Mon. Not. R. Astron. Soc. **000**, 1–18 (2014) Printed 18 November 2013 (MN LATEX style file v2.2)

Direct Minkowski Functional analysis of large redshift surveys: a new high–speed code tested on the LRG SDSS–DR7 catalogue

Alexander Wiegand[1], Thomas Buchert[2] and Matthias Ostermann[3,4⋆]
[1]*Max–Planck–Institut für Gravitationsphysik, Albert–Einstein–Institut, Am Mühlenberg 1, D–14476 Potsdam, Germany*
[2]*Université 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*
[3]*Arnold Sommerfeld Center, Ludwig–Maximilians–Universität, Theresienstraße 37, D–80333 München, Germany*
[4]*Oskar–Maria–Graf Gymnasium, Keltenweg 5, D–85375 Neufahrn, Germany*

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 LRG sample of SDSS DR7 data. We find significant (more than $2-$sigma) deviations from the ΛCDM mock catalogues on samples as large as $500h^{-1}$Mpc and $700h^{-1}$Mpc, respectively, and we investigate possible sources of these deviations.

Key words: Cosmology:large-scale structure of Universe, methods: statistical - analytical - data analysis, galaxy catalogues: morphology, higher–order correlations, Minkowski Functionals

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 et al. (2013); Whitbourn & Shanks (2013). However, to verify the reality of these structures, a robust statistical tool is mandatory Nadathur (2013); Nadathur & Hotchkiss (2013).

The most common tool for the characterisation 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 ΛCDM model are often based on these lower–order statistics. However, of course, they do not allow for a complete characterisation 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 characterise the distribution uniquely Carron & Neyrinck (2012). As is well–known, a smoothed–out nonlinear 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 the present paper provide a compact and transparent framework to completely characterise 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. We shall use this property in Section 5, to extract certain integrals over the three–point correlation function of the galaxy distribution.

Due to 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

⋆ E–mail:alexander.wiegand@aei.mpg.de, buchert@obs.univ-lyon.fr and mail@matthias-ostermann.de

with the family of Minkowski functionals in Mecke et al. (1994); Buchert (1995) (see Schmalzing et al. (1996) for a brief tutorial), and which will be briefly reviewed in Section 2.2, has been used for example for the IRAS 1.2Jy and PSCz surveys Kerscher et al. (1996, 1998, 2001a), and the Abell/ACO cluster catalogue Kerscher et al. (1997). For an early sample of the SDSS catalogue (data release DR 3), the Minkowski functionals have been determined for smoothed iso–density 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); Sathyaprakash et al. (1998); Sahni et al. (1998); Schmalzing et al. (1999); Nakagami et al. (2004); Blake et al. (2013); Choi et al. (2013). A further, recently proposed smoothing technique directly uses the Delauney tesselation of the point distribution Aragon-Calvo et al. (2010).

To robustly measure non–Gaussianity with the help of Minkowski Functionals is mostly discussed for iso–temperature contour maps of the Cosmic Microwave Background (CMB) Schmalzing & Gorski (1998); Schmalzing et al. (2000); Hikage & Matsubara (2012); Modest et al. (2013); Ducout et al. (2013); Munshi et al. (2013); Planck Collaboration et al. (2013), as well as for studies of the weak lensing convergence field Kratochvil et al. (2012); Petri et al. (2013).

One focus of the present paper is the determination of non-Gaussian features from Minkowski Functionals in three–dimensional galaxy data, which has been addressed in Pratten & Munshi (2012); 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 data release 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 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 optimised 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. 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.

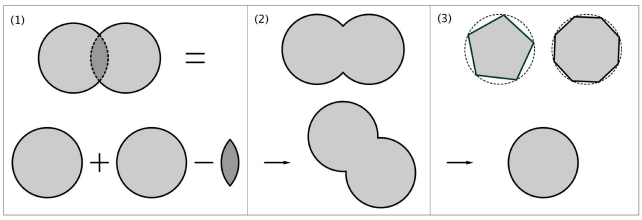

Figure 1. Properties of the scalar functionals: (1) Additivity, (2) Motion invariance, (3) Conditional continuity.

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 $V_1 - V_4$. In the normalization we use, they are related to geometrical properties of the body as follows:

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

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 Figure 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

$$v_\mu = V_\mu / |D| \;\; , \tag{2}$$

where $|D|$ 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 characterising 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., iso–density contours) and the *Germ–Grain model*.

In order to determine iso–density 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 we are going to analyse. The values of the four Minkowski Functionals (volume, surface, mean curvature, Euler characteristic) can then be determined as a function of the (over–)density that is used to determine the iso–density 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).

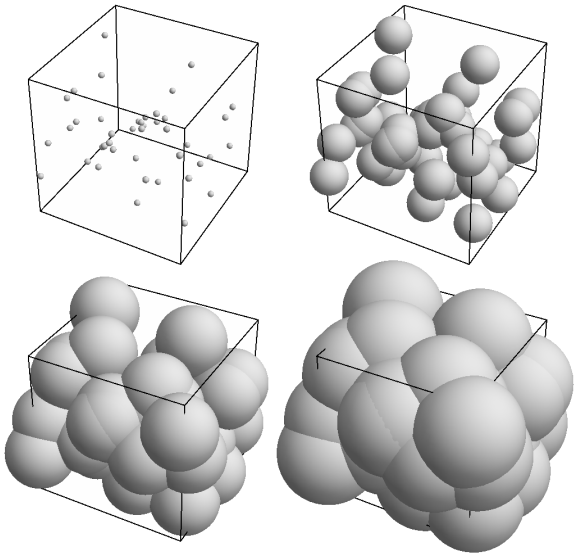

Figure 2. By increasing the radius of Balls around the points $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ up to the maximum radius R, more and more Balls intersect and a complex structure develops.

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 Figure 2. Also this method has been used quite extensively in the past to characterise galaxy and cluster distributions, e.g. in Kerscher et al. (1996); Platzöder & Buchert (1996); Kerscher et al. (1997, 1998, 2001a), 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 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 et al. (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 characterisation 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 realisation 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:

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

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

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

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

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

The integrals run over the positions of the centres $\mathbf{x}_i$ of the Balls $B_{\mathbf{x}_i}$. As it is the intersection of all Balls $B \cap B_{\mathbf{x}_1} \cap \ldots \cap B_{\mathbf{x}_n}$ 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 $\langle v_\mu \rangle$'s in terms of the statistical quantities describing the point process, one could

work with the dimensionfull connected correlation functions, without assuming that $\varrho_0 > 0$. We shall only need it 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 realisation of the full point process. We assume that it has a positive 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 $\overline{V}_\mu$, 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, $\xi_2 \geqslant 0$, and the average density ϱ_0 has to satisfy

$$\varrho_0 \int_A d\mathbf{y} \xi_2 \left(|\mathbf{y}|\right) \leqslant 1 \; . \tag{6}$$

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 $\overline{V}_\mu$'s. Equation (5) becomes

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

We already gave the explicit expressions for the Minkowski Functionals of the Balls $V_\mu \left(B\right)$ 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 $V_\mu \left(B \cap B_{\mathbf{x}_1}\right)$. 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 $V_\mu \left(B \cap B_{\mathbf{x}_1}\right)$ in a spherical coordinate system centred on B is then the following:

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

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

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

$$V_3 \left(r\right) = 1 \; , \tag{11}$$

where R is again the radius of the Balls. As they do not intersect if the separation of the two centres is larger than $2R$, $V_\mu \left(B \cap B_{\mathbf{x}_1}\right) = 0$ for $r > 2R$. Therefore, the integral (7) reduces to

$$\overline{V}_\mu = V_\mu \left(B\right) - 2\pi\varrho_0 \int_0^{2R} V_\mu \left(r\right) \xi_2 \left(r\right) r^2 \mathrm{d}r \; . \tag{12}$$

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

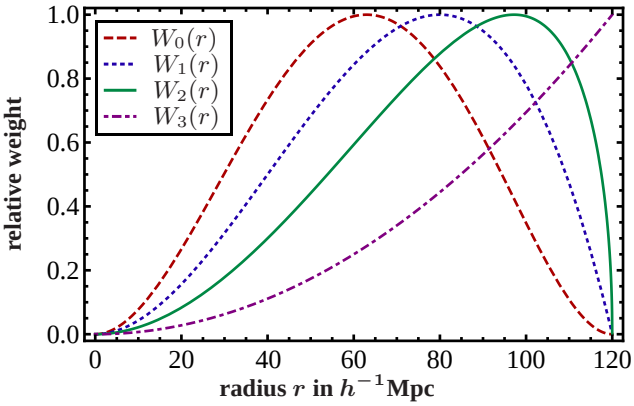

Figure 3. Form of the integration windows $W_\mu \left(r\right) = V_\mu \left(r\right) r^2$ in Equations (8)-(11) used in the integral (12), for a Ball radius of $R = 60 h^{-1}$Mpc. The functionals of higher index μ probe the correlation function at larger distances.

