

Cosmological model discrimination with weak lensing

S. Pires¹, J.-L. Starck¹, A. Amara^{1,2}, A. Réfrégier¹, and R. Teyssier¹

¹ Laboratoire AIM, CEA/DSM-CNRS-Université Paris Diderot, IRFU/SEDI-SAP, Service d'Astrophysique, CEA Saclay, Orme des Merisiers, 91191 Gif-sur-Yvette, France
e-mail: sandrine.pires@cea.fr

² Department of Physics, ETH Zürich, Wolfgang-Pauli-Strasse 16, 8093 Zürich, Switzerland

Received 2 December 2008 / Accepted 18 June 2009

ABSTRACT

Weak gravitational lensing provides a unique way of mapping directly the dark matter in the Universe. The majority of lensing analyses use the two-point statistics of the cosmic shear field to constrain the cosmological model, a method that is affected by degeneracies, such as that between σ_8 and Ω_m which are respectively the rms of the mass fluctuations on a scale of 8 Mpc/h and the matter density parameter, both at $z = 0$. However, the two-point statistics only measure the Gaussian properties of the field, and the weak lensing field is non-Gaussian. It has been shown that the estimation of non-Gaussian statistics for weak lensing data can improve the constraints on cosmological parameters. In this paper, we systematically compare a wide range of non-Gaussian estimators to determine which one provides tighter constraints on the cosmological parameters. These statistical methods include skewness, kurtosis, and the higher criticism test, in several sparse representations such as wavelet and curvelet; as well as the bispectrum, peak counting, and a newly introduced statistic called wavelet peak counting (WPC). Comparisons based on sparse representations indicate that the wavelet transform is the most sensitive to non-Gaussian cosmological structures. It also appears that the most helpful statistic for non-Gaussian characterization in weak lensing mass maps is the WPC. Finally, we show that the σ_8 - Ω_m degeneracy could be even better broken if the WPC estimation is performed on weak lensing mass maps filtered by the wavelet method, MRLens.

Key words. gravitational lensing – methods: data analysis – methods: statistical – cosmological parameters – dark matter

1. Introduction

Measurements of the image distortion of background galaxies caused by large-scale structures provide a direct way to study the statistical properties of the growth of structures in the Universe. Weak gravitational lensing measures the mass and can thus be directly compared to theoretical models of structure formation. Most lensing studies use the two-point statistics of the cosmic shear field because of its potential to constrain the power spectrum of density fluctuations in the late Universe (e.g., [Maoli et al. 2001](#); [Refregier et al. 2002](#); [Bacon et al. 2003](#); [Massey et al. 2005](#); [Dahle 2006](#)). Two-point statistics measure the Gaussian properties of the field. This represents a limited amount of information since it is well known that the low redshift Universe is highly non-Gaussian on small scales. Gravitational clustering is, indeed, a non linear process and on small scales, in particular, the mass distribution is highly non-Gaussian. Consequently, using only two-point statistics to place constraints on the cosmological model provides limited insight. Stronger constraints can be obtained using 3D weak lensing maps ([Bernardeau et al. 1997](#); [Pen et al. 2003](#); [Massey et al. 2007](#)). An alternative procedure is to consider higher-order statistics of the weak lensing shear field enabling a characterization of the non-Gaussian nature of the signal (see e.g., [Takada & Jain 2003](#); [Jarvis et al. 2004](#); [Kilbinger & Schneider 2005](#); [Hamana et al. 2004](#); [Donoho & Jin 2004](#)).

In this paper, we systematically compare a range of non-Gaussian statistics. For this purpose, we focus on the degeneracy between σ_8 and Ω_m which are respectively, the amplitude of the matter power spectrum and the matter density parameter, both at $z = 0$. We attempt to determine the most effective method for breaking this degeneracy that exists when only the two-point correlation is considered. A wide range of statistical

methods are systematically applied to a set of simulated data to characterize the non-Gaussianity present in the mass maps due to the growth of structures. Their ability to discriminate between different possible cosmological models are compared. For the analysis of CMB data, it has been proposed that statistics such as wavelet kurtosis or wavelet higher criticism can be used to detect clusters and curvelet kurtosis or curvelet higher criticism to detect anisotropic features such cosmic strings ([Jin et al. 2005](#)). Since weak lensing data may contain filamentary structures, we also considered statistical approaches based on sparse representations.

