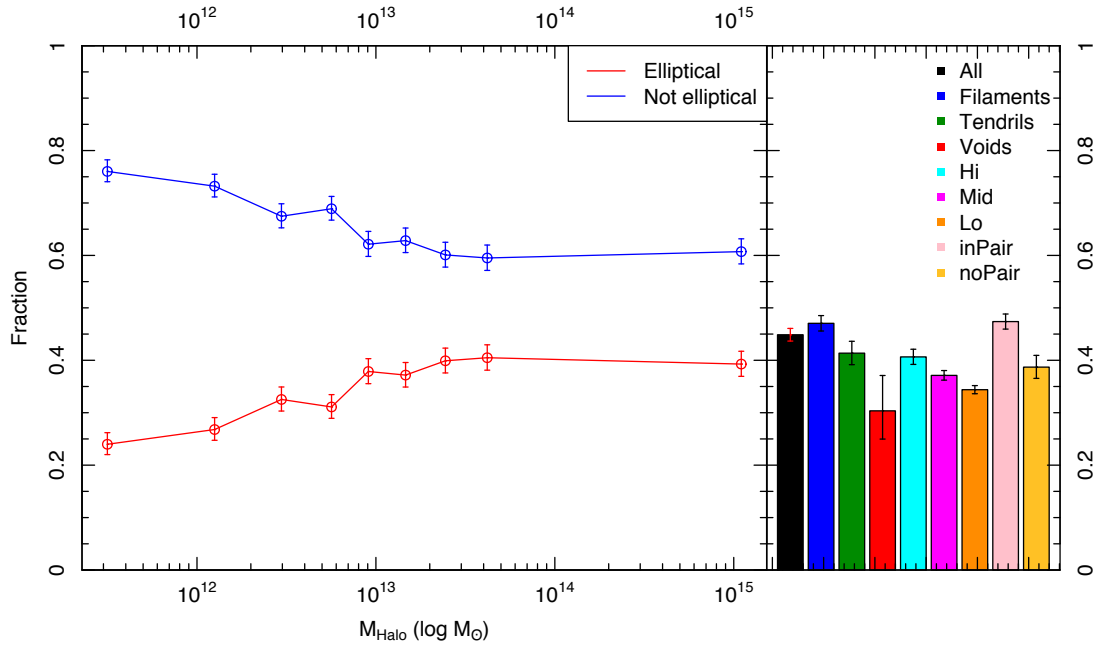


Figure 5.11: Elliptical fraction of all galaxies in VisualMorphologyv01 with $M_* \geq 10^{9.5} M_{\odot}$ as a function of halo mass shown in red. The fraction of non-elliptical galaxies is also shown, in blue (note that this is just the $1 -$ the elliptical fraction). The fraction of ellipticals rises steadily as a function of halo mass until approximately $3 \times 10^{13} M_{\odot}$, after which it remains relatively stable. Note that elliptical fractions are calculated for halo mass bins of equal size. The right hand panel shows the fraction of elliptical galaxies in a variety of environments. The error bars show 1σ uncertainties about the population fraction.

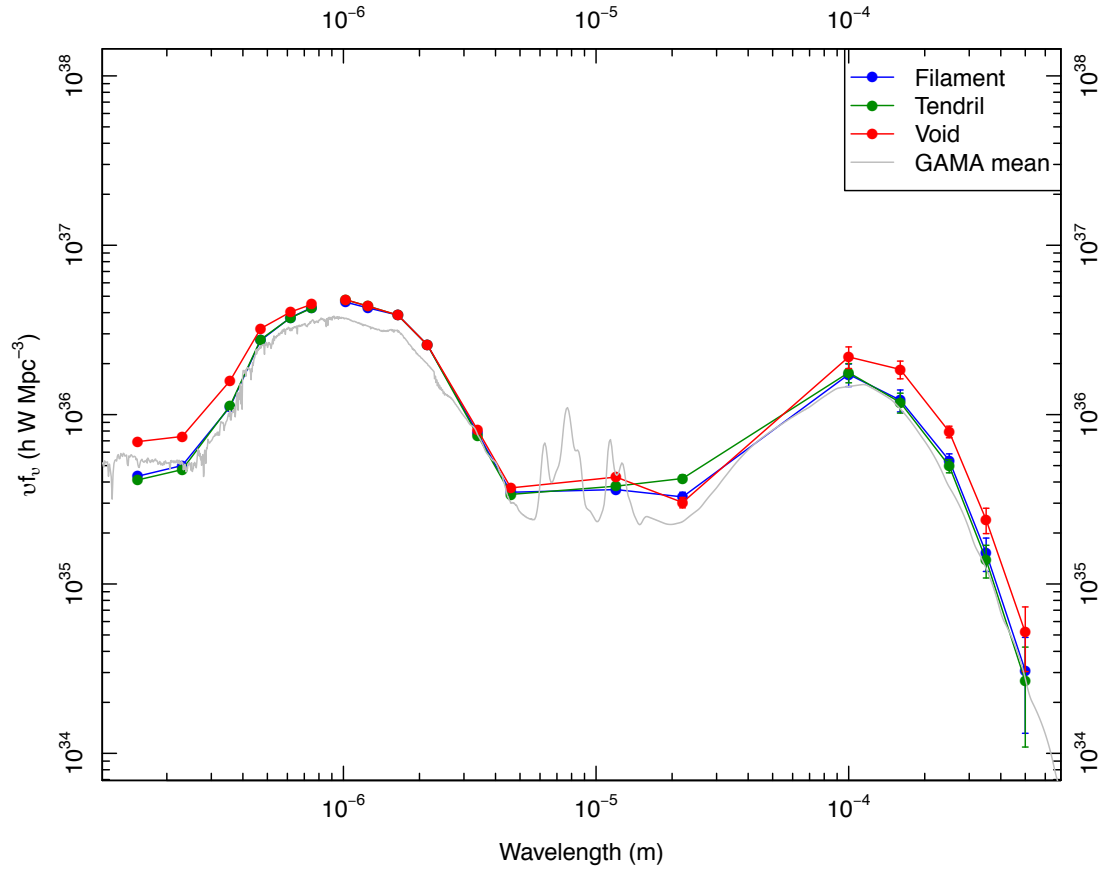


Figure 5.12: Total flux densities in 20 bands for mass normalised galaxies in filaments, tendrils, and voids, shown in blue, green and red respectively. These flux values are taken from the MultiBandPhotomv01 catalogue. The CSED shown in grey has been taken from Driver et al. (2012) and represents the mean GAMA energy output, scaled to lie above the points.

plotted. The grey CSED above the points is taken from (Driver et al., 2012) and represents the mean GAMA energy output, shown here for reference. It has been artificially scaled to lie above the points. This figure shows a slight increase in flux in *UV* regime for void galaxies, as well as slightly higher emission in the far infrared. This again suggests that the evolutionary processes in tendrils and filaments are similar; one of star-formation suppression (quenching) when compared to void systems.

5.3 Summary of results

Having access to such a complete catalogue of galaxies in pairs, groups, and large scale structure, as well as such a wide variety of uniformly generated catalogues of galaxy properties allows this analysis to be very broad. These results build on the existing body of knowledge

of galaxy environments, namely that galaxies are susceptible to their local environments. This chapter has looked for correlations in a wide variety of galaxy properties as a function of whether or not the galaxy is in a pair, group, void, tendril, or filament.

Galaxies in different environments show varying stellar mass functions, as shown in Figure 5.5, with void galaxies hosting a distinctly low stellar mass population. For this reason, in this chapter all comparisons of galaxy properties are made for a mass complete subset of galaxies with $z \leq 0.1$ and $M_* \geq 10^{9.5} M_\odot$. To further avoid any mass bias, each population of galaxies is resampled such that its mass distribution matches the mass distribution of voids.

Figure 5.8 establishes the basic properties of void galaxies: blue, low luminosity, late-type galaxies that tend to be smaller than filament and tendril galaxies, agreeing well with existing work (e.g. Kreckel et al., 2012; Pan et al., 2012). Filament galaxies contain the largest fraction of bright, red, early-type galaxies (this is expected, as filaments are the most dense environments in the Universe and we know that the brightest galaxies tend to be more clustered Zehavi et al., 2011) while tendril galaxies straddle between both populations, tending to be more similar to filament galaxies. The total CSED of galaxies, shown in Figure 5.12 shows a higher flux in the UV bands for void galaxies. Metallicity does not appear to be greatly affected by large scale structure (Figure 5.9).

Beyond local density and stellar mass, the greatest discriminant of galaxy evolution appears to be if a galaxy is part of a dynamically bound structure or not (a pair, a group, or a filament), as shown in Figure 5.10. These isolated galaxies are more likely to be blue, faint, and are more flattened (i.e. higher ellipticity) and have lower Sérsic index values. From Figures 5.11 and 5.10 one can estimate that the population of void galaxies appears to have the most extreme values; or in other words, selecting a galaxy from a void is more likely to select the most blue, faint, late-type galaxy compared to any other environment type.

The results in this chapter agree with recent work which asserts that the mass of the halo that a group resides in has little impact on the luminosity and shape of the galaxies that reside within it (for similar results, see Valentinuzzi et al., 2011; Lopes et al., 2013; Lacerna et al., 2013). This is particularly evident in Figure 5.10, where the variation in M_r and n_r vary least as a function of halo mass; a conclusion which is perhaps somewhat unexpected, and difficult to reconcile with similar studies done using weak lensing (Velandier et al., 2014). One important caveat of the work in this chapter is that the total CSED and many other DMUs on

which the results in this chapter are based on are still very preliminary and are likely to evolve, impacting on these quick look results. However, the figures indicate the merit of future studies in this field, especially now that the robust group and large scale structure catalogues are in place.

The preliminary analysis in this chapter shows some tentative trends of galaxy properties with large scale environment. While stellar mass is known to correlate most strongly with galaxy properties, halos and filaments are likely to be influencing galaxies during their infall into such structures. Moreover, large scale structure appears to drive stellar mass, and so may have some secondary impact on galaxy evolution via the mechanism driving mass growth in such structures (such as mergers). The role of large scale structure in galaxy evolution will expand by exploring the role of gas in driving the properties of galaxies in different environments, which will soon be possible with the advent of large radio surveys like ASKAP, followed by the SKA. These will provide excellent opportunities to study the role of gas in voids, tendrils, and filaments; focusing particularly on how the underlying gas distribution influences the associated galaxy population. Of particular interest is whether tendrils and filaments lie atop the underlying gas density and are therefore potential tracers of the gas flow from voids to cluster centres via tendrils into filaments. For this reason a detailed analysis of the role of environment may only become clear with the inclusion of HI data.

