

Direct Minkowski Functional analysis of large redshift surveys: a new high-speed code tested on the luminous red galaxy Sloan Digital Sky Survey-DR7 catalogue

Alexander Wiegand,^{1★} Thomas Buchert^{2★} and Matthias Ostermann^{3,4}

¹Max-Planck-Institut für Gravitationsphysik, Albert-Einstein-Institut, Am Mühlenberg 1, D-14476 Potsdam, Germany

²Université de Lyon, Observatoire de Lyon, Centre de Recherche Astrophysique de Lyon, CNRS UMR 5574; Université Lyon 1 and École Normale Supérieure de Lyon, 9 avenue Charles André, F-69230 Saint-Genis-Laval, France

³Arnold Sommerfeld Center, Ludwig-Maximilians-Universität, Theresienstraße 37, D-80333 München, Germany

⁴Oskar-Maria-Graf Gymnasium, Keltenweg 5, D-85375 Neufahrn, Germany

Accepted 2014 June 4. Received 2014 May 12; in original form 2013 November 14

ABSTRACT

As deeper galaxy catalogues are soon to come, it becomes even more important to measure large-scale fluctuations in the catalogues with robust statistics that cover all moments of the galaxy distribution. In this paper, we reinforce a direct analysis of galaxy data by employing the Germ–Grain method to calculate the family of Minkowski Functionals. We introduce a new code, suitable for the analysis of large data sets without smoothing and without the construction of excursion sets. We provide new tools to measure correlation properties, putting emphasis on explicitly isolating non-Gaussian correlations with the help of integral-geometric relations. As a first application, we present the analysis of large-scale fluctuations in the luminous red galaxy sample of Sloan Digital Sky Survey data release 7 data. We find significant deviations from the Λ cold dark matter mock catalogues on samples as large as $500 h^{-1}$ Mpc (more than 3σ) and slight deviations of around 2σ on $700 h^{-1}$ Mpc, and we investigate possible sources of these deviations.

Key words: methods: analytical – methods: data analysis – methods: statistical – cosmology: observations – large-scale structure of Universe.

1 INTRODUCTION

Over the past decade, huge progress has been made in accessing the galaxy distribution at larger and larger scales. At each step of this process, new and larger structures have been discovered, see e.g. Einasto et al. (2011), Sylos Labini (2011), Park et al. (2012), Clowes et al. (2013), Keenan, Barger & Cowie (2013), Whitbourn & Shanks (2014). However, to verify the reality of these structures, a robust statistical tool is mandatory (Nadathur 2013; Nadathur & Hotchkiss 2014).

The most common tool for the characterization of large-scale structure is based on two-point measures: the two-point correlation function of the galaxy distribution and the complementary power spectrum. They are particularly useful, as they can be related to the power spectrum determined from the physics of the Early Universe. Claims that structures on scales of several hundreds of megaparsecs are compatible with the Λ cold dark matter (CDM) model are often based on these lower order statistics. However, of course, they do not allow for a complete characterization of the distribution. This

needs higher order correlations that play an important role if the density field is not Gaussian, especially when probing stages after the formation of structure by gravitational collapse. Note, however, that also the full knowledge of all higher order correlations does not always characterize the distribution uniquely (Carron & Neyrinck 2012). As is well known, see e.g. Coles, Melott & Shandarin (1993), Melott, Buchert & Weiß (1995), a smoothed-out non-linear structure – even if smoothed over very large scales – is not described by structure described in linear gravitational instability, where in this latter, the distribution remains Gaussian, if it was so in the initial data.

For this strongly clustered regime in the Late Universe, the Minkowski Functionals that we are using in this paper provide a compact and transparent framework to completely characterize the galaxy distribution. As we shall see in Section 2, they include all higher N -point correlations in a power series in the sample density. We shall show in Section 5.3 that it is indeed not enough to include only the lowest order contributions of this series. This means that the values of the functionals that we determine significantly depend on higher order clustering. As is also well known, only higher order correlations are sensitive to the morphology of large-scale structure.

*E-mail: alex_wiegand@gmx.de (AW); buchert@obs.univ-lyon1.fr (TB)

Due to this interesting property of including higher correlations in a simple way, the Minkowski Functionals have been determined for many galaxy and cluster surveys. The specific Germ–Grain model that has been introduced into cosmology together with the family of Minkowski Functionals in Mecke, Buchert & Wagner (1994), Buchert (1995, see Schmalzing, Kerscher & Buchert 1996 for a brief tutorial), and which will be briefly reviewed in Section 2.2, has been used for example for the IRAS 1.2 Jy and PSCz surveys (Kerscher et al. 1996b, 1998, 2001a), and the Abell/ACO cluster catalogue (Kerscher et al. 1997). For an early sample of the Sloan Digital Sky Survey (SDSS) catalogue (data release DR 3), the Minkowski Functionals have been determined for smoothed isodensity contours of the galaxy distribution in Hikage et al. (2003), which is also the method employed in most work on Minkowski Functional analysis of galaxy and cluster catalogues as well as in simulations (Platzöder & Buchert 1996; Schmalzing & Buchert 1997; Sahni, Sathyaprakash & Shandarin 1998; Sathyaprakash, Sahni & Shandarin 1998; Schmalzing et al. 1999b; Nakagami et al. 2004; Blake, James & Poole 2013; Choi et al. 2013). A further, recently proposed smoothing technique directly uses the Delauney tessellation of the point distribution (Aragon-Calvo, Shandarin & Szalay 2010).

To robustly measure non-Gaussianity with the help of Minkowski Functionals is mostly discussed for isothermality contour maps of the cosmic microwave background (Schmalzing & Gorski 1998; Schmalzing, Takada & Futamase 2000; Hikage & Matsubara 2012; Ducout et al. 2013; Modest et al. 2013; Munshi et al. 2013; Planck Collaboration 2013), as well as for studies of the weak lensing convergence field (Kratochvil et al. 2012; Petri et al. 2013).

One focus of this paper is the determination of non-Gaussian features from Minkowski Functionals in three-dimensional galaxy data, which has been addressed in Pratten & Munshi (2012) and Codis et al. (2013). The other focus lies on reinforcing the Germ–Grain method in the three-dimensional case to calculate the Minkowski Functionals. We shall apply this method to the SDSS DR 7. This release was chosen due to its complete angular coverage of the SDSS survey region and the existence of a detailed standard analysis of the two-point correlation properties in Kazin et al. (2010). In order to be able to probe larger scales than before in Hikage et al. (2003), we specifically use the luminous red galaxy (LRG) sample of the DR 7 in the compilation of Kazin et al. (2010). Newer and upcoming data will be analysed in forthcoming work. Especially the full SDSS DR 12 catalogue, but also catalogues of after-Sloan programmes, are targets for our optimized code.

The paper is organized as follows: Section 2 recalls basic properties of the Minkowski Functionals and briefly reviews the Germ–Grain model for the direct analysis of the galaxy point process. We discuss the usefulness of this model by collecting the analytical results that are known for the Minkowski Functionals of this model, examine the Gauss–Poisson process, and introduce our method to extract information on higher order correlations from the Minkowski Functionals of the model. Section 3 describes the new code that we use in order to efficiently calculate the Minkowski Functionals in the Germ–Grain model for a large data set like the SDSS LRG catalogue. Section 4 presents and discusses the results for two different subsamples at different luminosity thresholds of this catalogue as a first application of our methods. In Section 5, we derive the values of some integrals over the two- and three-point correlation function and study the deviations of the point distribution from a Gauss–Poisson process and also from a lognormal distribution. We here explicitly demonstrate that the low-order correlations are actually not enough to describe the structure in the data set. We conclude in Section 6.

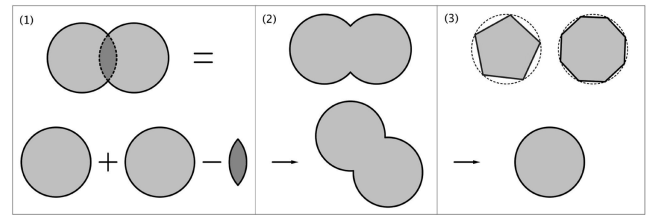


Figure 1. Properties of the scalar functionals: (1) Additivity, (2) Motion invariance, (3) Conditional continuity.

2 MINKOWSKI FUNCTIONALS OF THE GERM–GRAIN MODEL

Let us begin by a description of Minkowski Functionals in the Germ–Grain model that we shall use here. For a more complete description see Mecke & Stoyan (2000), Schmalzing (1999), Beisbart (2001) or Mecke (1994).

2.1 Minkowski Functionals

Minkowski Functionals are morphological descriptors of extended bodies that rely on well-developed results in integral geometry. In 3D Euclidean space, there are four of them that we shall label $V_0 - V_3$. In the normalization we use, they are related to geometrical properties of the body as follows:

$$V_0 = V; \quad V_1 = \frac{S}{6}; \quad V_2 = \frac{H}{3\pi}; \quad V_3 = \chi. \quad (1)$$

Here, V is the volume of the body, S is its surface area, H the integral mean curvature of the surface and χ the Euler characteristic (the integral Gaussian curvature of the surface).

The surprising fact, shown in Hadwiger (1957), is that every other scalar functional that can be defined to describe a given body and that fulfils the properties of motion invariance, additivity and conditional continuity (sketched in Fig. 1), can be expressed as a linear combination of the four functionals of equation (1).

Instead of working with the functionals V_μ , we will more often need the corresponding densities v_μ . They are simply defined by

$$v_\mu = V_\mu / |\mathcal{D}|, \quad (2)$$

where $|\mathcal{D}|$ is the volume within the survey boundary.

2.2 Germ–Grain model

The Minkowski Functionals as described in the previous section are only defined for extended bodies. To use them for characterizing the galaxy distribution, one has to define a procedure that transforms the point distribution into a collection of extended objects. The two major methods that are used so far to achieve this are the construction of *excursion sets* (e.g. isodensity contours) and the *Germ–Grain model*.

In order to determine isodensity contours, the point particle distribution is smoothed into a continuous density field. The surfaces of a given density threshold then provide the boundaries of the body that we are going to analyse. The values of the four Minkowski Functionals (volume, surface area, mean curvature, Euler characteristic) can then be determined as a function of the (over)density that is used to determine the isodensity contours (Schmalzing & Buchert 1997). This method is commonly employed in the community (see the reference list in the introduction), and it has also been used for the SDSS data in Hikage et al. (2003).

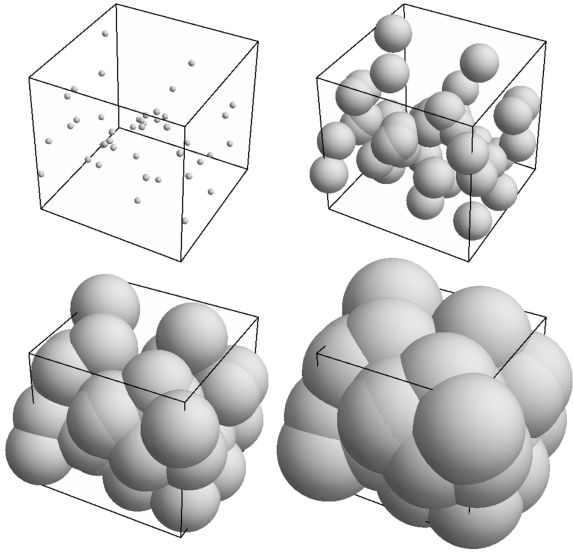


Figure 2. By increasing the radius of Balls around the points $\{x_1, \dots, x_N\}$ up to the maximum radius R , more and more Balls intersect and a complex structure develops.

In the Germ–Grain model, the point distribution of galaxies is converted into a set of extended bodies by decorating each galaxy with a Grain (here a Ball of radius R , but any shape of Grain could in principle be used to take into account internal morphologies). Instead of the (over)density, the (equal) radius of these Balls can be used as a diagnostic parameter, i.e. to present the values of the four functionals as a function of scale (the radius R of the Balls; Mecke et al. 1994). This results in quite complex structures as shown in Fig. 2. Also, this method has been used quite extensively in the past to characterize galaxy and cluster distributions, e.g. in Kerscher et al. (1996b, 1997, 1998, 2001a), Platzöder & Buchert (1996) and for other statistics in Kerscher et al. (1999), but it has not become a standard analysis tool in cosmology. With this paper, we emphasize the advantages of this direct method to analyse galaxy data.

Comparing the two methods, the Germ–Grain model has several important advantages over the construction of excursion sets:

(i) it can be implemented in an easy and robust manner. This simplicity also implies, that

(ii) we have an analytical understanding of the relation of the average Minkowski Functional (densities) to the connected correlation functions of the underlying point distribution, through which also the global contribution of Poisson noise is explicitly known;

(iii) the global functionals are represented by their local contributions (so-called *Partial Minkowski Functionals*). This local information can be used to extract subsamples with given environmental properties. The partial functionals form the basis of image analysis techniques, since they allow us to extract filamentary or cluster galaxies from a distribution, even if these morphological properties are strongly diffused by Poisson noise (Platzöder & Buchert 1996).¹

These useful relations, e.g. between the connected correlation functions and the average Minkowski Functional densities, as well as the image analysis properties of partial functionals have been demonstrated in Schmalzing (1999).

¹ See also Mantz, Jacobs & Mecke (2008) for a more recent application of Minkowski Functionals in image analysis.

2.3 Statistical interpretation

Thus far, the Germ–Grain Minkowski Functionals were mainly used for a comparison of individual samples. These comparisons made use of the property of the Minkowski Functionals to provide a morphological characterization of the galaxy distribution in the analysed sample: if the Germ–Grain Minkowski Functionals differ for two samples, these are morphologically distinct (Kerscher et al. 1998).