In Sect. 2, we review the major statistical methods used in the literature to constrain cosmological parameters from weak lensing data. Section 3 describes the simulations used in this paper, especially how 2D weak lensing mass maps of five different models have been derived from large statistical samples of 3D N -body simulations of density distribution. Section 4 is dedicated to the description of our analysis and we describe the different statistics that we studied along with the different multi-scale transforms investigated. We also present a new statistic that we call wavelet peak counting (WPC). In Sect. 5, we present our results and finally, Sects. 6 and 7, we present a discussion and summarize our conclusions.

2. Weak lensing statistics and cosmological models constraints: state of the art

2.1. Two-point statistics

The most common method for constraining cosmological parameters in weak lensing studies uses two-point statistics of the shear field calculated either in real or Fourier space. In general,

there is an advantage to using Fourier space statistics such as the power spectrum because the modes are independent. The power spectrum $P_\kappa(l)$ of the 2D convergence is defined as a function of the modes l by

$$\langle \hat{\kappa}(l)\hat{\kappa}(l') \rangle = (2\pi)^2 \delta(l-l') P_\kappa(l), \quad (1)$$

where hat symbols denote Fourier transforms, δ is the delta function, and the brackets denote an average over l . $P_\kappa(l)$ depends only on $l = |l|$ and κ is the convergence (i.e., which is proportional to the projected mass distribution).

This power spectrum $P_\kappa(l)$ can be expressed in terms of the 3D matter power spectrum $P(k, \chi)$ of the mass fluctuations $\delta\rho/\rho$ and of cosmological parameters as

$$P_\kappa(l) = \frac{9}{16} \left(\frac{H_0}{c} \right)^2 (\Omega_m)^2 \int d\chi \left[\frac{g(\chi)}{ar(\chi)} \right]^2 P\left(\frac{l}{r}, \chi\right), \quad (2)$$

where a is the expansion parameter, H_0 is the Hubble constant, and Ω_m is the matter density parameter. The correlation properties are expressed more conveniently in Fourier space, but for surveys with complicated geometry because of the removal of bright stars and artifacts, the missing data need proper handling (a review of the different existing methods can be found in Pires et al. 2008). Real space statistics are easier to estimate, but require more computational time. The following two-point statistics can be related to the underlying 3D matter power spectrum by means of the 2D convergence power spectrum $P_\kappa(l)$:

- The shear variance $\langle \gamma^2 \rangle$, which is an example of real space two-point statistic is defined as the variance in the average shear $\bar{\gamma}$ evaluated in circular patches of varying radius θ_s . The shear variance $\langle \gamma^2 \rangle$ is related to the power spectrum $P_\kappa(l)$ of the 2D convergence by

$$\langle \gamma^2 \rangle = \int \frac{dl}{2\pi} l P_\kappa(l) \frac{J_1^2(l\theta_s)}{(l\theta_s)^2}, \quad (3)$$

where J_n is a Bessel function of order n . This shear variance has been used in many weak lensing analysis to constrain cosmological parameters (Maoli et al. 2001; Hoekstra et al. 2006; Fu et al. 2008).

- The shear two-point correlation function, which is currently the most widely used statistic because it is easy to implement and can be estimated even in cases of complex geometry. It is defined to be

$$\xi_{i,j}(\theta) = \langle \gamma_i(\theta') \gamma_j(\theta' + \theta) \rangle, \quad (4)$$

where $i, j = 1, 2$ and the averaging is completed over pairs of galaxies separated by an angle $\theta = |\theta|$. In terms of parity, $\xi_{1,2} = \xi_{2,1} = 0$ and isotropy, $\xi_{1,1}$ and $\xi_{2,2}$ are functions only of θ . The shear two-point correlation functions can be related to the 2D convergence power spectrum by

$$\xi_+(\theta) = \xi_{1,1}(\theta) + \xi_{2,2}(\theta) = \int_0^\infty \frac{dl}{2\pi} l P_\kappa(l) J_0(l\theta). \quad (5)$$

These two-point correlation functions are the most popular statistical tools and have been used in the most recent weak lensing analyses (Benjamin et al. 2007; Hoekstra et al. 2006; Fu et al. 2008).

- The variance in the aperture mass M_{ap} was introduced by Schneider et al. (1998). It defines a class of statistics referred to as *aperture masses* associated with compensated filters. Several forms of filters have been suggested that balance locality in real space with locality in Fourier space.

Considering the filter defined in Schneider (1996) with a cut-off at some scale θ_s , the variance in the aperture mass can be expressed as a function of the 2D convergence power spectrum by

$$\langle M_{ap}^2(\theta_s) \rangle = \int \frac{dl}{2\pi} l P_\kappa(l) \frac{576 J_4^2(l\theta_s)}{(l\theta_s)^4}. \quad (6)$$

This statistic was used in Van Waerbeke et al. (2002); Semboloni et al. (2006); Hoekstra et al. (2006); Fu et al. (2008).

