

Galaxy filaments as pearl necklaces[★]

E. Tempel^{1,2}, R. Kipper^{1,3}, E. Saar^{1,4}, M. Bussov^{1,5}, A. Hektor², and J. Pelt¹

¹ Tartu Observatory, Observatooriumi 1, 61602 Tõravere, Estonia
e-mail: elmo.tempel@to.ee

² National Institute of Chemical Physics and Biophysics, Rävåla pst 10, 10143 Tallinn, Estonia

³ Institute of Physics, University of Tartu, 51010 Tartu, Estonia

⁴ Estonian Academy of Sciences, Kohtu 6, 10130 Tallinn, Estonia

⁵ Institute of Mathematical Statistics, University of Tartu, 50409 Tartu, Estonia

Received 17 June 2014 / Accepted 19 September 2014

ABSTRACT

Context. Galaxies in the Universe form chains (filaments) that connect groups and clusters of galaxies. The filamentary network includes nearly half of the galaxies and is visually the most striking feature in cosmological maps.

Aims. We study the distribution of galaxies along the filamentary network, trying to find specific patterns and regularities.

Methods. Galaxy filaments are defined by the Bisous model, a marked point process with interactions. We use the two-point correlation function and the Rayleigh Z-squared statistic to study how galaxies and galaxy groups are distributed along the filaments.

Results. We show that galaxies and groups are not uniformly distributed along filaments, but tend to form a regular pattern. The characteristic length of the pattern is around $7 h^{-1}$ Mpc. A slightly smaller characteristic length $4 h^{-1}$ Mpc can also be found, using the Z-squared statistic.

Conclusions. We find that galaxy filaments in the Universe are like pearl necklaces, where the pearls are galaxy groups distributed more or less regularly along the filaments. We propose that this well-defined characteristic scale could be used to test various cosmological models and to probe environmental effects on the formation and evolution of galaxies.

Key words. methods: numerical – methods: observational – large-scale structure of Universe

1. Introduction

Many cosmological probes have been developed and used to estimate the parameters of the cosmological models describing our Universe. Most of them rely on some aspects of the large-scale structure. The most common probe used to quantify galaxy clustering is the two-point correlation function which was first used decades ago (Davis & Geller 1976; Groth & Peebles 1977; Davis et al. 1988; Hamilton 1988; White et al. 1988; Boerner et al. 1989; Einasto 1991). Some recent examples include Connolly et al. (2002), Zehavi et al. (2005), Contreras et al. (2013), de Simoni et al. (2013), Wang et al. (2013), and the references in these papers. Other examples of cosmological probes are the three-point correlation functions (e.g. McBride et al. 2011; Marín et al. 2013; Guo et al. 2014a) and the power spectrum analysis (Hütsi 2006, 2010; Blake et al. 2010; Balaguera-Antolínez et al. 2011).

It is well known that galaxy filaments are visually the most dominant structures in the galaxy distribution, being part of the so-called cosmic network (Jöeveer et al. 1978; Bond et al. 1996). Presumably, nearly half (~40%) of the galaxies (or mass in simulations) are located in filaments. This number is based on morphological or dynamical classification of the cosmic web (e.g. Jasche et al. 2010; Tempel et al. 2014b; Cautun et al. 2014). Properties of the three-point correlation function also indicate that galaxies tend to populate filamentary structures (Guo et al. 2014a) and, indeed, filaments have been found between galaxy clusters (e.g. Dietrich et al. 2012). Tempel et al. (2014a) show

that filaments extracted from the spatial distribution of galaxies/haloes are also dynamical structures that are well connected with the underlying velocity field. Galaxy filaments that connect groups and clusters of galaxies also affect the evolution of galaxies (e.g. Tempel & Libeskind 2013; Tempel et al. 2013; Zhang et al. 2013) and the distribution of satellites around galaxies (Guo et al. 2014b).

In this paper we study the distribution of galaxies along galaxy filaments to search for regularities in galaxy and group distributions. As was shown decades ago (e.g. Jöeveer et al. 1978; Einasto et al. 1980), this clustering pattern is present in some filaments. In the current study, we use the two-point correlation function (see Peebles 1980) and the Rayleigh Z-squared statistics (see Buccheri 1992) to look for the regularity in the galaxy distribution. We suggest that the regularities we find in the galaxy distribution could be used as cosmological probes for dark energy and dark matter, the mysterious components in the dark energy dominated cold dark matter (Λ CDM) cosmological models.

Throughout this paper we assume the *Wilkinson* Microwave Anisotropy Probe (WMAP) cosmology: the Hubble constant $H_0 = 100 h \text{ km s}^{-1} \text{ Mpc}^{-1}$, with $h = 0.697$, the matter density $\Omega_m = 0.27$ and the dark energy density $\Omega_\Lambda = 0.73$ (Komatsu et al. 2011).

2. Data and methods

2.1. Galaxy and group samples for the SDSS data

The present work is based on the Sloan Digital Sky Survey (SDSS, York et al. 2000) data release 10 (DR10; Ahn et al. 2014). We use the galaxy and group samples as compiled in

[★] Appendices are available in electronic form at <http://www.aanda.org>

Tempel et al. (2014c) that cover the main contiguous area of the survey (the Legacy Survey, approximately 17.5% from the full sky). The flux-limited catalogue extends to the redshift 0.2 ($574 h^{-1}$ Mpc) and includes 588 193 galaxies and 82 458 groups with two or more members.

In Tempel et al. (2014c) the redshift-space distortions, the so-called finger-of-god (FoG) effects, are suppressed using the rms sizes of galaxy groups in the plane of the sky and their rms radial velocities as described in Liivamägi et al. (2012). We calculate the new radial distances for galaxies in groups in order to make the galaxy distribution in groups approximately spherical. We note that such a compression will remove the artificial line-of-sight filament-like structures as shown in Tempel et al. (2012).

2.2. Bisous model: extracting filaments from galaxy distribution

The detection of filaments is performed by applying an object/marked point process with interactions (the Bisous process; Stoica et al. 2005) to the distribution of galaxies. This algorithm provides a quantitative classification which agrees with the visual impression of the cosmic web and is based on a robust and well-defined mathematical scheme. A detailed description of the Bisous model is given in Stoica et al. (2007, 2010) and Tempel et al. (2014b); for convenience, a brief description is given below.

The marked point process we propose for filament detection is different from the ones already used in cosmology. In fact, we do not model galaxies, but the structure outlined by galaxy positions.

This model approximates the filamentary network by a random configuration of small segments (thin cylinders). We assume that locally galaxies may be grouped together inside a rather small cylinder, and such cylinders may combine to form a filament if neighbouring cylinders are aligned in similar directions. This approach has the advantage that it uses only positions of galaxies and does not require any additional smoothing to create a continuous density field.