In addition to this comparison of individual point sets, we here also want to extract some information on the statistical properties of the point distribution that underlie these individual galaxy data sets. We, therefore, interpret an observed or simulated galaxy sample, as usual, as a particular realization of a point process with a priori unknown statistical properties. This gives rise to an ensemble of Minkowski Functionals $\{V_\mu\}$ or their corresponding densities $\{v_\mu\}$, respectively. As made explicit in Appendix A, it can be shown that the ensemble average of the densities v_μ is related to the connected correlation functions ξ_n of the point process as follows:

$$\begin{aligned} \langle v_0 \rangle &= 1 - e^{-\varrho_0 \bar{V}_0} \\ \langle v_1 \rangle &= \varrho_0 \bar{V}_1 e^{-\varrho_0 \bar{V}_0}, \\ \langle v_2 \rangle &= \left(\varrho_0 \bar{V}_2 - \frac{3\pi}{8} \varrho_0^2 \bar{V}_1^2 \right) e^{-\varrho_0 \bar{V}_0}, \\ \langle v_3 \rangle &= \left(\varrho_0 \bar{V}_3 - \frac{9}{2} \varrho_0^2 \bar{V}_1 \bar{V}_2 + \frac{9\pi}{16} \varrho_0^3 \bar{V}_1^3 \right) e^{-\varrho_0 \bar{V}_0}. \end{aligned} \quad (3)$$

For a Poisson distribution, the quantities \bar{V}_μ are simply the Minkowski Functionals $V_\mu(B)$ of the Balls of common radius R that we use to obtain extended bodies:

$$\begin{aligned} V_0(B) &= \frac{4\pi}{3} R^3; \quad V_1(B) = \frac{2}{3} \pi R^2; \\ V_2(B) &= \frac{4}{3} R; \quad V_3(B) = 1. \end{aligned} \quad (4)$$

For a point distribution with structure, the \bar{V}_μ pick up contributions that depend on the dimensionless connected correlation function of order $n + 1$, ξ_{n+1} , as

$$\begin{aligned} \bar{V}_\mu &= V_\mu(B) + \sum_{n=1}^{\infty} \frac{(-\varrho_0)^n}{(n+1)!} \int_{\mathcal{D}} d^3x_1 \dots d^3x_n \\ &\quad \times \xi_{n+1}(0, \mathbf{x}_1, \dots, \mathbf{x}_n) V_\mu(B \cap B_{x_1} \cap \dots \cap B_{x_n}). \end{aligned} \quad (5)$$

The integrals run over the positions of the centres \mathbf{x}_i of the Balls B_{x_i} . As it is the intersection of all Balls $B \cap B_{x_1} \cap \dots \cap B_{x_n}$ that enters, the integrals vanish for configurations where the \mathbf{x}_i are separated by more than $2R$. Therefore, determining the Minkowski Functionals as a function of the Ball radius R probes the correlation of the point distribution up to a scale of $2R$. We shall exploit this property in Section 5.

The introduction of the dimensionless version of the connected correlation functions ξ_n requires the assumption that the global point process possesses a well-defined non-zero and stable (scale independent) average density $\varrho_0 = \langle \varrho(r) \rangle$, a requirement that is expected to hold for an existing homogeneity scale. Note, however, that this assumption is not required for the Minkowski Functional analysis itself. Also for expressing the $\langle v_\mu \rangle$ in terms of the statistical quantities describing the point process, one could work with the dimensionfull connected correlation functions, without assuming that $\varrho_0 \neq 0$. We shall only need this condition for the extraction procedure described in Section 2.5. The reasoning in that case is

then, that we interpret the analysed sample as being a representative realization of the full point process. We assume that it has a positive average density and estimate this background density ϱ_0 from the sample. Of course, this regional estimate can be biased relative to the true global value (assumed to exist): for an analysis of the correlation properties well inside the survey region, we consider this assumption as sufficiently accurate, since the estimation of the correlation properties would be most strongly influenced on the scale of the sample (on this scale the integral constraint requires vanishing of the fluctuations). This subtlety is worth to be kept in mind in future analyses.

2.4 Gauss–Poisson process

To get a better intuition about the influence of correlation functions on the modified Minkowski Functionals \bar{V}_μ , we shall first consider the case of a Gaussian distribution. For low enough average density ϱ_0 and certain correlations ξ_2 , a point distribution can be described by a Gauss–Poisson process. Kerscher (2001) shows that for this to be possible, the correlation function has to be non-negative, $\xi_2 \geq 0$, and the average density ϱ_0 has to satisfy

$$\varrho_0 \int_{\mathcal{D}} d\mathbf{y} \xi_2(|\mathbf{y}|) \leq 1, \quad (6)$$

on the domain \mathcal{D} of the point sample.

The resulting Gauss–Poisson process has the property that all higher connected correlation functions ξ_n for $n > 2$ are zero. This drastically simplifies the expressions for the \bar{V}_μ . Equation (5) becomes

$$\bar{V}_\mu = V_\mu(B) - \frac{\varrho_0}{2} \int_{\mathcal{D}} d^3x_1 \xi_2(|\mathbf{x}_1|) V_\mu(B \cap B_{x_1}). \quad (7)$$

We already gave the explicit expressions for the Minkowski Functionals of the Balls $V_\mu(B)$ in equation (4). For a known correlation function, it is in addition possible to calculate the contribution of the second term. To this end, we have to determine the Minkowski Functionals for the intersection of two Balls $V_\mu(B \cap B_{x_1})$. This intersection has the form of a convex lens and it is easy to figure out its volume and surface as a function of the distance r , separating the centres of the Balls. For the mean curvature, the result can be found in Mecke (1994). The form of the $V_\mu(B \cap B_{x_1})$ in a spherical coordinate system centred on B is then the following:

$$V_0(r) = \frac{1}{12} \pi (2R - r)^2 (r + 4R); \quad (8)$$

$$V_1(r) = \frac{1}{3} \pi R (2R - r); \quad (9)$$

$$V_2(r) = \frac{2}{3} (2R - r) + \frac{2}{3} R \sqrt{1 - \left(\frac{r}{2R}\right)^2} \arcsin\left(\frac{r}{2R}\right); \quad (10)$$

$$V_3(r) = 1, \quad (11)$$

where R is again the radius of the Balls. As they do not intersect if the separation of the two centres is larger than $2R$, $V_\mu(B \cap B_{x_1}) = 0$ for $r > 2R$. Therefore, the integral (7) reduces to

$$\bar{V}_\mu = V_\mu(B) - 2\pi\varrho_0 \int_0^{2R} V_\mu(r) \xi_2(r) r^2 dr. \quad (12)$$

To get a feeling what aspects of the correlation function these integrals probe, we show the form of the window functions $V_\mu(r)r^2$ in Fig. 3.

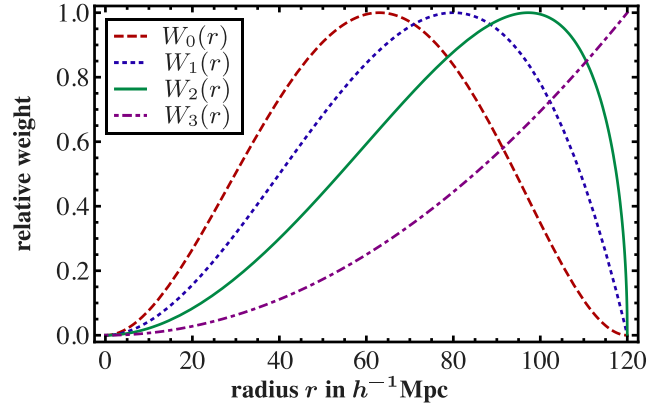


Figure 3. Form of the integration windows $W_\mu(r) = V_\mu(r)r^2$ in equations (8)–(11) used in the integral (12), for a Ball radius of $R = 60 h^{-1} \text{Mpc}$. The functionals of higher index μ probe the correlation function at larger distances.

To calculate these integrals for the standard Λ CDM structure, it is useful to express them directly in terms of the power spectrum. Inserting the Fourier transform of $\xi_2(r)$ gives

$$\bar{V}_\mu = V_\mu(B) - \frac{\varrho_0}{\pi} \int_0^\infty P(k) W_\mu(k, R) k^2 dk, \quad (13)$$

with the functions

$$W_0(k, R) = \frac{4\pi(\sin(kR) - kR \cos(kR))^2}{k^6}; \quad (14)$$

$$W_1(k, R) = -\frac{2\pi R(kR \sin(2kR) + \cos(2kR) - 1)}{3k^4}; \quad (15)$$

$$W_2(k, R) = -\frac{4(kR \sin(2kR) + \cos(2kR) - 1)}{3k^4} + \frac{2R}{3k} \int_0^{\pi/2} \sin(\phi) \sin(2kR \sin(\phi)) \phi d\phi; \quad (16)$$

$$W_3(k, R) = \frac{\sin(2kR) - 2kR \cos(2kR)}{k^3}. \quad (17)$$

As $W_0(k, R)$ is simply the square of the Fourier transform of a top hat window function, \bar{V}_0 can be related to another well-known statistical property of a point distribution, namely the matter variance in a sphere of radius R :

$$\sigma^2(R) = \frac{1}{(2\pi)^3} \int d^3k P(k) \left| \tilde{W}_{B(R)}(k) \right|^2. \quad (18)$$

This means that \bar{V}_0 and \bar{V}_3 are directly related to the two-point quantities by

$$\bar{V}_0 = \frac{4\pi}{3} R^3 \left(1 - \frac{4\pi}{3} R^3 \varrho_0 \sigma^2(R) \right), \quad (19)$$

and

$$\bar{V}_3 = 1 - 2\pi\varrho_0 \int_0^{2R} \xi_2(r) r^2 dr. \quad (20)$$

The other two modified functionals \bar{V}_1, \bar{V}_2 are not as directly related to other well-known quantities, but probe additional aspects of the form of the correlation function defined by their weights in Fig. 3.

2.5 Extracting higher order correlations

As we have seen in the previous section, we can directly derive certain integrals over the two-point correlation function from the Minkowski Functionals of a Gauss–Poisson point distribution. For a more general point distribution this is no longer straightforward, but we can use the fact that we know the exact form of the dependence of $\langle v_\mu \rangle$ on the average density of the point process ϱ_0 . For a single sample, of course, this density is fixed. By random subsampling of the original sample, however, we can create (noisier) samples of a lower average density. In this way, we can determine the $\langle v_\mu \rangle$ not only as a function of the radius of the Balls, but also as a function of the density of the point distribution. This allows us to extract the integrals over the correlation functions in (5) as follows.

Let us assume that we are able to measure the $\langle v_\mu \rangle$ accurately for a given density. By repeating this measurement for several densities ϱ_0 , we get an approximation to the functional dependence of v_μ on ϱ_0 . Then, by inverting the system (3), we can derive from the measured values of $\langle v_\mu(\varrho_0) \rangle$ the corresponding functional dependence of \bar{V}_μ on ϱ_0 . Calling this empirical function $\tilde{\bar{V}}_\mu(\varrho_0)$, we know that it can be written as a series expansion in ϱ_0 of precisely the form (5). This means that, if we can compute the coefficients of this series, we shall obtain the corresponding weighted integrals over the correlation functions. Taylor expanding $\tilde{\bar{V}}_\mu(\varrho_0)$ around $\varrho_0 = 0$, we obtain

$$\tilde{\bar{V}}_\mu(\varrho_0) = \sum_{n=0}^{\infty} \frac{\tilde{\bar{V}}_\mu^{(n)}(0)}{n!} \varrho_0^n,$$

where the exponent (n) stands for the n th derivative of $\tilde{\bar{V}}_\mu(\varrho_0)$ with respect to ϱ_0 . These derivatives can now be directly related to the coefficients of the expansion (5). Writing this expansion in short as

$$\bar{V}_\mu = \sum_{n=0}^{\infty} \frac{b_{n+1}^\mu}{(n+1)!} (-\varrho_0)^n,$$

with $b_1 = V_\mu(B)$, we deduce that

$$b_{n+1}^\mu = \int_{\mathcal{D}} \xi_{n+1}(0, \mathbf{x}_1, \dots, \mathbf{x}_n) V_\mu(B \cap B_{\mathbf{x}_1} \cap \dots \cap B_{\mathbf{x}_n}) \times d^3x_1 \dots d^3x_n = (n+1)(-1)^n \tilde{\bar{V}}_\mu^{(n)}(0). \quad (21)$$

This implies that we are able to quantitatively determine, how much the point distribution deviates from a pure Gauss–Poisson distribution. The ‘Gaussian part’ is related to the first derivative of $\tilde{\bar{V}}_\mu(\varrho_0)$ at zero density, and it especially allows us to compare the result to other independent measurements of $\sigma^2(R)$ and $\xi_2(R)$ via the relations (19) and (20).

To carry out this procedure in practice, we estimate $\langle v_\mu \rangle$ by the Minkowski Functional density v_μ of the given realization of the point process. For densities lower than the original density, we average v_μ of several random subsamples. To determine how accurate this estimate is, we use the average and fluctuations of the individual v_μ in an ensemble of mock samples produced from simulations. This allows us to test, whether the observed sample is consistent with the simulated cosmology.

For the three-point functions, the quantity with the simplest weight functions in the integral reads

$$b_3^0 = \int_{\mathcal{D}} d^3x_1 d^3x_2 d^3x_3 \zeta(|\mathbf{x}_1 - \mathbf{x}_3|, |\mathbf{x}_2 - \mathbf{x}_3|, |\mathbf{x}_1 - \mathbf{x}_2|) \times \theta(R - |\mathbf{x}_1|) \theta(R - |\mathbf{x}_2|) \theta(R - |\mathbf{x}_3|). \quad (22)$$

The other integrals of the three-point function are more complicated and we would not write them out explicitly. We calculate them numerically, using our code to determine $V_\mu(B \cap B_{\mathbf{x}_1} \cap B_{\mathbf{x}_2})$ for the cases where $\mu \neq 0$.

To characterize the deviation of the galaxy distribution in the SDSS from a Poisson and a Gauss–Poisson distribution, we shall determine the coefficients b_{n+1}^μ in Section 5 for $n \leq 2$.

3 THE NEW CODE PACKAGE: MINKOWSKI-4

As described in the previous section, we can learn a lot about the structure in the galaxy distribution in the Universe and especially about the magnitude of higher order correlations, if we are able to calculate the Minkowski Functional densities v_μ accurately.

With this paper, we provide the MINKOWSKI-4 package, built on the new code CHIMPINK (Code for High-speed Investigation of Partial Minkowski Functionals), which is a completely revised version of a code based on previous work by Jens Schmalzing and Andreas Rabus in 1998, see Rabus (1998). The package compiles modules to compute the Minkowski Functionals of a given point sample for the Germ–Grain model [which generalizes the Boolean Grain model – where the Germs are those of a Poisson process (Stoyan, Kendall & Mecke 1987) – to arbitrary point distributions]. It extracts correlation properties of the point set in the form of the Minkowski functional densities (3) and the modified Minkowski Functionals \bar{V}_μ . Optionally, it also delivers the full set of Partial Minkowski Functionals of the environmental morphology of every point in the sample.

3.1 Computation of the Germ–Grain model

The computational methods for the Germ–Grain model of the Minkowski Functionals (henceforth abbreviated as MFs) heavily rely on the work of Mecke, Buchert and Wagner in Mecke et al. (1994) and are therefore also strongly related to the works of Kerscher and collaborators (Kerscher et al. 1997, 1998, 2001a,b).

As outlined in Section 2.2, a sphere of radius r (the so-called Grain or Ball) is placed around each point of the sample, the Germ. The union of the Balls then forms the structure \mathcal{B}_r ,

$$\mathcal{B}_r = \bigcup_{i=1}^N \mathcal{B}(\mathbf{x}_i; r). \quad (23)$$

When we increase the Balls’ radius r up to a maximum radius R , a more and more complex structure develops, see Fig. 2. Thus, the radius $0 \leq r \leq R$ serves as a diagnostic parameter.

