MULTISCALE GEOMETRIC ANALYSIS OF THE 2DF DATA

JEAN-LUC STARCK
DAPNIA/SEDI-SAP, Service d'Astrophysique, CEA-Saclay, 91191 Gif-sur-Yvette, France
E-mail: jstarck@cea.fr

VICENT J. MARTíNEZ
Observatori Astronòmic, Universitat de València, Apartat de Correus 22085, E-46071 València, Spain
E-mail:martinez@uv.es

ENN SAAR
Tartu Observatoorium, Tõravere, 61602, Estonia
E-mail:saar@aai.ee

In order to investigate the recent discovery of a discrepancy between the 2DF data and the Λ CDM simulations, we have applied a Multiscale Geometric Analysis (MGA) on the 2DF data. We report in this paper the results of this study.

1 Introduction

The distribution of galaxies seen in the available galaxy redshift catalogues shows complex structures such as voids, filaments, walls, or clusters. In order to compare the data with the simulations resulting from the cosmological models, we need to extract statistical or morphological information from the data. The two-point correlation (2CF), extensively used by Peebles [1], is certainly the most popular indicator to describe the spatial clustering of the galaxy distribution. Many different 2CF estimators have been proposed in the past [2,3,4]. A detailed description of these estimators may be found in refs. [5,6] and they are compared in refs. [7,8]. The two-point correlation function can been generalized to the N-point correlation function [9,10]. Other statistical measures to characterize the spatial distribution of points have also been developed, such as the void probability function [11], the multifractal approach [12], the Minkowski functionals [13,14], the J function [15,16], the minimal spanning tree [17,18,19], or the wavelet [20,21,22,23,24]. The Sloan Digital Sky Survey (Early Data Release) has recently been analyzed using a 3D Genus Statistics [25] and results were consistent with that predicted by simulations of a Λ-dominated spatially-flat cold dark matter model. The Genus is calculated by (i) convolving the data by a kernel, generally a Gaussian, (ii) setting to zero all values under a threshold ν in the obtained distribution, and (iii) taking the difference D between the number of holes and the number of isolated regions. The Genus curve $G(\nu)$ is obtained by varying the threshold level ν. The first step of the algorithm, the convolution by a Gaussian, may be dramatic for the description of filaments, which are spread out along all directions [26]. Is has been shown that replacing the Gaussian smoothing by a wavelet denoising leads to much more reliable results [26]. The wavelet-Genus method has been applied to both the 2DF data and a set of 22 Λ CDM simulations and the 2DF genus curve is clearly not compatible with the simulations [26]. Figure 1 shows the wavelet genus function of the 2DF data. The solid line is the genus for the 2DF data and the crosses are the mean genus for 22 realizations of the λ-CDM simulations with the 3σ error bars.

Question: How to explain the discrepancy ? In [26], the discrepancy was attributed to the presence of a super cluster in the data, which was not in the simulation. Therefore, even if there is a discrepancy, the λ-CDM model is still considered as a good model for representing the 2DF data. In order to better investigate this difference between the 2DF and the λ-CDM simulations, we have achieved a Multiscale Geometric Analysis (MGA) [27] of the 2DF data. Section 2 presents the MGA approach and the data (simulations and 2DF data) are described in section 3. Results are given in section 4.

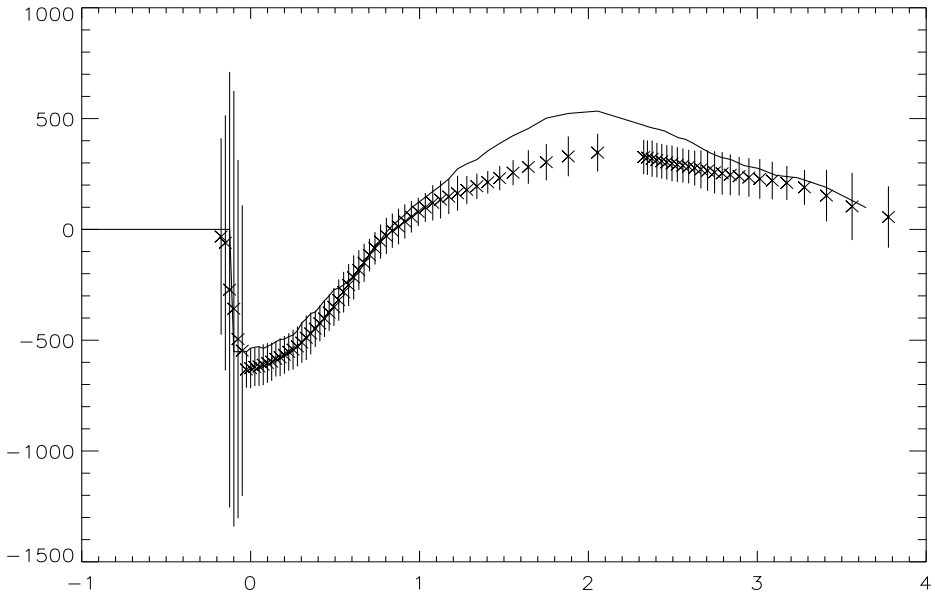

Figure 1. Wavelet Denoising+Genus: 2DF and Lambda-CDM Sim. The solid line is the genus for the wavelet denoised 2DF data and the crosses are the mean genus for 22 wavelet denoised realizations of the λ-CDM simulations with the 3σ error bars.