The solution provided by our model is stochastic. Therefore, we find some variation in the detected patterns for different Markov chain Monte Carlo runs of the model. The main advantage of using such a stochastic approach is the ability to give simultaneous morphological and statistical characterisation of the filamentary pattern.

In practice, after fixing the approximate scale of the filaments, the algorithm returns the filament detection probability field together with the filament orientation field. Based on these data, filament spines are extracted and a filament catalogue is built. Every filament in this catalogue is represented as a spine: a set of points that define the axis of the filament.

The spine detection we use is based on two ideas. First, filament spines are located at the highest density regions outlined by the filament probability maps. Second, in these regions of high probability for the filamentary network, the spines are oriented along the orientation field of the filamentary network. See Tempel et al. (2014b) for more details of the procedure.

2.3. Galaxy filament sample for the SDSS data

The catalogue of filaments was built by applying the Bisous process to the distribution of galaxies as outlined above. The method and parameters are exactly the same as in Tempel et al. (2014b),

where the Bisous model was applied to the SDSS DR8 (Aihara et al. 2011) data. Since the datasets in the SDSS DR8 and DR10 are basically the same, the extracted filaments in DR10 are statistically the same as those presented in Tempel et al. (2014b). The assumed scale (radius) for the extracted filaments is $0.5 h^{-1}$ Mpc. Because the survey is flux-limited, the sample is very diluted farther away. Hence, we were only able to detect filaments in this scale up to the redshift 0.15 ($\approx 450 h^{-1}$ Mpc). The longest filaments in our sample are up to $70 h^{-1}$ Mpc long.

The filaments are extracted from a flux-limited galaxy sample, hence, the completeness of extracted filaments decreases with distance. In Tempel et al. (2014b) we showed that the volume filling fraction of filaments is roughly constant with distance if filaments longer than $15 h^{-1}$ Mpc are considered. Therefore, in the current study we only used filaments longer than this limit. In addition, this choice is justified, since longer filaments allow us to study the distribution of galaxies along the filaments, which is the purpose of this paper. The remaining incompleteness of filaments in our sample is not a problem, because we analysed single filaments. The filaments we used are the strongest filaments (or segments of filaments) in the sample.

To suppress the flux-limited sample effect, we volume-limit the galaxy content for single filaments. To this end, we find for every filament the maximum distance (from the observer) of its galaxies and the corresponding magnitude limit and use only galaxies brighter than this value. Since the majority of filaments extend over a relatively narrow distance interval (always narrower than the length of the filament), the number of excluded galaxies in every filament is small.

To study the galaxy spacing along the filaments, we project every galaxy to the filament spine. Hence, the distance is measured along the filament spine. Figure 1 illustrates extracted galaxy filaments and their spines in the field of galaxies.

For our analysis, we use only filaments that contain at least ten galaxies. When studying galaxy groups, we use only filaments where the number of groups per filament is at least five. We use the groups from the catalogue by Tempel et al. (2014c). The upper panel in Fig. 2 shows the cumulative filament length and the cumulative number of galaxies in filaments as a function of the filament length.

The lower panel in Fig. 2 shows the filament length distribution. In this study, a galaxy is considered to belong to a filament if it is closer than $0.5 h^{-1}$ Mpc to the filament spine. This is also the scale of detected filaments in our Bisous model. In addition, we study the distribution of galaxies that are closer than $0.25 h^{-1}$ Mpc to the filament spine. In Fig. 2 we see that the total filament length in our study is $50\,000 h^{-1}$ Mpc and half of it comes from the filaments shorter than $23 h^{-1}$ Mpc. The numbers of filaments and galaxies in filaments for studied subsamples are given in Table 1.

2.4. Estimating the pair correlation function

To study galaxy correlations in filaments, we use the two-point correlation function $\xi(r)$ that measures the excess probability of finding two points separated by a vector r compared to that probability in a homogeneous Poisson sample (Peebles 1980; Martínez & Saar 2002). For galaxy filaments, the correlation function along the filament can be expressed in terms of the distances between the galaxies measured along the filament. We note that this is not exactly the case when studying a sample from a galaxy redshift survey. The line-of-sight component of the position of a galaxy is derived from the observed redshift, hence, the distance along the line of sight is influenced by redshift

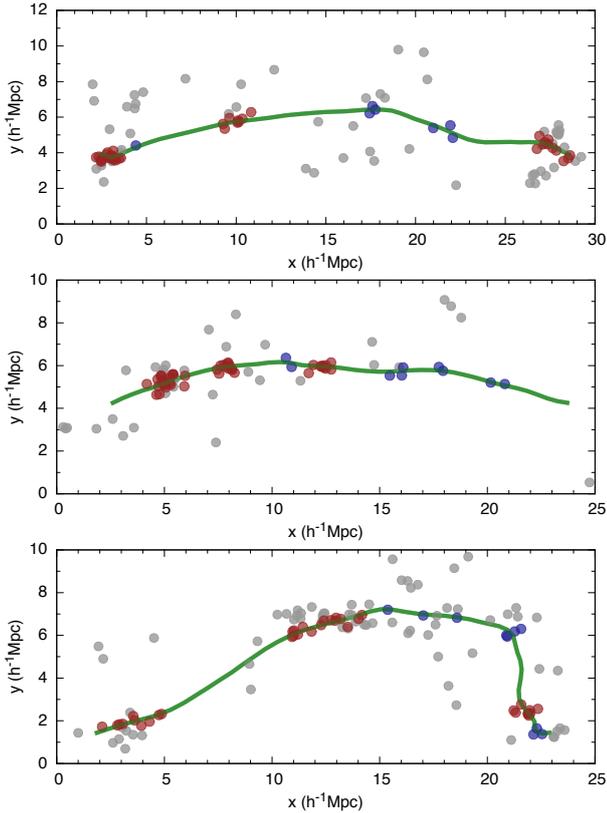


Fig. 1. Examples of filaments and their spines. Red points show galaxies in filaments (closer than $0.5 h^{-1} \text{ Mpc}$ to the filament axis) that are located in groups with five or more members. The group richness is taken from Tempel et al. (2014c). Blue points show other galaxies in filaments. Grey points are background galaxies that are not located in these filaments. We note that some of the background galaxies project to the filament spines, but are actually located farther than $0.5 h^{-1} \text{ Mpc}$ from the spine. The thick green line shows the spine of a filament. To show the scale of structures, Cartesian coordinates are shown around each filament.

distortions. However, since we are interested in scales larger than the group/cluster scale, we can ignore this effect. We measure the distances between galaxies along filament spines. Below (see Fig. 4) we divide our filament sample into two subsamples, the first parallel to the line of sight and second perpendicular to the line of sight, and show that redshift space distortions do not affect our results.