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

$$\overline{V}_\mu = V_\mu \left(B\right) - \frac{\varrho_0}{\pi} \int_0^\infty P \left(k\right) W_\mu \left(k, R\right) k^2 dk \; , \tag{13}$$

with the functions

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

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

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

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

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

$$\sigma^2 \left(R\right) = \frac{1}{\left(2\pi\right)^3} \int d^3 k P \left(\mathbf{k}\right) \left|\widetilde{W}_{\mathcal{B}(R)} \left(k\right)\right|^2 \; . \tag{18}$$

This means that $\overline{V}_0$ and $\overline{V}_3$ are directly related to the two–point quantities by

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

and

$$\overline{V}_3 = 1 - 2\pi\varrho_0 \int_0^{2R} \xi_2 \left(r\right) r^2 \mathrm{d}r \; . \tag{20}$$

The two other modified functionals $\overline{V}_1$, $\overline{V}_2$ then probe different aspects of the precise form of the correlation function by weighting the integral over it with a different function of r as shown in Figure 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 $\langle v_\mu \rangle$'s 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 to extract the integrals over the correlation functions in (5) as follows.

Let us assume that we are able to measure the $\langle v_\mu \rangle$'s 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 $\overline{V}_\mu$ on ϱ_0. Calling this empirical function $\tilde{\overline{V}}_\mu (\varrho_0)$, 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 $\tilde{\overline{V}}_\mu (\varrho_0)$ around $\varrho_0 = 0$ we obtain:

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

where the exponent (n) stands for the n–th derivative of $\tilde{\overline{V}}_\mu (\varrho_0)$ 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

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

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

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

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 $\tilde{\overline{V}}_\mu (\varrho_0)$ at zero density, and it especially allows 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 realisation 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's in an ensemble of mock samples produced from simulations. This allows 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:

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

The other integrals of the three–point function are more complicated and we won't write them out explicitly.

To characterise 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 $n \leqslant 2$.

3 THE NEW CODE PACKAGE: MINKOWSKI–4

As described in the previous section, we can learn a lot about the structure of 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 CHIPMINK (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 generalises the Boolean Grain model – where the Germs are those of a Poisson process (Stoyan et al. 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 $\overline{V}_\mu$. 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 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 $\mathcal{B}_r$,

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

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

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

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

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 Minkowski Functionals

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 application Schmalzing & Diaferio (2000),

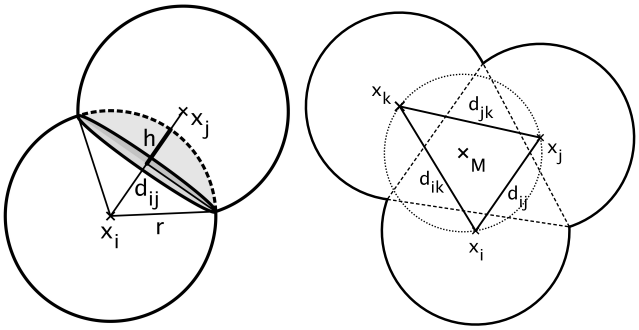

Figure 4. The left figure shows the covered surface area of a sphere around x_i when intersected with a second sphere as well as the intersection circle. The right figure illustrates the intersection of three spheres. The triple point can be found above the centre point of the circumcircle of the triangle generated by the three centre points $\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k$.

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

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 Figure 4. The MFs' volume densities are defined by

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

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 by 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):

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

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

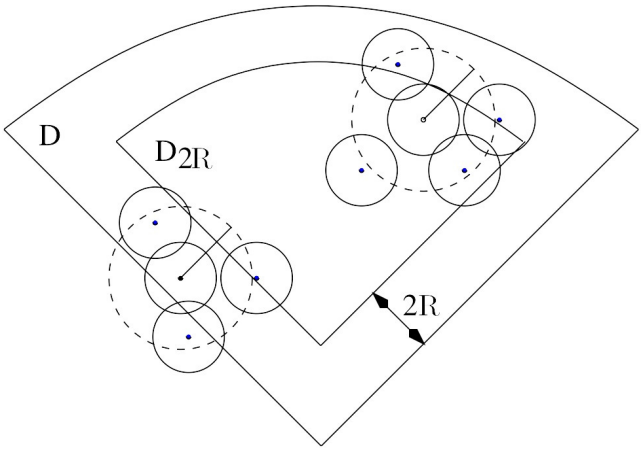

Figure 5. To avoid boundary effects when calculating the Partial MFs, we only take those points into account, which are more than two times the maximum Germ radius away from the survey mask.

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:

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

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 et al. (1996), 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 Figure 5. Thus, we create a shrunk "calculation window" D_{2R} and don't 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

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

where

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

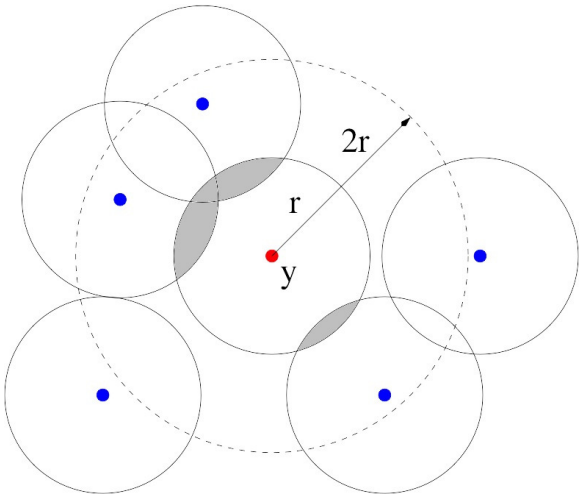

Figure 6. For any randomly thrown point $\mathbf{y}_i$, we generate the list of *real* neighbours, i.e. consisting of galaxies of the used sample. The Balls around all points within a distance of $2r$ around $\mathbf{y}_i$ intersect with the Ball $\mathcal{B}(\mathbf{y}_i; r)$. The volume fraction of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

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 localised 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; second, we determine neighbourlists for the random points. These neighbourlists consist of the thrown points themselves as well as of the *real* galaxies in their vicinity, i.e. points of the given sample, within a distance of twice the maximum radius, see Figure 6; third, 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 $v_0(\mathbf{y}_i; r)$ and normalising them by the number of random points, say M, for which we calculated the volume fraction,

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

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.

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 luminous red galaxy (LRG, Eisenstein et al. (2001)) sample of the Sloan Digital Sky Survey (SDSS, York et al. (2000)) Data release 7 (DR7, Abazajian et al. (2009)), and, second, to the mock catalogues drawn from ΛCDM simulations of the SDSS volume performed by the LasDamas[2] collaboration McBride et al. (2014).

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 a SDSS spectrum, is not in an area around bright stars, has a sector completeness of at least 60%, 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 105831 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 ra $\in [132^\circ, 235^\circ]$ and 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 off 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 off at -21.8 and a redshift in the range $z \in [0.16, 0.44]$. With these requirements, the "dim sample" contains 41375 galaxies and the "bright sample" 22386.

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 LasDamas[2] collaboration McBride et al. (2014). These authors simulated structure formation in a ΛCDM model with $\Omega_\Lambda = 0.75$ in boxes of $2.4h^{-1}$Gpc for 1280^3 particles. They identify halos 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

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

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 stars[3] and performing the same angular cut as for the SDSS data samples. This results in catalogues that contain on average 46710 galaxies for the "dim sample" and 22181 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. 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 ΛCDM model with $\Omega_\Lambda = 0.75$. An example of the galaxy distribution in the "dim sample" is shown in Figure 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 $507h^{-1}$Mpc. The thickness of z–shell of the "bright sample" $730h^{-1}$Mpc. The largest cube that fits into our "dim sample" region has a side length of $452h^{-1}$Mpc; Figure 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 Minkowski Functionals throughout the sample.

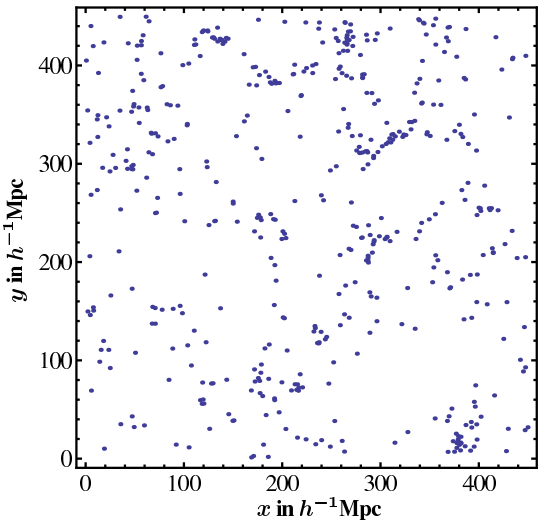

Figure 7. Example of the galaxy distribution in the "dim sample". The projection is extracted from a slice of a thickness of $22.5h^{-1}$Mpc from the maximally fitting cube.

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 Minkowski Functional 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 calculate the error bars from random subsampling *jack–knife* realisations drawn from the data and consisting of 80% 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 percent. So, for a first estimate of the errors already the output of the code is quite useful.