In the Germ–Grain model, the global MFs – apart from the volume – are localized on the surface of the structure and can be determined by means of the so-called partition formula (see e.g. Schmalzing & Diaferio 2000),

$$V_\mu(\mathcal{B}_r) = \sum_{i=1}^N V_\mu^{(i)} + \frac{1}{2} \sum_{i,j=1}^N V_\mu^{(ij)} + \frac{1}{6} \sum_{i,j,k=1}^N V_\mu^{(ijk)}, \quad (24)$$

where $V_\mu^{(i)}$ is the contribution of the Ball around \mathbf{x}_i (at given radius r), and where $V_\mu^{(ij)}$ and $V_\mu^{(ijk)}$ are those of its intersection with one, respectively, two neighbours.

3.2 Partial MFs

The global MFs can be calculated by adding up Partial MFs assigned to each Grain, see for example Mecke et al. (1994) and in an

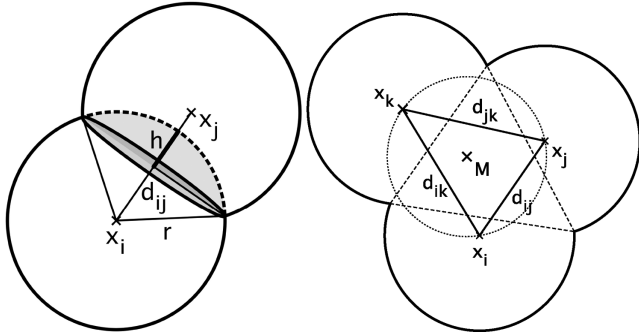


Figure 4. The left-hand figure shows the covered surface area of a sphere around x_i when intersected with a second sphere as well as the intersection circle. The right-hand figure illustrates the intersection of three spheres. The triple point can be found above the centre point of the circumcircle of the triangle generated by the three centre points x_i, x_j, x_k .

application Schmalzing & Diaferio (2000),

$$V_\mu(\mathcal{B}_r) = \sum_{i=1}^N V_\mu(\mathbf{x}_i; r), \quad (25)$$

where $V_\mu(\mathbf{x}_i; r)$ are the Partial MFs of the Ball around \mathbf{x}_i with radius r . These can be determined by the local intersections of the Balls. Since only neighbours within $2r$ around a point contribute to its Partial MFs, we determine a *neighbourlist* for each point of the sample before the actual calculation, which consists of the points within a distance of two times the maximum radius (as well as the point itself).

The statistical weight of intersections of more than three Balls is zero, see Mecke et al. (1994); therefore, we only take into account intersection circles of two Balls and intersection points of three Balls, the so-called triple points, see Fig. 4. The MFs' volume densities are defined by

$$v_\mu(\mathcal{B}_r) = \frac{1}{|D|} \sum_{i=1}^N V_\mu(\mathbf{x}_i; r), \quad (26)$$

where $|D|$ denotes the volume of the sample mask.

In summary: for any given point \mathbf{x}_i of the sample, we calculate (for each radius r up to the maximum radius R)

- (i) the uncovered surface area A_i of the Ball around that point,
- (ii) the intersection circles of the Ball around that point with any of its neighbours; here, ℓ_{ij} is the uncovered arc length, i.e. the uncovered segment of the intersection circle of the Balls around \mathbf{x}_i and \mathbf{x}_j ,
- (iii) the triple points of the intersection with the Balls around any two neighbours, where ϵ_{ijk} is called spherical excess; it can be calculated using the formula of l'Huilier and denotes the contribution of the triple points to the Partial Euler characteristic.

With these quantities, the Partial MFs read (see Mecke et al. 1994; Mecke & Stoyan 2000 for more details)

$$\begin{aligned} V_1(\mathbf{x}_i; r) &= \frac{A_i}{6}; \\ V_2(\mathbf{x}_i; r) &= \frac{A_i}{3\pi r} - \frac{1}{2} \sum_j \frac{d_{ij} \ell_{ij}}{6\pi}; \\ V_3(\mathbf{x}_i; r) &= \frac{A_i}{4\pi r^2} - \frac{1}{2} \sum_j \frac{d_{ij} \ell_{ij}}{4\pi r \cdot \rho_{ij}} + \frac{1}{3} \sum_{j < k} \frac{\epsilon_{ijk}}{4\pi}, \end{aligned} \quad (27)$$

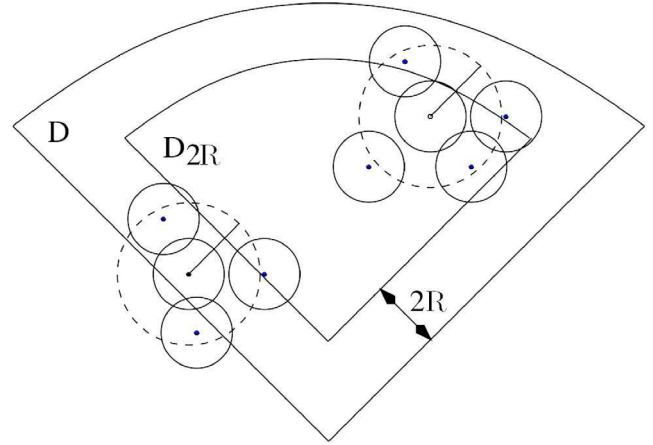


Figure 5. To avoid boundary effects when calculating the Partial MFs, we only take those points into account, which are more than two times the maximum Germ radius away from the survey mask.

where $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\| \leq 2r$ denotes the distance of two points, and $\rho_{ij} = \sqrt{r^2 - (d_{ij}/2)^2}$ the radius of its intersection circle.

The use of the Partial MFs also has the advantage that one can obtain an error estimate for the fluctuations of $V_\mu(\mathcal{B}_r)$ by calculating the variance of the values of the Partial MFs.

3.3 Treatment of boundaries

The family of MFs allows a complete deconvolution of the boundary, based on the principal kinematical formula, see for example the review of Kerscher (Kerscher 2000, and references therein). Note that this formula involves all the functionals of the family; for individual functionals, the boundary cannot be corrected with this powerful tool. Most of the previously cited papers refer to this method for the boundary correction. The principal kinematical formula reads

$$m_\mu(\mathcal{B}_r) = \frac{M_\mu(\mathcal{B}_r \cap W)}{M_0(W)} - \sum_{\nu=0}^{\mu-1} \binom{\mu}{\nu} m_\nu(\mathcal{B}_r) \frac{M_{\mu-\nu}(W)}{M_0(W)}, \quad (28)$$

where $M_\mu(\mathcal{B}_r)$ are the MFs, and $m_\mu(\mathcal{B}_r)$ their mean volume densities as defined in Kerscher (2000). W denotes the boundaries, i.e. the survey mask or window. For an example illustrating these boundary corrections, we recommend Kerscher, Schmalzing & Buchert (1996a), for an application to a galaxy catalogue see e.g. Kerscher et al. (1997).

Unlike in these previous papers, we here calculate the Partial MFs only for points more than *twice* the maximum radius away from the boundary, i.e. the sample mask, see Fig. 5. Thus, we create a shrunk 'calculation window' D_{2R} and do not have to take into account any boundary effects. Naturally, if the survey mask is full of holes, we neglect a lot of galaxies this way, so this approach is better suited for modern galaxy catalogues like the SDSS and after-SDSS surveys.

However, it is important to note that the neglected points do count when it comes to calculating Partial MFs, since their Balls intersect with Balls inside of the window. Therefore, they have to be part of the neighbourlists.

The MFs volume densities (26) now take the form

$$v_\mu = \frac{1}{|D_{2R}|} \sum_{i=1}^N \chi_{D_{2R}}(\mathbf{x}_i) V_\mu(\mathbf{x}_i; r), \quad (29)$$

where

$$\chi_{D_{2R}}(\mathbf{x}_i) = \begin{cases} 1 & \text{if } \mathbf{x}_i \in D_{2R} \\ 0 & \text{if } \mathbf{x}_i \notin D_{2R} \end{cases} \quad (30)$$

is the characteristic function of the shrunk window. As mentioned in Schmalzing & Diaferio (2000), these quantities are minus estimators for the MF's volume densities. Minus estimators only provide unbiased estimates if applied to stationary point processes, as investigated by Kerscher (1999). Hence, we use volume-limited subsamples of the catalogues when carrying out the structure analysis.

3.4 The structure volume

Since the volume of the structure is not localized on its surface, we cannot calculate it in the way outlined above. However, to achieve analogy to the three other functionals, and in view of the possibility of parallel computing, our goal was to determine the volume by means of adding up the partial functionals.

We do this as follows: first, we throw a number of randomly distributed points \mathbf{y}_i into the shrunk mask D_{2R} of the sample; secondly, we determine neighbourlists for the random points. These neighbourlists consist of the thrown point as the centre and the *real* galaxies in its vicinity, i.e. points of the sample we analyse, within a distance of twice the maximum radius, see Fig. 6; thirdly, in a second Poisson process, we throw random points into the Ball around \mathbf{y}_i , i.e. $\mathcal{B}(\mathbf{y}_i; r)$, and determine whether the random point is covered by a Ball around any of the *real* neighbours or not. This way, we calculate the fraction of volume covered by the structure in that local area. Hence, we defined a Partial MF $v_0(\mathbf{y}_i; r)$ for the volume similar to the other three (strictly speaking, we defined the volume density of the Partial MF); the last step for obtaining the global volume density of the structure consists of adding up the $v_0(\mathbf{y}_i; r)$

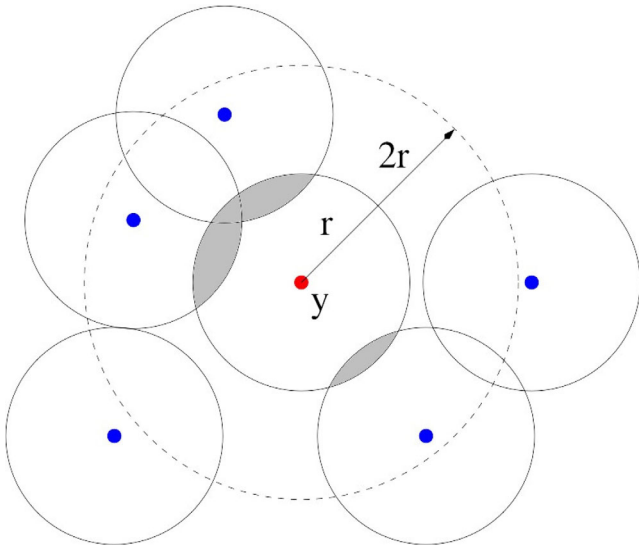


Figure 6. For any randomly thrown point \mathbf{y}_i , we generate the list of real neighbours, i.e. consisting of galaxies of the used sample. The Balls around all points within a distance of $2r$ around \mathbf{y}_i intersect with the Ball $\mathcal{B}(\mathbf{y}_i; r)$. The volume fraction of the structure within $\mathcal{B}(\mathbf{y}_i; r)$ is then calculated by a Monte Carlo integration.

and normalizing them by the number of random points, say M , for which we calculated the volume fraction,

$$v_0 = \frac{1}{M} \sum_{i=1}^M v_0(\mathbf{y}_i; r). \quad (31)$$

If we are interested in the absolute value of the structure's volume within the shrunk windows, we obtain it by multiplying the global volume fraction by the window's volume. Note: only the third and fourth step of the volume fraction calculation are executed by the CHIPMINK code itself, whereas the primary steps are data preparation. So instead of throwing points into the shrunk window D_{2R} in the first step, one can throw them into the original survey mask and create neighbourlists for all of them. Thus, calculations for different maximum radii or specific areas of the survey mask can be carried out with subsamples of this set of neighbourlists.

For consistency checks, we also provide two other methods for the calculation of the structure volume. However, these methods only calculate the global volume fraction and they do not use the Partial MFs defined in this section.

4 APPLICATION: THE SDSS LRG SAMPLE

We shall now apply the code described in the previous section to two different data sets. First, to the LRG (Eisenstein et al. 2001) sample of the SDSS (York et al. 2000) DR7 (Abazajian et al. 2009), and, secondly, to the mock catalogues drawn from Λ CDM simulations of the SDSS volume performed by the LasDamas² collaboration McBride et al. (2009).

4.1 The data

From the SDSS DR7 LRG data described in Abazajian et al. (2009), we use in particular the samples extracted by Kazin et al. (2010). For selecting them the authors used the following criteria: the galaxy has an SDSS spectrum, is not in an area around bright stars, has a sector completeness of at least 60 percent, a redshift in the range 0.16–0.47 and a colour- and k -corrected magnitude between -21.2 and -23.1 . The details of the selection can be found in Kazin et al. (2010). After this pre-selection, the sample contains 105 831 LRGs.

We neglect the small amount of area in the southern galactic regions from the DR7 sample, as our code requires contiguous regions (note that boundaries can be corrected by integral-geometrical means; this property is exploited in our previous codes). In addition, in order to have simpler boundaries, we choose to restrict ourselves to a region with $\text{RA} \in [132^\circ, 235^\circ]$ and $\text{Dec.} \in [-1^\circ, 60^\circ]$.

With the radial selection we have to make sure that the sample we obtain is volume limited. According to Kazin et al. (2010), the sample is volume limited up to a redshift of 0.36 for a magnitude of -21.2 and up to a redshift of 0.44, if we select galaxies with a magnitude brighter than -21.8 . This implies that we shall analyse two different samples derived from the pre-selected galaxies: a first one that we shall refer to as the 'dim sample' with a magnitude cut at -21.2 and a redshift in the range $z \in [0.16, 0.35]$, and a second one to which we refer to as the 'bright sample' with a magnitude cut at -21.8 and a redshift in the range $z \in [0.16, 0.44]$. With these requirements, the 'dim sample' contains 41 375 galaxies and the 'bright sample' 22 386.

² See <http://lss.phy.vanderbilt.edu/lasdamas/> for information on the project and for downloading the samples.

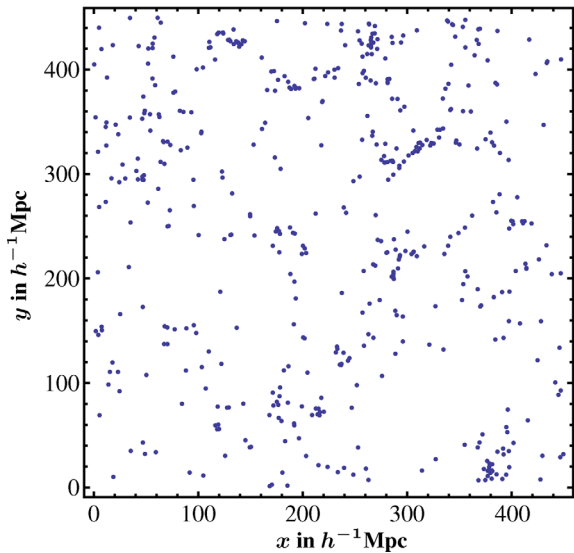


Figure 7. Example of the galaxy distribution in the ‘dim sample’. The projection is extracted from a slice of a thickness of $22.5 h^{-1} \text{ Mpc}$ from the maximally fitting cube.