Two-point cosmic shear measurements yield interesting constraints on the amplitude of the matter power spectrum σ_8 on which the lensing strongly depends. However, when deriving joint constraints, a degeneracy exists between σ_8 and Ω_m (see e.g., Maoli et al. 2001; Hoekstra et al. 2002; Semboloni et al. 2006). This degeneracy between σ_8 and Ω_m is typical of cosmic shear measurements that use only two-point statistics.

Two-point statistics are inadequate for characterizing non-Gaussian features. Non-Gaussianity produced by the non-linear evolution of the Universe is of great importance for the understanding of the underlying physics, and it may help us to differentiate between cosmological models.

2.2. Non-Gaussian statistics

In the standard structure formation model, initial random fluctuations are amplified by non-linear gravitational instability to produce a final distribution of mass that is highly non-Gaussian. The weak lensing field is thus highly non-Gaussian. On small scales, we observe structures such as galaxies and clusters of galaxies, and on larger scales, some filament structures. Detecting these non-Gaussian features in weak lensing mass maps can be very useful for constraining the cosmological model parameters.

The three-point correlation function $\xi_{i,j,k}$ is the lowest-order statistics that can be used to detect non-Gaussianity, and is defined to be

$$\xi_{i,j,k}(\theta) = \langle \kappa(\theta_1) \kappa(\theta_2) \kappa(\theta_3) \rangle. \quad (7)$$

In Fourier space, it is called bispectrum and depends only on the distances $|l_1|$, $|l_2|$, and $|l_3|$, i.e.,

$$B(|l_1|, |l_2|, |l_3|) \propto \langle \hat{\kappa}(|l_1|) \hat{\kappa}(|l_2|) \hat{\kappa}^*(|l_3|) \rangle. \quad (8)$$

It has been shown that tighter constraints can be obtained with the three-point correlation function (Bernardeau et al. 1997; Cooray & Hu 2001; Bernardeau et al. 2003; Takada & Jain 2003, 2004; Schneider & Lombardi 2003; Schneider et al. 2005; Benabed & Scoccimarro 2006).

A simpler quantity than the three-point correlation function is provided by measuring the third-order moment (skewness) of the smoothed convergence κ (Bernardeau et al. 1997) or of the aperture mass M_{ap} (Jarvis et al. 2004; Kilbinger & Schneider 2005).

Another approach to searching for non-Gaussianity is to perform a statistical analysis directly on non-Gaussian structures such as clusters. Galaxy clusters are the largest virialized cosmological structures in the Universe. They provide a unique way to focus on non-Gaussianity present on small scales. One interesting statistic is *peak counting*, which searches the number of peaks detected on the 2D convergence corresponding to the cluster abundance (see e.g., Hamana et al. 2004).

The methods to search for non-Gaussianity in the weak lensing literature concentrate mainly on the higher-order correlation function. For the analysis of the CMB, the skewness

Table 1. Parameters of the five cosmological models that have been chosen along the (σ_8, Ω_m) -degeneracy.

| Model | Box | Ω_m | Ω_L | h | σ_8 |
|---------|-------|------------|------------|-------|------------|
| model 1 | 165.8 | 0.23 | 0.77 | 0.594 | 1 |
| model 2 | 159.5 | 0.3 | 0.7 | 0.7 | 0.9 |
| model 3 | 152.8 | 0.36 | 0.64 | 0.798 | 0.8 |
| model 4 | 145.7 | 0.47 | 0.53 | 0.894 | 0.7 |
| model 5 | 137.5 | 0.64 | 0.36 | 0.982 | 0.6 |

The box side is in Mpc/h, the simulations have 256^3 particles, Ω_m is the matter density parameter, Ω_L is the dark energy density parameter, h is equal to $H_0/100$ where H_0 is the Hubble constant, and σ_8 the amplitude of the matter power spectrum.

and kurtosis of the wavelet coefficients are also standard tools to check the CMB Gaussianity (Vielva et al. 2004; Starck et al. 2006a; Vielva et al. 2006; Wiaux et al. 2008), and it was shown that curvelets (Starck et al. 2003) were useful in detecting anisotropic feature such as cosmic strings in the CMB (Starck et al. 2004; Jin et al. 2005).

In this following, we perform a comparison between most existing methods to find the most suitable higher order statistic for constraining cosmological parameters from weak lensing data. To explore the effectiveness of a non-Gaussian measure, we use a battery of N -body simulations. By choosing five models whose two-point correlation statistics are degenerate, we attempt to determine which statistics are able to distinguish between these models.