There are two possible reasons for the Minkowski Functional densities (3) to fluctuate between different realisations. First, the $N-$point correlation functions of the point distribution in different realisations 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's. 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's. This means that we have to ensure that all the realisations 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 realisations have the same density ϱ_0, we implement a random choice of $\approx 80\%$ 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%, but is adjusted to give the same average density for the mocks and the data.

4.2.1 *The "dim sample"*

Figure 8 shows the MF densities obtained from this procedure and the code described in Section 3. We plot the average from 244 random 80%–realisations of the "dim sample" as the (red) line with the dots. The 80% of selected points correspond to an average density of $\varrho_0 = 7.7\times10^{-5}h^3\text{Mpc}^{-3}$. 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. 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 Figure 8 indicate that the determination of the MF densities is quite robust. The one–sigma 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 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 Figure 8 show a zoom into these curves. It is obtained by subtracting the average of the mocks. In this zoom we find that the mock catalogues and the observed galaxy sample are marginally consistent, but only by allowing for $2-$sigma fluctuations.

4.2.2 *The "bright sample"*

We applied the same procedure to the "bright sample". In this case, randomly choosing 80% of the points corresponds to a density of

[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).

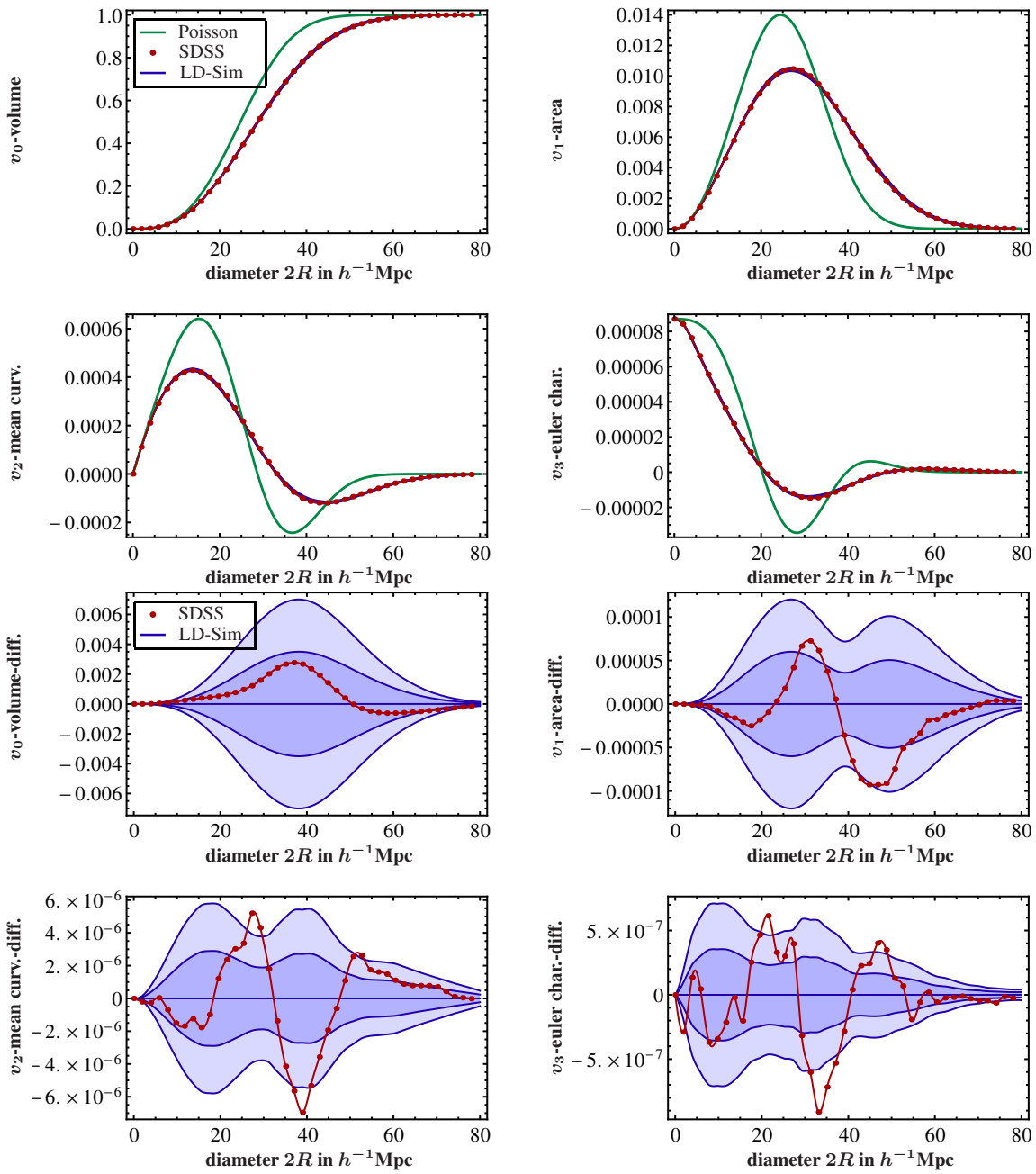

Figure 8. Top four: the four Germ–Grain MF densities for the SDSS LRG "dim sample" compared to those of the corresponding LasDamas mock galaxies and those of a Poisson distribution. The errors and average for the mock samples are obtained taking 160 different mock realisations.
Bottom four: the same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

$\varrho_0 = 2.1 \times 10^{-5} h^3 \mathrm{Mpc}^{-3}$. The resulting differences of the MF densities of the mocks to those of the SDSS "bright sample" are shown in Figure 9. The reduced galaxy density allows 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. From these plots we see a discrepancy between the mock galaxies and the observed ones even by allowing for 2−sigma fluctuations.

To show how this discrepancy is transferred to the modified MFs $\overline{V}_\mu$, we show them in Figure 10. The upper four plots show them together with the Poisson case. This latter is very simple for $\overline{V}_\mu$ as it consists of the functionals of a Ball, Equation (4). The values and error bars have been obtained by calculating the $\overline{V}_\mu$'s for every realisation and taking the average and variance of these values. As the errors grow rapidly beyond $60h^{-1}$Mpc (in diameter), we only plot the $\overline{V}_\mu$'s up to this scale. The reason for this growth is that around $60h^{-1}$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 $e^{-\varrho_0 \overline{V}_\mu}$. But also on these scales, where the

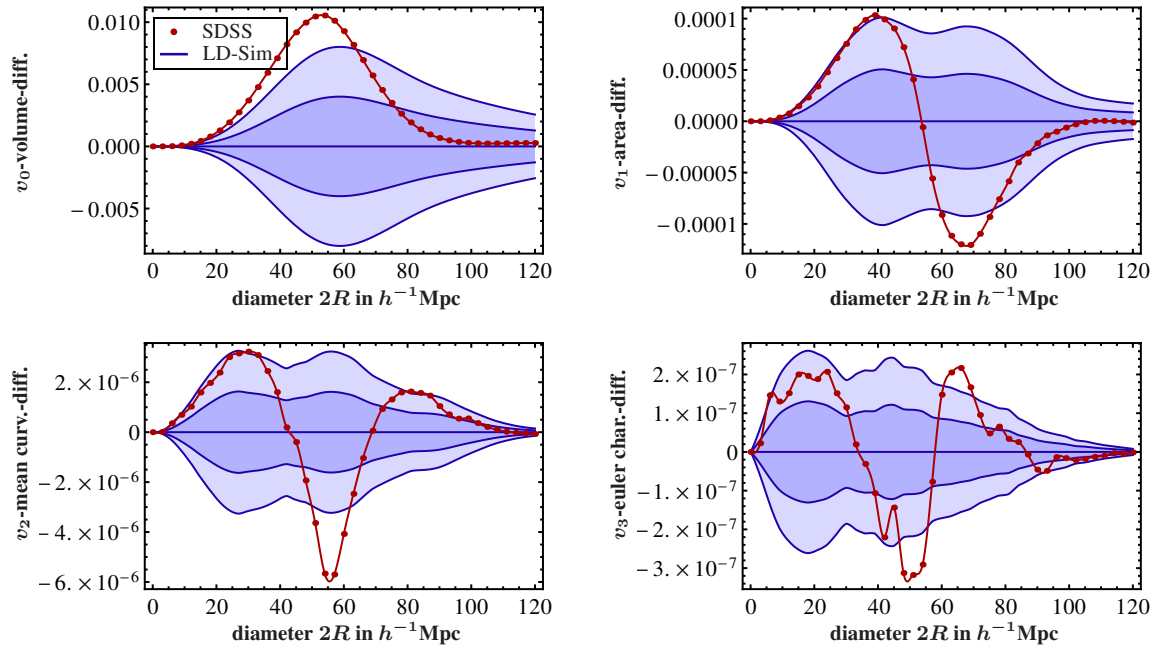

Figure 9. The same quantities as in the lower four plots of Figure 8, but for the bright "sample". The dark shaded regions are the 1–σ, the light shaded regions 2–σ error bands.

errors are still controllable, and for this way to present the results, there is a clear discrepancy between mocks and SDSS data.