In order to compare the structure in the galaxy data to the model of gravitational structure formation, we use mock data samples provided by the LasDamas² Collaboration (McBride et al. 2009). These authors simulated structure formation in a Λ CDM model with $\Omega_{\Lambda} = 0.75$ in boxes of $2.4 h^{-1} \text{ Gpc}$ for 1280^3 particles. They identify haloes with a friends-of-friends algorithm, and populate them with mock galaxies using a halo occupation distribution (HOD). The HOD parameters are chosen as to reproduce small-scale clustering of the observed LRG sample. From 40 independent N -body simulations, LasDamas provides 160 sky-based mock galaxy catalogues for the northern SDSS region that we use.

We further modify the basic catalogues they provide by also removing regions around bright stars³ and performing the same angular cut as for the SDSS data samples. This results in catalogues that contain on average 46 710 galaxies for the ‘dim sample’ and 22 181 galaxies for the ‘bright sample’.

Our treatment of various issues encountered in data taking, which should be included in the mock galaxy selection, is not complete. It does not take problems like e.g. sector completeness or fibre collisions into account. To deal with these issues, the necessary weights could probably be included into equation (25), but we leave it for a more thorough analysis in future work to figure out the exact form of this weighting. For the purpose of this paper, testing the code and a general analysis of the influence of higher clustering we do not need this precision.

Instead of considering the MFs in redshift space, we convert all redshifts into comoving distances using the distance redshift relation of a Λ CDM model with $\Omega_{\Lambda} = 0.75$. An example of the galaxy distribution in the ‘dim sample’ is shown in Fig. 7. It also helps to recall the dimensions of the sample. The thickness of the z -shell of the ‘dim sample’ $z \in [0.16, 0.35]$ is $507 h^{-1} \text{ Mpc}$. The thickness of z -shell of the ‘bright sample’ $730 h^{-1} \text{ Mpc}$. The largest

cube that fits into our ‘dim sample’ region has a side length of $452 h^{-1} \text{ Mpc}$; Fig. 7 presents a slice of this cube. In Wiegand (2012), two independent cubes of this size have been used to demonstrate the stability of the MFs throughout the sample.

4.2 The functionals on different scales

We now turn to the analysis of the samples defined in the previous section. In this analysis, we compare the structure in the observed samples to the structure in the mock samples. For this comparison it is crucial to know, how precise our results for the MF densities are. To estimate these errors, we determine the MF densities for each of the 160 mock samples and calculate the error bars from the resulting fluctuation. For comparison, we also calculated the error bars from random subsampling *jackknife* realizations drawn from the data and consisting of 80 per cent of the points of the samples. They turn out to be of the same magnitude. Finally, we also compared them to the error estimate that the CHIPMINK code determines directly from the fluctuations of the Partial MFs. Also in this case, the error bars are close to those determined from the mocks, even though systematically smaller by a few per cent. So, for a first estimate of the errors already the output of the code is quite useful.

There are two possible reasons for the MF densities (3) to fluctuate between different realizations. First, the N -point correlation functions of the point distribution in different realizations may be different. Then, the integrals (21) and therefore the coefficients in the expansion (5) vary and lead to fluctuations in the measured v_{μ} . But the series (5) indicates that also a different average density ϱ_0 of the sample will lead to a change of the measured v_{μ} . This means that we have to ensure that all the realizations approximately have the same density, if we really want to compare the structure of the point distribution. Otherwise, the analysis of the influence of (higher order) correlations in the point distribution would be spoiled by a fluctuation in ϱ_0 .

To ensure that all realizations have the same density ϱ_0 , we implement a random choice of ≈ 80 per cent of the sample that discards all configurations that do not have the desired density. Due to slightly different average densities of the mock and data samples this fraction is not exactly 80 per cent, but is adjusted to give the same average density for the mocks and the data.

4.2.1 The ‘dim sample’

Fig. 8 shows the MF densities obtained from this procedure and the code described in Section 3.

We plot the average from 244 random 80 per cent realizations of the ‘dim sample’. The functionals were evaluated for Balls of 40 different radii with a spacing of about $1 h^{-1} \text{ Mpc}$. They are shown as the (red) points in Fig. 8.

The selection of 80 per cent of the points results in an average density of $\varrho_0 = 7.7 \times 10^{-5} h^3 \text{ Mpc}^{-3}$. For the mock catalogues, we also extract 244 configurations of this average density ϱ_0 from each of the 160 simulated samples and calculate the average value for the MFs of each mock with the same radii of the Balls. The mean value and the error bars are then calculated from the mean and the variance of these 160 averages.

The upper four plots in Fig. 8 indicate that the determination of the MF densities is quite robust. The 1σ error bands around the average are barely visible. The values for a Poisson distribution with the same density lie far away from both mocks and observed

² See <http://lss.phy.vanderbilt.edu/lasdams/> for information on the project and for downloading the samples.

³ We use the software *mangle* Swanson et al. (2008) to apply the mask that can be found in the NYU value-added catalogue (Blanton et al. 2005; Padmanabhan et al. 2008).

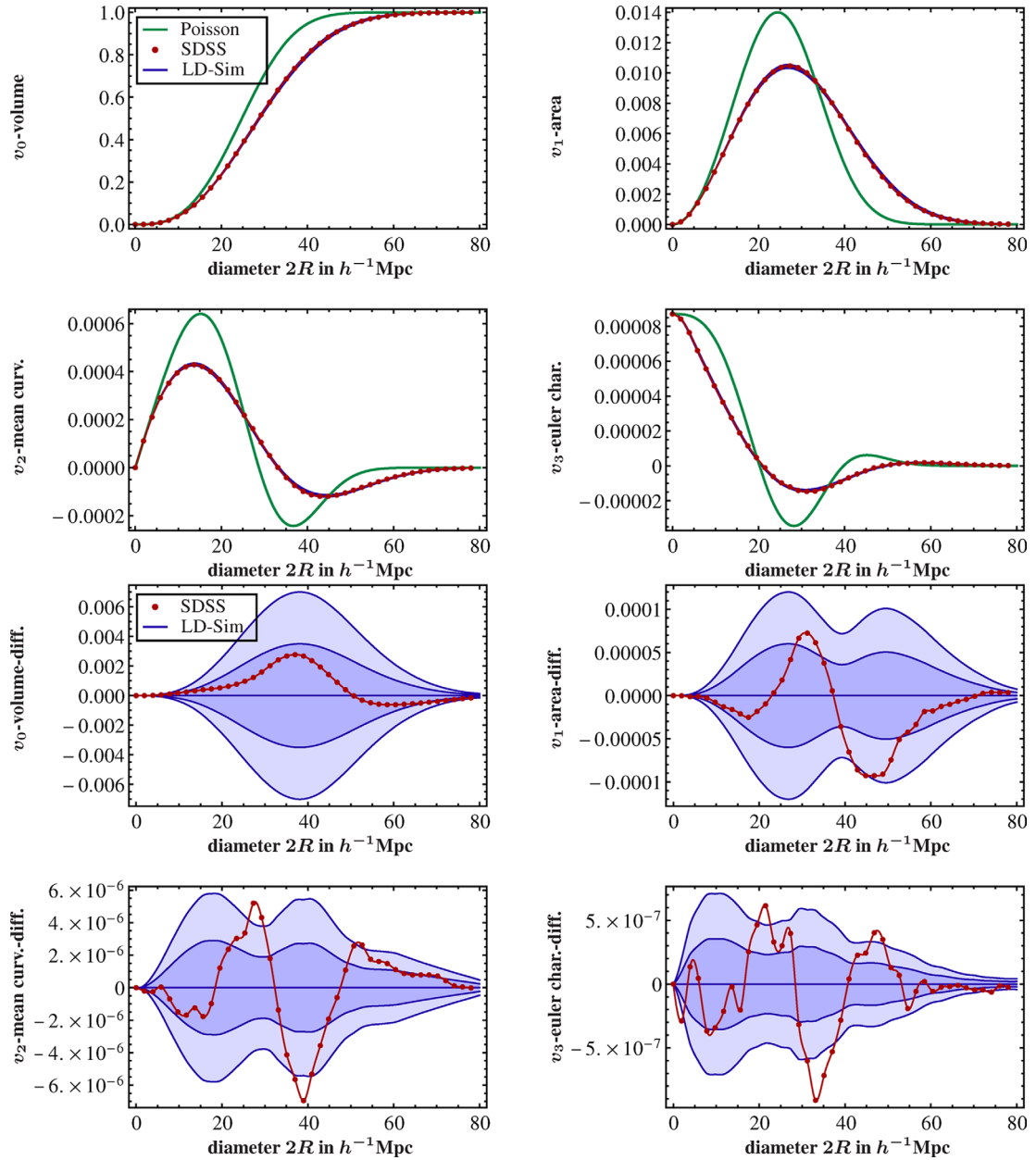


Figure 8. Top four: the four Germ–Grain MF densities for the SDSS LRG ‘dim sample’ compared to those of the corresponding LasDamas mock galaxies and those of a Poisson distribution. The errors and average for the mock samples are obtained taking 160 different mock realizations. Bottom four: the same quantities, but with the average of the mock sample subtracted to make the error bars more visible. The dark shaded regions are the 1σ , the light shaded regions 2σ error bands.

galaxies, a clear indication of the presence of structure (which is of course not surprising). For the curves of the mocks and observed galaxies, however, it is harder to distinguish them.

In order to facilitate the comparison of the observed and mock results, the lower four plots of Fig. 8 show the residuals obtained by subtracting the average of the mocks. Most of the 40 points lie in the 1σ and 2σ bands around the average of the mocks, but some points deviate more strongly.

In order to quantify how significant the resulting deviation actually is, we calculate the χ^2 values using the standard relation,

$$\chi^2 := \sum_{ij} (v_{\mu}^d(r_i) - \bar{v}_{\mu}^m(r_i)) \hat{C}_{ij}^{-1} (v_{\mu}^d(r_j) - \bar{v}_{\mu}^m(r_j)), \quad (32)$$

where $v_{\mu}^d(r_i)$ are the functionals measured from the SDSS, $\bar{v}_{\mu}^m(r_i)$ is the average of all mocks and \hat{C}_{ij}^{-1} is the inverse covariance matrix of the 40 points, estimated from the 160 mock realizations. We use the unbiased estimator of Hartlap, Simon & Schneider (2007) for \hat{C}^{-1} , which takes into account that the estimate is based on a finite number of mock samples.

The resulting χ^2 values are shown in the first column of Table 1. They are significantly higher than the average of the χ^2 distribution of 40 points. In order to quantify the deviation, we convert the p -value corresponding to the value of χ^2 into standard deviations of a Gaussian random variable, where the p -value is taken to correspond to the two sided deviation. The deviation is larger than 3σ for all of the four MFs. As this is surprising regarding the plots, and to

Table 1. χ^2 values for the deviation of the ‘dim sample’ from the LasDamas mocks for the MF densities v_μ and the modified functionals \bar{V}_μ . σ_G quantifies the p -value that corresponds to this χ^2 , in terms of standard deviations of a Gaussian. χ_I^2 quantifies the deviation, if we would assume statistical independence of the points. The data for v_μ are plotted in the lower panel of Fig. 8. The χ^2 is done over 40 degrees of freedom for v_μ , and over 20 degrees of freedom for \bar{V}_μ .

	χ^2	σ_G	χ_I^2	σ_{GI}		χ^2	σ_G	χ_I^2	σ_{GI}
v_0	78.7	3.66	21	10^{-2}	\bar{V}_0	43.9	3.17	11	0.1
v_1	111.0	5.65	33	0.3	\bar{V}_1	68.7	5.13	15	0.2
v_2	81.2	3.83	53	1.7	\bar{V}_2	65.1	4.87	17	0.4
v_3	93.3	4.62	68	2.9	\bar{V}_3	62.9	4.71	30	1.8

illustrate the amount of correlation in the data, we also list the χ^2 values we would obtain, if the data points were independent. We do this by taking the diagonal of the covariance matrix before inverting it. The resulting χ_I^2 values are in good agreement with the visual impression given the residuals in Fig. 8.

4.2.2 The ‘bright sample’

We applied the same procedure to the ‘bright sample’. In this case, randomly choosing 80 per cent of the points corresponds to a density of $\varrho_0 = 2.1 \times 10^{-5} h^3 \text{Mpc}^{-3}$. The resulting differences of the MF densities of the mocks to those of the SDSS ‘bright sample’ are shown in Fig. 9. The reduced galaxy density allows us to go to larger scales, due to the larger volume and a more restrictive selection. This is possible, because it needs a larger radius of the Balls to fill the observed volume completely. Therefore, the structure saturates for larger values of R only. The downside is, however, that we have less points and therefore a less precise determination of the average functionals. The plots in Fig. 9 show a similar deviation as those for the ‘dim sample’. When we evaluate the significance of this discrepancy, however, we find that it is less pronounced. The χ^2 values calculated with equation (32) and shown in Table 2, are much

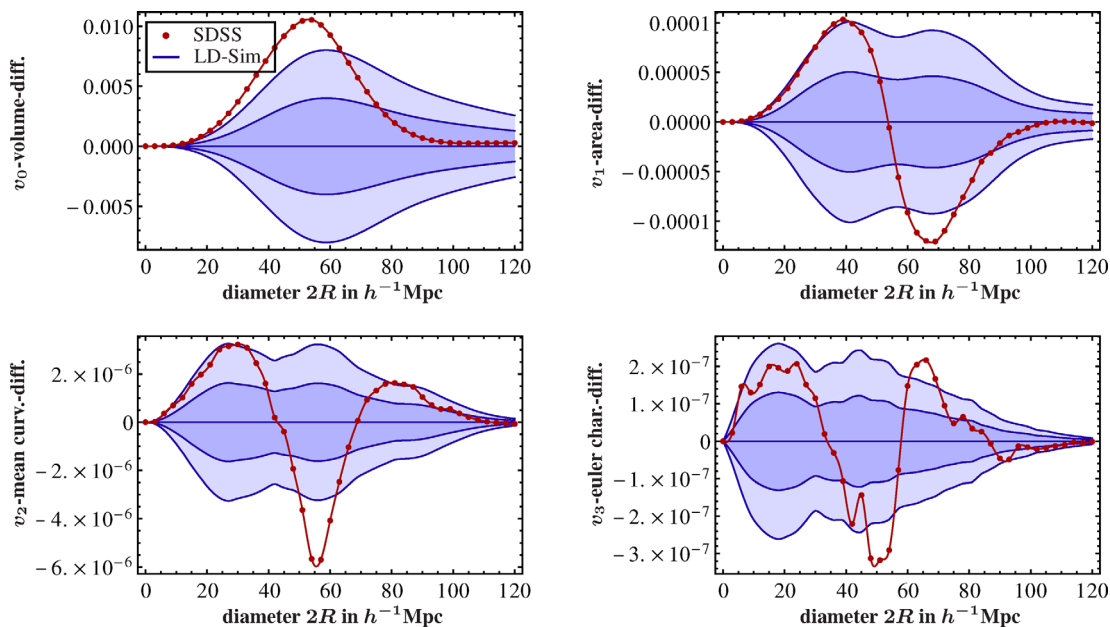


Figure 9. The same quantities as in the lower four plots of Fig. 8, but for the bright ‘sample’. The dark shaded regions are the 1σ , the light shaded regions 2σ error bands.