2 Multiscale Geometric Analysis

As the data contain clusters, filaments and sheets, it has recently been proposed to analyze the data with three multiscale transforms, each of them being well adapted for representing only one kind of feature [27]. Wavelets represent well isotropic features (i.e. clusters in 3D), while more recent geometric multiscale methods such the beamlet and the ridgelet represent well data containing respectively filaments and sheets.

For each $a > 0$, $b_1, b_2, b_3 \in \mathbf{R}^3$, the *wavelet* is defined by

$$\psi_{a,b_1,b_2,b_3} : \mathbf{R}^3 \rightarrow \mathbf{R}$$

$$\psi_{a,b_1,b_2,b_3}(x_1,x_2,x_3) = a^{-3/2} \cdot \psi(\frac{x_1-b_1}{a}, \frac{x_2-b_2}{a}, \frac{x_3-b_3}{a})$$

The *ridgelet* function is defined by:

$$\psi_{a,b,\theta_1,\theta_2} : \mathbf{R}^3 \rightarrow \mathbf{R}$$

$$\psi_{a,b,\theta_1,\theta_2}(x_1,x_2,x_3) = a^{-1/2} \cdot \psi((x_1\cos\theta_1\cos\theta_2 + x_2\sin\theta_1\cos\theta_2 + x_3\sin\theta_2 - b)/a)$$

and the *beamlet* function is defined by:

$$\psi_{a,b_1,b_2,\theta_1,\theta_2} : \mathbf{R}^3 \rightarrow \mathbf{R}$$

$$\psi_{a,b_1,b_2,\theta_1,\theta_2}(x_1,x_2,x_3) = a^{-1/2} \cdot \psi((-x_1\sin\theta_1 + x_2\cos\theta_1 + b_1)/a, (x_1\cos\theta_1\cos\theta_2 + x_2\sin\theta_1\cos\theta_2 - x_3\sin\theta_2 + b_2)/a)$$

Figure 2 shows an example of 3D wavelet function and Figure 3 shows respectively examples of ridgelet function (left) and beamlet function (right). The ridgelet function is a wavelet function in the direction defined by the line (θ_1, θ_2), and it is constant along the orthogonal plane to this line. The beamlet function is constant along lines of direction (θ_1, θ_2), and a 2D wavelet function along plane orthogonal to this direction. More details about the implementation of these 3D multiscale transforms can be found in ref [27].

Local 3D Ridgelet and Beamlet Transform

The ridgelet (resp. beamlet) transform is optimal to find sheets (resp. filaments) of the size of the cube. To detect smaller sheets (resp. filaments), a partitioning must be introduced [28]. The cube c is decomposed into blocks of lower side-length b so that for a $N \times N \times N$ cube, we count N/b blocks in each direction. After the block partitioning, the transform is tuned for sheets (resp. filaments) of size $b \times b$ (resp. b) and of thickness a_j, a_j corresponding to the different dyadic scales used in the transformation.

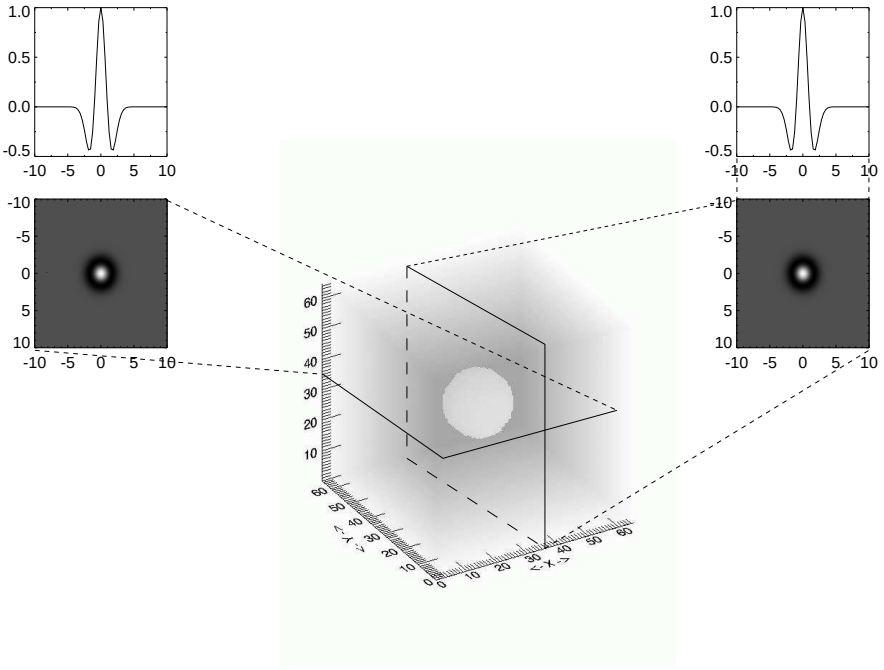

Figure 2. Example of wavelet function.

MGA statistic

Hence, we have three different multiscale transforms, and for the beamlet and the ridgelet transform, we can also use several block sizes when analyzing the data. In the following, we will use the following seven decompositions:

1. 3D Isotropic Wavelet Transform with 4 dyadic scales.
2. 3D Ridgelet Transform using a block size of 8 Mpc and two scales. Here the scale is related to the width of the ridgelet function, its length being fixed by the block size.
3. 3D Ridgelet Transform using a block size of 16 Mpc and three scales.
4. 3D Ridgelet Transform using a block size of 32 Mpc and three scales.
5. 3D Beamlet Transform using a block size of 8 Mpc and two scales. Here the scale is related to the width of the beamlet function, its length being fixed by the block size.
6. 3D Beamlet Transform using a block size of 16 Mpc and three scales.
7. 3D Beamlet Transform using a block size of 32 Mpc and three scales.