6

Conclusions and future work

In this thesis I have described the work that I have done on investigating the large scale structure of the nearby Universe using data from the GAMA survey.

As galaxy groups form the foundation of filaments and the most dense environments of large scale structure, I began by verifying the masses of the groups in the GAMA Galaxy Group Catalogue (G^3C). This was done using the caustic mass measurement (Diaferio, 1999), which is a complementary method to the dynamical masses measurements made for the G^3C . By calibrating the mass measurement algorithm on the GAMA mock galaxy catalogues, I calculated caustic masses for galaxies in the G^3C and found there to be excellent agreement between dynamical and caustic masses. While the caustic mass method was originally developed to be used only on populous galaxy clusters with > 200 members, I was able to show that it can accurately measure the mass of a group of galaxies with down to 5 members. Both the caustic mass and dynamical mass measurements perform equally poorly when applied to groups with 2 or 3 galaxies. Caustic mass measurements in particular appear to be very susceptible to accurate velocity dispersion measurements.

Having confirmed the accurate mass measurements of galaxy groups, I produced the GAMA Large Scale Structure Catalogue, creating a new algorithm to do this in the process. This algorithm differs from existing methods in that it is designed to work with observed data, and to be mathematically and computationally simple; as well as identifying galaxies that reside in underdense regions as well as filaments. The structure finder uses a three pass approach, whereby groups of galaxies are first used to construct filaments using minimal spanning trees, followed by the association of nearby field galaxies to their nearest filament within $\sim 4.5 h^{-1}$ Mpc. All galaxies in and around filaments are then removed, and a new minimal spanning tree is run on the remaining galaxies in order to identify underlying structures formed by galaxies in the field; as well as extremely isolated galaxies that reside in voids. Comparing the large scale structure identified in the observed GAMA data to the structures found by running the algorithm on GAMA mock galaxy catalogues reveals that these simulations agree remarkably well with observations. This provides a probe of higher order distribution statistics not captured by the popularly used two-point correlation function. The GLSSC is the largest ever observational catalogue of large scale structure to date, and will be available for public use alongside the second data release of GAMA.

I have identified a new, third population of galaxies, straddling between filaments and voids, that are referred to as ‘tendrils’ galaxies. Tendrils are coherent structures formed by isolated field galaxies that span, on average, $10 h^{-1}$ Mpc, emerging from filaments and penetrating into voids, or connecting to other filaments. Tendrils are both visually morphologically distinct from filaments, and are also shown to have less linear correlation than filaments, by measuring their line correlation function (Figure 4.16, Obreschkow et al., 2013). This measurement reveals that galaxies in voids show very little line correlation, however it is unclear if deeper measurements of the same region of space would reveal further underlying, unseen structures around these void galaxies. What is certain, however, is that we must update catalogues of known voids to reflect surveys that have taken place after SDSS: examining GAMA galaxies within SDSS voids shows that even when looking at only the inner two thirds of a void, only 25% of the galaxies within it are actually considered to be void galaxies in GAMA. This is an important result that highlights the necessity of deeper, more complete surveys of low density regions, and the careful consideration of survey selection limitations when studying the properties of voids.

The final portion of this thesis is a preliminary study of the properties of galaxies as a

function of large scale structure and other large scale environments. Such a study has not been possible in the past due to the lack of sufficiently complete catalogues of the diverse structures in the Universe. To a first approximation, 72.6% of stellar mass of galaxies is contained in filaments, while 23.8% and 8.1% are contained in tendrils and void galaxies respectively. In fact, galaxies in a variety of environments have distinct galaxy stellar mass functions as shown in Figure 5.5. The CSED of void galaxies shows higher flux in UV and IR bands. By examining a set of galaxies in different environments that are mass complete and normalised to have matching mass PDFs, I show that void galaxies tend to be blue, faint, less elliptical, and have lower Sérsic indices when compared to galaxies in tendrils and filaments. Figure 5.10 shows that galaxies in different environments behave in similar ways: objects in voids, outside of groups, and outside of pairs are all overwhelmingly blue and faint, while halo mass is not a great discriminator of colour and morphology. All of these results paint the picture of a Universe where, secondary to the effects of stellar mass, local density and membership within a pair, a galaxy's presence within a large structure is more relevant to its properties than the nature of the structure itself. It is important to note that many of these results do rely on GAMA data catalogues that are preliminary and are subject to change, but do pave the way for future studies.

6.1 Future work

My current research interests are to understand the impact of large scale structure on galaxy evolution, and to further study my sample of void galaxies. I aim to continue my research by gaining access to the wealth of deeper, greater resolution photometric data currently being delivered by the VST KIDS and VISTA VIKING surveys as well as future data from the ASKAP DINGO survey and upcoming deeper GAMA catalogues. I am particularly interested in revisiting the final chapter of the thesis once these more complete catalogues are in place. Future avenues of research I am interested in are as follows:

Existing void galaxy surveys (e.g. Kreckel et al., 2012) show that void galaxies are more faint and more blue than their counterparts in more dense environments. For the final chapter of my Ph.D. thesis, I am exploring the effects of large scale structure on properties such as luminosity, Sérsic index, specific star formation rate, and colour. Using this deeper data and combining it with existing work such as the GAMA Group Catalogue (Robotham et al., 2011), I will be able to determine if it is the nearby halo environment, or large scale structure that has

the greatest effect on galaxy properties. Figure 5.12 is a preliminary calculation of the cosmic spectral energy density (CSED; Driver et al., 2012) of filament, tendril and void galaxies with $z \leq 0.1$ and $M_* \geq 10^{9.5} M_\odot$ from the GLSSC. This early work indicates at some differences in the ultraviolet and far-IR energy outputs of void galaxies, and is based on an early version of a 20 band GAMA photometry catalogue. I hope to generate similar CSEDs for galaxies in different environments (different halo masses, central galaxies of groups, and within pairs) to further understand what scale environment affects galaxies most. The structural catalogues to do this work are already in place within GAMA and the improved full MAGPHYS DMU is soon to be in place.

My sample of well-defined, isolated void galaxies presents an unprecedented opportunity for the study of gas and dust in these low density environments. This data will provide new insights into how stellar populations and star formations of galaxies are impacted by the dynamics of gas and dust in low density environments, especially when compared to galaxies in filaments. Understanding the gas content of voids is crucially important to shed light on the impact of galaxy evolution in underdense environments.

While measuring the excess line correlation function (Obreschkow et al., 2013) of GLSSC void galaxies shows that they are not embedded in any kind of linear structure, it is important to determine if a deeper survey would indeed reveal further underlying structure in voids, and around void galaxies. It is therefore crucial to not only understand and quantify the impact of observational survey parameters on the sizes of voids, but also to preemptively calculate line correlation functions for galaxies in voids within simulations that mimic different observational depths for surveys.

One of the most compelling results of the structure catalogue is the discovery of tendrils of galaxies. Not only do these galaxies exist in a previously unexplored environment, but they also challenge the current understanding of the sizes and shapes of voids, as they are largely undetected in other galaxy surveys. By rerunning my filament-finding algorithm on existing GAMA data with a lower magnitude and redshift cut, I will be able to detect filamentary structures at lower mass regimes close to M_* , providing an insight into testing formation processes of galaxies at the threshold of high density and low density environments when combined with detailed simulations on the cluster scale. Figure 5.2 displays 500×500 kpc/h postage stamps for galaxies in tendrils across their range of $u-r$ colour and stellar mass, and highlights

the large variety in morphologies and colours that these galaxies possess. New photometric data from the KiDS survey will also help to underpin the nature of these objects, and H_I data from the ASKAP DINGO survey will help to reveal if these structures carry gas from voids into filaments and clusters.



IOTA and the aperture matched photometry catalogue

A.1 Introduction

With GAMA being a large collaboration spanning multiple institutions and a large number of individual researchers, it is very important to be able to maintain a central database containing all of the research output of the team. This not only allows for a standardisation of all of our data, but also makes it easier for collaborations within the team to be formed. To this end, GAMA's data has been organised into Data Management Units (DMUs), with one person being in charge of each DMU. These range from stellar masses to a group catalogue, and many of these data catalogues depend explicitly on the r -band aperture matched photometry catalogue, which I was responsible for constructing and maintaining. Every DMU used in calculating the observational parameters examined in Chapter 5 depend on this photometry, so it is important to understand how it was put together. The work described in this appendix has been done in

conjunction with Edward Taylor, Aaron Robotham, and Simon Driver.

Aperture matched photometry provides a reliable method for delivering photometry across a large number of wavelengths for the same source. By fixing the aperture on the object in the best possible image, one ensures that image degradation in other bands will have a lesser effect on the quality of the photometry. It also ensures uniformity on parameters based on aperture size, most notably measurements of object size and certain types of flux and consequently, magnitude measurements. For the case of GAMA, whose imaging data at the time of the making of this photometry catalogue is from the SDSS Seventh Data Release (*ugriz* Abazajian et al., 2009) and UKIDSS LAS Early Data Release (EDR) (Dye et al., 2006) and First and Second Data Releases (Lawrence et al., 2007; Warren et al., 2007; Warren et al., 2007 *YJHK*), the *r*-band imaging data has been used to generate the survey’s input catalogue. Therefore all of our sources are *r*-band selected, motivating the photometry to be matched to the *r*-band as well.

GAMA brings together the results of a number of different surveys conducted using various instruments and data reduction pipelines. A natural requirement of preparing this collection of data for any further analysis is to homogenise it as much as possible. These data are provided to the survey in their most raw form possible; as individual data frames typically consisting of 2048×1489 pixels for SDSS and 2072×2072 pixels for UKIDSS LAS with pixel scales of $0.396''/\text{pixel}$ and $0.4''/\text{pixel}$ respectively. Note that the entirety of the UKIDSS *J* and EDR fields of the *H* and *K* bands are microstepped, with sizes of 4103×4103 pixels and $0.2''/\text{pixel}$ scales. The initial reprocessing of this data is described in great detail in Hill et al. (2011) and a summary is given below.