Table 2. Same as Table 1, but for the ‘bright sample’. The data for v_μ are plotted in Fig. 9 and those for \bar{V}_μ in the lower panel of Fig. 10. The χ^2 is done over 40 degrees of freedom for v_μ , and over 20 degrees of freedom for \bar{V}_μ .

	χ^2	σ_G	χ_I^2	σ_{GI}		χ^2	σ_G	χ_I^2	σ_{GI}
v_0	56.6	2.02	105	5.4	\bar{V}_0	35.3	2.36	112	7.8
v_1	52.6	1.71	74	3.3	\bar{V}_1	26.0	1.38	80	5.9
v_2	50.8	1.56	82	3.9	\bar{V}_2	31.7	1.99	66	4.9
v_3	36.0	0.46	70	3.0	\bar{V}_3	28.4	1.65	48	3.5

lower than for the ‘dim sample’. This is somewhat surprising but, apparently, the deviations follow more closely the form imposed by the correlation pattern.

As this correlation pattern is different for the modified MFs \bar{V}_μ , we also want to study their deviations from the mock samples and show them in Fig. 10. The upper four plots show them together with the Poisson case. This latter is very simple for \bar{V}_μ as it consists of the functionals of a Ball, equation (4). The values and error bars have been obtained by calculating the \bar{V}_μ for every realization and taking the average and variance of these values. As the errors grow rapidly beyond $60 h^{-1} \text{Mpc}$ (in diameter), we only plot the \bar{V}_μ up to this scale. The reason for this growth is that around $60 h^{-1} \text{Mpc}$, the volume becomes largely filled with the Balls, and therefore, the measurement has to become more and more accurate to give correct values after the removal of the exponential damping factor $e^{-\varrho_0 \bar{V}_\mu}$. Considering the scales where the errors are still controllable leaves us with the 20 points in Fig. 10. The χ^2 values of their deviation are summarized in the second part of Table 2. The reduced χ^2 values are a bit higher than for the densities v_μ , but also not significant.

5 NON-GAUSSIAN CORRELATIONS

Inspecting Figs 9 and 10, it is not clear whether the origin of the slight deviation of the data from the mocks is already present for

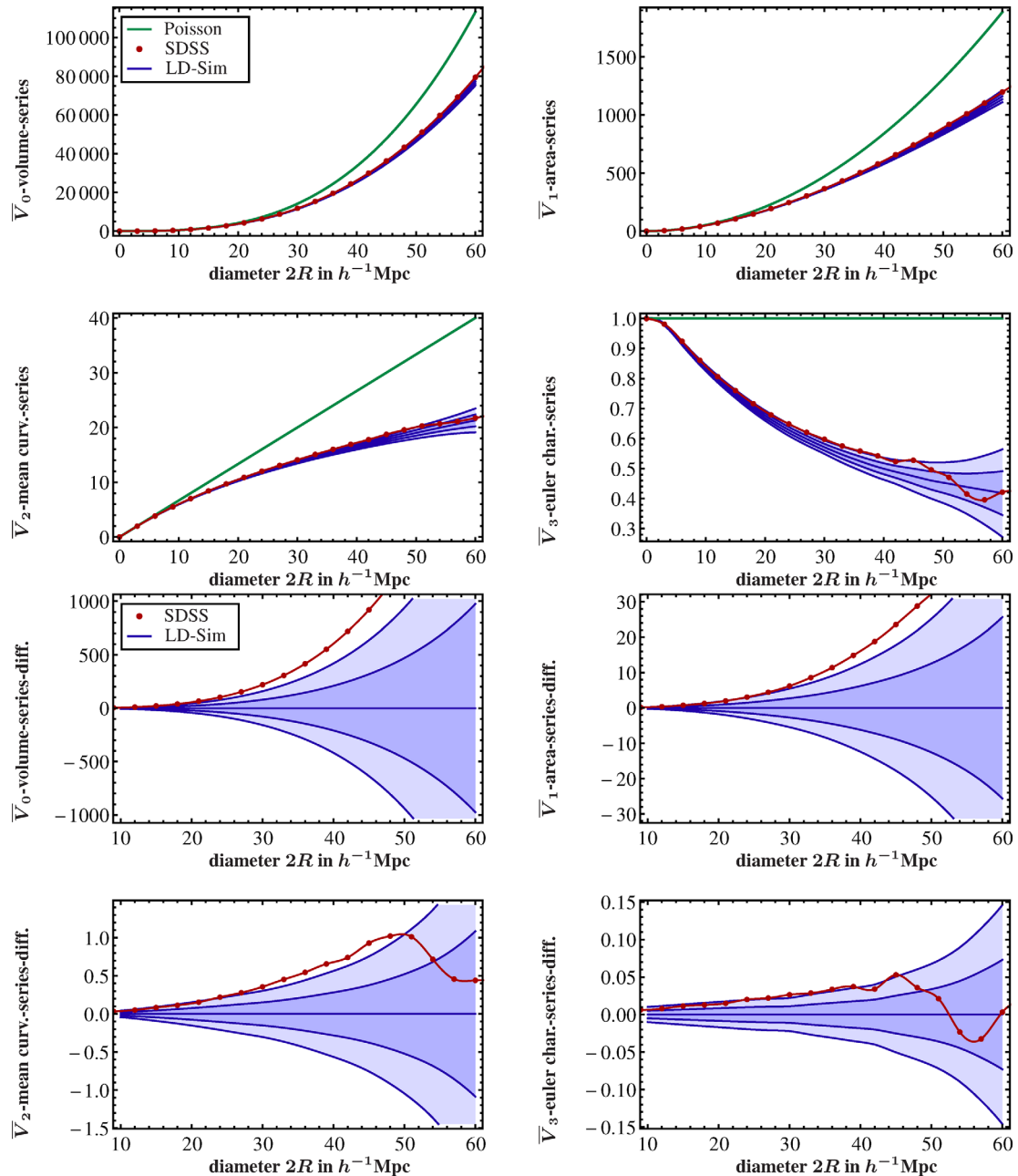


Figure 10. Top four: the four modified Germ–Grain MFs \bar{V}_μ of the SDSS LRG ‘bright sample’ and the LasDamas mock galaxies. The errors and average for the mock samples are obtained taking 160 different mock realizations. The Poisson case is simply given by the MFs of a single ball (see equation 4). Bottom four: the same quantities, but with the average of the mocks subtracted to make the error bars more visible. The dark shaded regions are the 1σ error bands, the light regions correspond to 2σ .

the two-point statistics, or whether it is due to a difference in the non-Gaussian properties showing up in the three- and N -point contributions to the series (5). To address this question more thoroughly, we use the method introduced in Section 2.5. This method requires to measure the MFs as a function of ϱ_0 , and to extract for every scale R the function $\tilde{V}_\mu(\varrho_0)$.

The procedure we use is the following: first, we choose 24 different densities corresponding to a fraction f between 0.05 and 0.8 of the full density of $\varrho_0 = 2.6 \times 10^{-5} h^3 \text{Mpc}^{-3}$. For each of these densities, we generate a large number of realizations (from about 15 000 for $f = 0.05$ down to 244 for $f = 0.8$). For each of these realizations, we determine the modified MFs \bar{V}_μ as a function of

the Ball radius R . For each R , we take the average over all the realizations and arrive at an $\bar{V}_\mu(\varrho_0)$ evaluated at 24 points. We add the value of \bar{V}_μ at $\varrho_0 = 0$ as the 25th point, which is simply the MF of a Ball (see equation 4) with the respective radius R . From these 25 points, we want to derive the first coefficients in the expansion (5) which, by equation (21), is equivalent to the determination of the components of a polynomial fit to $\tilde{V}_\mu(\varrho_0)$.

As the resampling of the points introduces an important amount of correlation between the realizations and as the r -dependence of the v_μ also contains some correlation, we use the 160 mock samples to estimate the covariance matrix for the points of the empirical

function $\tilde{V}_\mu(r, \varrho_0)$. We use $\tilde{V}_\mu(\varrho_0)$ at four different distances r and use its values at 20 of the 25 ϱ_0 points. This gives us a covariance matrix of dimension 80×80 . This is still quite large, given that we only have 160 mock samples: there is an important uncertainty in the estimation of the covariance matrix. In addition to the use of the unbiased estimator of Hartlap et al. (2007), as in Section 4.2, we, therefore, need to take into account the propagation of the errors in

the estimated covariance matrix to the values we want to measure. For doing so, we use the prescription in Percival et al. (2014). This considerably increases our error bars. We use the resulting covariance matrix for a polynomial fit of fourth order to the average $\tilde{V}_\mu(r, \varrho_0)$ for all the mock samples and for the SDSS. With equation (21) this gives us the coefficients b_2 – b_4 for both the mock samples and the SDSS.

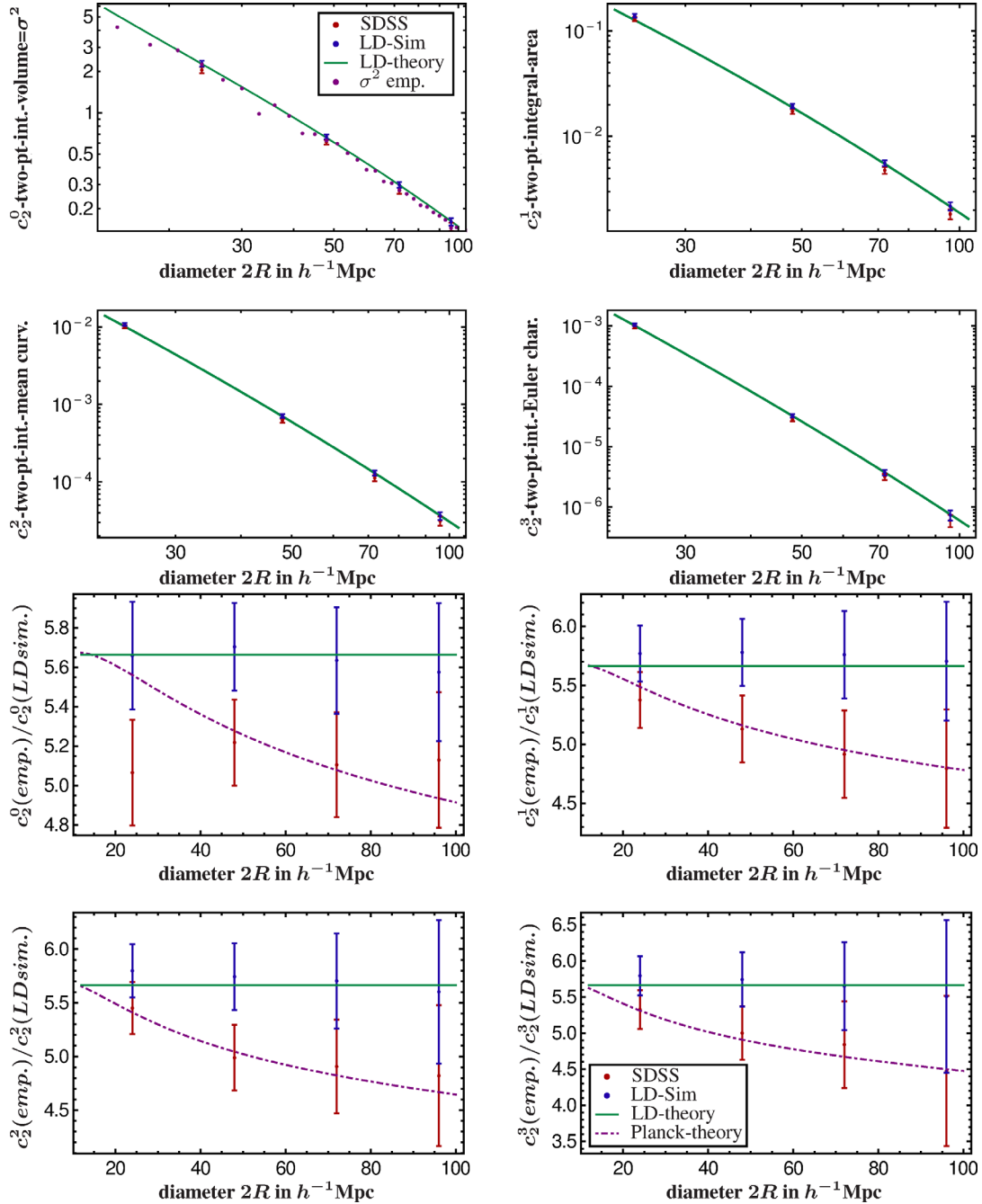


Figure 11. Top four: the four Minkowski integrals of the two-point correlation function, i.e. c_2^μ from equation (33), for the SDSS LRG ‘bright sample’ (lower points, red) and one of the corresponding mock samples (upper points, blue). The error bars are 1σ errors for the parameters from the fit $\tilde{V}_\mu(\varrho_0)$ with a polynomial of fourth order in ϱ_0 . The theoretical curves are calculated from the matter power spectrum corresponding to the parameters of the simulation using equation (13). As c_2^0 corresponds to σ^2 , we also plot the result of a standard estimation of σ^2 for the observed galaxies. Bottom four: the same quantities but all divided by the theoretical prediction of the LasDamas power spectrum. The value of the straight line indicates the bias between the linear correlation function and the correlations in the ‘bright sample’. The line following the trend of the SDSS data corresponds to the theoretical prediction for a power spectrum with the *Planck* parameters $\Omega_m \approx 0.32$. The χ^2 values are given in Table 3.

We finally divide these coefficients by the corresponding power of the volume and arrive at an estimate of the integrals

$$c_{n+1}^\mu = \frac{b_{n+1}^\mu}{V_0^{n+1}(B)} = \frac{1}{V_0^{n+1}(B)} \int_{\mathcal{D}} d^3x_1 \dots d^3x_n \times \xi_{n+1}(0, \mathbf{x}_1, \dots, \mathbf{x}_n) V_\mu(B \cap B_{x_1} \cap \dots \cap B_{x_n}), \quad (33)$$

where $V_0(B)$ is the volume of a Ball of radius R .