5 NON–GAUSSIAN CORRELATIONS

From Figures 9 and 10 it is not clear, whether the origin of the discrepancy between mocks and data is already present for 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 answer this question, 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 $\tilde{\overline{V}}_\mu(\varrho_0)$.

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 $\varrho_0 = 2.6 \times 10^{-5} h^3 \mathrm{Mpc}^{-3}$. For each of these densities we generate a large number of realisations (from about 15000 for $f = 0.05$ down to 244 for $f = 0.8$). For each of these realisations we determine the modified MFs $\overline{V}_\mu$ as a function of the Ball radius R. For each R we take the average over all the realisations and arrive at an $\tilde{\overline{V}}_\mu(\varrho_0)$ evaluated at 24 points. We add the value of $\overline{V}_\mu$ 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 $\tilde{\overline{V}}_\mu(\varrho_0)$.

As the resampling of the points introduces an important amount of correlation between the realisations 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 empiric function $\tilde{\overline{V}}_\mu(r, \varrho_0)$. We use $\tilde{\overline{V}}_\mu(\varrho_0)$ at 7 different distances r and use its values at 20 of the 25 ϱ_0–points. This gives us a covariance matrix of dimension 140×140. We use this combined covariance matrix to do a polynomial fit of fourth order to the average $\tilde{\overline{V}}_\mu(r, \varrho_0)$ for all the mocks and for the SDSS. With Equation (21) this gives us the coefficients b_2–b_4 for both the mocks and the SDSS.

We finally divide these integrals by the corresponding power of the volume,

$$c^\mu_{n+1} = \frac{b^\mu_{n+1}}{V_0^{n+1}(B)} = \frac{1}{V_0^{n+1}(B)} \int_{\mathcal{D}} \mathrm{d}^3x_1 \ldots \mathrm{d}^3x_n \times \\ \times \xi_{n+1}(0, \mathbf{x}_1, \ldots \mathbf{x}_n)\, V_\mu(B \cap B_{\mathbf{x}_1} \cap \ldots \cap B_{\mathbf{x}_n}), \quad (32)$$

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^μ_2, i.e. those involving an integral over the two–point correlation function, are shown in the upper four plots of Figure 11. As described in Section 2.4, the first quantity plotted, c^0_2, is related to the matter variance in spheres of radius R. In fact, from Equation (32) and (19), one can see that $c^0_2 = \sigma^2$. So, the first plot of Figure 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 $\sigma^2_{emp.}$ are calculated from the observed sample with the usual estimator for the matter variance in spheres of radius R,

$$\sigma^2(R) = \frac{\mathbb{E}\left[M(R)^2\right] - \mathbb{E}\left[M(R)\right]^2}{\mathbb{E}\left[M(R)\right]^2}, \quad (33)$$

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^0_2, directly becomes Equation (18). For the form of $P(k)$, we use the parametrisation of Eisenstein & Hu (1998) that includes the effects of baryons. However, we use the form without their oscillations. We normalise the amplitude of the power spectrum to match the

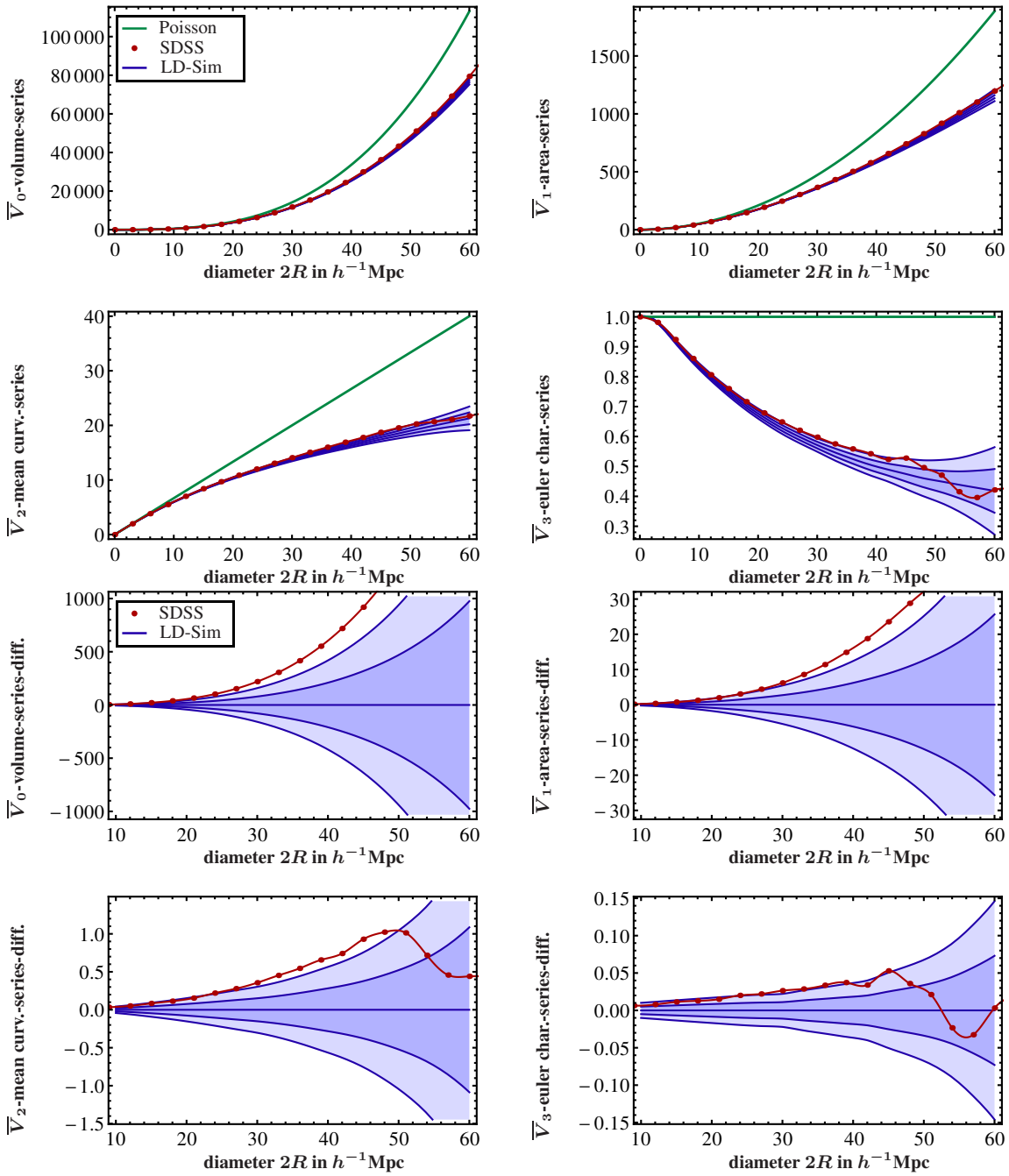

Figure 10. Top four: the four modified Germ–Grain MFs $\overline{V}_\mu$ of the SDSS LRG "bright sample" and the LasDamas mock galaxies. The errors and average for the mock samples are obtained taking 160 different mock realisations. The Poisson case is simply given by the MFs of a single ball (see Equation (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 1–σ error bands, the light regions correspond to 2–σ.

amplitude of the mocks. This gives a linear bias of $b \approx 2.37$, i.e. $\sigma^2_{mock} = b^2\sigma^2_{lin.}$.

In the V_1–V_3 plots of Figure 11, we use the power spectrum with the same normalisation and Equation (13) for the theory prediction. The first plot of Figure 11 shows that all three ways of calculating σ^2 are overall in agreement. This is also true for the plots derived from $\overline{V}_1$–$\overline{V}_3$, but there is also a clear deviation of the coefficients c^μ_2 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 calculation. This shows that the simulations and the extraction procedure appear to give a consistent picture. The observed galaxies, however, seem to deviate from the simulated cosmology. This means that, even though the overall normalisation 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 Figure 11 show the same quantities, but we divided them by the theoretical prediction for the simulated