A.2 Data processing and SWarps

After being downloaded from their respective databases (SDSS DR7 and ROE/WAFU), the images are rescaled to a single uniform zero-point of 30 magnitudes. Following renormalisation, the data are convolved to a common seeing of $2''$. This is done to ensure a uniform image quality across all images, and the value of two arcseconds is chosen such that 95% of images experience a degradation of their PSFs (as increasing the quality of the PSF would be impossible). This process is done via initially measuring the seeing values directly from each frame using PSFEX (Bertin, 2011); each frame is then convolved, using FGAUSS from the

FTOOLS package¹ with a Gaussian whose full-width half-maximum is such that the measured PSF is converted to $2''$. The seeing of each convolved frame is measured once again after the convolution process, along with the pixel scale using WCS information from the header (this is particularly important for determining which frames are microstepped when convolving UKIDSS frames). The results of this process can be seen in Figure A.1. There is a slight offset from the required value of 2 arcseconds; this is caused by virtue of the fact that we convolve the seeing—which we assume is Gaussian—with another Gaussian, when in practice the profile of the seeing is closer to a Lorentzian, rendering this assumption somewhat invalid. Due to this, the final PSF of the convolved frames isn't exactly centred on 2 arcseconds. Log files are generated and kept for each step of this process, to ensure maximum reproducibility.

As part of the process of improving the overall quality of the GAMA mosaics, a visual inspection of all frames whose original PSF values were deemed to be unrealistic ($> 3''$) was undertaken. In the *YJHK* bands, we respectively discovered 33, 13, 49 and 48 frames (alternatively, we retain 95.2% of fields in the *Y* and *J* bands, and 97.5% in the *H* and *K* bands) whose image quality had degraded past the point of acceptability, due to either star trails or improper focusing. Examples of both of these are shown in Figure A.2. The origin or cause of these poor images is not known, however they had been present in earlier versions of the mosaics, so their removal has increased the overall image quality of the new GAMA mosaics. Maps showing regions where frames are removed are shown in Figure A.3.

Once the remaining data have all been convolved, SWARP (Bertin et al., 2002) is used to stitch frames together into large so-called *mosaics*. SWARP is an image resampling and co-addition software that is designed to work primarily with FITS images, with the additional capacity to stitch images together using a varied number of different projections. When running SWARP the user provides the program with a list of images to put together, as well as a configuration file containing values for a number of parameters used in making the output images. In addition to generating co-added images, it can also output weight-maps which are analogous to the output images, but provide information on how accurate the flux reading at each pixel is. Another important property of SWARP is background subtraction: this is to account for systematic large-scale gradients across individual frames caused by instrumentation. Given that the flux at each pixel is some combination of signal and noise, it is important to be able to accurately remove any noise components from the data. Without background sub-

¹<http://heasarc.gsfc.nasa.gov/lheasoft/ftools/>

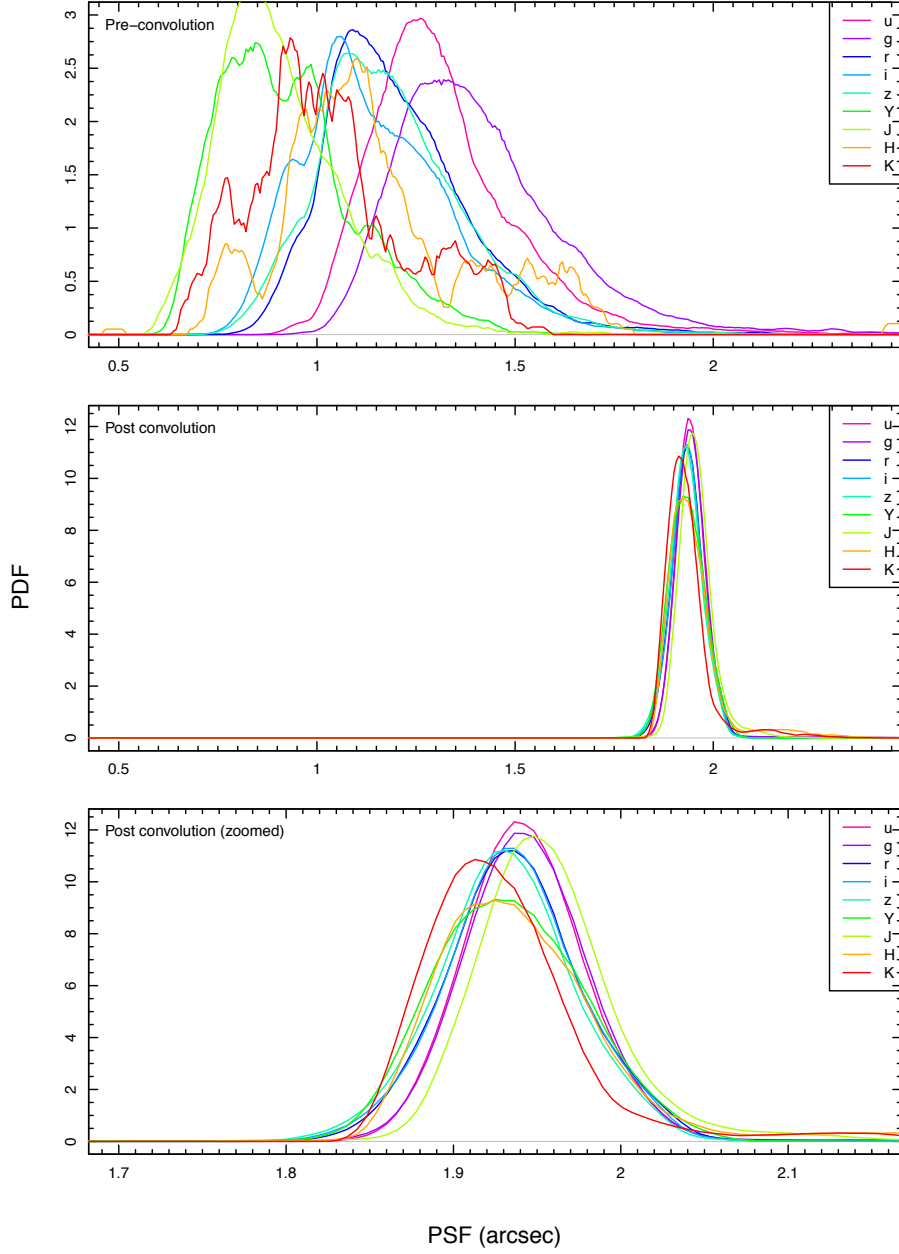


Figure A.1: Three plots showing the initial distribution of PSF values for all frames in each band before convolution (top panel) and after convolution (middle and lower panels). Note that the colour legend remains unchanged in all three frames. Also note the drastic increase in sharpness of the distribution of PSFs post-convolution, as shown by the scale of the y-axis.

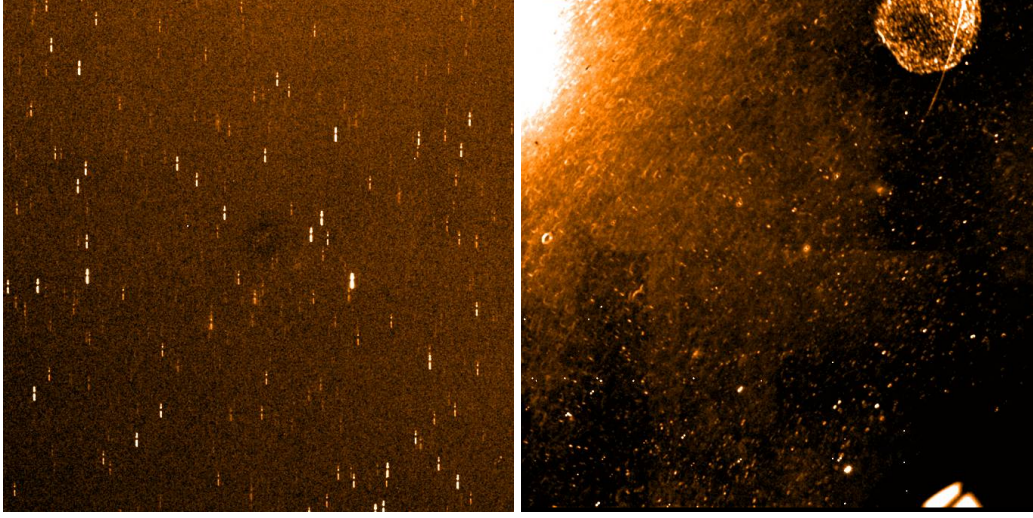


Figure A.2: Two examples of bad imaging quality in the raw UKIDSS frames. The left image, in the *H*-band shows the commonly seen effect of star trails. On the right panel, this time in the *K*-band is a distinctive ‘donut’-like shape that is present in almost all unfocused images. Note that these images have not been altered past their original state except for having been renormalised to a zeropoint of 30 magnitudes.

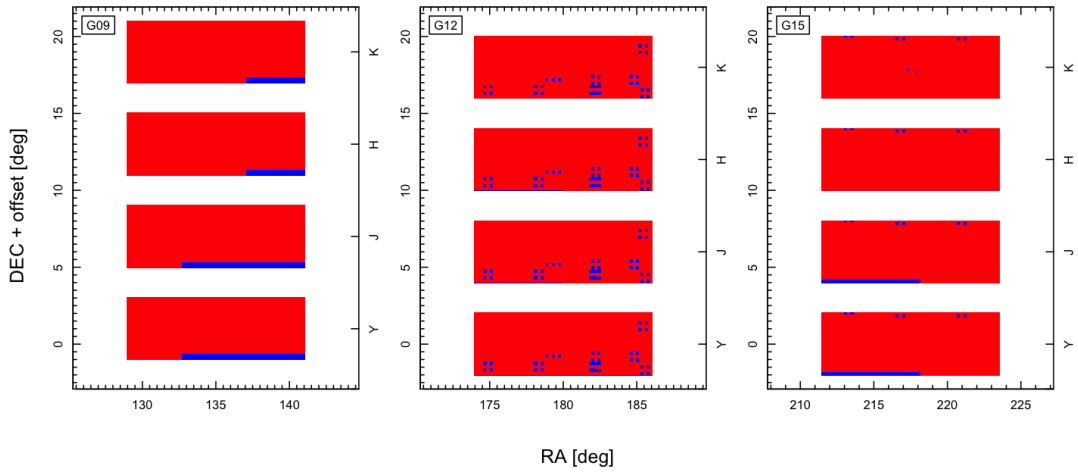


Figure A.3: Coverage for the v2 mosaics in the *YJHK* bands, where each band is plotted with a declination offset. Blue areas show regions where there is no flux in the mosaics, indicating a removed frame or missing data. We have not plotted this information for the *ugriz* data, as the SDSS mosaics are 100% complete saved for masked regions.

traction, a co-added image often displays a ‘patchwork’ style pattern where different frames meet.