5.1 Integrals of the two-point correlation function

The results for the coefficients c_2^μ , i.e. those involving an integral over the two-point correlation function, are shown in the upper four plots of Fig. 11. As described in Section 2.4, the first quantity plotted, c_2^0 , is related to the matter variance in spheres of radius R . In fact, from equations (33) and (19), one can see that $c_2^0 = \sigma^2$. So, the first plot of Fig. 11 compares different ways of calculating σ^2 . The data and mock points and their error bars are derived from the MFs using the method just described. The points of σ_{emp}^2 are calculated from the observed sample with the usual estimator for the matter variance in spheres of radius R ,

$$\sigma^2(R) = \frac{\mathbb{E}[M(R)^2] - \mathbb{E}[M(R)]^2}{\mathbb{E}[M(R)]^2}, \quad (34)$$

where $M(R)$ is the integrated matter density of the sphere and $\mathbb{E}[X]$ is the average over all spheres.

The third way of calculating σ^2 is direct integration of the theoretical power spectrum using equation (13) which, for c_2^0 , directly becomes equation (18). For the form of $P(k)$, we use the parametrization of Eisenstein & Hu (1998) that includes the effects of baryons. However, we use the form without their oscillations. We normalize the amplitude of the power spectrum to match the amplitude of the mocks. This gives a linear bias of $b \approx 2.37$, i.e. $\sigma_{\text{mock}}^2 = b^2 \sigma_{\text{lin}}^2$.

In the V_1 – V_3 plots of Fig. 11, we use the power spectrum with the same normalization and equation (13) for the theory prediction. The first plot of Fig. 11 shows that all three ways of calculating σ^2 are overall in agreement. This is also true for the plots derived from \bar{V}_1 – \bar{V}_3 , but there is also a deviation of the coefficients c_2^μ of the observed galaxies from those of the simulated mock galaxies. On small scales they start quite close, but on larger scales the deviation becomes more important. It is interesting to see that the mocks are indeed well described by the theoretical power spectrum that entered into their simulation. This shows that the simulations and the extraction procedure give a consistent picture. The observed galaxies, however, deviate slightly from the simulated cosmology. This means that, even though the overall normalization of the correlation function is correct by the consistency of the σ^2 results, other features of the correlation function are not captured equally well.

The lower four plots of Fig. 11 show the same quantities, but divided by the theoretical prediction for the simulated cosmology. This removes the general trend and allows a more detailed comparison. Thus, in these plots, the scales on the y-axis have only a relative meaning: the value $y = 1$ marks the (scale-dependent) integrated mock power spectrum for $\Omega_m = 0.25$. The value of ≈ 5.6 for the straight mock line corresponds to the squared bias of the mocks b^2 . The line going through the points of the SDSS data, represents the c_2^μ calculated from a *Planck* $\Omega_m \approx 0.32$ power spectrum for comparison. Fig. 11 shows that the mild deviations of the MFs for the data and the mock galaxies, as found in Figs 9 and 10, already occur at the level of the first correction to the leading Poisson term in the expansion \bar{V}_μ , equation (5) (but they are a bit less significant

Table 3. χ^2 values for the SDSS data points in the lower panel of Fig. 11 from the theoretical curves. The table on the left compares the (red) SDSS points to the straight line of the LasDamas cosmology. The table on the right compares them to the dashed line of the *Planck* cosmology. The χ^2 values are over 4 degrees of freedom.

	χ^2	σ_G	χ_I^2	σ_{GI}		χ^2	σ_G	χ_I^2	σ_{GI}
c_2^0	11.4	2.30	16	3.0	c_2^0	8.24	1.73	3.7	0.76
c_3^1	5.11	1.09	12	2.4	c_3^1	0.75	0.07	0.25	0.01
c_3^2	8.84	1.84	10	2.0	c_3^2	1.34	0.18	0.15	0.003
c_3^3	3.22	0.64	7.7	1.6	c_3^3	0.35	0.02	0.14	0.003

here without the higher order information, see Table 3). We shall turn to the influence of the higher orders in the next section.

5.2 Integrals of the three-point correlation function

In addition to the integrals over the two-point correlation function of the previous section, which are completely describing a Gauss–Poisson point process, also the higher order terms are important for a general point distribution. In this sense, the first corrections to the Gaussian point distribution are the integrals c_3^μ .

These integrals c_3^μ are shown in Fig. 12. In comparison with Fig. 11, we recognize that the amplitude of the c_3^μ integrals is larger by a factor of 5 than the integrals c_2^μ . Like in Fig. 11, we also have a slight but again less significant deviation for these quantities.

The lines in these plots are calculated from a tree-level bispectrum using equation (22) and the bias of $b = 2.37$ as found from the two-point normalization.

As described in the beginning of this section, we use a fit up to ϱ_0^4 and, therefore, we have also determined the coefficients c_4^μ . However, the quality of the determination of those coefficients becomes even worse than already for the c_3^μ coefficients and therefore we do not plot them here.

5.3 Importance of higher order correlations

To get an impression of how well the first four terms in the series (5) already describe the MFs of the ‘bright’ galaxy sample, we use the coefficients $c_2^\mu(R) - c_4^\mu(R)$, as obtained from the fit to $\bar{V}_\mu(\varrho_0)$ in the two previous sections, to calculate the \bar{V}_μ of equation (5) up to $n = 3$.

This decomposition allows us to address another interesting question: how non-Gaussian the point distribution actually is. For a pure Gauss–Poisson point process, we have seen in Section 2.4 that a truncation after $n = 1$ is exact. The comparison of the contribution of $c_2^\mu(R)$ with the other components $c_{n+1}^\mu(R)$ provides a measure of non-Gaussianity. This is only a meaningful test, however, if the distribution can be Gaussian in the first place. Due to the positive definite matter density field ϱ there are some restrictions, as discussed around equation (6). For a Gauss–Poisson process to exist, the density must be low enough for a given amount of two-point correlation. The precise relation between ϱ_0 and ξ_2 is that of equation (6):

$$\varrho_0 \int_{\mathcal{D}} d\mathbf{y} \xi_2(|\mathbf{y}|) \leq 1.$$

For the strongly clustered SDSS LRGs, this requirement is not fulfilled for the full density of the sample. Also for 80 per cent of the density like in Figs 9 and 10, where the density was $\varrho_0 = 2.1 \times 10^{-5} h^3 \text{Mpc}^{-3}$, the two-point amplitude is too high to allow

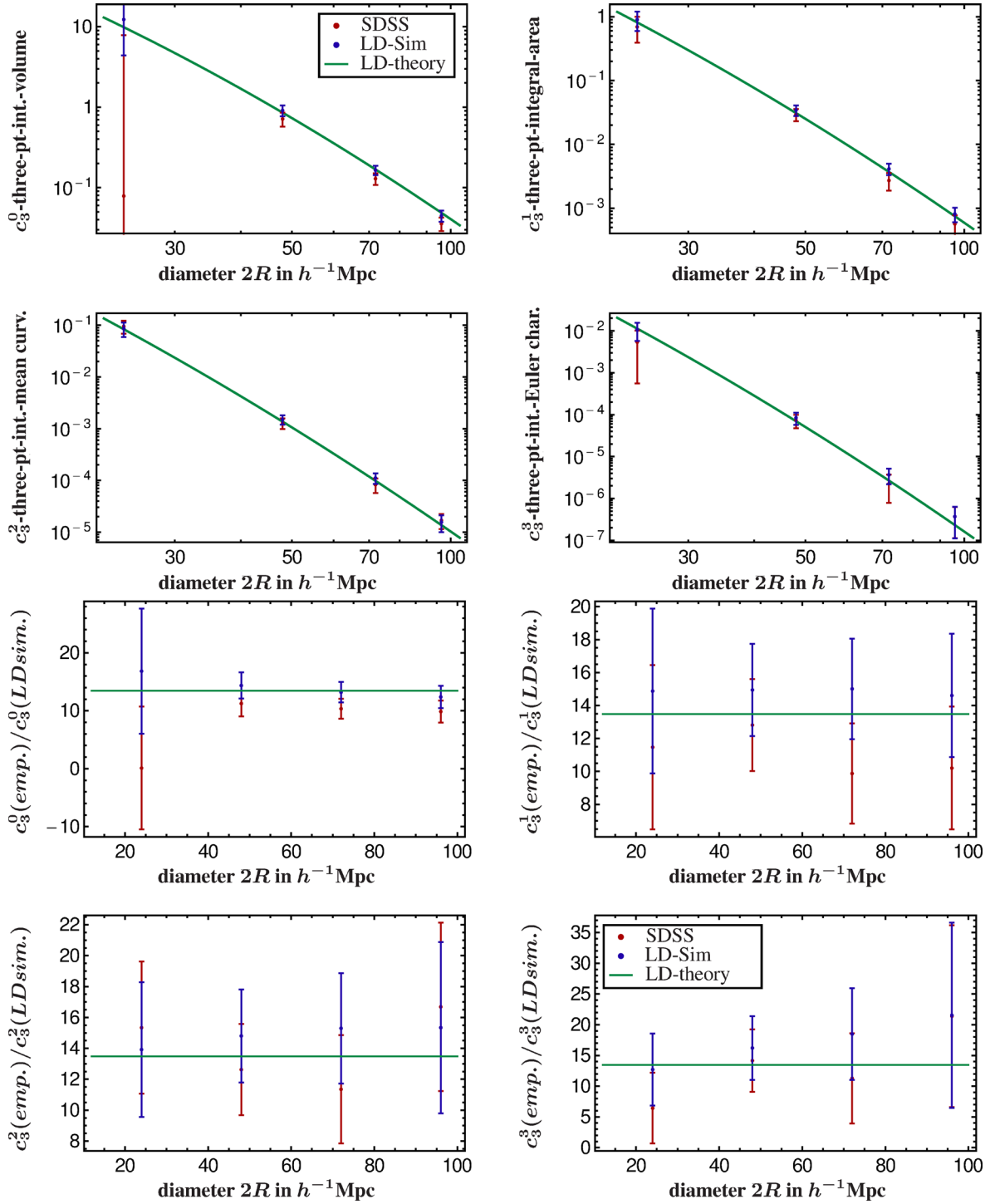


Figure 12. Top four: the four Minkowski integrals of the three-point correlation function, i.e. c_3^μ from equation (33), for the SDSS LRG ‘bright sample’ (lower points, in red) and the average of the corresponding mock samples (upper points, blue). The error bars are the 1σ errors from the diagonal of the covariance matrix of the fit of $\bar{V}_\mu(\varrho_0)$ with a polynomial of fourth order in ϱ_0 . The lines are the integrals of equation (22) evaluated for the LasDamas cosmology. Bottom four: the same quantities but all divided by the theoretical prediction of the LasDamas power spectrum. The value of the straight line indicates the (linear) bias between the linear three-point correlation function and the correlations in the ‘bright sample’.

for a Gaussian approximation. So already from this condition, we know that the point sample is not a Gaussian distribution. However, for a sample having 30 per cent of the full density, i.e. $\varrho_0 = 0.78 \times 10^{-5} h^3 \text{Mpc}^{-3}$, the condition is (marginally) satisfied. Therefore, we calculated the modified MFs for a large number of realizations

of 30 per cent of the ‘bright sample’ mocks in the same way as in Section 4.2.

Fig. 13 shows the influence of the first three components $c_{n+1}^\mu(R)$ on the series of the modified MFs \bar{V}_μ . We confirm that the mock point catalogues are not a realization of a

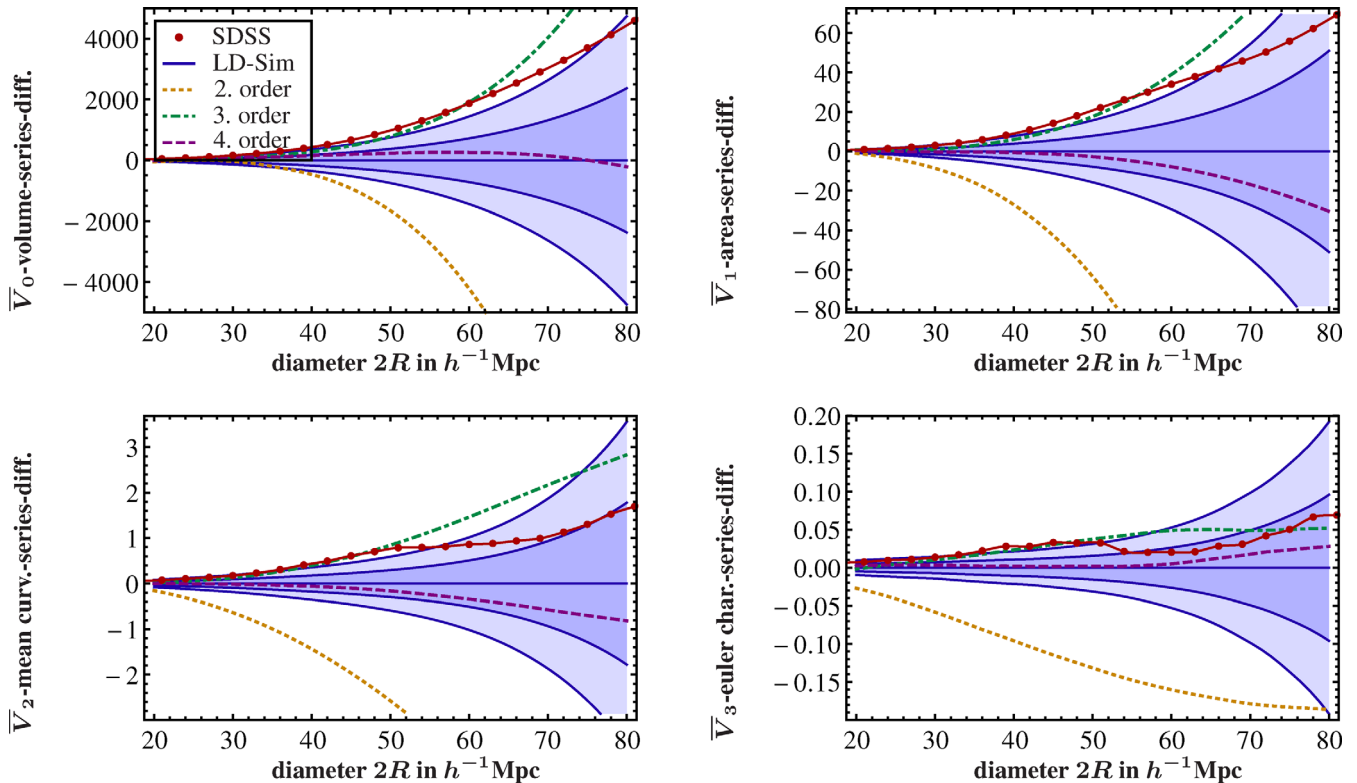


Figure 13. The four modified MFs \bar{V}_μ of 30 per cent of the ‘bright sample’, after subtraction of the average of the mocks. The (yellow) dotted line shows the theoretical expectation for a pure Gauss–Poisson process of this density and this amount of two-point correlation. The (purple) dashed line gives the functionals as derived from the expansion (5) truncated at ϱ_0^3 with the coefficients $c_2^\mu - c_4^\mu$ as found in Sections 5.1 and 5.2.