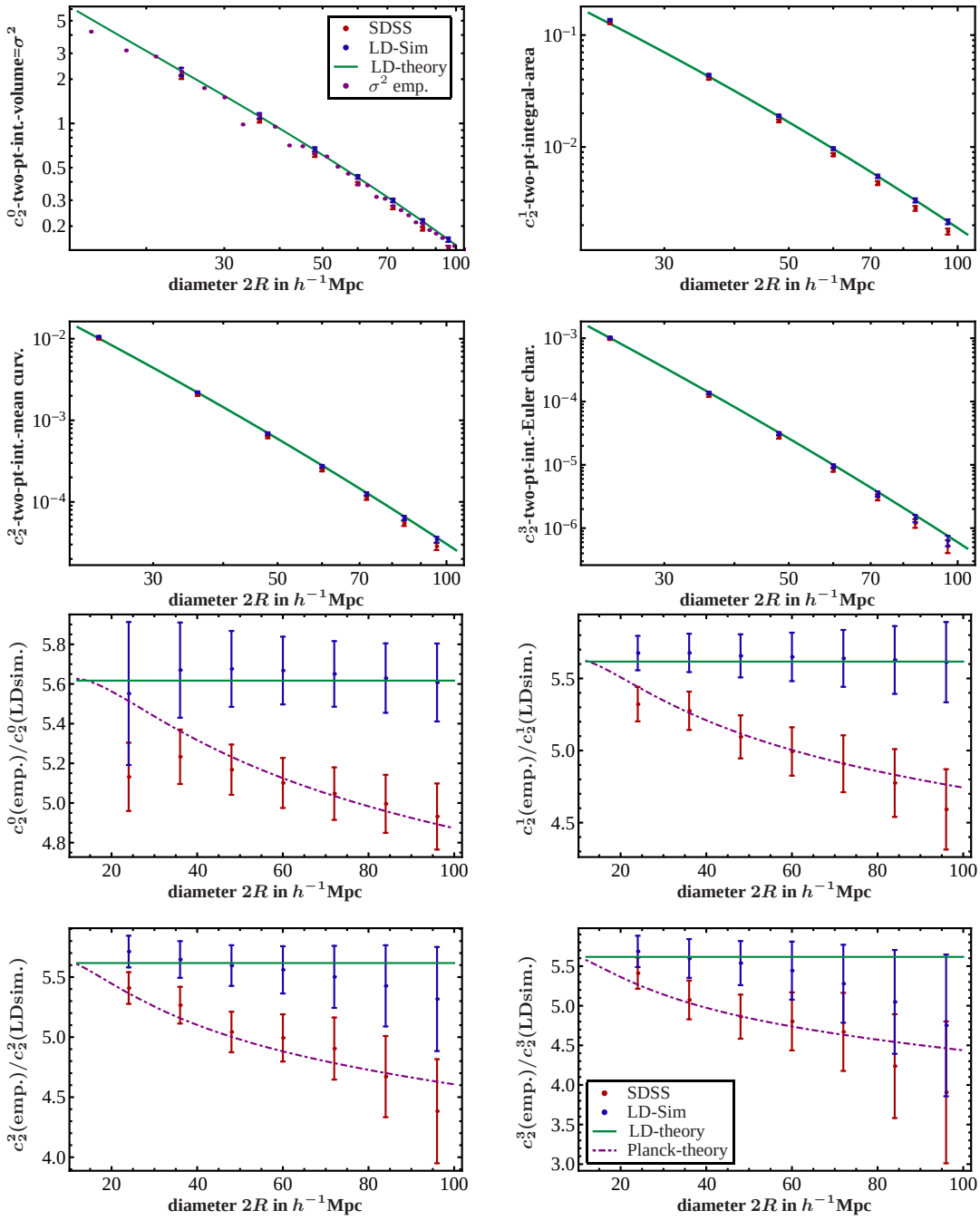

Figure 11. Top four: the four Minkowski integrals of the two–point correlation function, i.e. c_2^μ from Equation (32), for the SDSS LRG "bright sample" and one of the corresponding mock samples. The error bars are 1–σ errors for the parameters from the fit $\tilde{\overline{V}}_\mu(\varrho_0)$ with a polynomial of fourth order in ϱ_0. The lines for V_1–V_3 are empirical fits to the data points. 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 determination of σ^2 for the observed galaxies.
Bottom four: the same quantities but all divided by the theoretical prediction of the mock power spectrum. This allows a more detailed comparison. 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 $\Omega_m \approx 0.32$.

cosmology. This divides out 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's calculated from a *Planck* $\Omega_m \approx 0.32$ power spectrum for comparison. Figure 11 shows that the deviations of the MFs for the data and the mock galaxies, as found in Figure 9 and 10, already occur at the level of the first correction to the leading Poisson term in the expansion $\overline{V}_\mu$, Equation (5). We shall return 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 Figure 12. In comparison with Figure 11 we recognise that the amplitude of the c_3^μ integrals is larger by a factor of 5 than the integrals c_2^μ. Like in Figure 11, we also have a slight but less significant deviation for these quantities.

The line in the upper left plot is calculated from a tree–level bispectrum using Equation (22) and the bias of $b = 2.37$ as found from the two point normalisation.

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

In order to get an impression of how accurately 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 $\widetilde{\overline{V}}_\mu(\varrho_0)$ in the two previous sections, to calculate the $\overline{V}_\mu$'s of Equation (5) up to $n = 3$. Using Equations (3), this allows to derive the densities $v_\mu(R)$ within this approximation.

For comparison, we can then plot the approximative functionals in Figure 13. However, as the series $\overline{V}_\mu$ of Equation (5) is a power series in the density, the approximation of the function by a truncated version of this series will be better when ϱ_0 is lower. In addition we would like to compare the series expansion to the case of a pure Gauss–Poisson point distribution. It has been shown in Kerscher (2001) that for a Gauss–Poisson process to exist, the density must be low enough for a given amount of two–point correlation. To be more precise:

$$\varrho_0 \int_A d\mathbf{y}\xi_2(|\mathbf{y}|) \leqslant 1 \ .$$

For the strongly clustered SDSS LRGs this requirement is not fulfilled for the full density of the sample. Also for 80% of the density like in Figures 9 and 10, where the density was $\varrho_0 = 2.1 \times 10^{-5} h^3\mathrm{Mpc}^{-3}$, the two point amplitude is too high to allow 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% of the full density, i.e. $\varrho_0 = 0.78 \times 10^{-5} h^3\mathrm{Mpc}^{-3}$, the condition is (marginally) satisfied. Therefore, we calculated the modified MFs for a large number of realisations of 30% of the "bright sample" mocks in the same way as in Section 4.2.

Figure 13 shows the approximating series directly for the modified MFs $\overline{V}_\mu$ of the Mocks divided by their average. We confirm that the mock point catalogues are not a realisation of a Gauss–Poisson process and that higher–order corrections are crucial for the MFs.

Instead of being Gaussian it has been shown that the logarithm of the density field can have a Gaussian distribution. This implies that a Log–normal distribution can describe the two–point properties of galaxy samples quite well. We compare the distribution found in the mock catalogues with that of a Log–normal distribution in Appendix B

The effect of including these higher correction terms up to ϱ_0^3 in (5) is shown by the two other lines. This truncated model is quite good in describing the modified MF densities of the data up to a scale of around $60h^{-1}$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 $\varrho_0 = 0.78 \times 10^{-5} h^3\mathrm{Mpc}^{-3}$, 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.

6 CONCLUSION

In this paper we provided new analysis tools based on the Germ–Grain model to calculate the family of Minkowski Functionals of point sets. 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 Minkowski Functionals 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 Minkowski Functionals 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 Minkowski Functionals 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 Minkowski Functionals of the observed galaxies to a grid of ΛCDM N–body simulations of the galaxy distribution, for two different luminosity thresholds. In both cases, the Minkowski Functionals of the observed and the simulated galaxy distributions show a significant disagreement. For the galaxies with the higher luminosity, this difference is more 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. For the two–point function we were able to compare 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 that, therefore, higher–order correlations are important to describe the observed structure. As previously often emphasized, it is crucial 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, whether larger and deeper surveys will confirm these results.