For each scale of each transform, we calculate the Kurtosis value (i.e. $K = \frac{1}{N\sigma^4}\sum_k (x_k - \bar{x})^4 - 3 = \frac{1}{\sigma^4}(\bar{x^4} - 4\bar{x}\bar{x^3} + 6\bar{x^2}\bar{x}^2 - 3\bar{x}^4) - 3)$. Positive K implies a higher peak and larger wings than the Gaussian distribution with the same mean and variance. Negative K means a wider peak and shorter wings.

3 2DF Data

The best available redshift catalog to study morphology of the galaxy distribution at present is the 2dF Galaxy Redshift Survey (2dFGRS) [29]. It fills large compact volume(s) in space and includes more than a quarter of million of galaxies. This is a flux-limited catalog and therefore the density of galaxies decreases with distance. For statistical analysis of such of surveys, a weighting scheme that compensates for the missing galaxies at large distances, has to be used. Usually, each galaxy is weighted by the inverse of the selection function [6]. However, the resulting densities will have different resolution at different locations, and will not be suitable for morphological studies.

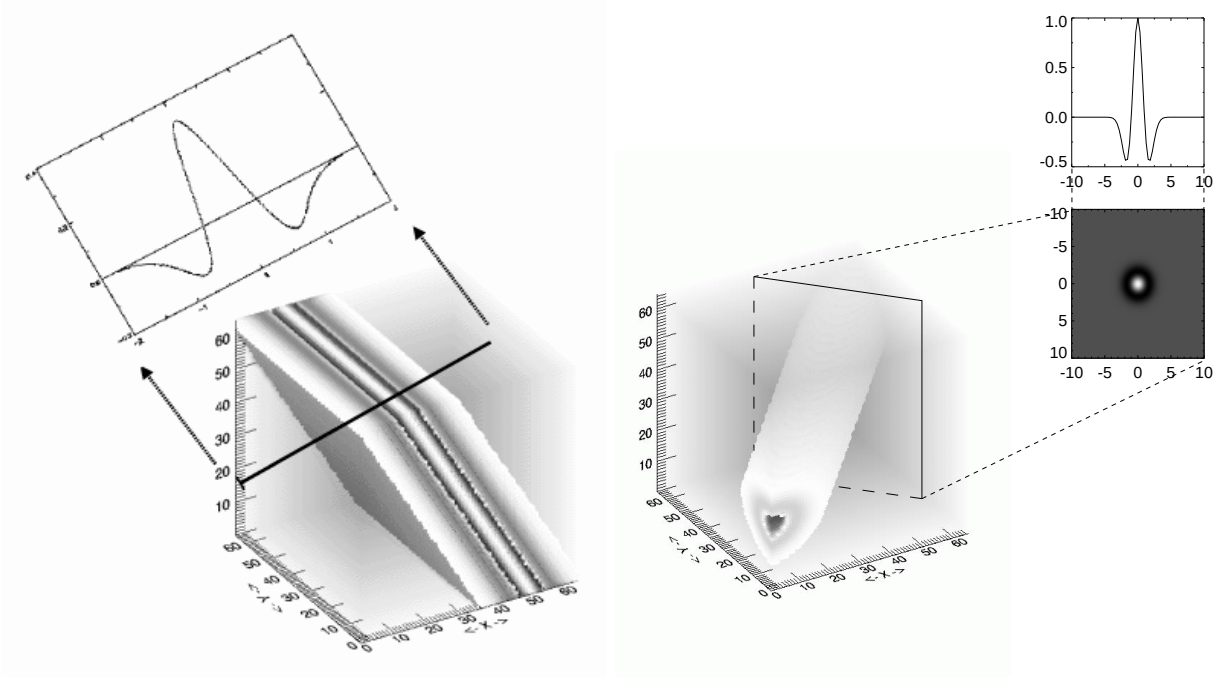

Figure 3. Examples of ridgelet function (left) and beamlet function (right).

At the cost of discarding many surveyed galaxies, one can alternatively use volume-limited samples. In this case, the variation in density at different locations depends only on the fluctuations of the galaxy distribution itself. We have used the volume-limited samples prepared by the 2dF team for scaling studies [30,31], and kindly sent to us by Darren Croton. As our basic sample, we chose the catalog with absolute luminosities in the range $-19 > M_{B_J} - 5\log_{10} h > -20$ (the type dependent $k + e$ correction [32] has been applied to the magnitudes). This sample contains galaxies with luminosity around L_*. This catalog is the largest of the 2dF volume-limited catalogs, and as [33] point out, it provides optimal balance between the surveyed volume and the number density of galaxies. Although the catalog does not suffer from luminosity incompleteness, it is slightly spectroscopically incomplete, mainly due to missing galaxies because of fiber collisions. The incompleteness parameter has been determined by every galaxy by the 2dF team; when calculating densities, each galaxy can be weighted by the inverse of this parameter.