Gauss–Poisson process and that higher order corrections are crucial for the MFs.

With the inclusion of c_3^μ and c_4^μ , this truncated series is quite good in describing the modified MF densities of the data up to a scale of around $60 h^{-1}$ Mpc. For larger scales, the coefficients b_{n+1}^μ become too big (even though the c_{n+1}^μ decay with R , they decay slower than $V_0^{n+1}(B)$ and, therefore, the b_{n+1}^μ grow), and the quality of the approximation becomes worse. This deviation from the approximated function shows that, even for densities as low as the present $\varrho_0 = 0.78 \times 10^{-5} h^3 \text{ Mpc}^{-3}$, the MFs include contributions from galaxy correlations way beyond the standard two-point correlations. This confirms the claim made in the introduction that they are sensitive to higher order correlations. For higher densities these contributions become even more important as \bar{V}_μ , equation (5), is a power series in ϱ .

5.4 Coefficients of a lognormal distribution

The non-Gaussianity investigated in the previous section is, of course, not surprising. It is well known that the non-linear evolution of the density field leads to deviations from the initial Gaussian distribution. A better model is provided by a lognormal distribution, proposed by Coles & Jones (1991) and investigated in Colombi (1994), Wang et al. (2011). Comparing it to simulations, it has been shown to reproduce well the one-point function (see e.g. Kofman et al. 1994) and has been used recently to enhance the extraction of information from the two-point function (Neyrinck, Szapudi & Szalay 2009; Seo et al. 2012). In comparison with a Gaussian distribution, however, it has the inconvenience not to be described by a simple truncation of the \bar{V}_μ series. This is why we concentrated on

the Gaussian reference case in the previous section. Nevertheless, the lognormal distribution has the interesting property that, even though it contains an infinite number of higher order correlations, they are not independent of the two-point properties; the knowledge of the two-point function fixes the form of the higher order terms. We shall exploit this property in this section to compare the c_{n+1}^μ coefficients we extracted from the mock samples with the lognormal prediction.

We first have to choose the two-point properties for which we use the standard two-point correlation function for the LasDamas cosmology. Then, we can calculate, in principle, all higher connected correlation functions for a multivariate lognormal distribution. In practice, however, this quickly leads to long expressions, a combinatorial increase in the number of two-point functions. Therefore, we only use connected correlation functions up to ξ_4 .

Using the ξ constructed in this way, we can evaluate the series (5) up to an index $n = 4$. However, we restrict ourselves to V_0 , because the necessary integral can be expressed for any n in a rather compact way.

Fig. 14 then shows the result. The two-point contribution matches quite well, which is not surprising as it is the same as in the previous case. The integrals over the higher order correlation functions, however, deviate quite strongly from the data measured from the mock samples. This shows that higher order correlations can also be crucial to test improved reference models.

6 CONCLUSION

In this paper, we provided new analysis tools based on the Germ–Grain model to calculate the family of MFs of point sets.

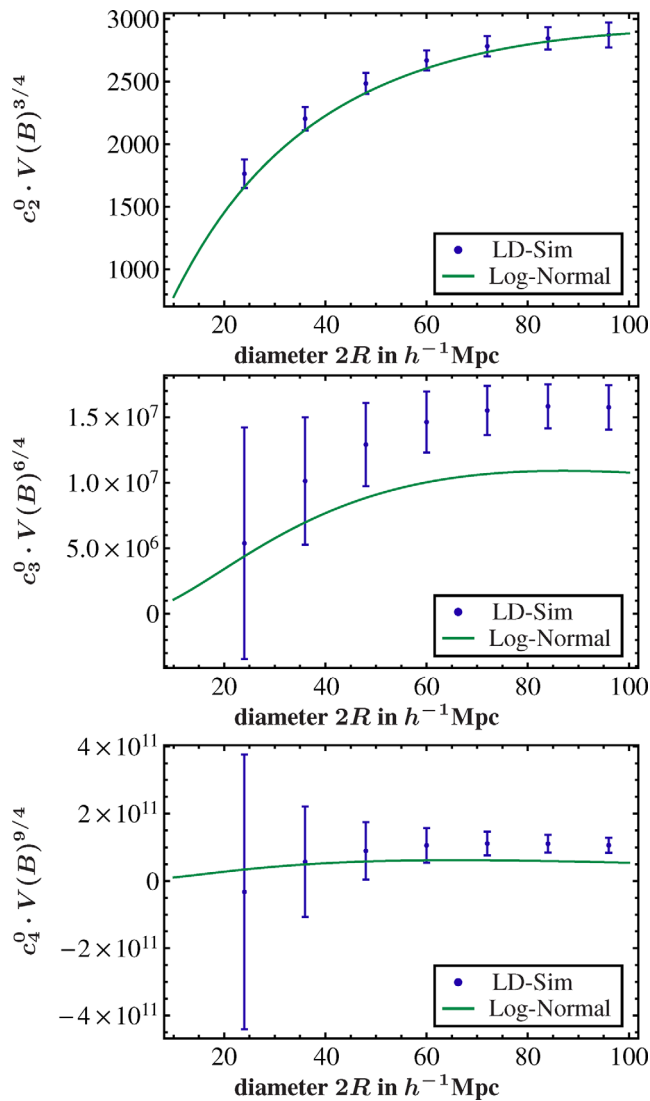


Figure 14. Comparison of the coefficients c_2^0 – c_4^0 defined in equation (33) to those of a lognormal distribution with the same two-point function. For a better visibility, the coefficients c_n^0 have been multiplied by $V(B)^{3/4(n-1)}$ in these plots.

The code is made available to the community with this paper that complements the existing Germ–Grain codes. The advantage of the former codes is still the possibility to deconvolve the boundaries, which is especially needed for sparse catalogues with complicated mask structure. The advantage of the new code has to be seen in the fast performance for large samples, the possibility to study the Partial MFs locally and in explicit relation to correlation properties of the data sets.

As a first application of the new code, we also performed an extensive analysis of the MFs of the SDSS LRG sample for the Germ–Grain model. We favour this model because it provides a direct way to analyse the data without additional smoothing, and because it provides analytical relations between the MFs and the N -point correlation functions of the galaxy distribution. We especially had a detailed look at those correlation properties to locate the deviations found between the mock samples and the SDSS data.

In Section 4, we compared the MFs of the observed galaxies to a grid of Λ CDM N -body simulations of the galaxy distribution, for two different luminosity thresholds. In both cases, the MFs of

the observed and the simulated galaxy distributions show significant disagreement. For the galaxies with the higher luminosity, this difference is statistically less important than for the ‘dim’ LRG sample.

Making use of the analytic relation to the correlation functions provided in this paper, we derived in Section 5 some integrals over the two- and three-point function of the galaxy distribution. We compared the results to the prediction for structure in a perturbed Friedmannian universe. We found that this prediction describes quite well the simulated galaxy distribution, but fails to describe the observed one.

We also showed that the galaxy distribution is clearly different from a Gauss–Poisson distribution (and for higher orders also from a lognormal distribution), which demonstrates how higher order correlations are crucial to describe the observed structure. As previously often emphasized, it is necessary to address higher order correlations for the purpose of determining morphological fluctuations. It is not sufficient, if the density or the two-point measures agree on selected samples, in order to conclude on the absence of significant fluctuations or the reality of large structures.

It will be interesting to see, which aspects of these results will become more significant in larger and deeper surveys.

ACKNOWLEDGEMENTS

We thank Martin Kerscher, Carlo Schimd, Francesco Sylos Labini and Herbert Wagner for helpful discussions. AW thanks Lars Andersson, Steffen Hess and Cameron McBride for interesting remarks. The work of AW was partially supported by the DFG under grant GRK 881. The work by TB was conducted within the ‘Lyon Institute of Origins’, Grant No. ANR-10-LABX-66. The Minkowski code CHIPMINK is a completely revised version by MO, with additions by AW, based on a former code architecture (including a parallel architecture) written by Jens Schmalzing and Alexander Rabus. TB and MO acknowledge many fruitful discussions with them. The MINKOWSKI-4 package is available on request by sending email to TB.

The simulations used for the MF analysis were carried out by the Large Suite of Dark Matter Simulations (LasDamas) project. The data are publicly available at <http://lss.phy.vanderbilt.edu/lasdamas/>.

Funding for the SDSS has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the US Department of Energy, the Japanese Monbukagakusho, and the Max Planck Society. The SDSS website is <http://www.sdss.org/>.

The SDSS is managed by the Astrophysical Research Consortium (ARC) for the Participating Institutions. The Participating Institutions are The University of Chicago, Fermilab, the Institute for Advanced Study, the Japan Participation Group, The Johns Hopkins University, Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, University of Pittsburgh, Princeton University, the United States Naval Observatory, and the University of Washington.

REFERENCES

- Abazajian K. N. et al., 2009, *ApJS*, 182, 543
- Aragon-Calvo M. A., Shandarin S. F., Szalay A., 2010, preprint ([arXiv:1006.4178](https://arxiv.org/abs/1006.4178))
- Beisbart C., 2001, PhD thesis, Ludwig-Maximilians-Universität München

- Blake C., James J. B., Poole G. B., 2013, *MNRAS*, 437, 2488
- Blanton M. R. et al., 2005, *AJ*, 129, 2562
- Buchert T., 1995, in Mücke J. P., Gottlöber S., Müller V., eds, *Large Scale Structure in the Universe*. World Scientific Press, Singapore, p. 156
- Carron J., Neyrinck M. C., 2012, *ApJ*, 750, 28
- Choi Y.-Y., Kim J., Rossi G., Kim S. S., Lee J.-E., 2013, *ApJS*, 209, 19
- Clowes R. G., Harris K. A., Raghunathan S., Campusano L. E., Söchting I. K., Graham M. J., 2013, *MNRAS*, 429, 2910
- Codis S., Pichon C., Pogosyan D., Bernardeau F., Matsubara T., 2013, *MNRAS*, 435, 531
- Coles P., Jones B., 1991, *MNRAS*, 248, 1
- Coles P., Melott A. L., Shandarin S. F., 1993, *MNRAS*, 260, 765
- Colombi S., 1994, *ApJ*, 435, 536
- Ducout A., Bouchet F. R., Colombi S., Pogosyan D., Prunet S., 2013, *MNRAS*, 429, 2104
- Einasto M. et al., 2011, *ApJ*, 736, 51
- Eisenstein D. J., Hu W., 1998, *ApJ*, 496, 605
- Eisenstein D. J., Annis J., Gunn J. E., Szalay A. S., Connolly A. J., Nichol R. C., 2001, *AJ*, 122, 2267
- Hadwiger H., 1957, *Vorlesungen über Inhalt, Oberfläche und Isoperimetrie. Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen mit besonderer Berücksichtigung der Anwendungsgebiete*. Springer, Berlin, p. 93
- Hartlap J., Simon P., Schneider P., 2007, *A&A*, 464, 399
- Hikage C., Matsubara T., 2012, *MNRAS*, 425, 2187
- Hikage C. et al., 2003, *PASJ*, 55, 911
- Kazin E. A. et al., 2010, *ApJ*, 710, 1444
- Keenan R. C., Barger A. J., Cowie L. L., 2013, *ApJ*, 775, 62
- Kerscher M., 1999, *A&A*, 343, 333
- Kerscher M., 2000, in Mecke K. R., Stoyan D., eds, *Lecture Notes in Physics, Vol. 554, Statistical Physics and Spatial Statistics. The Art of Analyzing and Modeling Spatial Structures and Pattern Formation*. Springer-Verlag, Berlin, p. 36
- Kerscher M., 2001, *Phys. Rev. E*, 64, 056109
- Kerscher M., Schmalzing J., Buchert T., 1996a, in Coles P., Martínez V., Pons-Borderia M.-J., eds, *ASP Conf. Ser. Vol. 94, Mapping, Measuring, and Modelling the Universe*. Astron. Soc. Pac., San Francisco, p. 247
- Kerscher M., Schmalzing J., Buchert T., Wagner H., 1996b, in Weiss A., Raffelt G., Hillebrandt W., von Feilitzsch F., Buchert T., eds, *Astro-Particle Physics. The significance of the fluctuations in the IRAS 1.2 Jy Galaxy Catalogue*. Technische Universität München, p. 83
- Kerscher M. et al., 1997, *MNRAS*, 284, 73
- Kerscher M., Schmalzing J., Buchert T., Wagner H., 1998, *A&A*, 333, 1
- Kerscher M., Pons-Bordería M. J., Schmalzing J., Trasarti-Battistoni R., Buchert T., Martínez V. J., Valdarnini R., 1999, *ApJ*, 513, 543
- Kerscher M., Mecke K., Schmalzing J., Beisbart C., Buchert T., Wagner H., 2001a, *A&A*, 373, 1
- Kerscher M. et al., 2001b, *A&A*, 377, 1
- Kofman L., Bertschinger E., Gelb J. M., Nusser A., Dekel A., 1994, *ApJ*, 420, 44
- Kratochvil J. M., Lim E. A., Wang S., Haiman Z., May M., Huffenberger K., 2012, *Phys. Rev. D*, 85, 103513
- Mantz H., Jacobs K., Mecke K., 2008, *J. Stat. Mech.: Theory Exp.*, 12, 15
- McBride C., Berlind A., Scocimarro R., Wechsler R., Busha M., Gardner J., van den Bosch F., 2009, *Am. Astron. Soc. Meeting Abstr.*, 213, 425.06
- Mecke K., 1994, *Reihe Physik 25: Integralgeometrie in der Statistischen Physik: Perkolations, komplexe Flüssigkeiten und die Struktur des Universums*. Deutscher, Thun Frankfurt/Main
- Mecke K., Stoyan D., 2000, in Mecke K. R., Stoyan D., eds, *Lecture Notes in Physics, Vol. 554, Statistical Physics and Spatial Statistics: The Art of Analyzing and Modeling Spatial Structures and Pattern Formation*. Springer-Verlag, Berlin
- Mecke K., Wagner H., 1991, *J. Stat. Phys.*, 64, 843
- Mecke K. R., Buchert T., Wagner H., 1994, *A&A*, 288, 697
- Melott A. L., Buchert T., Weiß A. G., 1995, *A&A*, 294, 345
- Modest H. I. et al., 2013, *MNRAS*, 428, 551
- Munshi D., Smidt J., Cooray A., Renzi A., Heavens A., Coles P., 2013, *MNRAS*, 434, 2830
- Nadathur S., 2013, *MNRAS*, 434, 398
- Nadathur S., Hotchkiss S., 2014, *MNRAS*, 440, 1248
- Nakagami T., Matsubara T., Schmalzing J., Jing Y., 2004, preprint ([astro-ph/0408428](http://arxiv.org/abs/astro-ph/0408428))
- Neyrinck M. C., Szapudi I., Szalay A. S., 2009, *ApJ*, 698, L90
- Padmanabhan N. et al., 2008, *ApJ*, 674, 1217
- Park C., Choi Y.-Y., Kim J., Gott J. R., III, Kim S. S., Kim K.-S., 2012, *ApJ*, 759, L7
- Percival W. J. et al., 2014, *MNRAS*, 439, 2531
- Petri A., Haiman Z., Hui L., May M., Kratochvil J. M., 2013, *Phys. Rev. D*, 88, 123002
- Planck Collaboration, 2013, *A&A*, preprint ([arXiv:1303.5083](http://arxiv.org/abs/1303.5083))
- Platzöder M., Buchert T., 1996, in Weiss A., Raffelt G., Hillebrandt W., von Feilitzsch F., Buchert T., eds, *Astro-Particle Physics, Applications of Minkowski-Functionals to the Statistical Analysis of Dark Matter Models*. Technische Universität München, p. 251
- Pratten G., Munshi D., 2012, *MNRAS*, 423, 3209
- Rabus A., 1998, Diploma thesis (German), Ludwig Maximilians Universität München
- Sahni V., Sathyaprakash B. S., Shandarin S. F., 1998, *ApJ*, 495, L5
- Sathyaprakash B. S., Sahni V., Shandarin S., 1998, *ApJ*, 508, 551
- Schmalzing J., 1999, PhD thesis, Ludwig Maximilians Universität München
- Schmalzing J., Buchert T., 1997, *ApJ*, 482, L1
- Schmalzing J., Diaferio A., 2000, *MNRAS*, 312, 638
- Schmalzing J., Gorski K. M., 1998, *MNRAS*, 297, 355
- Schmalzing J., Kerscher M., Buchert T., 1996, in Bonometto S., Primack J. R., Provenzale A., eds, *Dark Matter in the Universe*. IOS Press, Oxford, p. 281
- Schmalzing J., Gottlöber S., Klypin A. A., Kravtsov A. V., 1999a, *MNRAS*, 309, 1007
- Schmalzing J., Buchert T., Melott A. L., Sahni V., Sathyaprakash B. S., Shandarin S. F., 1999b, *ApJ*, 526, 568
- Schmalzing J., Takada M., Futamase T., 2000, *ApJ*, 544, L83
- Seo H.-J., Sato M., Takada M., Dodelson S., 2012, *ApJ*, 748, 57
- Stoyan D., Kendall W. S., Mecke K., 1987, *Stochastic Geometry and its Applications*, Vol. 8. Wiley, Chichester
- Swanson M. E. C., Tegmark M., Hamilton A. J. S., Hill J. C., 2008, *MNRAS*, 387, 1391
- Sylos Labini F., 2011, *Europhys. Lett.*, 96, 59001
- Wang X., Neyrinck M., Szapudi I., Szalay A., Chen X., Lesgourgues J., Riotto A., Sloth M., 2011, *ApJ*, 735, 32
- Whitbourn J. R., Shanks T., 2014, *MNRAS*, 437, 2146
- Wiegand A., 2012, PhD thesis, Universität Bielefeld
- York D. G. (SDSS Collaboration) et al., 2000, *AJ*, 120, 1579