We estimate $\xi(r)$ following the Landy-Szalay border-corrected estimator (Landy & Szalay 1993). We generated a random distribution of points for each filament considered, and estimated the correlation function $\xi(r)$:

$$\widehat{\xi}(r) = 1 + \frac{DD(r)}{RR(r)} - 2 \frac{DR(r)}{RR(r)}, \quad (1)$$

where $DD(r)$, $RR(r)$, and $DR(r)$ are the probability densities for the galaxy-galaxy, random-random, and galaxy-random pairs, respectively, for a pair distance r . The random catalogue is built separately for each filament assuming an uniform distribution of points along the spine of the filament. In this work, we fix the size of the random point set for each filament by $N_{\text{rd}} = 50N$ (N is the number of galaxies in each filament). We tested that our results do not change if we increase the number of random points used to $N_{\text{rd}} = 100N$.

We estimate the probability densities by the kernel method, summing the box spline $B_3(\cdot)$ kernels (Saar et al. 2007) centred

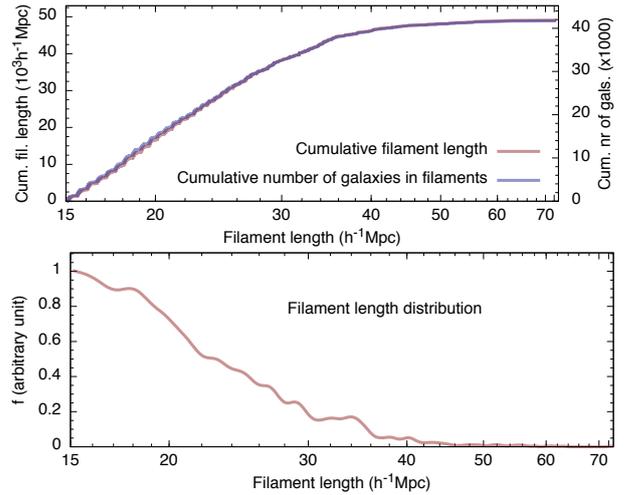


Fig. 2. Upper panel: cumulative filament length and cumulative number of galaxies in filaments as a function of filament length. The lines show the sum of the filament lengths (red line) and the number of galaxies in filaments (blue line) summed over filaments shorter than indicated in the abscissa. On average there is slightly less than one galaxy per $h^{-1} \text{ Mpc}$. Lower panel: filament length distribution in our sample.

Table 1. The numbers of filaments (N_{fil}) and galaxies in filaments (N_{gal}) in various subsamples.

Sample	N_{fil}	N_{gal}
All galaxies ($d_{\text{fil}} < 0.5 h^{-1} \text{ Mpc}$)	2150	41 543
All galaxies ($d_{\text{fil}} < 0.25 h^{-1} \text{ Mpc}$)	1752	26 830
Groups ($N_{\text{rich}} \geq 1$; $d_{\text{fil}} < 0.5 h^{-1} \text{ Mpc}$)	1943	15 874
Groups ($N_{\text{rich}} \geq 1$; $d_{\text{fil}} < 0.25 h^{-1} \text{ Mpc}$)	1493	10 492
Groups ($N_{\text{rich}} \geq 2$; $d_{\text{fil}} < 0.5 h^{-1} \text{ Mpc}$)	759	4667
Groups ($N_{\text{rich}} \geq 2$; $d_{\text{fil}} < 0.25 h^{-1} \text{ Mpc}$)	401	2323
Groups ($N_{\text{rich}} \geq 3$; $d_{\text{fil}} < 0.5 h^{-1} \text{ Mpc}$)	211	1196
Groups ($N_{\text{rich}} \geq 3$; $d_{\text{fil}} < 0.25 h^{-1} \text{ Mpc}$)	111	606

Notes. Only filaments longer than $15 h^{-1} \text{ Mpc}$ and containing at least ten galaxies (or five groups) are considered. N_{rich} denotes the number of galaxies in groups (richness) and d_{fil} is the galaxy/group distance from the filament spine.

at each pair distance, and sampling the distributions at smaller intervals than the kernel width. The kernel width we use is $1.0 h^{-1} \text{ Mpc}$ (for groups it is twice as large). In Martínez et al. (2009) the same method is applied to detect baryon acoustic peaks. Such a method does not require binning the data and with a good choice of the kernel width, can optimally recover the probability distribution; it does not introduce any bias in the correlation estimator. The kernel width influences the estimate the same way as the bin width – if the kernel is too wide, the signal is lost, and in the contrary, if the kernel is too narrow, noise dominates. See Appendix A for the details of the kernel method and for comparison with binning the data.

Clustering in configuration space is subject to the integral constraint which introduces a bias (i.e. a shift in the overall shape) in the estimates of the correlation function. Since the lengths of the filaments are comparable with the studied scale of the correlation function and the length of each filament is different (hence the bias is different), we cannot ignore the integral constraint. The first-order approximation for the integral constraint can be estimated as explained in Roche et al. (1999) and Labatie et al. (2012).

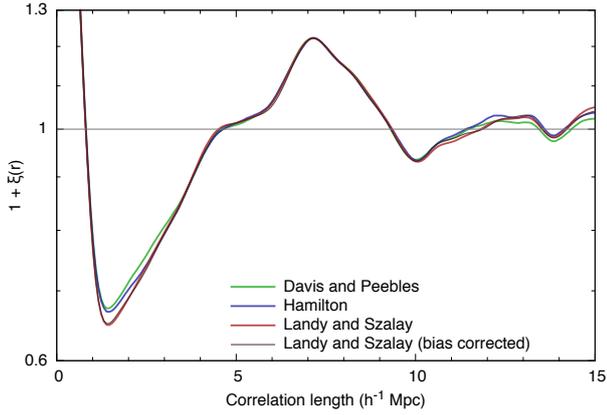


Fig. 3. Galaxy correlation function along filaments. Various correlation function estimates are shown: [Davis & Peebles \(1983, green line\)](#), [Hamilton \(1993, blue line\)](#), and [Landy & Szalay \(1993, red line\)](#). The black line shows the correlation function corrected for integral constraint computed using the Landy-Szalay estimator.

In all cases studied below, we found that the integral constraint correction does not change the features of our measured correlation function, it mostly affects only the general scaling of the correlation function. Figure 3 shows the Landy-Szalay correlation function estimator without (red) and with (black) integral constraint correction. The differences are irrelevant, hence, our results are not sensitive to the details of the estimation of the integral constraint.