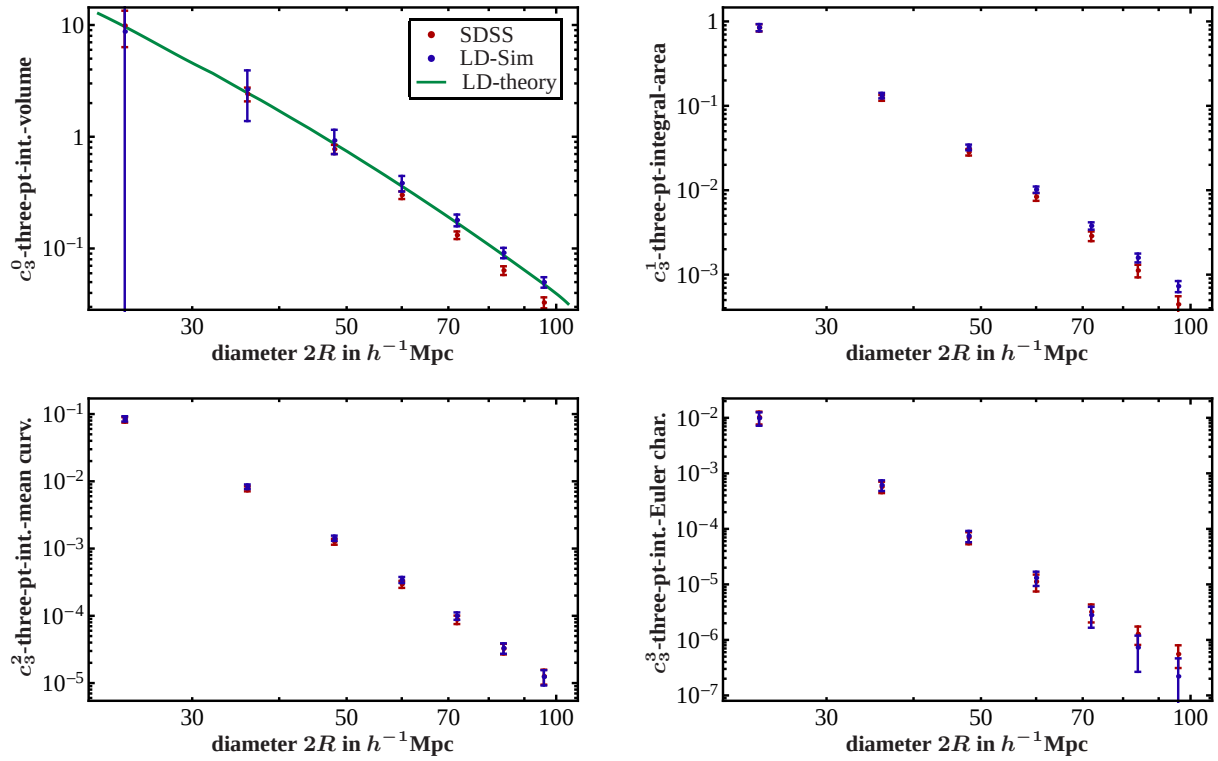

Figure 12. The four Minkowski integrals of the three–point correlation function, i.e. c_3^μ from Equation (32), for the SDSS LRG "bright sample" and the average of the corresponding mock samples. The error bars are the 1–σ errors from the fit of $\tilde{\overline{V}}_\mu(\varrho_0)$ with a polynomial of fifth order in ϱ_0. The line in the first figure is the integral of Equation (22) evaluated for the LasDamas cosmology.

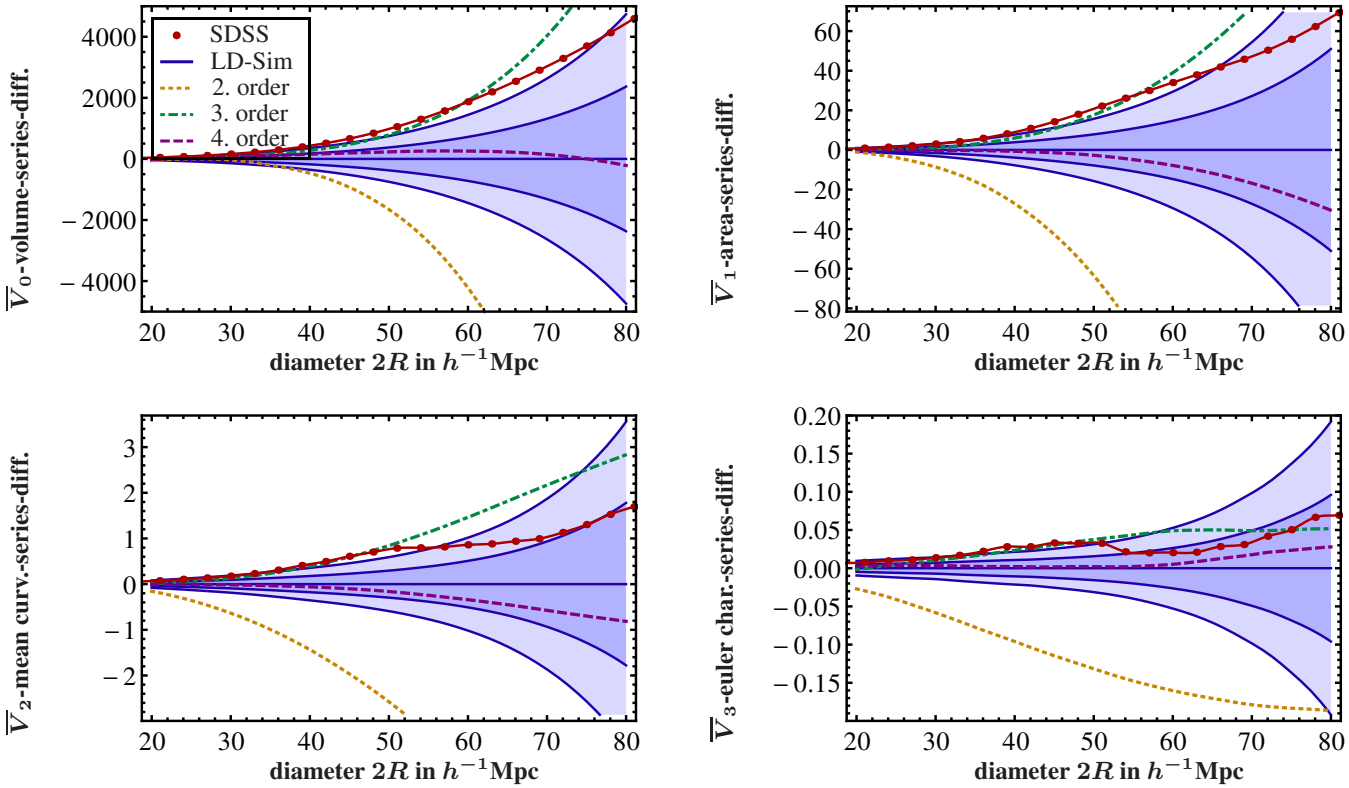

Figure 13. The four modified MFs $\overline{V}_\mu$ of 30% of the "bright sample", after subtraction of the average of the mocks. The (yellow) dotted line shows the theoretical expectation for a pure Gauss–Poisson process of this density and this amount of two–point correlation. The (purple) dashed line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

ACKNOWLEDGMENTS

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 Minkowski Functional 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 Sloan Digital Sky Survey (SDSS) has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the U.S. Department of Energy, the Japanese Monbukagakusho, and the Max Planck Society. The SDSS Web site 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., Adelman-McCarthy J. K., Agüeros M. A., Allam S. S., Allende Prieto C., An D., Anderson K. S. J., Anderson S. F., Annis J., Bahcall N. A., et al. 2009, ApJS, 182, 543, 0812.0649, ADS

Aragon-Calvo M. A., Shandarin S. F., Szalay A., 2010, ArXiv e-prints, 1006.4178, ADS

Beisbart C., 2001, PhD thesis, Ludwig-Maximilians-Universität München

Blake C., James J. B., Poole G. B., 2013, ArXiv e-prints, 1310.6810, ADS

Blanton M. R., Schlegel D. J., Strauss M. A., Brinkmann J., Finkbeiner D., Fukugita M., Gunn J. E., Hogg D. W., Ivezić Ž., Knapp G. R., Lupton R. H., Munn J. A., Schneider D. P., Tegmark M., Zehavi I., 2005, AJ, 129, 2562, arXiv:astro-ph/0410166, ADS

Buchert T., 1995, in Mücket J. P., Gottlöber S., Müller V., eds, Large Scale Structure in the Universe Robust Morphological Measures for Large Scale Structure. p. 156, arXiv:astro-ph/9412061, ADS

Carron J., Neyrinck M. C., 2012, ApJ, 750, 28, 1201.1444, ADS

Choi Y.-Y., Kim J., Rossi G., Kim S. S., Lee J.-E., 2013, ApJS, 209, 19, 1309.4381, ADS

Clowes R. G., Harris K. A., Raghunathan S., Campusano L. E., Söchting I. K., Graham M. J., 2013, MNRAS, 429, 2910, 1211.6256, ADS

Codis S., Pichon C., Pogosyan D., Bernardeau F., Matsubara T., 2013, MNRAS, 435, 531, 1305.7402, ADS

Ducout A., Bouchet F. R., Colombi S., Pogosyan D., Prunet S., 2013, MNRAS, 429, 2104, 1209.1223, ADS

Einasto M., Liivamägi L. J., Tempel E., Saar E., Tago E., Einasto P., Enkvist I., Einasto J., Martínez V. J., Heinämäki P., Nurmi P., 2011, ApJ, 736, 51, 1105.1632, ADS

Eisenstein D. J., Annis J., Gunn J. E., Szalay A. S., Connolly A. J., Nichol R. C., 2001, AJ, 122, 2267, astro-ph/0108153, ADS

Eisenstein D. J., Hu W., 1998, ApJ, 496, 605, arXiv:astro-ph/9709112, ADS

Hadwiger H., 1957, Vorlesungen über Inhalt, Oberfläche und Isoperimetrie. Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen mit besonderer Berücksichtigung der Anwendungsgebiete ; 93, Springer, ISBN: 3-540-02151-5,

Hikage C., Matsubara T., 2012, MNRAS, 425, 2187, 1207.1183, ADS

Hikage C., Schmalzing J., Buchert T., Suto Y., Kayo I., Taruya A., Vogeley M. S., Hoyle F., Gott III J. R., Brinkmann J., 2003, PasJ, 55, 911, arXiv:astro-ph/0304455, ADS

Kazin E. A., Blanton M. R., Scoccimarro R., McBride C. K., Berlind A. A., Bahcall N. A., Brinkmann J., Czarapata P., Frieman J. A., Kent S. M., Schneider D. P., Szalay A. S., 2010, ApJ, 710, 1444, 0908.2598, ADS