We split the volume-limited sample into the Northern and Southern subsamples, and cut off the numerous whiskers in the plane of the sky to obtain compact volumes. The geometry of the Northern sample is similar to a flat slice, while the Southern sample is enclosed between two cones of opening angles of 64.5° and 55.5°. When we tried to cut cuboidal volumes (bricks) from the Southern sample cone, we ended up with small brick volumes. We report in this paper only the analysis of the Northern subsamples.

In order to obtain a compact volume, we choose the angular limits for the Northern sample as $-4.5° \leq \delta \leq 2.5°$ and $149.0° \leq \alpha \leq 209.0°$. The slice lies between two cones defined by the δ limits. The right ascension limits cut the cones by planes from both sides, and there are two additional cuts by two spheres. The radii of the spheres are fixed by the original data, and depend only on the chosen absolute magnitude limits (and on the cosmological model). For our sample they are: $R_1 = 61.1\ h^{-1}$ Mpc, $R_2 = 375.6\ h^{-1}$ Mpc.

As this sample is pretty flat, we cut from it a maximal volume cuboidal window, a "brick" with dimensions of $254.0 \times 133.1 \times 31.1\ h^{-1}$ Mpc, with 8487 galaxies (see Fig. 4). This gives for the per-particle-volume size $d = 5.0\, h^{-1}$ Mpc.

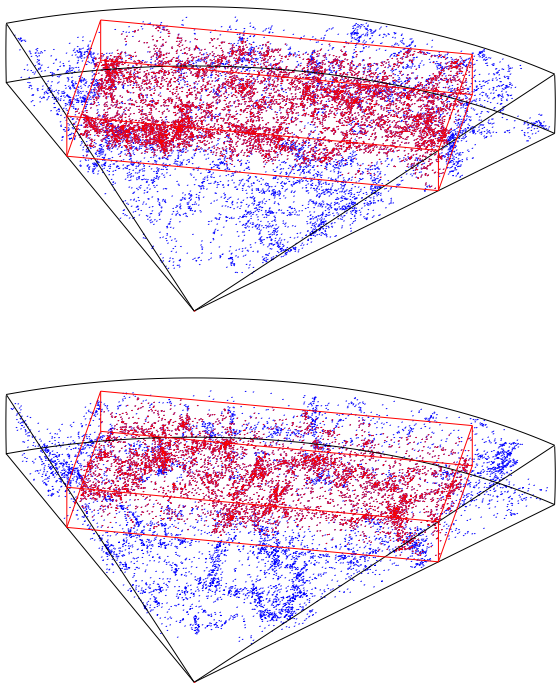

Figure 4. The volume-limited cuboidal sample analyzed in this paper drawn from the Northern slice of the 2dFGRS (top) and from a mock realization.

3.1 Mock catalogs

In order to estimate sample errors of the Minkowski functionals, we use mock catalogs, provided by the 2dF team. [32] created 22 mock catalogs for the 2dFGRS that have been used by the 2dfGRS team to measure the influence of cosmic variance of different statistics, as correlation functions, counts-in-cells, the void probability function, clustering of groups, etc. [30,31,33,34]. The mock catalogs were extracted from the Virgo Consortium ΛCDM Hubble volume simulation, and a biasing scheme described in [35] was used to populate the dark matter distribution with galaxies. The catalogs were created by placing observers in the Hubble volume, applying the radial and angular selection functions of the 2dFGRS, and translating the positions and velocities of galaxies into redshift space. No luminosity clustering dependence is present in the mock catalogs.

The mock catalogs represent typical volumes of space. The real 2dF catalog, however, includes two superclusters, one in the Northern, another in the Southern subsample (see a thorough discussion in [31]). The Northern supercluster is especially prominent in our $M \in [-19, -20]$ catalog; all mock samples for this catalog have less galaxies than the 2dF sample, as the mocks were normalized by the total number of galaxies in both subsamples. We cut mock bricks from the mock samples, too, as we did for the real 2dF data; the mean number of galaxies in the mock bricks is 1.36 times smaller than in the 2dF brick.

4 MGA and the 2DF

We have applied the seven decompositions to the 2DF data and the 22 mock catalogs.

Figure 5 shows the Kurtosis for each scale of the wavelet transform. Crosses with the 2σ (continuous line) and 3σ (dotted line) error bars represent the mean kurtosis for the 22 simulation and the triangle represents the kurtosis relative to the wavelet scale of the 2DF data. Figure 7 shows the kurtosis for the three beamlet transforms and Figure 6 shows the kurtosis for the three ridgelet transforms.

We can see that a strong discrepancy between the data and the simulations appears at the finest scale of the wavelet transform. It is also visible at

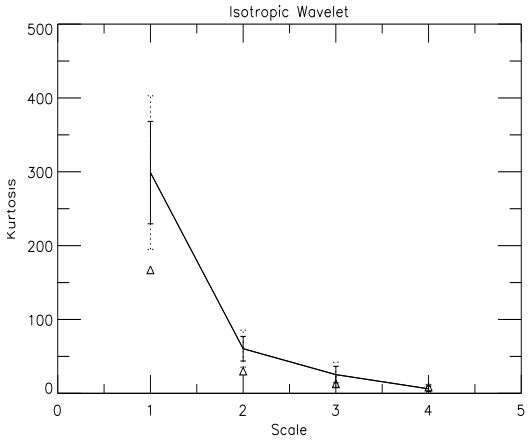

Figure 5. Isotropic WT Kurtosis versus scale. Crosses with the 2σ (continuous line) and 3σ (dotted line) error bars represent the mean kurtosis for the 22 simulation and the triangle represents the kurtosis relative to the wavelet scale of the 2DF data. The x-axis is the wavelet scale, corresponds to the analysis of isotropic structures of size 2^x Mpc.