Additionally, we checked how the estimated correlation function depends on the used estimator. In Fig. 3 we show the correlation function estimated using the [Davis & Peebles \(1983\)](#) and [Hamilton \(1993\)](#) estimators. The differences between the estimates are very small, and they do not affect the features in the correlation function that we are studying.

Using Eq. (1), we estimate the correlation function for each filament separately. The final pair correlation function, averaged over all filaments, is estimated as a weighted sum

$$\xi^{\text{fil}}(r) = \frac{\sum_{i=1}^{N_{\text{fil}}} \mathbb{1}\{L_{\text{fil},i} > r\} L_{\text{fil},i} \xi_i(r)}{\sum_{i=1}^{N_{\text{fil}}} \mathbb{1}\{L_{\text{fil},i} > r\} L_{\text{fil},i}}, \quad (2)$$

where summation is over all filaments, ξ_i is the correlation function for a single filament, $L_{\text{fil},i}$ is the length of the i th filament, and $\mathbb{1}\{\cdot\}$ is the indicator function that selects only filaments longer than r .

We estimate the statistical error on our $\xi^{\text{fil}}(r)$ measurements with the standard jackknife method (see e.g. [Norberg et al. 2009](#)), where we omit one filament under consideration at a time. In our figures we show the 95% confidence intervals.

2.5. Rayleigh (Z-squared) statistic

To test whether the galaxy distribution might have regularity along the filaments we use the Rayleigh (or Z-squared) statistic (see e.g. [Buccheri et al. 1983](#); [Muno et al. 2003](#)). It is an excellent method when the event rate (in our case, the number of galaxies per filament) is low. The method has been used to detect periodicity in time series for the data in the form of discrete events (photon arrival times) and can be applied to detect periodicity in the galaxy distribution along filaments, where the galaxy positions can be considered as events.

The algorithm works as follows. For each filament, we produce a periodogram using the Z_1^2 (Rayleigh statistic),

$$Z_1^2 = \frac{2}{N} \left[\left(\sum_{j=1}^N \cos \phi_j \right)^2 + \left(\sum_{j=1}^N \sin \phi_j \right)^2 \right], \quad (3)$$

where N is the number of galaxies in a filament and $\phi_j = 2\pi l_j/d$ is the phase value for a galaxy j for a fixed period d ; l_j is a distance of the galaxy j along the filament spine from the beginning of the filament.

To measure the Z-squared statistic for a period d , we use only filaments longer than $2d$. This assures that there are at least two periods for each filament. For a signal resulting purely from a Poisson noise, Z_1^2 has a χ^2 distribution with two degrees of freedom. However, this is only in the case if the number of events is high enough. If the number of events is lower than 100 (as usual in our case), Z_1^2 does not have a χ^2 probability distribution. So, we derive a null-hypothesis probability function using Monte Carlo simulations with N (the number of galaxies in a filament) data points assuming a uniform distribution of points along the spine of the filament. This also allows us to estimate the confidence intervals for our measured signal.

We compute the Z_1^2 statistic as a function of a period d for every filament and then we find the average signal using the filament length L_{fil} as the weight

$$Z_{1,\text{fil}}^2(d) = \frac{\sum_{i=1}^{N_{\text{fil}}} \mathbb{1}\{L_{\text{fil},i} > 2d\} L_{\text{fil},i} Z_{1,i}^2(d)}{\sum_{i=1}^{N_{\text{fil}}} \mathbb{1}\{L_{\text{fil},i} > 2d\} L_{\text{fil},i}}, \quad (4)$$

where $Z_{1,i}^2$ is the Z-squared statistic for the i th filament. The confidence limits are found using a jackknife technique, as we did for the correlation functions.

Examples of how the Z-squared statistic works are given in Appendix B.

3. Results

3.1. Correlations along galaxy filaments

Figure 4 shows the galaxy correlation function along filaments. Three specific features are seen in this correlation function: a maximum near the zero pair distance, a minimum that follows it, and a bump close to $7 h^{-1}$ Mpc. The first maximum is caused by galaxy groups. It shows that galaxies are not distributed uniformly in the space, they form groups and clusters, as it is well known. The minimum next to the first maximum shows that the groups themselves are not distributed uniformly along filaments. It shows that two groups cannot be located extremely close to each other (merging groups are exceptions) and there is a preferred minimum distance between galaxy groups. This is also expected since matter is falling into groups and there is not enough matter in a nearby neighbourhood of groups to form another group. The most interesting feature is the small maximum close to $7 h^{-1}$ Mpc. It shows that galaxies (and also groups) have a tendency to be separated by this distance along the filaments. We will analyse the nature of this bump in more detail below.

Since we are using a galaxy sample where FoG effects are suppressed, we checked whether and how this influences our main results. For that we divided our filament sample into two subsamples: mostly parallel to the line of sight ($\cos i > 0.5$) and mostly perpendicular to the line of sight ($\cos i < 0.5$). Here

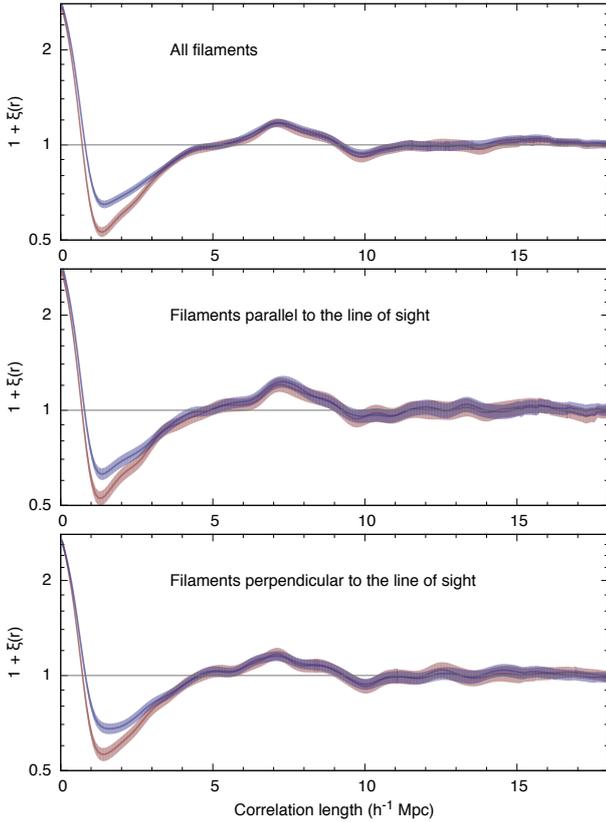


Fig. 4. Galaxy correlation function along filaments. The *upper panel* shows the correlation function for all filaments, the *middle panel* shows this function for filaments that are parallel to the line of sight ($\cos i > 0.5$), and the *lower panel* shows the correlation function for filaments that are perpendicular to the line of sight ($\cos i < 0.5$): i denotes the angle between the line of sight and the filament direction. Blue and red lines with filled regions show the correlation function together with its 95% confidence limits (based on a jackknife estimate) for galaxies closer than 0.5 and $0.25 h^{-1}$ Mpc to the filament axis, respectively.