Keenan R. C., Barger A. J., Cowie L. L., 2013, ApJ, 775, 62, 1304.2884, ADS

Kerscher M., 1999, A&A, 343, 333, arXiv:astro-ph/9811300, ADS

Kerscher M., 2000, in Mecke K. R., Stoyan D., eds, Statistical Physics and Spatial Statistics. The Art of Analyzing and Modeling Spatial Structures and Pattern Formation Vol. 554 of Lecture Notes in Physics, Berlin Springer Verlag, Statistical Analysis of Large-Scale Structure in the Universe. p. 36, ADS

Kerscher M., 2001, PRE, 64, 056109, arXiv:astro-ph/0102153, ADS

Kerscher M., Mecke K., Schmalzing J., Beisbart C., Buchert T., Wagner H., 2001, A&A, 373, 1, arXiv:astro-ph/0101238, ADS

Kerscher M., Mecke K., Schücker P., Böhringer H., Guzzo L., Collins C. A., Schindler S., De Grandi S., Cruddace R., 2001, A&A, 377, 1, arXiv:astro-ph/0105150, ADS

Kerscher M., Schmalzing J., Buchert T., 1996, in Coles P., Martinez V., Pons-Borderia M.-J., eds, Mapping, Measuring, and Modelling the Universe Vol. 94 of Astronomical Society of the Pacific Conference Series, Analyzing Galaxy Catalogues with Minkowski Functionals. p. 247, arXiv:astro-ph/9512105, ADS

Kerscher M., Schmalzing J., Buchert T., Wagner H., 1996, 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.. pp 83–98, ADS

Kerscher M., Schmalzing J., Buchert T., Wagner H., 1998, A&A, 333, 1, arXiv:astro-ph/9704028, ADS

Kerscher M., Schmalzing J., Retzlaff J., Borgani S., Buchert T., Gottlöber S., Müller V., Plionis M., Wagner H., 1997, MNRAS, 284, 73, arXiv:astro-ph/9606133, ADS

Kratochvil J. M., Lim E. A., Wang S., Haiman Z., May M., Huffenberger K., 2012, PRD, 85, 103513, 1109.6334, ADS

Mantz H., Jacobs K., Mecke K., 2008, Journal of Statistical Mechanics: Theory and Experiment, 12, 15, ADS

McBride, C. et al., 2014, SDSS mock galaxies

Mecke K., 1994, Integralgeometrie in der Statistischen Physik: Perkolation, komplexe Flüssigkeiten und die Struktur des Universums. Reihe Physik ; 25, Deutsch, ISBN: 3-8171-1380-3,
Mecke K., Stoyan D., 2000, Statistical Physics and Spatial Statistics: The Art of Analyzing and Modeling Spatial Structures and Pattern Formation. Lecture Notes in Physics, Springer, ISBN: 9783540677505,
Mecke K., Wagner H., 1991, Journal of Statistical Physics, 64, 843
Mecke K. R., Buchert T., Wagner H., 1994, A&A, 288, 697, arXiv:astro-ph/9312028, ADS
Modest H. I., Räth C., Banday A. J., Rossmanith G., Sütterlin R., Basak S., Delabrouille J., Górski K. M., Morfill G. E., 2013, MNRAS, 428, 551, 1209.5106, ADS
Munshi D., Smidt J., Cooray A., Renzi A., Heavens A., Coles P., 2013, MNRAS, 434, 2830, 1011.5224, ADS
Nadathur S., 2013, MNRAS, 434, 398, 1306.1700, ADS
Nadathur S., Hotchkiss S., 2013, ArXiv e-prints, 1310.2791, ADS
Nakagami T., Matsubara T., Schmalzing J., Jing Y., 2004, ArXiv Astrophysics e-prints, arXiv:astro-ph/0408428, ADS
Padmanabhan N., Schlegel D. J., Finkbeiner D. P., Barentine J. C., Blanton M. R., Brewington H. J., Gunn J. E., Harvanek M., Hogg D. W., Ivezić Ž., Johnston D., Kent S. M., Kleinman S. J., Knapp G. R., Krzesinski J., Long D., Neilsen Jr. E. H., Nitta A., Loomis C., Lupton R. H., Roweis S., Snedden S. A., Strauss M. A., Tucker D. L., 2008, ApJ, 674, 1217, arXiv:astro-ph/0703454, ADS
Park C., Choi Y.-Y., Kim J., Gott III J. R., Kim S. S., Kim K.-S., 2012, ApJ, 759, L7, 1209.5659, ADS
Petri A., Haiman Z., Hui L., May M., Kratochvil J. M., 2013, ArXiv e-prints, 1309.4460, ADS
Planck Collaboration Ade P. A. R., Aghanim N., Armitage-Caplan C., Arnaud M., Ashdown M., Atrio-Barandela F., Aumont J., Baccigalupi C., Banday A. J., et al. 2013, ArXiv e-prints, 1303.5083, ADS
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. p. 251, arXiv:astro-ph/9509014, ADS
Pratten G., Munshi D., 2012, MNRAS, 423, 3209, 1108.1985, ADS
Rabus A., 1998, Diploma thesis (german), Ludwig Maximilians Universität München
Sahni V., Sathyaprakash B. S., Shandarin S. F., 1998, ApJ, 495, L5, arXiv:astro-ph/9801053, ADS
Sathyaprakash B. S., Sahni V., Shandarin S., 1998, ApJ, 508, 551, arXiv:astro-ph/9805285, ADS
Schmalzing J., 1999, PhD thesis, Ludwig Maximilians Universität München
Schmalzing J., Buchert T., 1997, ApJ, 482, L1, arXiv:astro-ph/9702130, ADS
Schmalzing J., Buchert T., Melott A. L., Sahni V., Sathyaprakash B. S., Shandarin S. F., 1999, ApJ, 526, 568, arXiv:astro-ph/9904384, ADS
Schmalzing J., Diaferio A., 2000, MNRAS, 312, 638, arXiv:astro-ph/9910228, ADS
Schmalzing J., Gorski K. M., 1998, MNRAS, 297, 355, arXiv:astro-ph/9710185, ADS
Schmalzing J., Kerscher M., Buchert T., 1996, in Bonometto S., Primack J. R., Provenzale A., eds, Dark Matter in the Universe Minkowski Functionals in Cosmology. p. 281, arXiv:astro-ph/9508154, ADS
Schmalzing J., Takada M., Futamase T., 2000, ApJ, 544, L83, ADS
Stoyan D., Kendall W. S., Mecke J., 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, 0711.4352, ADS
Sylos Labini F., 2011, EPL (Europhysics Letters), 96, 59001, 1110.4041, ADS
Whitbourn J. R., Shanks T., 2013, ArXiv e-prints, 1307.4405, ADS
Wiegand A., 2012, PhD thesis, Universität Bielefeld
York D. G., et al. SDSS Collaboration 2000, AJ, 120, 1579, astro-ph/0006396, ADS

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) states the general case without explicit derivation.

To derive the relation for all four Minkowski Functionals, we first define the Minkowski polynomial as follows:

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

from which we can extract the functionals by taking the derivatives

$$V_{\mu}\left(K\right)=\frac{1}{\alpha_{\mu}}\left.\frac{\partial^{\mu}M\left(t;K\right)}{\partial t^{\mu}}\right|_{t=0}.$$

These Minkowski polynomials obey the additivity relation of the Minkowski functionals, $\forall\,\mathcal{B}_1,\mathcal{B}_2\in\mathcal{R}$:

$$M\left(\mathcal{B}_1\cup\mathcal{B}_2\right)=M\left(\mathcal{B}_1\right)+M\left(\mathcal{B}_2\right)-M\left(\mathcal{B}_1\cap\mathcal{B}_2\right)\ . \qquad \text{(A1)}$$

So, for a collection of Balls, we have:

$$M\left(\cup_{i=1}^{N}B_i\right)=\sum_{i=1}^{N}M\left(B_i\right)-\sum_{i<j}M\left(B_i\cap B_j\right)+\ldots$$

$$+\left(-1\right)^{N+1}M\left(B_1\cap\ldots\cap B_N\right).$$

To calculate the average Minkowski Functionals for a structure with a given correlation, we weight this configuration with its probability

$$p\propto\varrho_N\left(x_1,x_2,\ldots,x_N\right)dV_1dV_2\ldots dV_N\ ,$$

where $\varrho_N\left(x_1,x_2,\ldots,x_N\right)$ is the complete N–point correlation function that is related to the probability of finding particles at the N positions x_n simultaneously. Calling the integration measure

$$\int d\tau_n=\int d^3x_1d^3x_2\ldots d^3x_n\ ,$$

we therefore find for the average

$$\langle M\rangle=\frac{1}{N^N}\int d\tau_N\varrho_N\left(x_1,x_2,\ldots,x_N\right)M\left(\cup_{i=1}^{N}B_i\right)=$$

$$=\sum_{n=1}^{N}\left(\begin{array}{c}N\\n\end{array}\right)N^{-n}\left(-1\right)^{n+1}\int_{\mathcal{D}}\mathrm{d}^3\tau_n\varrho_n\left(x_1,x_2,\ldots,x_n\right)\times$$

$$\times M\left(B_{x_1}\cap B_{x_2}\cap\ldots\cap B_{x_n}\right).$$

In the limit of an infinite structure $N \to \infty$, $\left(\begin{array}{c} N \\ n \end{array} \right) N^{-n} \to \frac{1}{n!}$, and so

$$\langle M \rangle = 1 - \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \int_{\mathcal{D}} \mathrm{d}^3\tau_n \varrho_n\left(x_1, x_2, \ldots, x_n\right) \times$$
$$\times M\left(B_{x_1} \cap B_{x_2} \cap \ldots \cap B_{x_n}\right).$$