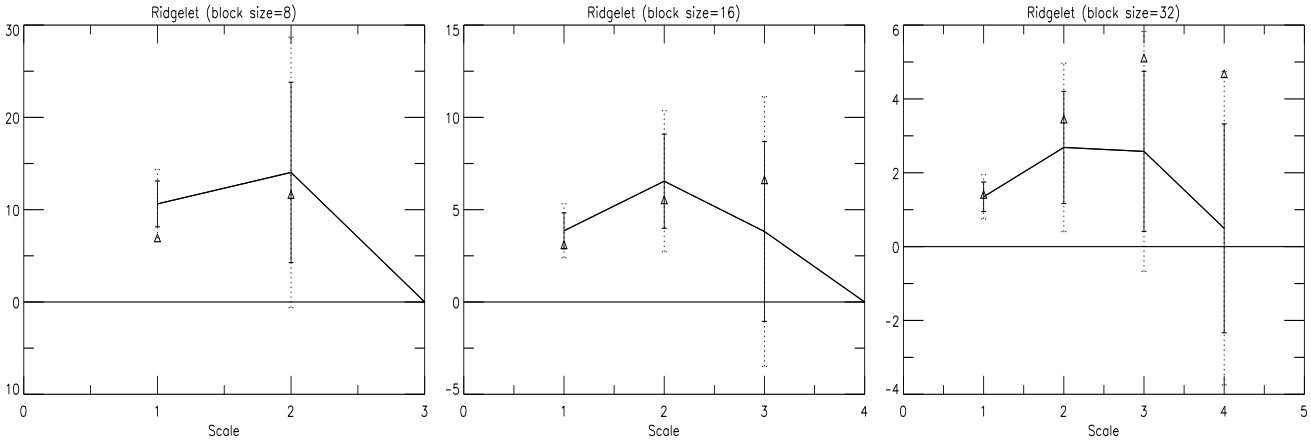

Figure 6. Ridgelet Kurtosis. The x-axis is the ridgelet scale, and corresponds to the analysis of structures of size $8 \times 8 \times 2^x$ Mpc (left), $16 \times 16 \times 2^x$ Mpc (middle) and $32 \times 32 \times 2^x$ Mpc (right).

the finest scale of the beamlet transform (block size 8). The last scale of the ridgelet transform (block size =32) shows also a difference, however not at a 3σ level. The main difference between the data and the simulations is clearly related to the smallest scales. As it is the wavelets which detect this difference, it is certainly the distribution of clusters (and not the distribution of filaments and walls) which is different.

A kurtosis exess in the simulated data can be due to a larger number of clusters and/or a few clusters with a larger amplitude. The first possibility would be in contradiction with the genus curve (see Figure 1) which indicates that the real high-density haloes (galaxy groups) are more concentrated than the mock galaxy groups. The second hypothesis, also supported by a visual inspection of the first and second wavelet scales, seems more adequate. If the real data contains more faint clusters, the simulations contain a few more prominent clusters which create a kurtosis exess.

The supercluster has been suspected to be responsible of the genus curve difference between the simulation and the data in [26], but the MGA analysis leads to the conclusion that the problem is more likely related to the non-linear regime in the simulation which does not reflect the observations. It is probably due to individual simulated dark-matter haloes badly populated with galaxies, but it could

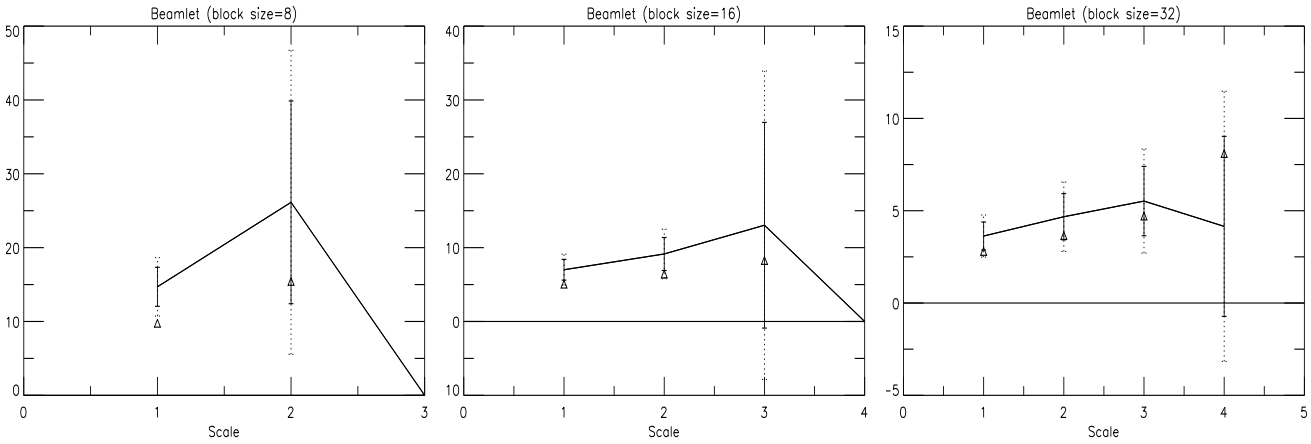

Figure 7. Beamlet Kurtosis. The x-axis is the beamlet scale, and corresponds to the analysis of structures of size $8 \times 2^x \times 2^x$ Mpc (left), $16 \times 2^x \times 2^x$ Mpc (middle) and $32 \times 2^x \times 2^x$ Mpc (right).

also be due to some departure from the simplest models of primordial fluctuations of dark energy.