For each input image, prior to co-addition, SWARP begins the background subtraction process by overlaying a grid onto the image, whose size is specified by the user. Within this grid, the mean and σ of the pixel values is computed, then iteratively clipped by discarding outliers until convergence at $\pm 3\sigma$ around the mean. If, during this entire process, the σ value drops less than 20%, the image is considered to be ‘not crowded’ and the background value is considered to be the mean. Alternatively, the background is set to be $2.5\times$ the median $-1.5\times$ the mean. The full background map is then computed by interpolating with a bi-cubic spline over all grids covering the image. Optionally, the background map can be further smoothed with a median filter by a second, smaller mesh. It is recommended to set a grid size that is greater than the average size of the objects in each image; within GAMA we use background sizes of 512×512 pixels, with a smoothing size of the map of 3×3 pixels; the default value. These parameters are specified using the `BACK_SIZE` and `BACK_FILTERSIZE` parameters for SWARP.

We generate one mosaic (consisting of a composite image and its corresponding weight-map) per band per field, resulting in a total of 27 (9×3) frames for the current GAMA I fields. These mosaics have resolutions of $0.339''/\text{pixel}$, sizes of at least 193900×79700 pixels, and are put together using a TAN projection centred on the given GAMA field being shown in the mosaic. Alongside these SWARPed images we also create a set of corresponding weight maps, which are primarily used to determine which sources are found in which frame.

A.3 IOTA

The Input/Output Tiling Analysis (IOTA) forms the backbone of the creation process for the aperture matched photometry catalogue. Originally created using code written for the SIGMA algorithm by Kelvin et al. (2012), this algorithm has since been heavily modified by the author of this thesis. In brief, IOTA is a script and wrapper written with the R statistical programming language built primarily around preparing imaging data to be fed into the Source Extractor (SEXTRACTOR) software package (Bertin & Arnouts, 1996) and collating the output of the program into a compact catalogue. Overall, the algorithm has two main parts: the first focussing on creating cutouts from the master GAMA mosaics centred around the galaxy to be analysed, and the second on running SEXTRACTOR in matched-aperture mode on all 9 bands and amal-

gamating the outputs into a single catalogue. When running IOTA from the command line, the user can specify a number of options, including how many processors the script runs on, how the results are presented, if the code should only generate a subset of results, and which input catalogue to use.

IOTA is designed to run through an input catalogue of objects and measure the photometry for each one individually. The code is designed to work around objects within the GAMA database (which contain a unique GAMA ID), or from the SDSS catalogue (using their unique OBJID identifier), however any input catalogue containing objects named with a unique identifier will do. Besides this identifying information, all IOTA needs are the RA and Dec of each object, as well as the directories and filenames of the mosaics where it must perform the photometry.

The input catalogue used to generate the new aperture matched photometries was originally put together from GAMA's TILINGCATV16 (Baldry et al., 2010), which contains science targets for our spectroscopic observations, as well as other information from the SDSS survey. This catalogue contains objects from the first phase of the survey; however the mosaics used to measure the photometry are from the second phase, and therefore have better overall imaging quality and are correctly processed to the right seeing. Furthermore, they encompass a larger area. From TILINGCATV16 we pick out all objects whose SURVEY_CLASS parameter is ≥ 3 (this is an internal GAMA variable that determines if an object belongs to the main survey or not), giving 152,742 galaxies down to an r -band magnitude of 19.8 and an average redshift of 0.22.

A.3.1 Source Extractor

SEXTRACTOR is an automated object detection and photometry program originally developed as part of the TERAPIX pipeline for image processing of MEGACAM data. Nowadays it is broadly used for performing high-speed photometric measurements for large datasets and FITS files. It is a highly controllable piece of software, and is almost entirely user-configurable; this carries the benefit that it will almost always work, but its results can often be nonsensical without the proper configuration files. Besides its speed, the primary draw of SEXTRACTOR for IOTA is its ability to detect sources in one image and do photometry in the other.

When running SEXTRACTOR the user specifies the location of a configuration file, within

which one has previously specified the values and options for a large number of configuration parameters, each of which governs one aspect of the program’s function; from object detection to background subtraction. The user must also specify a file which contains a list of the output parameters desired—these are also wide ranging, measuring various kinds of fluxes and magnitudes as well as a myriad of other photometric values.

For SEXTRACTOR to consider something to be an object, it must satisfy the following requirements:

1. All pixels in the object must have flux values above a minimum threshold, specified by the DETECT_THRESH parameter.
2. All these pixels must be adjacent to each other.
3. There must be more than a certain number of pixels in the object, specified by DETECT_MINAREA.

The detection threshold parameter can be given in units of surface brightness, ADUs, or relative to the background RMS; this final option is the default one and the one used in this script. There is also the option of applying a filter to the image prior to detection. This has the effect of smoothing the image, and is not used in IOTA. Following object detection, SEXTRACTOR examines groups of pixels to determine if they are indeed one object, or a blend of multiple objects; this process is called deblending and is described in greater detail below. SEXTRACTOR then performs its photometric measurements on what it has determined to be objects, and outputs these variables as a catalogue in a number of possible formats.

A.3.2 Image preparation and CUTTERPIPE

Given the size of the input catalogue at roughly 150,000 galaxies, it would be prohibitively computationally intensive to load up every single mosaic for each band into SEXTRACTOR for every galaxy. Given that each mosaic is roughly 60GB in size and measures hundreds of thousands of pixels in size, approaching the analysis this way would be highly inefficient. As a workaround to this, IOTA utilises the CUTTERPIPE script first developed by Lee Kelvin for use in the SIGMA pipeline (Kelvin et al., 2012) to generate cutouts of each galaxy in all 9 bands from the master GAMA mosaics.

Following sample definition and preparation, described above, IOTA begins to loop over

each galaxy in the sample. Initially, a tree of folders is created to house the outputs of CUTTERPIPE and SEXTRACTOR are created for the galaxy in question. IOTA then runs CUTTERPIPE and generates a cutout from the master GAMA mosaics in all 9 bands, which are stored in the folders that have been created. By using the WCS information contained in the header of the mosaics to convert the RA and δ of the target into x and y pixel coordinates, CUTTERPIPE is able to create a 400×400 pixel cutout centred on the galaxy in each band; this corresponds to a region of approximately $135'' \times 135''$, which is sufficiently large to encompass even the largest galaxies in GAMA. Transferring these cutouts into SEXTRACTOR is considerably less resource-intensive, and this workaround forms the single most time-saving feature of this routine. The r -band cutout is referred to as the *master* cutout.

A.3.3 Source extraction and photometry

The cutout images are sequentially fed into SEXTRACTOR operating in matched aperture mode. In this setting, the program takes two images; the first is used for object detection and the placement of the aperture, and the same aperture in the same location is used on the second image to take photometric measurements. As discussed previously, this method is preferred over single band photometry as GAMA's entire survey sample is defined on r -band imaging from the SDSS. For each galaxy, IOTA runs SEXTRACTOR 9 times; first in single band mode for the master cutout, followed by dual band mode for all other bands (*ugizYJHK*). The catalogue output of each *SExtractor* run is saved as a separated comma separated variable (CSV) table, from which the target object is extracted at the end of the run. Finally, all information from each CSV table is collated into a single row and appended onto the master catalogue. Given that our cutouts have areas of approximately $135'' \times 135''$, it is to be expected that SEXTRACTOR will detect and analyse other nearby objects to the target galaxy. To ensure that IOTA only catalogues results pertaining to the target galaxy, the results of the SEXTRACTOR object whose position is the shortest distance away from the target galaxy's position in x and y co-ordinates is transferred to the master catalogue.

SEXTRACTOR is a highly customisable source extraction software that relies on a large number of input parameters to obtain its results. Of these, the most relevant to the work done by IOTA are the parameters concerned with the issue of deblending objects that are visually close to each other or overlapping. Following object detection (governed by parameters concerned with thresholding), SEXTRACTOR goes on to break up the detections into different objects by

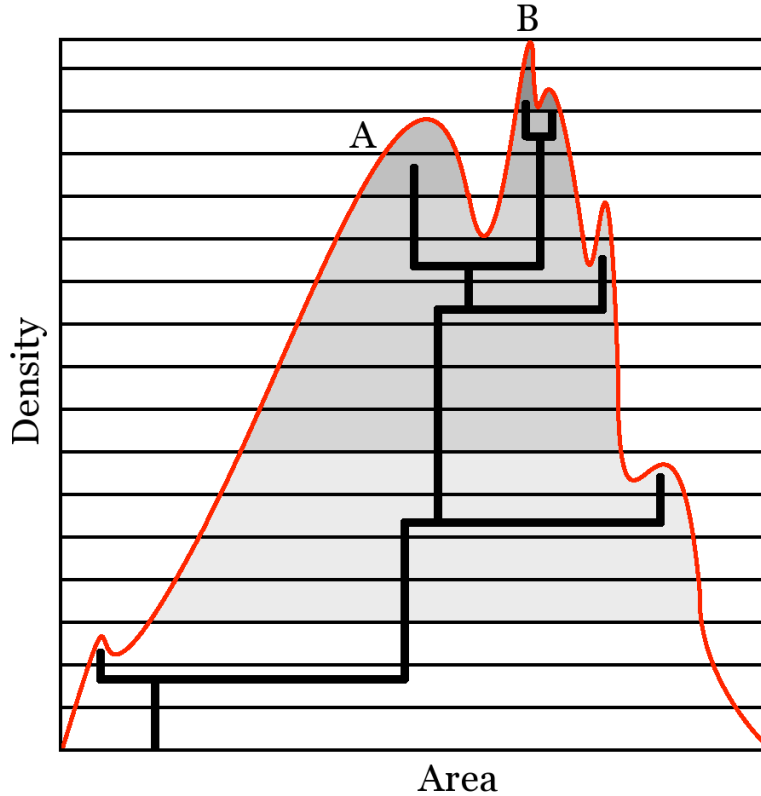


Figure A.4: An illustration of a one-dimensional representation of how SETRACTOR goes through the deblending process. Here the horizontal subdivisions in density represent the different levels determined by DEBLEND_MINCONT, with the greyscale highlighting under the red curve showing the DEBLEND_MINCONT fraction. Here, A is considered to be a branch because B is another branch at a higher level, with a higher fraction of the total flux.

deblending them.