i denotes the angle between the direction of the filament and the line of sight. The correlation functions for these two subsamples are shown in the middle and lower panels in Fig. 4, respectively. We note that the bump around $7 h^{-1}$ Mpc is visible in both subsamples, hence, it is a real feature. However, the bump is slightly weaker for the filaments perpendicular to the line of sight. So, we also calculated the correlation function for filaments where $\cos i < 0.25$, and the bump remained in the same place. We conclude that the $7 h^{-1}$ Mpc maximum is not affected by the FoG effect. The most noticeable difference between the middle and lower panels in Fig. 4 is that the minimum is slightly stronger for filaments parallel to the line of sight. This is probably because of our FoG suppression. When suppressing FoG effects, we also suppress the line-of-sight scatter of some of the background and foreground galaxies in the group neighbourhood that creates a lack of galaxies close to the groups. To test the effect of FoG suppression, we calculated the correlation signal for galaxies closer than 0.5 and $0.25 h^{-1}$ Mpc to the filament axis (blue and red lines in Fig. 4, respectively). For the smaller radius, the effect of finger-of-god suppression should be smaller. As expected, the prominent features are visible in both subsamples.

In Fig. 5 we show the correlation function for galaxy groups. In this figure, the smoothing scale is twice as large as used for the galaxy correlation functions and we analyse only filaments

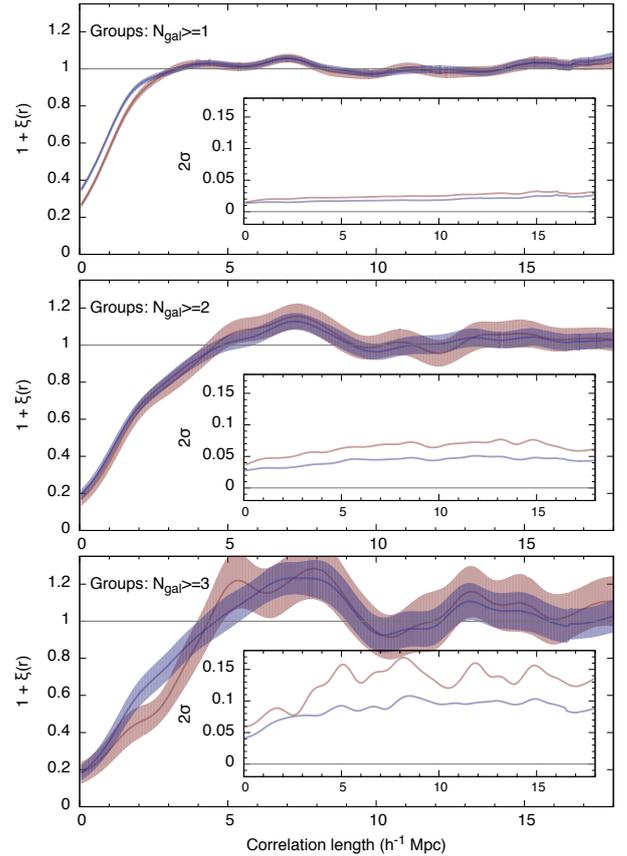


Fig. 5. Correlation function along filaments for groups. The *upper panel* shows the correlation function for all groups and isolated galaxies together, the *middle panel* shows the correlation function only for groups with two or more members, and the *bottom panel* for groups with at least three member galaxies. The group richness (number of galaxies in a group) is taken from Tempel et al. (2014c). Blue lines show the correlation function for groups closer than $0.5 h^{-1}$ Mpc to the filament axis and red lines show the correlation function for groups closer than $0.25 h^{-1}$ Mpc. The shaded regions are the 95% confidence intervals. The insets show the width of the 95% confidence regions as a function of the pairwise distance.

that contain at least five groups. Figure 5 shows the correlation function for groups with different minimum group richness. We see that the bump around $7 h^{-1}$ Mpc is present in every subsample, indicating that groups themselves have a preferred distance between them. Interestingly, the upper panel in Fig. 5 shows a small bump around $4 h^{-1}$ Mpc as well. This scale is also seen when using the Z-squared statistic (see Sect. 3.2).

The biggest difference when comparing the correlation function for galaxies (Fig. 4) and for groups (Fig. 5) is the fact that there is no zero-distance maximum for groups. This shows that groups themselves do not form clusters, which is expected. The minimum at zero distance in the group correlation function shows that there are practically no other groups around each group. The radius of the group influence extends to $4 h^{-1}$ Mpc, after that the correlation function is mostly flat.

Figure 5 also shows that the strength of the bump at $7 h^{-1}$ Mpc is higher if isolated galaxies ($N_{\text{gal}} = 1$) are excluded. This indicates that the bump is related to galaxy clustering. Unfortunately, the current dataset is too small to adequately quantify the dependence of the amplitude and significance of the bump on group richness.

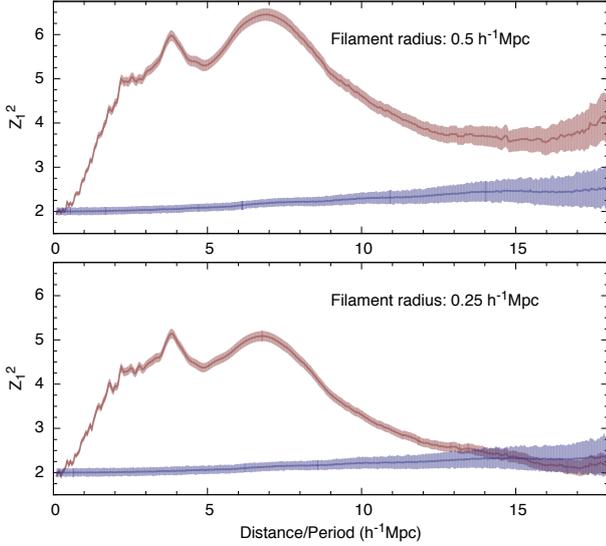


Fig. 6. Dependence of the Rayleigh (Z -squared) statistic Z_1^2 on distance (period). The *upper panel* shows the results for galaxies closer than $0.5 h^{-1}$ Mpc to the filament axis and the *lower panel* shows the results for galaxies closer than $0.25 h^{-1}$ Mpc. The red line shows the Z_1^2 statistic together with the jackknife 95% confidence estimate. The blue line shows the results from Monte Carlo simulations for the null hypothesis together with the 95% confidence limits.