APPENDIX A: DERIVATION OF EQUATION (3)

As we make heavy use of the analytic relation (3), we recall here the derivation of this formula in the general formulation that we need here. The version for the Poisson case has been shown in Mecke & Wagner (1991). Schmalzing (1999) and Schmalzing et al. (1999a) state the general case, but only sketch the derivation.

We shall prove equation (3) for a statistically homogeneous point process on a three-dimensional support $\mathcal{D} \in \mathbb{E}^3$. The key ingredients for the proof will be the additivity and motion invariance of the MFs, the principal kinematical formula and statistical homogeneity (i.e. the motion invariance of statistical quantities characterizing the point process).

To derive the relation for all four MFs in one go, we first define the Minkowski polynomial as follows:

$$M(t; K) = \sum_{\mu=0}^3 \frac{t^\mu}{\mu!} \alpha_\mu V_\mu(K), \quad (\text{A1})$$

with $\alpha_\mu = \frac{\omega_{d-\mu}}{\omega_d}$ and $\omega_j = \frac{\pi^{j/2}}{\Gamma(j/2+1)}$. To recover the functionals, we can take the derivatives

$$V_\mu(K) = \frac{1}{\alpha_\mu} \frac{\partial^\mu M(t; K)}{\partial t^\mu} \Big|_{t=0}. \quad (\text{A2})$$

These Minkowski polynomials obey the additivity relation of the MFs, $\forall \mathcal{B}_1, \mathcal{B}_2 \in \mathcal{R}$ (\mathcal{R} is the ring of polyconvex bodies):

$$M(\mathcal{B}_1 \cup \mathcal{B}_2) = M(\mathcal{B}_1) + M(\mathcal{B}_2) - M(\mathcal{B}_1 \cap \mathcal{B}_2). \quad (\text{A3})$$

So, for a collection of N Balls located at the positions r_i [in short $B(r_i) = B_i$ in the following], we have

$$M\left(\bigcup_{i=1}^N B_i\right) = \sum_{i=1}^N M(B_i) - \sum_{i<j} M(B_i \cap B_j) + \dots + (-1)^{N+1} M(B_1 \cap \dots \cap B_N), \quad (\text{A4})$$

where $M(B_1 \cap \dots \cap B_N)$ is the Minkowski polynomial of the intersection of the N Balls, which is 0 if they do not intersect.

Equation (3) describes the statistical average over an ensemble of such realizations of the point process. To calculate these average MFs for a point process described by a given set of N -point correlation functions up to $n = N$, we weight such a configuration of Balls with its probability

$$p \propto \varrho_N(x_1, x_2, \dots, x_N) dV_1 dV_2 \dots dV_N. \quad (\text{A5})$$

$\varrho_N(x_1, x_2, \dots, x_N) = \langle \varrho(x_1) \dots \varrho(x_N) \rangle$ is the complete N -point correlation function that is related to the probability of finding particles at the N positions x_n simultaneously. Abbreviating the integration measure by

$$\int_{\mathcal{D}} d\tau_n = \int_{\mathcal{D}} d^3x_1 d^3x_2 \dots d^3x_n, \quad (\text{A6})$$

we, therefore, find for the average

$$\begin{aligned} \langle M \rangle &= \frac{1}{N^N} \int_{\mathcal{D}} d\tau_N \varrho_N(x_1, x_2, \dots, x_N) M\left(\bigcup_{i=1}^N B_i\right) \\ &= \sum_{n=1}^N \binom{N}{n} N^{-n} (-1)^{n+1} \int_{\mathcal{D}} d\tau_n \varrho_n(x_1, x_2, \dots, x_n) \\ &\quad \times M(B_{x_1} \cap B_{x_2} \cap \dots \cap B_{x_n}). \end{aligned} \quad (\text{A7})$$

In the limit of an infinite structure $N \rightarrow \infty$, $\binom{N}{n} N^{-n} \rightarrow \frac{1}{n!}$, and so

$$\begin{aligned} \langle M \rangle &= 1 - \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \int_{\mathcal{D}} d\tau_n \varrho_n(x_1, x_2, \dots, x_n) \\ &\quad \times M(B_{x_1} \cap B_{x_2} \cap \dots \cap B_{x_n}), \end{aligned} \quad (\text{A8})$$

which is a result already obtained in Mecke et al. (1994) for the individual functionals.

To pass from this expression which involves the complete N -point correlations to the formulation in (3) with the connected N -point correlation functions ξ_n we need three ingredients:

(i) The principal kinematical formula

$$\int_{\mathcal{D}} d^3x M(A \cap B_x) \sim_{\text{mod}(t^4)} M(A) M(B) \quad (\text{A9})$$

that connects the integral of the Minkowski polynomial of the intersection $A \cap B_x$ with a product of Minkowski polynomials when dropping all terms involving powers of t larger than t^3 . Here, $A, B \in \mathcal{R}$ and it is important that there is no function of x in the integral.

(ii) The motion invariance of the MFs

$$M(A_x \cap B_{x+y}) = M(A_0 \cap B_y).$$

(iii) The motion invariance of the correlation functions of a statistically homogeneous distribution

$$\xi_n(\mathbf{y}, \mathbf{y} + \mathbf{x}_1, \dots, \mathbf{y} + \mathbf{x}_{n-1}) = \xi_n(\mathbf{0}, \mathbf{x}_1, \dots, \mathbf{x}_{n-1}).$$

With these ingredients we can now show the following formal cumulant relation:

$$1 + \sum_{n=1}^{\infty} m_n t^n / n! = \exp\left(\sum_{n=1}^{\infty} \kappa_n t^n / n!\right), \quad (\text{A10})$$

connecting the ‘moments’

$$m_n = \frac{1}{V_{\mathcal{D}}} \int_{\mathcal{D}} d\tau_n \varrho_n(x_1, \dots, x_n) M(B_{x_1} \cap \dots \cap B_{x_n}) \quad (\text{A11})$$

to the ‘cumulants’

$$\kappa_n = \frac{1}{V_{\mathcal{D}}} \int_{\mathcal{D}} d\tau_n \varrho^n \xi_n(x_1, \dots, x_n) M(B_{x_1} \cap \dots \cap B_{x_n}). \quad (\text{A12})$$

The relation (A10) is satisfied if

$$m_n = \sum_{\pi} \prod_{B \in \pi} \kappa(X_i : i \in B),$$

where π is the set of all partitions of $\{1, \dots, n\}$. We already know that there is such a relation connecting ϱ_n and the ξ_n , because the latter are the joint cumulants of the former, up to a factor involving the background density $\varrho = \varrho_1(x_1) = \langle \varrho(x_1) \rangle$ which is non-zero and constant on the support \mathcal{D} ,

$$\varrho_n(x_1, x_2, \dots, x_n) = \sum_{\pi} \prod_{B \in \pi} \varrho^n \xi(\varrho(x_i) : i \in B).$$

To show that m_n and κ_n as defined above also fulfil this condition, we do a weighted integral on both sides

$$\begin{aligned} &\frac{1}{V_{\mathcal{D}}} \int_{\mathcal{D}} d\tau_n \varrho_n(x_1, \dots, x_n) M(B_{x_1} \cap \dots \cap B_{x_n}) \\ &= \sum_{\pi} \prod_{B \in \pi} \frac{1}{V_{\mathcal{D}}} \int_{\mathcal{D}} d\tau_n \varrho^n \xi(\varrho(x_i) : i \in B) M(B_{x_1} \cap \dots \cap B_{x_n}) \end{aligned} \quad (\text{A13})$$

and find m_n on the left-hand side. What we need to demonstrate is that the products of ξ_i become products of the related κ_i .

This means that we have to break up the term $M(D \cap B_{x_1} \cap B_{x_2} \cap \dots \cap B_{x_n})$ into products of M corresponding to the ξ . We show how this is done for one generic product.

Let $B \in \pi$ be the partition $\{\{1, \dots, i\}, \{i+1, \dots, n\}\}$ and $j = n - i$, then one of the products in equation (A13) consists of two terms which are

$$\begin{aligned} &\frac{1}{V_{\mathcal{D}}} \int_{\mathcal{D}} d\tau_n \xi_i(x_1, \dots, x_i) \xi_j(x_{i+1}, \dots, x_n) \\ &\times M(B_{x_1} \cap B_{x_2} \dots \cap B_{x_i} \cap B_{x_{i+1}} \cap B_{x_{i+2}} \cap \dots \cap B_{x_n}). \end{aligned}$$

We can now make use of our assumption of a statistically homogeneous point distribution and make a change of coordinates from x_1, \dots, x_i to x_1, y_1, \dots, y_{i-1} , where the new coordinates y_k are simply $y_k = x_{k+1} - x_1$

$$\begin{aligned} &\frac{1}{V_{\mathcal{D}}} \int_{\mathcal{D}} d^3x_1 d^3y_1 \dots d^3y_{i-1} \xi_i(y_1, \dots, y_{i-1}) \int_{\mathcal{D}} d\tau_j \xi_j(x_{i+1}, \dots, x_n) \\ &\times M(B_{x_1} \cap B_{x_1+y_1} \dots \cap B_{x_1+y_{i-1}} \cap B_{x_{i+1}} \cap B_{x_{i+2}} \cap \dots \cap B_{x_n}). \end{aligned}$$

Due to statistical homogeneity, ξ_i does not depend on x_1 and we can use the principal kinematical formula (A9) to split the M into

$$\frac{1}{V_{\mathcal{D}}} \int_{\mathcal{D}} d^3 y_1 \dots d^3 y_{j-1} \xi_i(y_1, \dots, y_{j-1}) M(B_0 \cap B_{y_1} \dots \cap B_{y_{j-1}}) \\ \times \int_{\mathcal{D}} d\tau_j \xi_j(x_{i+1}, \dots, x_n) M(B_{x_{i+1}} \cap B_{x_{i+2}} \cap \dots \cap B_{x_n}).$$

Formally, we can reinsert the integral over x_1 , $\frac{1}{V_{\mathcal{D}}} \int d^3 x_1$, and when we change the variables back to the redundant x_1, \dots, x_i we have

$$\frac{1}{V_{\mathcal{D}}} \int_{\mathcal{D}} d\tau_i \xi_i(x_1, \dots, x_i) M(B_{x_1} \cap B_{x_2} \dots \cap B_{x_i}) \\ \times \frac{1}{V_{\mathcal{D}}} \int_{\mathcal{D}} d\tau_j \xi_j(x_{i+1}, \dots, x_n) M(B_{x_{i+1}} \cap B_{x_{i+2}} \cap \dots \cap B_{x_n}) = \kappa_i \kappa_j.$$

Iterating this procedure in case of multiple terms of ξ_i in the product finally establishes the formal cumulant relation (A10).

So, inserting equations (A11) and (A12) into (A10), we find that the average Minkowski polynomial of equation (A8) becomes

$$m := \frac{\langle M \rangle}{V_{\mathcal{D}}} = 1 - \exp \sum_{n=1}^{\infty} \frac{(-Q)^n}{n! V_{\mathcal{D}}} \int_{\mathcal{D}} d\tau_n \xi_n(x_1, x_2, \dots, x_n) \\ \times M(B_{x_1} \cap B_{x_2} \cap \dots \cap B_{x_n}). \quad (\text{A14})$$

It can be reconnected to the single MF densities of the structure by taking the derivative

$$v_{\mu} = \frac{1}{\alpha_{\mu}} \left. \frac{\partial^{\mu} m(t)}{\partial t^{\mu}} \right|_{t=0},$$

which then leads directly to equation (3), when, in addition, we use again motion invariance in the exponent to recover the form of Equation (5).

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.