Deblending in SETRACTOR is based around a ‘tree’ analysis. For a given object (a collection of adjacent pixels whose flux value is above the one specified by the DETECT_NTHRESH parameter), the program constructs such a ‘tree.’ A number of levels, whose quantity is determined by the DEBLEND_NTHRESH parameter are defined between the maximum flux count of the object and the detection threshold. The tree’s branches are then regions of pixels whose flux is above a certain fraction of the total flux count of the entire object, given by DEBLEND_MINCONT. If there is at least one other branch above this level that is also above this fraction, then both objects are deblended. This point is clarified in Figure A.4, where the deblending process is illustrated in one dimension.

The deblending parameters chosen to be used in IOTA are such that SETRACTOR makes a large number of deblending levels (DEBLEND_NTHRESH = 32) and the DEBLEND_MINCONT

parameter is set to 0.00005 (where 0.001 correspond to 1% of flux). This ensures that when objects are blended together, it is very likely that SExtractor will deblend them – these settings would be rather ungainly for a blind source detection run as it would yield far too many false deblends, but in the case of IOTA where we are seeking photometry on targeted sources, it is ideal. More importantly, we must enforce a rigorous set of deblending parameters to compensate for convolving our mosaics to a seeing of 2 arcseconds, which in many cases will cause tightly clustered sources that would otherwise be separable in the higher quality original data to overlap. To ensure optimal results, a vast number of different combinations of these parameters were tested on a busy SDSS *r*-band field using the Graphical Astronomy Image and Analysis Tool (GAIA)². GAIA is an image viewer that is designed to work with FITS files, but is also able to run its own internal version of SExtractor with the added benefit of displaying the apertures placed around each object within the image viewer; this makes it ideal to experiment with different configuration parameters.

Following source extraction in all bands with proper deblending parameters, IOTA collects all 9 output catalogues and, from each, selects the row containing the object that is geometrically closest (i.e. the target i for which $\sqrt{(x_i - x_{\text{cen}})^2 + (y_i - y_{\text{cen}})^2}$ is minimal) to the SDSS target object on the SWARPED mosaic, as all other detections will be for untargeted objects. The average pixel separation is ≈ 0.736 pixels. A full distribution of pixel separations for all of IOTA is shown in Figure A.5 These 9 rows are then stitched together into a single catalogue entry (where all self-similar entries such as object position are removed) and added to the master IOTA catalogue. A summary of results for the particular object is printed both onto the user's console screen and into a log file for record-keeping and investigation should an error occur.

As IOTA is designed to run through an input catalogue and analyses objects from it in order, it is a rather simple process to parallelize. The option to split IOTA into any number of sub-processes can be specified by the user as a start-up command. In generating the matched aperture photometry catalogue, IOTA was run on a 16 processor computer at the University of St Andrews, allowing us to spawn up to 16 IOTA sub-processes, greatly reducing the total run time for the script.

²<http://star-www.dur.ac.uk/~pdraper/gaia/gaia.html>

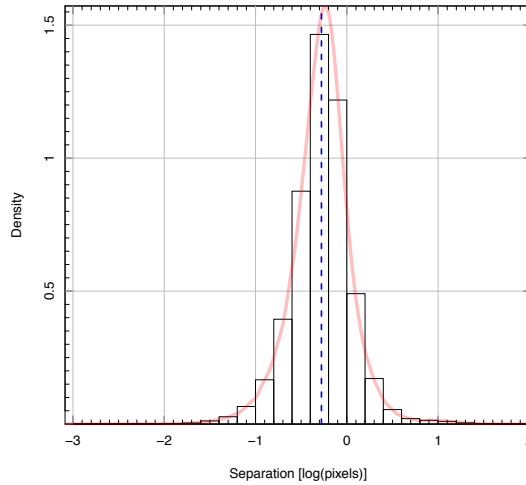


Figure A.5: The distribution of pixel separation values for all objects in IOTA. The transparent line represents the smoothed density distribution of the values shown in the histogram, and the dashed blue line lies on the mean value.

A.4 Matched aperture photometry catalogue

IOTA's final result is the second version of the matched aperture photometry catalogue, dubbed `APMATCHEDCATV2`, containing Kron and Petrosian photometry in 9 bands for all 152,742 objects in the input catalogue. Alongside these photometric parameters, the catalogue also contains a number of other `SEXTRACTOR` outputs; most notably estimators on position angle and object size (in pixels). Any `SEXTRACTOR` error flags are also kept.

The most immediate comparison one can make using this data is to plot it against the photometry from the first version of the catalogue; this is shown in Figure A.6. In most cases, particularly for the optical bands, there is excellent agreement between both versions of the catalogue, however this is less the case for the NIR bands; this is to be expected due to the reconstruction of the mosaics in these bands, from which the photometry is measured. Any object falling *above* the green line is brighter (i.e. $v_2 > v_1$) in the `v1` catalogue, while the opposite is true for objects below the green line. The density plots shown on the right panels are particularly useful for identifying this change, as they are not centred on 0.

Another visualisation of this comparison is to examine the relationship between photometry in different bands in both catalogues. In an ideal scenario, a PDF describing the distribution of $r - u$ magnitudes, for example, would be a Gaussian distribution. Introducing a source of well understood, ideal errors, we would expect this Gaussian distribution to be spread out. It

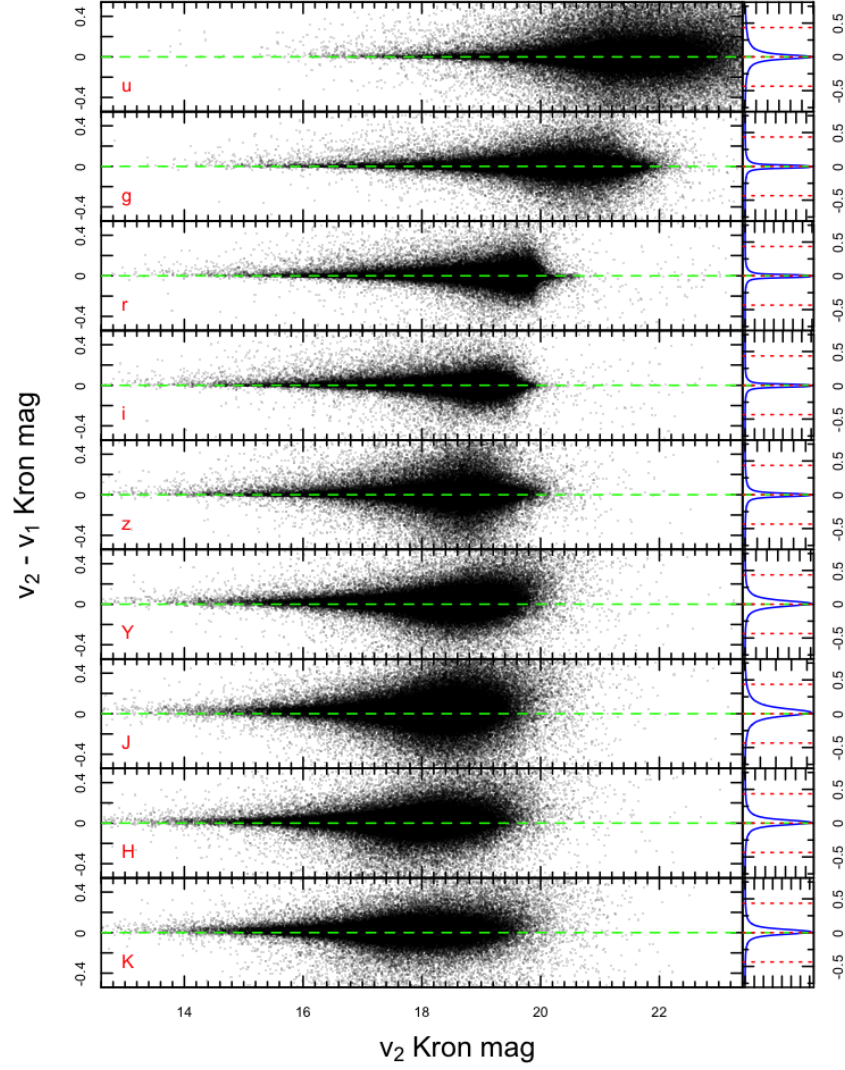


Figure A.6: Comparison of photometry between v2 and v1 of the matched aperture photometry catalogue in all 9 bands. The rotated density plots shown to the right of the scatter plots show a projected distribution of all the data for that band, and are particularly useful for identifying upwards or downwards shifts in the photometry. The green dashed lines are drawn on the $\Delta \text{mag} = 0$ line and the red lines show means and standard deviations for the density plots.