Additionally, we checked how the correlation signal depends on the galaxy luminosity, colour or morphology. We found a weak dependency on colour and luminosity, so that the bump at $7 h^{-1}$ Mpc is slightly stronger for redder and more luminous galaxies. The interpretation of this is unclear. On one hand, it could be related to galaxy evolution and in this case could provide information on galaxy bias. On the other hand, it might just be the result of galaxy clustering. More luminous and redder galaxies are more clustered and the bump at $7 h^{-1}$ Mpc is also stronger for groups.

3.2. Regularity of the galaxy distribution along filaments

Figure 6 shows the Z -squared statistic for all galaxies closer than $0.5 h^{-1}$ Mpc (upper panel) and $0.25 h^{-1}$ Mpc (lower panel) to the filament axis. The Z -squared statistic based on galaxies is shown with red lines, where the shaded region shows the 95% confidence limits. The blue line shows the statistic for the null hypothesis using Monte Carlo simulations for a Poisson sample. The shaded region shows the 95% confidence limits for this case. For Monte Carlo simulation, the filaments and numbers of galaxies per filament are the same as for the real sample, but galaxies are Poisson distributed. Since we have averaged filaments with different lengths, the value of Z -squared statistic for the Poisson sample is not 2 as it is in Fig. B.1. Since the deviation is small, it does not affect our conclusions.

The maximum around $7 h^{-1}$ Mpc that was visible in the correlation functions is also visible in Fig. 6, confirming that galaxies are distributed along filaments in a regular pattern. Interestingly, the Z_1^2 statistic shows that there is also a small maximum around $4 h^{-1}$ Mpc. This indicates that between two groups that are separated by $7 h^{-1}$ Mpc there is quite often another group.

Figure 7 shows the Z -squared statistic for groups with different richness limits (at least one, two, or three galaxies per group). The blue region shows the Monte Carlo simulation results for a

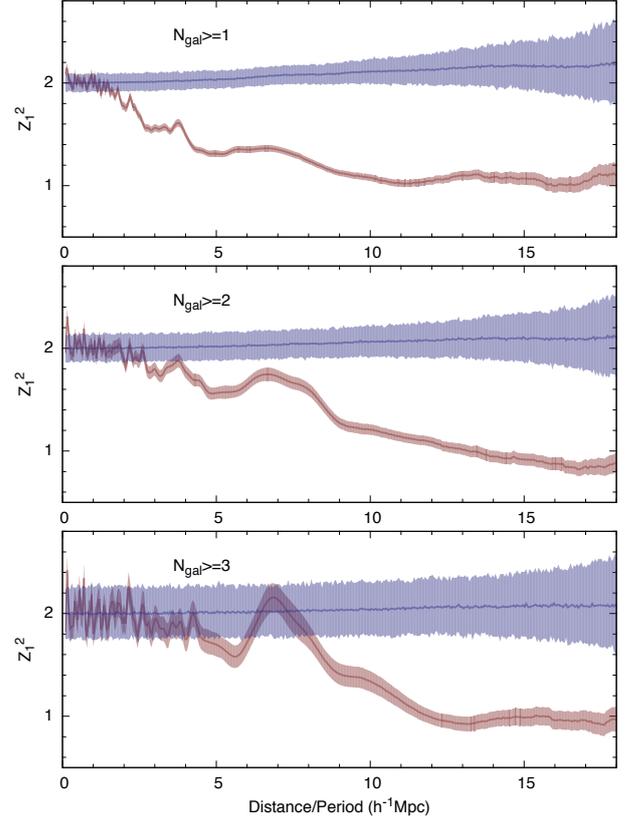


Fig. 7. Z -squared statistic for groups, where the minimum group richness is one (*upper panel*), two (*middle panel*), or three (*lower panel*). The designation of lines is the same as in Fig. 6.

Poisson sample. The Z_1^2 statistic for groups lies considerably below of that for the Poisson sample. This indicates that galaxy groups are distributed along filaments more regularly than in the Poisson case. This is expected, since galaxy groups cannot be located extremely close to each other, there is a minimum distance between the groups that is also visible in the two-point correlation functions.

In Fig. 7 we see that the maxima around 4 and $7 h^{-1}$ Mpc that define the preferred scales for the galaxy distribution are also visible in the group distributions. Interestingly, the $7 h^{-1}$ Mpc scale is the strongest for groups with at least three member galaxies. It shows that the regularity is stronger if the weakest groups and single galaxies are excluded.

4. Concluding remarks

Using the Bisous model (marked point process with interactions) we extracted the galaxy filaments from the SDSS spectroscopic galaxy survey. The diameter of the extracted filaments is roughly $1 h^{-1}$ Mpc and the catalogue of filament spines is built as described in Tempel et al. (2014b). Using the galaxies and groups in filaments (with a distance from the filament axis less than $0.5 h^{-1}$ Mpc) we studied how the galaxies/groups are distributed along the filament axis. The main results of our study can be summarised as follows.

- The galaxy and group distributions along filaments show a regular pattern with a preferred scale around $7 h^{-1}$ Mpc. A weaker regularity is also visible at a scale of $4 h^{-1}$ Mpc. The regularity of the distribution of galaxies along filaments is a

new result that might help to understand structure formation in the Universe.

- The pair correlation functions of galaxies and groups along filaments show that around each group, there is a region where the number density of galaxies/groups is smaller than on average.
- Galaxy groups in the Universe are more uniformly distributed along filaments than in the Poisson case.

The clustering pattern of galaxies and groups along filaments tells us that galaxy filaments are like pearl necklaces, where the pearls are galaxy groups that are distributed along the filaments in a regular pattern.

We suggest that the measured regularity of the galaxy distribution along filaments could be used as a cosmological probe to discriminate between various dark energy and dark matter cosmological models. Additionally, it can be used to probe environmental effects in the formation and evolution of galaxies. We plan to test these hypothesis in our following analysis using N -body simulations and deep redshift surveys like the Galaxy And Mass Assembly (GAMA, Driver et al. 2011) and the VIMOS Public Extragalactic Survey (VIPERS, Garilli et al. 2014).