4.1 MGA:EARLY-LATE Type Galaxies

In this section, we have separated the 2DF data set into two parts, the early type galaxies (ETG) (3826 galaxies) and the late type galaxies (3913 galaxies) (LTG). It is well known that ETG are more clustered than LTG and this has also already be seen in the 2DF data [31]. However, we can wonder if the ELT-LTG clustering property is dependent on the type of structure. For instance, is this ELT-LTG property more important in cluster than in filaments or walls ?

In order to answer this question, we have performed a separate MGA analysis of the two catalogs, i.e. we have applied the seven transformations to both catalogs and calculated the kurtosis in the different scales. In order to normalize all values, we have also computed the kurtosis for 100 simulated cubes which contains each around 3870 galaxies randomly distributed (Poisson noise). A mean kurtosis and a standard deviation has been calculated for each transform and for each scales. The kurtosis derived from the two catalogs have been normalized using the mean values and the standard deviation values.

Figures 8,9 and 10 shows the kurtosis for the two catalogs and the different transforms. The solid lines correspond to the ETG kurtosis and the dashed line to the LTG kurtosis. It is clear that for all transforms and all scales, the ETG presents a higher kurtosis than the LTG. This suggests that in all types of structures (filaments, clusters, walls), the ETG are more concentrated than the LTG. Both the ridgelet kurtosis and the beamlet kurtosis increase when the block size increases. This suggest that filaments and walls of the size of at least 32 Mpc exists in the data. Filaments with a width between 2 and 4 Mpc dominates (second scale of the beamlet transform) while walls seems to be thinner (1 or 2 Mpc) as the first scale of the ridgelet transform always presents a higher kurtosis.

In order to evaluate if a difference in the respective concentration exits between the three kind of features, we keep for each of the seven transformations only the higher normalized kurtosis along the scales. Hence, we built the variable $K_{max}^{(ETG)}(t)$, t being the transform number (i.e. $t = 1, ..7$ for respectively the wavelet transform, the ridgelet transform for a block size equals to 8,16,32, and the beamlet transform for a block size equals to 8,16,32). Figure 11 shows the ratio $R_K(t) = \frac{K_{max}^{(ETG)}(t)}{K_{max}^{(LTG)}(t)}$. For all the transforms, the ratio $R_K(t)$ is between two and three, which shows a remarkable stability in the clustering properties. It seems that the rates of ETG and LTG are relatively the same in all kinds of structures (i.e. filaments, walls and clusters).

Acknowledgments

We thank Darren Croton for providing us with the 2dF volume-limited data and explanations, David

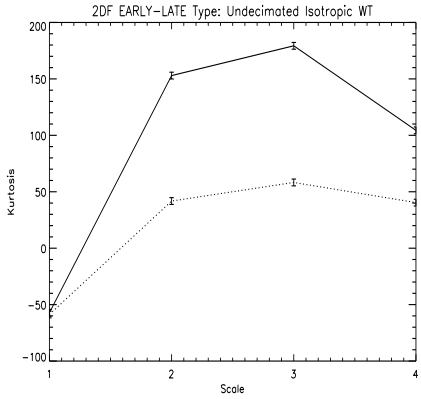

Figure 8. Isotropic WT Kurtosis of both the ETG (solid line) and the LTG (dotted line).

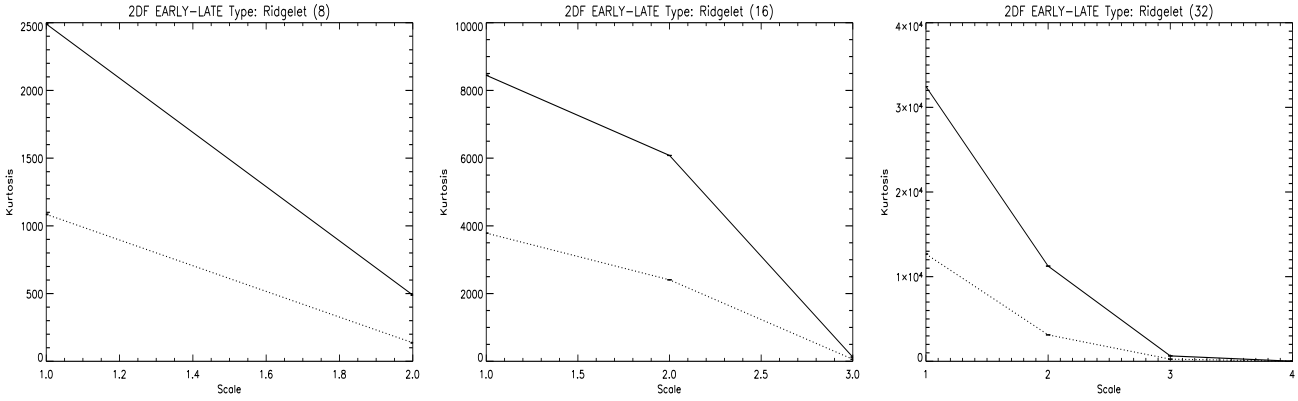

Figure 9. Ridgelet Kurtosis of both the ETG (solid line) and the LTG (dotted line).