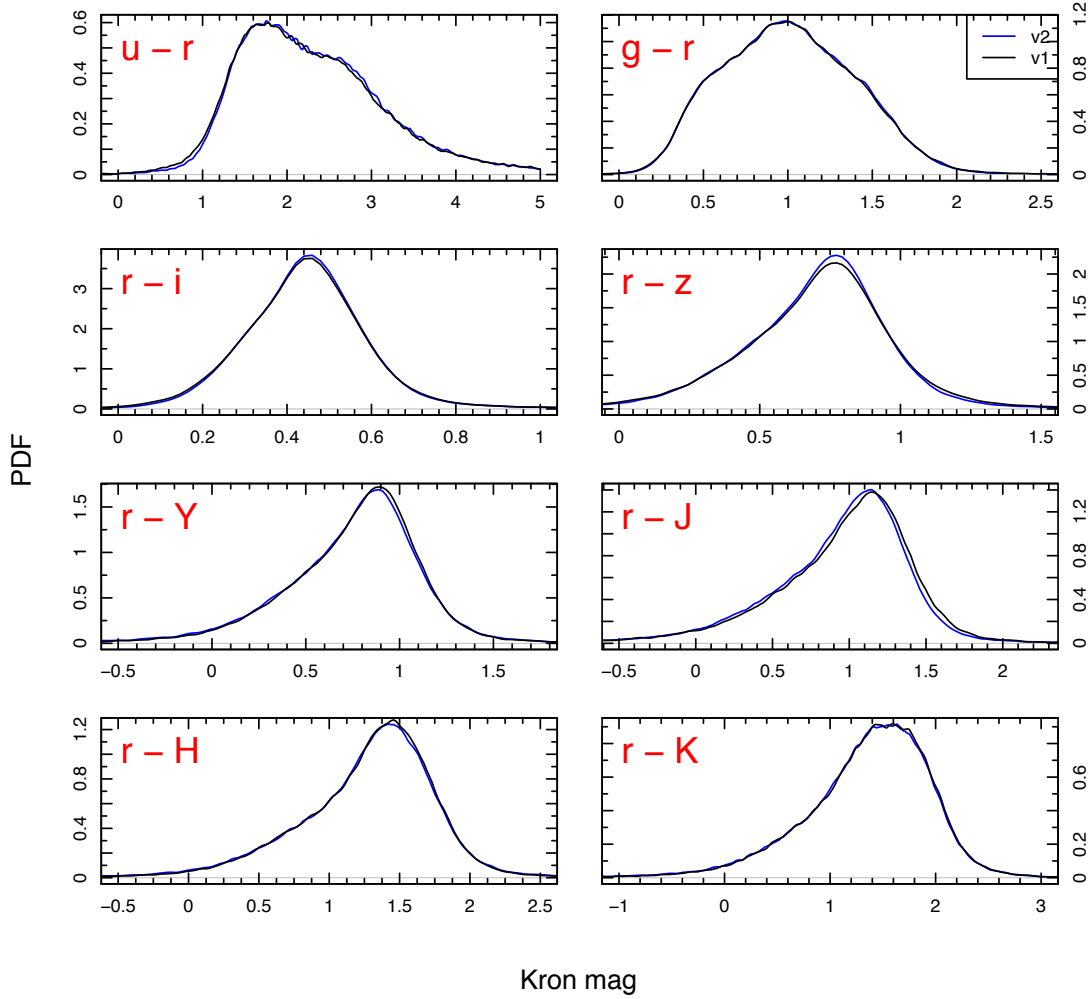


Figure A.7: Colour magnitude plots, with the red text in each panel displaying the bands being compared. The blue and black lines show the distribution for v2 and v1 photometry catalogues respectively. In a small number of cases, most notably for the z -band this distribution sharpens in the new version.

follows therefore that smaller errors would tighten up this Gaussian once again. By comparing the PDFs of $r - X$ band magnitudes (where X refers to all other bands) between v2 and v1 of the catalogues and looking for changes in the spread of both distributions, we can make an assessment as to the presence of any improvement in our measurements and errors. This comparison is plotted in Figure A.7.

In many cases, however, this difference is not seen. A tightening of the distribution is most notable in the $r - z$ case, where the version 2 photometry has a visibly higher central peak. To a lesser degree, this is also visible in the $r - i$ panel. The overall shift in the distributions shown in the $r - Y$ and $r - J$ reflects the off-centre PDF for these bands shown on the right

hand side of Figure A.6 and is a reflection on the reconstruction of the mosaics in these bands – the effect is less noticeable in the H and K bands because this data contains fewer frames that were microstepped. That the distribution has only shifted for the $r - Y$ and $r - J$ bands indicates that correcting the mosaics has not changed the associated magnitude errors as much as it has the photometry itself.

A.4.1 Errors

Much of the work in this subsection has been done in conjunction with Edward Taylor. The errors for the catalogue come straight from SExtractor’s photometric errors, which themselves are calculated from flux errors and the noise measurements of each pixels. These errors are often underestimates due to correlations of flux between neighbouring pixels; but if we assume that the errors can be trusted to map variations in photometric noise as a function of position or aperture size, then we can calibrate our errors. Assuming the SDSS and UKIDSS-LAS surveys before us have adequately estimated photometric errors in their imaging data, we can apply some corrections to our errors in order to match them to this data, as all catalogues come from the same imaging data. In the case of the optical data, this scaling has been done via the following expression:

$$\Delta m_f = (A^2 \Delta m_i - b^2)^{0.5} \quad (\text{A.1})$$

where Δm_f and Δm_i refer to the scaled and unscaled magnitude errors; and A and b are to parameters that describe the transformation from SExtractor errors to SDSS-calibrated errors. Had GAMA used the same aperture sizes as the SDSS, a mere object-object match would have sufficed for error calibration. Instead, we must compare objects between SDSS and GAMA whilst controlling aperture size and flux. We can therefore rescale IOTA uncertainties so that they match SDSS errors within fixed bins of magnitude and aperture size, arriving at A and b . A similar analysis for UKIDSS-LAS data is currently impossible, as this survey does not report aperture sizes. We are therefore limited to rescaling our near-IR errors by a single factor to match the two distributions of errors as closely as possible.

Where SExtractor fails to calculate a magnitude (for whatever reason) the default value returned is 99. Any nonsensical results for flux, background measurements or the full-width half-maximum of the image are usually caused by failures in SExtractor; in these cases

one can consult the flags returned by the program, which are also present in the photometry catalogue. In cases where objects are located in the frames removed by visual inspection, their magnitude is set to -99, their fluxes to 0 and all associated errors to 1E99.

Acknowledgements

It would be remiss of me to not thank a few key individuals without whom I would not have been able to complete this body of work. To my supervisors, Simon Driver and Aaron Robotham, I extend my most heartfelt gratitude: not simply for furnishing me with the tools to successfully start my scientific career, but also for their tremendous support and friendship. To my parents, for their constant and undying enthusiasm and pride in what I do. Finally, of course, thank you to all my friends to all my friends in the UK and Australia: you always made sure to remind me to keep my feet on the ground and let my hair down from time to time.

Bibliography

- Abazajian, K. N. et al. 2009, *Astrophysical Journal*, Supplement, 182, 543
- Adelman-McCarthy, J. K. et al. 2006, *Astrophysical Journal*, Supplement, 162, 38
- Alcock, C., & Paczynski, B. 1979, *Nature*, 281, 358
- Alpaslan, M. et al. 2014, *Monthly Notices of the Royal Astronomical Society*, 438, 177
- Angulo, R. E., Springel, V., White, S. D. M., Jenkins, A., Baugh, C. M., & Frenk, C. S. 2012, *Monthly Notices of the Royal Astronomical Society*, 426, 2046
- Aragon-Calvo, M. A., van de Weygaert, R., Araya-Melo, P. A., Platen, E., & Szalay, A. S. 2010, *Monthly Notices of the Royal Astronomical Society: Letters*, 404, L89
- Aragón-Calvo, M. A., van de Weygaert, R., & Jones, B. J. T. 2010, *Monthly Notices of the Royal Astronomical Society*, 408, 2163
- Arnold, V. I., Shandarin, S. F., & Zeldovich, Y. B. 1982, *Geophysical & Astrophysical Fluid Dynamics*, 20, 111
- Bahcall, N. 1988, *Astronomy and Astrophysics*, 26, 631
- Baldry, I. K., Balogh, M. L., Bower, R. G., Glazebrook, K., Nichol, R. C., Bamford, S. P., & Budavari, T. 2006, *Monthly Notices of the Royal Astronomical Society*, 373, 469
- Baldry, I. K. et al. 2012, *Monthly Notices of the Royal Astronomical Society*, 421, 621
- Baldry, I. K. et al. 2010, *Monthly Notices of the Royal Astronomical Society*, 404, 86
- Balogh, M. L. 2007, in *Astronomical Society of the Pacific Conference Series*, Vol. 379, *Cosmic Frontiers*, ed. N. Metcalfe & T. Shanks, 109
- Barrow, J. D., Bhavsar, S. P., & Sonoda, D. H. 1985, *Monthly Notices of the Royal Astronomical Society*, 216, 17
- Beers, T. C., Flynn, K., & Gebhardt, K. 1990, *Astrophysical Journal*, 100, 32
- Benítez-Llambay, A., Navarro, J. F., Abadi, M. G., Gottlöber, S., Yepes, G., Hoffman, Y., & Steinmetz, M. 2013, *Astrophysical Journal*, 763, L41
- Berlind, A. A., & Weinberg, D. H. 2002, *Astrophysical Journal*, 575, 587
- Bertin, E. 2011, 442, 435