Acknowledgements. We thank the referee for the useful and constructive report that helped to improve the presentation of our results. This work was supported by institutional research funding IUT26-2, IUT40-2 and TK120 of the Estonian Ministry of Education and Research, and by the Estonian Research Council grant MJD272. Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III web site is <http://www.sdss3.org/>. SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofísica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

References

- Ahn, C. P., Alexandroff, R., Allende Prieto, C., et al. 2014, *ApJS*, 211, 17
 Aihara, H., Allende Prieto, C., An, D., et al. 2011, *ApJS*, 193, 29
 Balaguera-Antolínez, A., Sánchez, A. G., Böhringer, H., et al. 2011, *MNRAS*, 413, 386
 Blake, C., Brough, S., Colless, M., et al. 2010, *MNRAS*, 406, 803
 Boerner, G., Mo, H., & Zhou, Y. 1989, *A&A*, 221, 191
 Bond, J. R., Kofman, L., & Pogosyan, D. 1996, *Nature*, 380, 603
 Buccheri, R. 1992, in *Statistical Challenges in Modern Astronomy*, eds. E. D. Feigelson, & G. J. Babu, 381
 Buccheri, R., Bennett, K., Bignami, G. F., et al. 1983, *A&A*, 128, 245
 Cautun, M., van de Weygaert, R., Jones, B. J. T., & Frenk, C. S. 2014, *MNRAS*, 441, 2923
 Connolly, A. J., Scranton, R., Johnston, D., et al. 2002, *ApJ*, 579, 42
 Contreras, C., Blake, C., Poole, G. B., et al. 2013, *MNRAS*, 430, 924
 Davis, M., & Geller, M. J. 1976, *ApJ*, 208, 13
 Davis, M., & Peebles, P. J. E. 1983, *ApJ*, 267, 465
 Davis, M., Meiksin, A., Strauss, M. A., da Costa, L. N., & Yahil, A. 1988, *ApJ*, 333, L9
 de Simoni, F., Sobreira, F., Carnero, A., et al. 2013, *MNRAS*, 435, 3017
 Dietrich, J. P., Werner, N., Clowe, D., et al. 2012, *Nature*, 487, 202
 Driver, S. P., Hill, D. T., Kelvin, L. S., et al. 2011, *MNRAS*, 413, 971
 Einasto, M. 1991, *MNRAS*, 252, 261
 Einasto, J., Jöeveer, M., & Saar, E. 1980, *Nature*, 283, 47
 Fadda, D., Slezak, E., & Bijaoui, A. 1998, *A&AS*, 127, 335
 Feigelson, E. D., & Jogesh Babu, G. 2012, *Modern Statistical Methods for Astronomy* (Cambridge University Press)
 Ferdosi, B. J., Buddelmeijer, H., Trager, S. C., Wilkinson, M. H. F., & Roerdink, J. B. T. M. 2011, *A&A*, 531, A114
 Garilli, B., Guzzo, L., Scodreggio, M., et al. 2014, *A&A*, 562, A23
 Groth, E. J., & Peebles, P. J. E. 1977, *ApJ*, 217, 385
 Guo, H., Li, C., Jing, Y. P., & Börner, G. 2014a, *ApJ*, 780, 139
 Guo, Q., Tempel, E., & Libeskind, N. I. 2014b, *ApJ*, submitted [[arXiv:1403.5563](https://arxiv.org/abs/1403.5563)]
 Hamilton, A. J. S. 1988, *ApJ*, 331, L59
 Hamilton, A. J. S. 1993, *ApJ*, 417, 19
 Hütsi, G. 2006, *A&A*, 459, 375
 Hütsi, G. 2010, *MNRAS*, 401, 2477
 Jöeveer, M., Einasto, J., & Tago, E. 1978, *MNRAS*, 185, 357
 Jasche, J., Kitaura, F. S., Li, C., & Enßlin, T. A. 2010, *MNRAS*, 409, 355
 Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, *ApJS*, 192, 18
 Labatie, A., Starck, J. L., Lachièze-Rey, M., & Arnalte-Mur, P. 2012, *Stat. Methodol.*, 9, 85
 Landy, S. D., & Szalay, A. S. 1993, *ApJ*, 412, 64
 Liivamägi, L. J., Tempel, E., & Saar, E. 2012, *A&A*, 539, A80
 Marín, F. A., Blake, C., Poole, G. B., et al. 2013, *MNRAS*, 432, 2654
 Martínez, V. J., & Saar, E. 2002, *Statistics of the Galaxy Distribution* (Chapman & Hall)
 Martínez, V. J., Arnalte-Mur, P., Saar, E., et al. 2009, *ApJ*, 696, L93
 McBride, C. K., Connolly, A. J., Gardner, J. P., et al. 2011, *ApJ*, 739, 85
 Muno, M. P., Baganoff, F. K., Bautz, M. W., et al. 2003, *ApJ*, 599, 465
 Norberg, P., Baugh, C. M., Gaztañaga, E., & Croton, D. J. 2009, *MNRAS*, 396, 19
 Peebles, P. J. E. 1980, *The large-scale structure of the universe* (Princeton University Press)
 Roche, N., Eales, S. A., Hippelein, H., & Willott, C. J. 1999, *MNRAS*, 306, 538
 Saar, E., Martínez, V. J., Starck, J.-L., & Donoho, D. L. 2007, *MNRAS*, 374, 1030
 Silverman, B. W. 1986, *Density estimation for statistics and data analysis* (Chapman & Hall)
 Stoica, R. S., Gregori, P., & Mateu, J. 2005, *Stochastic Proc. and their Applications*, 115, 1860
 Stoica, R. S., Martínez, V. J., & Saar, E. 2007, *J. Royal Stat. Soc. Ser. C*, 56, 459
 Stoica, R. S., Martínez, V. J., & Saar, E. 2010, *A&A*, 510, A38
 Tempel, E., & Libeskind, N. I. 2013, *ApJ*, 775, L42
 Tempel, E., Tago, E., & Liivamägi, L. J. 2012, *A&A*, 540, A106
 Tempel, E., Stoica, R. S., & Saar, E. 2013, *MNRAS*, 428, 1827
 Tempel, E., Libeskind, N. I., Hoffman, Y., Liivamägi, L. J., & Tamm, A. 2014a, *MNRAS*, 437, L11
 Tempel, E., Stoica, R. S., Martínez, V. J., et al. 2014b, *MNRAS*, 438, 3465
 Tempel, E., Tamm, A., Gramann, M., et al. 2014c, *A&A*, 566, A1
 Vio, R., Fasano, G., Lazzarin, M., & Lessi, O. 1994, *A&A*, 289, 640
 Wall, J. V., & Jenkins, C. R. 2012, *Practical Statistics for Astronomers* (Cambridge University Press)
 Wand, M. P., & Jones, M. C. 1995, *Kernel Smoothing* (Chapman & Hall)
 Wang, Y., Brunner, R. J., & Dolence, J. C. 2013, *MNRAS*, 432, 1961
 White, S. D. M., Tully, R. B., & Davis, M. 1988, *ApJ*, 333, L45
 York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, *AJ*, 120, 1579
 Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2005, *ApJ*, 630, 1
 Zhang, Y., Dietrich, J. P., McKay, T. A., Sheldon, E. S., & Nguyen, A. T. Q. 2013, *ApJ*, 773, 115