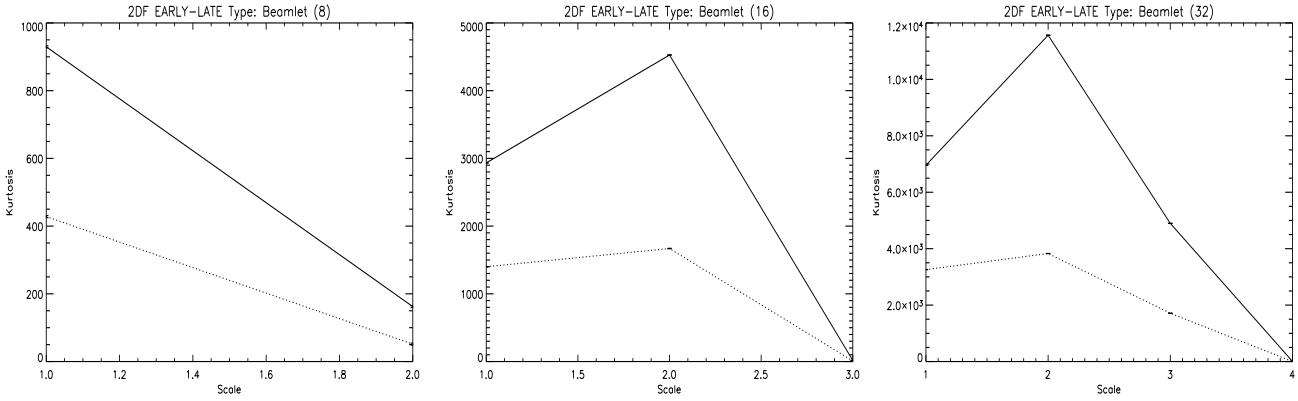

Figure 10. Beamlet Kurtosis of both the ETG (solid line) and the LTG (dotted line).

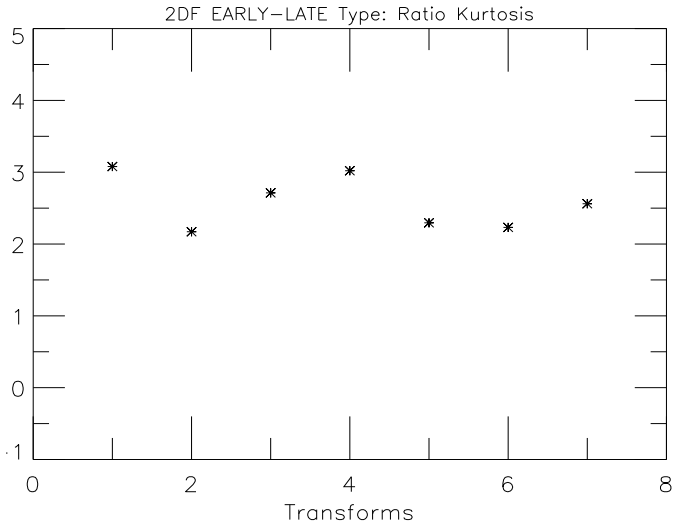

Figure 11. Kurtosis ratio $R_K = K_{max}^{(ETG)}/K_{max}^{(LTG)}$ between ETG and LTG versus the transformation. x-axis corresponds respectively for $x = 1$ to 7 to the wavelet transform, the ridgelet transform with block size equals 8, 16, 32 and the beamlet transform with block size equals 8, 16, 32.

Donoho and Joe Silk for discussions. This work has been supported by the University of Valencia through a visiting professorship for Enn Saar, by the Spanish MCyT project AYA2003-08739-C02-01 (including FEDER), by the Generalitat Valenciana project GRUPOS03/170, and by the National Science Foundation grant DMS-01-40587 (FRG), and by the Estonian Science Foundation grant 4695.

References

1. P. Peebles, *The Large-Scale Structure of the Universe* (Princeton University Press, 1980).
2. M. Davis and P. Peebles, *Astrophysical Journal* **267**, 465(April 1983).
3. A. Hamilton, *Astrophysical Journal* **417**, p. 19(November 1993).
4. S. Landy and A. Szalay, *Astrophysical Journal* **412**, 64(July 1993).
5. M. Kerscher, *Astronomy and Astrophysics* **343**, 333(March 1999).
6. V. J. Martínez and E. Saar, *Statistics of the Galaxy Distribution* (Chapman and Hall/CRC press, Boca Raton, 2002).
7. M. Pons-Bordería, V. Martínez, D. Stoyan, H. Stoyan and E. Saar, *Astrophysical Journal* **523**, 480 (1999).
8. M. Kerscher, I. Szapudi and A. S. Szalay, *Astrophysical Journal Letters* **535**, L13(May 2000).
9. S. Szapudi and A. S. Szalay, *Astrophysical Journal Letters* **494**, p. L41(February 1998).
10. P. Peebles, The galaxy and mass N-point correlation functions: a blast from the past, in *Historical Development of Modern Cosmology*, eds. V. Martínez, V. Trimble and M. Pons-Bordería (ASP Conference Series, Astronomical Society of the Pacific, 2001).
11. S. Maurogordato and M. Lachieze-Rey, *Astrophysical Journal* **320**, 13 (1987).
12. V. J. Martínez, B. J. T. Jones, R. Domínguez-Tenreiro and R. van de Weygaert, *Astrophysical Journal* **357**, 50 (1990).
13. K. R. Mecke, T. Buchert and H. Wagner, *Astronomy and Astrophysics* **288**, 697(August 1994).
14. M. Kerscher, Statistical analysis of large-scale structure in the Universe, in *Statistical Physics and Spatial Satistics: The Art of Analyzing and Modeling Spatial Structures and Pattern Formation*, eds. K. Mecke and D. Stoyan (Lecture Notes in Physics 554, 2000).
15. M. V. Lieshout and A. Baddeley, *Statistica Neerlandica* **50**, 344 (1996).
16. M. Kerscher, M. J. Pons-Bordería, J. Schmalzing, R. Trasarti-Battistoni, T. Buchert, V. J. Martínez and R. Valdarnini, *Astrophysical Journal* **513**, 543 (1999).