- Bertin, E., & Arnouts, S. 1996, *Astronomy and Astrophysics, Supplement*, 117, 393
- Bertin, E., Mellier, Y., Radovich, M., Missonnier, G., Didelon, P., & Morin, B. 2002, in , 228
- Beygu, B., Kreckel, K., van de Weygaert, R., van der Hulst, J. M., & van Gorkom, J. H. 2013, *Astrophysical Journal*, 145, 120
- Bharadwaj, S., Bhavsar, S. P., & Sheth, J. V. 2004, *Astrophysical Journal*, 606, 25
- Blake, C. et al. 2011, *The WiggleZ Dark Energy Survey: the growth rate of cosmic structure since redshift $z=0.9$* , 1104.2948
- Blanton, M. R., & Roweis, S. 2007, *Astrophysical Journal*, 133, 734
- Blumenthal, G. R., da Costa, L. N., Goldwirth, D. S., Lecar, M., & Piran, T. 1992, *Astrophysical Journal*, 388, 234
- Blumenthal, G. R., Faber, S. M., Primack, J. R., & Rees, M. J. 1984, *Nature*, 311, 517
- Bond, J. R., & Efstathiou, G. 1984, *Astrophysical Journal*, 285, L45
- Bond, J. R., Kofman, L., & Pogosyan, D. 1996, *Nature*, 380, 603
- Bond, N. A., Strauss, M. A., & Cen, R. 2010, *Monthly Notices of the Royal Astronomical Society*, 409, 156
- Bower, R. G., Benson, A. J., Malbon, R., Helly, J. C., Frenk, C. S., Baugh, C. M., Cole, S., & Lacey, C. G. 2006, *Monthly Notices of the Royal Astronomical Society*, 370, 645
- Butcher, H., & Oemler, Jr., A. 1984, *Astrophysical Journal*, 285, 426
- Cameron, E. 2011, *Publications of the Astronomical Society of Australia*, 28, 128
- Cannon, R. et al. 2006, *Monthly Notices of the Royal Astronomical Society*, 372, 425
- Carlberg, R. G., Yee, H. K. C., Ellingson, E., Abraham, R., Gravel, P., Morris, S., & Pritchet, C. J. 1996, *Astrophysical Journal*, 462, 32
- Cautun, M., van de Weygaert, R., & Jones, B. J. T. 2013, *Monthly Notices of the Royal Astronomical Society*, 429, 1286
- Cen, R. 2013, *ArXiv e-prints*, 1311.5916
- Chabrier, G. 2003, *Publications of the Astronomical Society of the Pacific*, 115, 763
- Choi, E., Bond, N. A., Strauss, M. A., Coil, A. L., Davis, M., & Willmer, C. N. A. 2010, *Monthly Notices of the Royal Astronomical Society*, 406, 320
- Colberg, J. M. 2007, *Monthly Notices of the Royal Astronomical Society*, 375, 337
- Colberg, J. M., Krughoff, K. S., & Connolly, A. J. 2005, *Monthly Notices of the Royal Astronomical Society*, 359, 272
- Colless, M. et al. 2001, *Monthly Notices of the Royal Astronomical Society*, 328, 1039

- Conroy, C. et al. 2005, *Astrophysical Journal*, 635, 990
- Croton, D. J. et al. 2005, *Monthly Notices of the Royal Astronomical Society*, 356, 1155
- da Costa, L. N. et al. 1998, *The Astronomical Journal*, 116, 1
- Davé, R. et al. 2001, *Astrophysical Journal*, 552, 473
- Davis, M. et al. 2003, in , 161–172
- de Lapparent, V., Geller, M. J., & Huchra, J. P. 1986, *Astrophysical Journal, Letters*, 302, L1
- Dekel, A., & Rees, M. J. 1994, *Astrophysical Journal, Letters*, 422, L1
- Diaferio, A. 1999, *Monthly Notices of the Royal Astronomical Society*, 309, 610
- Diaferio, A. 2009, Measuring the mass profile of galaxy clusters beyond their virial radius, 0901.0868
- Diaferio, A., & Geller, M. J. 1997, *Astrophysical Journal*, 481, 633
- Diaferio, A., Geller, M. J., & Rines, K. J. 2005, *Astrophysical Journal*, 628, L97
- Doroshkevich, A., Tucker, D. L., Allam, S., & Way, M. J. 2004, *Astronomy and Astrophysics*, 418, 7
- Dressler, A. 1980, *Astrophysical Journal*, 236, 351
- Dressler, A. et al. 1997, *Astrophysical Journal*, 490, 577
- Drinkwater, M. J. et al. 2010, *Monthly Notices of the Royal Astronomical Society*, 401, 1429
- Driver, S. P. et al. 2011, *Monthly Notices of the Royal Astronomical Society*, 413, 971
- . 2009, *A&G*, 50, 12
- Driver, S. P. et al. 2012, *Monthly Notices of the Royal Astronomical Society*, 427, 3244
- Dye, S. et al. 2006, *Monthly Notices of the Royal Astronomical Society*, 372, 1227
- Einasto, Ä., Jöeveer, M., & Saar, E. 1980, *Monthly Notices of the Royal Astronomical Society*, 193, 353
- Eisenstein, D. J. et al. 2001, *The Astronomical Journal*, 122, 2267
- . 2005, *Astrophysical Journal*, 633, 560
- Eke, V. R. et al. 2004, *Monthly Notices of the Royal Astronomical Society*, 348, 866
- Eke, V. R., Cole, S., & Frenk, C. S. 1996, *Monthly Notices of the Royal Astronomical Society*, 282, 263
- El-Ad, H., & Piran, T. 1997, *Astrophysical Journal*, 491, 421
- Elbaz, D. et al. 2007, *Astronomy and Astrophysics*, 468, 33

- Ellis, R. S., Colless, M., Broadhurst, T., Heyl, J., & Glazebrook, K. 1996, *Monthly Notices of the Royal Astronomical Society*, 280, 235
- Erdogdu, P et al. 2006, *Monthly Notices of the Royal Astronomical Society*, 368, 1515
- Fadda, D., Biviano, A., Marleau, F. R., Storrie-Lombardi, L. J., & Durret, F. 2008, *Astrophysical Journal*, 672, L9
- Falco, E. E. et al. 1999, *Publications of the Astronomical Society of the Pacific*, 111, 438
- Fèvre, O. L. et al. 2005, *Astronomy and Astrophysics*, 439, 845
- Forero-Romero, J. E., Hoffman, Y., Gottlöber, S., Klypin, A., & Yepes, G. 2009, *Monthly Notices of the Royal Astronomical Society*, 396, 1815
- Fry, J. N. 1984, *Astrophysical Journal*, 279, 499
- Fukugita, M., Hogan, C. J., & Peebles, P. J. E. 1998, *Astrophysical Journal*, 503, 518
- Garilli, B. et al. 2008, *Astronomy and Astrophysics*, 486, 683
- Geller, M. J., & Huchra, J. P. 1989, *Science*, 246, 897
- Girardi, M., Giuricin, G., Mardirossian, F., Mezzetti, M., & Boschini, W. 1998, *Astrophysical Journal*, 505, 74
- González, R. E., & Padilla, N. D. 2010, *Monthly Notices of the Royal Astronomical Society*, 407, 1449
- Gottlöber, S., Łokas, E. L., Klypin, A., & Hoffman, Y. 2003, *Monthly Notices of the Royal Astronomical Society*, 344, 715, astro-ph/0305393
- Graham, M. J., Clowes, R. G., & Campusano, L. E. 1995, *Monthly Notices of the Royal Astronomical Society*, 275
- Gregory, S. A., & Thompson, L. A. 1978, *Astrophysical Journal*, 222, 784
- Gunawardhana, M. L. P et al. 2013, *Monthly Notices of the Royal Astronomical Society*, 433, 2764
- Gurbatov, S. N., Saichev, A. I., & Shandarin, S. F. 2012, *Physics-Uspekhi*, 55, 223
- Habib, S. et al. 2012, *ArXiv e-prints*, 1211.4864
- Hahn, O., Carollo, C. M., Porciani, C., & Dekel, A. 2007a, *Monthly Notices of the Royal Astronomical Society*, 381, 41
- Hahn, O., Porciani, C., Carollo, C. M., & Dekel, A. 2007b, *Monthly Notices of the Royal Astronomical Society*, 375, 489
- Hill, D. T. et al. 2011, *Monthly Notices of the Royal Astronomical Society*, 412, 765

- Hoffman, Y., Metuki, O., Yepes, G., Gottlöber, S., Forero-Romero, J. E., Libeskind, N. I., & Knebe, A. 2012, *Monthly Notices of the Royal Astronomical Society*, 425, 2049
- Hoyle, F., & Vogeley, M. S. 2002, *Astrophysical Journal*, 566, 641
- . 2004, *Astrophysical Journal*, 607, 751
- Huang, J., Glazebrook, K., Cowie, L. L., & Tinney, C. 2003, *Astrophysical Journal*, 584, 203
- Hughes, J. P. 1989, *Astrophysical Journal*, 337, 21
- Iyanaga, S. 1980, *Encyclopedic dictionary of mathematics*, 2nd edn. (Cambridge Mass.: MIT Press)
- Jenkins, A., Frenk, C. S., White, S. D. M., Colberg, J. M., Cole, S., Evrard, A. E., Couchman, H. M. P., & Yoshida, N. 2001, *Monthly Notices of the Royal Astronomical Society*, 321, 372
- Joeveer, M., Einasto, J., & Tago, E. 1978, *Monthly Notices of the Royal Astronomical Society*, 185, 357
- Jones, D. H. et al. 2009, *Monthly Notices of the Royal Astronomical Society*, 399, 683
- Kaiser, N. 1987, *Monthly Notices of the Royal Astronomical Society*, 227, 1
- Kaiser, N., Squires, G., & Broadhurst, T. 1995, *Astrophysical Journal*, 449, 460
- Kelvin, L. S. et al. 2012, *Monthly Notices of the Royal Astronomical Society*, 421, 1007
- Kirshner, R. P., Oemler, A., J., Schechter, P. L., & Shethman, S. A. 1981, *Astrophysical Journal*, 248, L57
- Komatsu, E. et al. 2011, *Astrophysical Journal Supplements*, 192, 18
- Kormendy, J., & Kennicutt, R. C. 2004, *Annual Review of Astronomy and Astrophysics*, 42, 603
- Kreckel, K., Platen, E., Aragón-Calvo, M. A., van Gorkom, J. H., van de Weygaert, R., van der Hulst, J. M., & Beygu, B. 2012, *Astrophysical Journal*, 144, 16
- Lacerna, I., Rodríguez-Puebla, A., Avila-Reese, V., & Hernández-Toledo, H. M. 2013, *ArXiv e-prints*, 1311.6473
- Landy, S. D., & Szalay, A. S. 1993, *Astrophysical Journal*, 412, 64
- Lavaux, G., & Wandelt, B. D. 2010, *Monthly Notices of the Royal Astronomical Society*, 403, 1392
- Lavaux, G., & Wandelt, B. D. 2012, *Astrophysical Journal*, 754, 109
- Lawrence, A. et al. 2007, *Monthly Notices of the Royal Astronomical Society*, 379, 1599
- Lewis, I. et al. 2002, *Monthly Notices of the Royal Astronomical Society*, 334, 673
- Lilly, S. J., Fevre, O. L., Crampton, D., Hammer, F., & Tresse, L. 1995, *Astrophysical Journal*, 455, 50