To pass from the complete N–point correlations to the connected N–point correlation functions ξ_n one needs the principal kinematical formula

$$\int_{\mathcal{D}} d^3x M\left(A \cap B_x\right) \sim_{mod\left(t^4\right)} M\left(A\right) M\left(B\right),$$

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. This is then used to show the following formal cumulant relation

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

connecting the "moments"

$$m_n = \frac{1}{\varrho^n} \int d\tau_n \varrho_n\left(x_1, x_2, \ldots, x_n\right) M\left(B_{x_1} \cap B_{x_2} \cap \ldots \cap B_{x_n}\right)$$

to the "cumulants"

$$\kappa_n = \int d\tau_n \xi_n\left(x_1, x_2, \ldots, x_n\right) M\left(B_{x_1} \cap B_{x_2} \cap \ldots \cap B_{x_n}\right) .$$

This formal cumulant relation leads to the expression of the average Minkowski polynomial m of the form

$$m = 1 - \exp \sum_{n=1}^{\infty} \frac{(-\varrho)^n}{n!} \int_{\mathcal{D}} \mathrm{d}^3\tau_n \xi_n\left(x_1, x_2, \ldots, x_n\right) \times$$
$$\times M\left(B_{x_1} \cap B_{x_2} \cap \ldots \cap B_{x_n}\right) . \quad \text{(A2)}$$

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

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

which then leads directly to Equation (3).

APPENDIX B: COEFFICIENTS OF A LOG–NORMAL DISTRIBUTION

Here we want to check how well the expansion coefficients can be predicted, if we chose a Log–normal distribution as hypothesis. We plot the comparison of the values obtained from the mock catalogues, to those calculated for the Log–normal distribution in Figure B1. This latter resembles the Gaussian distribution, but has as probability distribution the form

$$p\left(\left\{\varrho\left(\mathbf{r}_i; \Delta V\right)/\varrho_0\right\}\right) = B\left(\varrho\left(\mathbf{r}_i; \Delta V\right)/\varrho_0\right) \times$$
$$\times e^{-\frac{1}{2}\sum_{i,j}\left(\log\left(\varrho\left(\mathbf{r}_i;\Delta V\right)/\varrho_0\right)-\mu_i\right)A_{ij}\left(\log\left(\varrho\left(\mathbf{r}_j;\Delta V\right)/\varrho_0\right)-\mu_j\right)}. \quad \text{(B1)}$$

For a single cell, without the correlation to other cells encoded in A_{ij}, the distribution is

$$p\left(\varrho/\varrho_0\right) = \frac{1}{\left(\varrho/\varrho_0\right)\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}\left(\log\left(\varrho/\varrho_0\right)-\mu\right)^2}, \quad \text{(B2)}$$

with mean

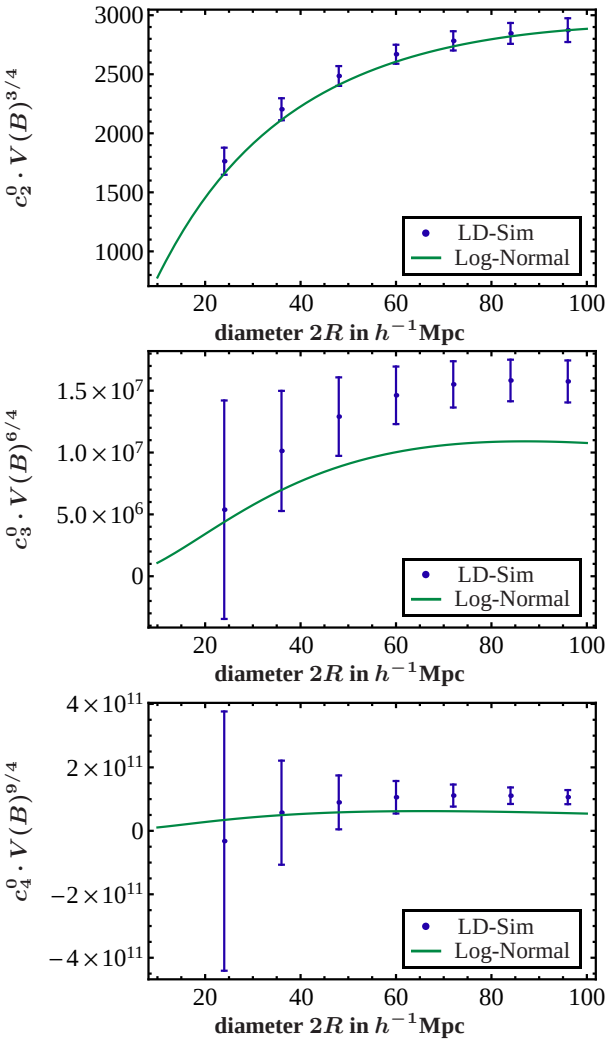

Figure B1. Comparison of the coefficients c_2^0–c_4^0 defined in (32) to those of a Log–normal distribution with the same two–point function. For a better visibility the coefficients c_n^0 have been multiplied by $V\left(B\right)^{3/4(n-1)}$ in these plots.

$$\mathbb{E}\left[\varrho/\varrho_0\right] = e^{\mu+\frac{\sigma^2}{2}} \quad \text{(B3)}$$

and variance

$$\mathbb{E}\left[\left(\varrho/\varrho_0\right)^2\right] - \mathbb{E}\left[\varrho/\varrho_0\right]^2 = \left(e^{\sigma^2} - 1\right) e^{2\mu+\sigma^2} . \quad \text{(B4)}$$

As we need the cumulants, we have to determine the joint expectation values of several variables. This gives

$$\mathbb{E}\left[\prod_{i\in B} X_i\right] = e^{\sum_i \mu_i + \frac{1}{2}\sum_{ij} A_{ij}^{-1}}, \quad \text{(B5)}$$

where the X_i are again the distributions of the i–th cell, $X_i = \hat{\varrho}\left(\mathbf{x}_i\right)/\varrho_0$, and A_{ij}^{-1} are the elements of the inverse matrix to the matrix A. The sums in the exponents run over the indices corresponding to the X_i in the product $\prod_{i\in B} X_i$. For the first two expectation values this means

$$\mathbb{E}\left[X_1\right] = e^{\mu_1 + \frac{1}{2}A_{11}^{-1}}, \quad \text{(B6)}$$

$$\mathbb{E}\left[X_1 X_2\right] = e^{\mu_1+\mu_2+\frac{1}{2}A_{11}^{-1}+A_{12}^{-1}+\frac{1}{2}A_{22}^{-1}} . \quad \text{(B7)}$$

This leads to the second cumulant $\xi\left(\mathbf{x}_{12}\right) =$

$$= \kappa\left(X_1, X_2\right) = e^{\mu_1+\mu_2+\frac{1}{2}A^{-1}_{11}+\frac{1}{2}A^{-1}_{22}}\left(-1+e^{A^{-1}_{12}}\right). \qquad \text{(B8)}$$

This implies that, unlike in the Gaussian case, the relation of the matrix elements A^{-1}_{ij} to the correlation function of the density field is now given by

$$A^{-1}_{ij} = \log\left(1+\xi\left(\mathbf{r}_i, \mathbf{r}_j\right)\right) \ . \qquad \text{(B9)}$$

This modified dependence on ξ comes from the fact that the over–density field $X_i = \hat{\varrho}\left(\mathbf{x}_i\right)/\varrho_0$ has as its average the value 1. So $\mathbb{E}\left[X_1\right] = 1$ and (B6) fixes the parameter μ_i in the distribution (B1) to be $\mu_i = -\frac{1}{2}A^{-1}_{ii}$. This directly leads to (B9) for the components of A^{-1}_{ij}.

With these two conditions all parameters of the distribution (B1) are fixed. Using (B5) we can now calculate, in principle, all higher connected correlation functions for a multivariate log–normal distribution having a given two–point correlation function. In practice, however, this quickly leads to long expressions, because of the rapid increase of the combinatorial component. Therefore, we only use connected correlation functions up to ξ_4.

Using the ξ's constructed in this way we can evaluate the series (5) up to an index $n = 4$. However, we do not calculate the coefficients for all four Minkowski Functionals. For V_1 and V_2 the problem is to determine the values of the MFs of an intersection of three Balls $V_\mu\left(B \cap B_{\mathbf{x}_1} \cap B_{\mathbf{x}_2}\right)$. While this is straightforward for V_0 and V_3 for the other two functionals, this would imply the search for suitable parametrisations of the body resulting from the intersection. Especially for V_3 this would be hard in view of the complicated expression (10) for V_2 for two Balls. This is why we restrict ourselves to the calculation of the higher–order terms for V_0.