17. S. P. Bhavsar and R. J. Splinter, *Monthly Notices of the Royal Astronomical Society* **282**, 1461(Octtober 1996).
18. L. G. Krzewina and W. C. Saslaw, *Monthly Notices of the Royal Astronomical Society* **278**, 869(February 1996).
19. A. G. Doroshkevich, D. L. Tucker, R. Fong, V. Turchaninov and H. Lin, *Monthly Notices of the Royal Astronomical Society* **322**, 369(April 2001).
20. E. Escalera, E. Slezak and A. Mazure, *Astronomy and Astrophysics* **264**, 379(Octtober 1992).
21. E. Slezak, V. de Lapparent and A. Bijaoui, *Astrophysical Journal* **409**, 517 (1993).
22. V. J. Martínez, S. Paredes and E. Saar, *Monthly Notices of the Royal Astronomical Society* **260**, 365 (1993).
23. A. Pagliaro, V. Antonuccio-Delogu, U. Becciani and M. Gambera, *Monthly Notices of the Royal Astronomical Society* **310**, 835(December 1999).
24. T. Kurokawa, M. Morikawa and H. Mouri, *Astronomy and Astrophysics* **370**, 358(May 2001).
25. C. Hikage, Y. Suto, I. Kayo, A. Taruya, T. Matsubara, M. S. Vogeley, F. Hoyle, J. R. I. Gott and J. Brinkmann, *Publications of the Astronomical Society of the Japan* **54**, 707(Octtober 2002).
26. V. J. Martínez, J.-L. Starck, E. Saar, D. L. Donoho, S. C. Reynolds, P. de la Cruz and S. Paredes, *Astrophysical Journal* **634**, 744(December 2005).
27. J.-L. Starck, V. Martinez, D. Donoho, O. Levi, P. Querre and E. Saar, *Eurasip Journal* **15**, 2455 (2005).
28. E. J. Candès, *Applied and Computational Harmonic Analysis* **6**, 197 (1999).
29. M. Colless *et al.*, *The 2dF Galaxy Redshift Survey: Final Data Release*, tech. rep., astro-ph/0306581 (2003).
30. D. J. Croton *et al.*, *Monthly Notices of the Royal Astronomical Society* **352**, 1232(August 2004).
31. D. J. Croton *et al.*, *Monthly Notices of the Royal Astronomical Society* **352**, 828(August 2004).
32. P. Norberg, S. Cole, C. M. Baugh, C. S. Frenk, I. Baldry, J. Bland-Hawthorn, T. Bridges, R. Cannon, M. Colless, C. Collins, W. Couch, N. J. G. Cross, G. Dalton, R. De Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, K. Glazebrook, C. Jackson, O. Lahav, I. Lewis, S. Lumsden, S. Maddox, D. Madgwick, J. A. Peacock, B. A. Peterson, W. Sutherland and K. Taylor, *Monthly Notices of the Royal Astronomical Society* **336**, 907(November 2002).
33. C. M. Baugh, D. J. Croton, E. Gaztañaga, P. Norberg, M. Colless, I. K. Baldry, J. Bland-Hawthorn, T. Bridges, R. Cannon, S. Cole, C. Collins, W. Couch, G. Dalton, R. De Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, C. S. Frenk, K. Glazebrook, C. Jackson, O. Lahav, I. Lewis, S. Lumsden, S. Maddox, D. Madgwick, J. A. Peacock, B. A. Peterson, W. Sutherland and K. Taylor, *Monthly Notices of the Royal Astronomical Society* **351**, L44(June 2004).
34. N. D. Padilla, C. M. Baugh, V. R. Eke, P. Norberg, S. Cole, C. S. Frenk, D. J. Croton, I. K. Baldry, J. Bland-Hawthorn, T. Bridges, R. Cannon, M. Colless, C. Collins, W. Couch, G. Dalton, R. De Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, K. Glazebrook, C. Jackson, O. Lahav, I. Lewis, S. Lumsden, S. Maddox, D. Madgwick, J. A. Peacock, B. A. Peterson, W. Sutherland and K. Taylor, *Monthly Notices of the Royal Astronomical Society* **352**, 211(July 2004).
35. S. Cole, S. Hatton, D. H. Weinberg and C. S. Frenk, *Monthly Notices of the Royal Astronomical Society* **300**, 945(November 1998).