- Lilly, S. J. et al. 2007, *Astrophysical Journal*, Supplement, 172, 70
- Liske, J., Lemon, D. J., Driver, S. P., Cross, N. J. G., & Couch, W. J. 2003, *Monthly Notices of the Royal Astronomical Society*, 344, 307
- Lopes, P. A. A., Ribeiro, A. L. B., & Rembold, S. B. 2013, *ArXiv e-prints*, 1310.6309
- Loveday, J. et al. 2012, *Monthly Notices of the Royal Astronomical Society*, 420, 1239
- Loveday, J., Peterson, B. A., Efstathiou, G., & Maddox, S. J. 1992, *Astrophysical Journal*, 390, 338
- Lukash, V. N., Mikheeva, E. V., & Malinovsky, A. M. 2011, *Physics-Uspekhi*, 54, 983
- Masaki, S., Lin, Y.-T., & Yoshida, N. 2013, *Monthly Notices of the Royal Astronomical Society*, 436, 2286
- Merson, A. I. et al. 2012, *Monthly Notices of the Royal Astronomical Society*, 429, 556
- Meyer, M. 2009, in *Panoramic Radio Astronomy: Wide-field 1-2 GHz Research on Galaxy Evolution*, 0912.2167
- Murphy, D. N. A., Eke, V. R., & Frenk, C. S. 2011, *Monthly Notices of the Royal Astronomical Society*, 413, 2288
- Murray, S. G., Power, C., & Robotham, A. S. G. 2013, *Monthly Notices of the Royal Astronomical Society*, 434, L61
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, *Astrophysical Journal*, 490, 493
- Nelder, J. A., & Mead, R. 1965, *The Computer Journal*, 7, 308
- Norberg, P. et al. 2002, *Monthly Notices of the Royal Astronomical Society*, 332, 827
- Novikov, D., Colombi, S., & Doré, O. 2006, *Monthly Notices of the Royal Astronomical Society*, 366, 1201
- Obreschkow, D., Power, C., Bruderer, M., & Bonvin, C. 2013, *Astrophysical Journal*, 762, 115
- Oort, J. H. 1983, *Annual Review of Astronomy and Astrophysics*, 21, 373
- Pan, D. C., Vogeley, M. S., Hoyle, F., Choi, Y.-Y., & Park, C. 2012, *Monthly Notices of the Royal Astronomical Society*, 421, 926
- Park, C., Choi, Y.-Y., Kim, J., Gott III, J. R., Kim, S. S., & Kim, K.-S. 2012, *Astrophysical Journal*, 759, L7
- Paz, D. J., Sgró, M. A., Merchán, M., & Padilla, N. 2011, *Monthly Notices of the Royal Astronomical Society*, 414, 2029
- Peacock, J. A., & Smith, R. E. 2000, *Monthly Notices of the Royal Astronomical Society*, 318, 1144

- Peebles, P. J. E. 2001, *Astrophysical Journal*, 557, 495
- Peebles, P. J. E., & Yu, J. T. 1970, *Astrophysical Journal*, 162, 815
- Pettini, M., & Pagel, B. E. J. 2004, *Monthly Notices of the Royal Astronomical Society*, 348, L59
- Pimbblet, K. A. 2005, *Publications of the Astronomical Society of Australia*, 22, 136
- Pisani, A. 1993, *Monthly Notices of the Royal Astronomical Society*, 265, 706
- Porter, S. C., Raychaudhury, S., Pimbblet, K. A., & Drinkwater, M. J. 2008, *Monthly Notices of the Royal Astronomical Society*, 388, 1152
- Postman, M., & Geller, M. J. 1984, *Astrophysical Journal*, 281, 95
- Press, W. H., & Davis, M. 1982, *Astrophysical Journal*, 259, 449
- Press, W. H., & Schechter, P. 1974, *Astrophysical Journal*, 187, 425
- Prim, R. 1957, *Bell System Technology Journal*, 36, 1389
- Primack, J. R. 2009, in *American Institute of Physics Conference Series*, Vol. 1192, American Institute of Physics Conference Series, ed. F. Roig, D. Lopes, R. de La Reza, & V. Ortega, 101–137, 0909.2021
- Quinn, P. J., Hernquist, L., & Fullagar, D. P. 1993, *Astrophysical Journal*, 403, 74
- R Development Core Team. 2011, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria
- Ratcliffe, A., Shanks, T., Broadbent, A., Parker, Q. A., Watson, F. G., Oates, A. P., Fong, R., & Collins, C. A. 1996, *Monthly Notices of the Royal Astronomical Society*, 281, L47
- Reddick, R. M., Wechsler, R. H., Tinker, J. L., & Behroozi, P. S. 2013, *Astrophysical Journal*, 771, 30
- Regós, E., & Geller, M. J. 1989, *Astrophysical Journal*, 98, 755
- Rieder, S., van de Weygaert, R., Cautun, M., Beygu, B., & Portegies Zwart, S. 2013, *Monthly Notices of the Royal Astronomical Society*, 435, 222
- Rines, K., & Diaferio, A. 2006, *Astrophysical Journal*, 132, 1275
- Rines, K., Geller, M. J., Kurtz, M. J., & Diaferio, A. 2003, *Astrophysical Journal*, 126, 2152
- Robotham, A. S. G. et al. 2010, *Publications of the Astronomical Society of Australia*, 27, 76
- . 2013, *Monthly Notices of the Royal Astronomical Society*, 431, 167
- . 2011, *Monthly Notices of the Royal Astronomical Society*, 416, 2640
- Ryden, B. S. 1995, *Astrophysical Journal*, 452, 25

- Sahni, V., Sathyaprakash, B. S., & Shandarin, S. F. 1998, *Astrophysical Journal, Letters*, 495, L5
- Saunders, W. et al. 2000, *Monthly Notices of the Royal Astronomical Society*, 317, 55
- Schechter, P. 1976, *Astrophysical Journal*, 203, 297
- Schechter, S., Landy, S., Oemler, A., Tucker, D., Kirshner, R., Lin, H., & Schechter, P. 1995, in *Wide Field Spectroscopy and the Distant Universe*, ed. S. J. Maddox & A. Aragon-Salamanca, 98
- Serra, A. L., Diaferio, A., Murante, G., & Borgani, S. 2011, *Monthly Notices of the Royal Astronomical Society*, 412, 800
- Schechter, S. A., Landy, S. D., Oemler, A., Tucker, D. L., Lin, H., Kirshner, R. P., & Schechter, P. L. 1996, *Astrophysical Journal*, 470, 172
- Sheth, R. K., & van de Weygaert, R. 2004, *Monthly Notices of the Royal Astronomical Society*, 350, 517
- Silverman, B. W. 1986, *Density estimation for statistics and data analysis*
- Smith, A. G., Hopkins, A. M., Hunstead, R. W., & Pimbblet, K. A. 2012, *Monthly Notices of the Royal Astronomical Society*, 422, 25
- Sousbie, T. 2011, *Monthly Notices of the Royal Astronomical Society*, 414, 350
- Sousbie, T., Colombi, S., & Pichon, C. 2009, *Monthly Notices of the Royal Astronomical Society*, 393, 457
- Sousbie, T., Pichon, C., Colombi, S., Novikov, D., & Pogosyan, D. 2008, *Monthly Notices of the Royal Astronomical Society*, 383, 1655
- Springel, V. et al. 2005, *Nature*, 435, 629
- Steidel, C. C., Adelberger, K. L., Shapley, A. E., Pettini, M., Dickinson, M., & Giavalisco, M. 2003, *Astrophysical Journal*, 592, 728
- Stoica, R. S., Martínez, V. J., & Saar, E. 2010, *Astronomy and Astrophysics*, 510, A38
- Strauss, M. A. et al. 2002, *The Astronomical Journal*, 124, 1810
- Sutter, P. M., Lavaux, G., Wandelt, B. D., & Weinberg, D. H. 2012, *Astrophysical Journal*, 761, 187
- Tanaka, M., Hoshi, T., Kodama, T., & Kashikawa, N. 2007, *Monthly Notices of the Royal Astronomical Society*, 379, 1546
- Taylor, E. N. et al. 2011, *Monthly Notices of the Royal Astronomical Society*, 418, 1587
- Tempel, E., Stoica, R. S., Saar, E., Martinez, V. J., Liivamägi, L. J., & Castellán, G. 2013, *ArXiv e-prints*, 1308.2533

- Thompson, L. A., & Gregory, S. A. 2011, An Historical View: The Discovery of Voids in the Galaxy Distribution
- Toomre, A., & Toomre, J. 1972, *Astrophysical Journal*, 178, 623
- Tucker, D. L. et al. 2000, *Astrophysical Journal*, Supplement, 130, 237
- Valentinuzzi, T. et al. 2011, *Astronomy and Astrophysics*, 536, A34
- van de Weygaert, R., & van Kampen, E. 1993, *Monthly Notices of the Royal Astronomical Society*, 263, 481
- Velander, M. et al. 2014, *Monthly Notices of the Royal Astronomical Society*, 437, 2111
- Vettolani, G. et al. 1997, *Astronomy and Astrophysics*, 325, 954
- Warren, S. J. et al. 2007, *ArXiv Astrophysics e-prints*
- Warren, S. J. et al. 2007, *Monthly Notices of the Royal Astronomical Society*, 375, 213
- Watson, C. R. et al. 2009, *Astrophysical Journal*, 696, 2206
- Weinmann, S. M., van den Bosch, F. C., Yang, X., & Mo, H. J. 2006, *Monthly Notices of the Royal Astronomical Society*, 366, 2
- White, S. D. M. 1979, *Monthly Notices of the Royal Astronomical Society*, 186, 145
- White, S. D. M., & Rees, M. J. 1978, *Monthly Notices of the Royal Astronomical Society*, 183, 341
- Williams, R. J., Mulchaey, J. S., Kollmeier, J. A., & Cox, T. J. 2010, *Astrophysical Journal*, Letters, 724, L25
- Yee, H. K. C. et al. 2000a, *Astrophysical Journal*, Supplement, 129, 475
- . 2000b, *Astrophysical Journal*, Supplement, 129, 475
- Zappacosta, L., Nicastro, F., Maiolino, R., Tagliaferri, G., Buote, D. A., Fang, T., Humphrey, P. J., & Gastaldello, F. 2010, *Astrophysical Journal*, 717, 74
- Zehavi, I. et al. 2011, *Astrophysical Journal*, 736, 59
- . 2005, *Astrophysical Journal*, 630, 1
- Zel'dovich, Y. 1970, *Astronomy and Astrophysics*, 5