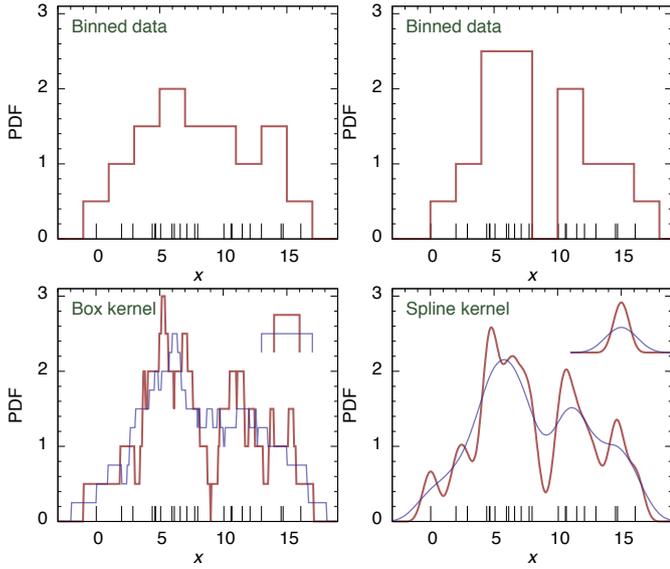


Fig. A.1. Upper row: probability density estimation using binning. In the left and right panels, the bin width is the same but the centre of the bin is shifted. Bottom row: density estimation using kernel smoothing. In the bottom left panel the simplest box kernel is used, in the bottom right panel the B_3 spline kernel is used. The kernel shape and size is shown in the upper-right corner of the figures. The rug plot on the bottom axis shows the data points that were used for density estimation.

Appendix A: Kernel density estimation

In the simplest case, the probability density can be estimated as the binned density histogram. However, this estimate depends both on the bin widths and the location of the bin edges. A better way is to use kernel smoothing (e.g. Silverman 1986; Wand & Jones 1995; Feigelson & Jogesh Babu 2012), where the density is represented by a sum of kernels centred at the data points:

$$f(x) = \frac{1}{h} \sum_i K\left(\frac{x - x_i}{h}\right). \quad (\text{A.1})$$

The kernels $K(x)$ are distributions ($K(x) > 0$, $\int K(x)dx = 1$) of zero mean and of a typical width h . The width h is an analogue of the bin width, but there are no bin edges to worry about.

To create our density estimation we use the popular B_3 spline function

$$B_3(x) = \frac{|x - 2|^3 - 4|x - 1|^3 + 6|x|^3 - 4|x + 1|^3 + |x + 2|^3}{12}. \quad (\text{A.2})$$

This kernel is well suited for estimating densities – it is compact, differing from zero only in the interval $x \in [-2, 2]$, and it conserves mass: $\sum_i B_3(x - i) = 1$ for any x .

In many papers (e.g. Vio et al. 1994; Fadda et al. 1998; Ferdosi et al. 2011) it has been shown that kernel smoothing is the best and recommended choice for density estimation. It is more robust and reliable than other simpler methods and provides comparable (or better) results than methods with high computational costs (e.g. the maximum likelihood technique).

To illustrate the use of kernel smoothing, in Fig. A.1 we show the estimated probability density function using either binning the data or kernel smoothing with two types of kernels (the box and B_3 spline kernels). The upper row in Fig. A.1 demonstrates density estimation using binning, where the bin widths are the same but the bin centres are slightly shifted. We see that the resulting density estimate may depend strongly on the chosen bin locations.

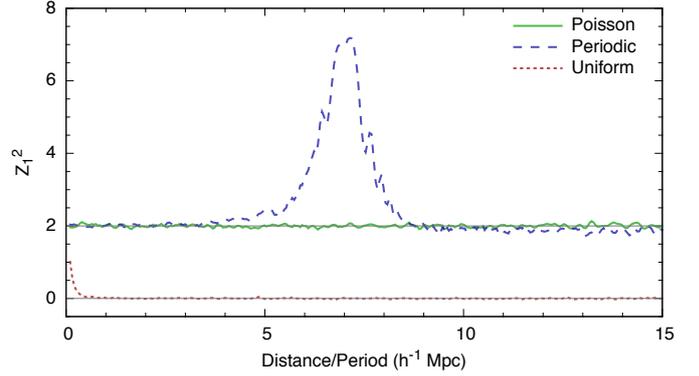


Fig. B.1. Z-squared statistics for three cases. The green line shows the Z-squared statistic for a Poisson sample. The blue line shows the statistic for a periodic signal, where the period for each datapoint is drawn from a Gaussian distribution centred at $7 h^{-1}$ Mpc with a standard deviation of $0.5 h^{-1}$ Mpc. The red line shows the statistic for data points with a uniform point distribution – see text for more information.

The lower-left panel in Fig. A.1 demonstrates density estimation using the simplest box kernel. The kernel width is the same as box width for binned density estimates above. Even the simplest box kernel reveals the details in density distribution; however, the resulting distribution function is not smooth. To get a smooth density distribution, a smooth kernel should be used. The kernel density estimates using B_3 spline kernels are shown in the lower-right panel of Fig. A.1. To show that the results of kernel smoothing are quite robust, we doubled the kernel width and compared the probability density estimates as shown in Fig. A.1.

Appendix B: Examples of the Rayleigh (Z-squared) statistic

To illustrate how the Z-squared statistic works, we generated three datasets and calculated the Rayleigh statistics for them. The results are shown in Fig. B.1. In the first case, we generated a Poisson distribution (green line). For a Poisson sample, the statistic gives an average value of 2. In the second case, we added some periodicity to the sample (blue line). We generated point distributions with the same (zero) phase and with periods chosen from a Gaussian distribution centred at $7 h^{-1}$ Mpc with a standard deviation of $0.5 h^{-1}$ Mpc, and added these together. In Fig. B.1 we see that the Z-squared statistic recovers the period well. In the third case, we generated a uniform distribution of points (red line). For that, we divided the test filament into N (the number of points) equal regions and in each region we put one point, selected from a uniform distribution. This ensures that we get points that are more homogeneously distributed along the filament and the Z-squared statistic gives the value zero. We can conclude that if the value of the statistic lies above 2, there is some periodicity in the data. Conversely, if it is below 2, it describes a more homogeneous distribution. Additionally, the peaks in the Z-squared statistic show that there is preferred periodicity in the data with the scale of the peak position.

It is known that for the uniform distribution of points, the Z-squared statistic is distributed as the chi-square with two degrees of freedom (see e.g. Wall & Jenkins 2012). So its expected value is 2, as seen in Fig. B.1.