3. Simulations of weak lensing mass maps

3.1. 3D N -body cosmological simulations

We obtained realistic realizations of convergence mass maps from N -body cosmological simulations using the RAMSES code (Teyssier 2002). The cosmological models were assumed to be in concordance with the Λ CDM model. We limited the model parameters to within a realistic range (see Table 1) choosing five models along the (σ_8, Ω_m) -degeneracy discussed in Sect. 2.1.

For each of our five models, we run 21 N -body simulations, each containing 256^3 particles. We refined the base grid of 256^3 cells when the local particle number had exceeded 10. We further similarly refined each additional levels up to a maximum level of refinement of 6, corresponding to a spatial resolution of 10 kpc h^{-1} .

3.2. 2D weak lensing mass map

In the N -body simulations that are commonly used in cosmology, the dark matter distribution is represented by discrete massive particles. The simplest way of treating these particles is to map their positions onto a pixelized grid. In the case of multiple sheet weak lensing, we do this by taking slices through the 3D simulations. These slices are then projected into 2D mass sheets.

The effective convergence can subsequently be calculated by stacking a set of these 2D mass sheets along the line of sight, using the lensing efficiency function. This is a procedure that used before by Vale and White (2003), where the effective 2D mass distribution κ_e is calculated by integrating the density fluctuation along the line of sight. Using the Born approximation, which neglects that the light rays do not follow straight lines, the

convergence can be numerically expressed by

$$\kappa_e \approx \frac{3H_0^2 \Omega_m L}{2c^2} \sum_i \frac{\chi_i(\chi_0 - \chi_i)}{\chi_0 a(\chi_i)} \left(\frac{n_p R^2}{N_i s^2} - \Delta r_{f_i} \right), \quad (9)$$

where H_0 is the Hubble constant, Ω_m is the density of matter, c is the speed of light, L is the length of the box, and χ are comoving distances where χ_0 is the comoving distance to the source galaxies. The summation is performed over the i th box. The number of particles associated with a pixel of the simulation is n_p , the total number of particles within a simulation is N_i , and $s = L_p/L$, where L_p is the length of the plane responsible for the lensing. R is the size of the 2D maps and $\Delta r_{f_i} = \frac{r_2 - r_1}{L}$, where r_1 and r_2 are comoving distances.

Using the previous 3D N -body simulations, we derived 100 realizations of the five models. Figure 1 shows one realization of the convergence for each of the 5 models. In all cases, the field is $3.95^\circ \times 3.95^\circ$, downsampled to 1024^2 pixels and we assume that the sources lie at exactly $z = 1$. On large scale, the map clearly shows a Gaussian signal. On small scales, the signal, in contrast, is clearly dominated by clumpy structures (dark matter halos) and is therefore highly non-Gaussian.

3.3. 2D weak lensing noisy mass map

In practice, the observed shear γ_i is obtained by averaging over a finite number of galaxies and is therefore noisy. The noise arises in both the measurement errors and from the intrinsic ellipticities dispersion of galaxies. As a good approximation, we modeled the noise as an uncorrelated Gaussian random field with variance

$$\sigma_{\text{noise}}^2 = \frac{\sigma_\epsilon^2}{A n_g}, \quad (10)$$

where A is the pixel size in arcmin 2 , n_g the average number of galaxies per arcmin 2 , and σ_ϵ the rms of the shear distribution. We assume that $\sigma_\epsilon \simeq 0.3$ per shear component and $n_g = 100$ gal/arcmin 2 as found approximately for space-based weak lensing surveys (Massey et al. 2004). An estimate of the noisy convergence κ_n is obtained using the least squares estimator defined in Starck et al. (2006b).

4. Cosmological model discrimination framework

In identifying the most suitable statistic for breaking the (σ_8, Ω_m) -degeneracy, we are interested in comparing different statistics estimated in different representations using the set of simulated data described in the previous section.

For each statistic, we need to characterize, the discrimination obtained between each couple of models. The optimal statistic will be the one that maximizes the discrimination for all model pairs.

4.1. Characterization of the discrimination

To find the most suitable statistic, we need to quantitatively characterize for each statistic the discrimination between two different models m_1 and m_2 . One way to proceed is to define a discrimination efficiency that expresses the ability of a statistic to discriminate in percentage. We then need to define, for each individual statistic, two different thresholds (see Fig. 2):

– a threshold \mathcal{T}_1 : if the estimation of a given statistic in a map κ_i is below \mathcal{T}_1 , the map κ_i is classified to belong to the model m_1 , and not if it is above;

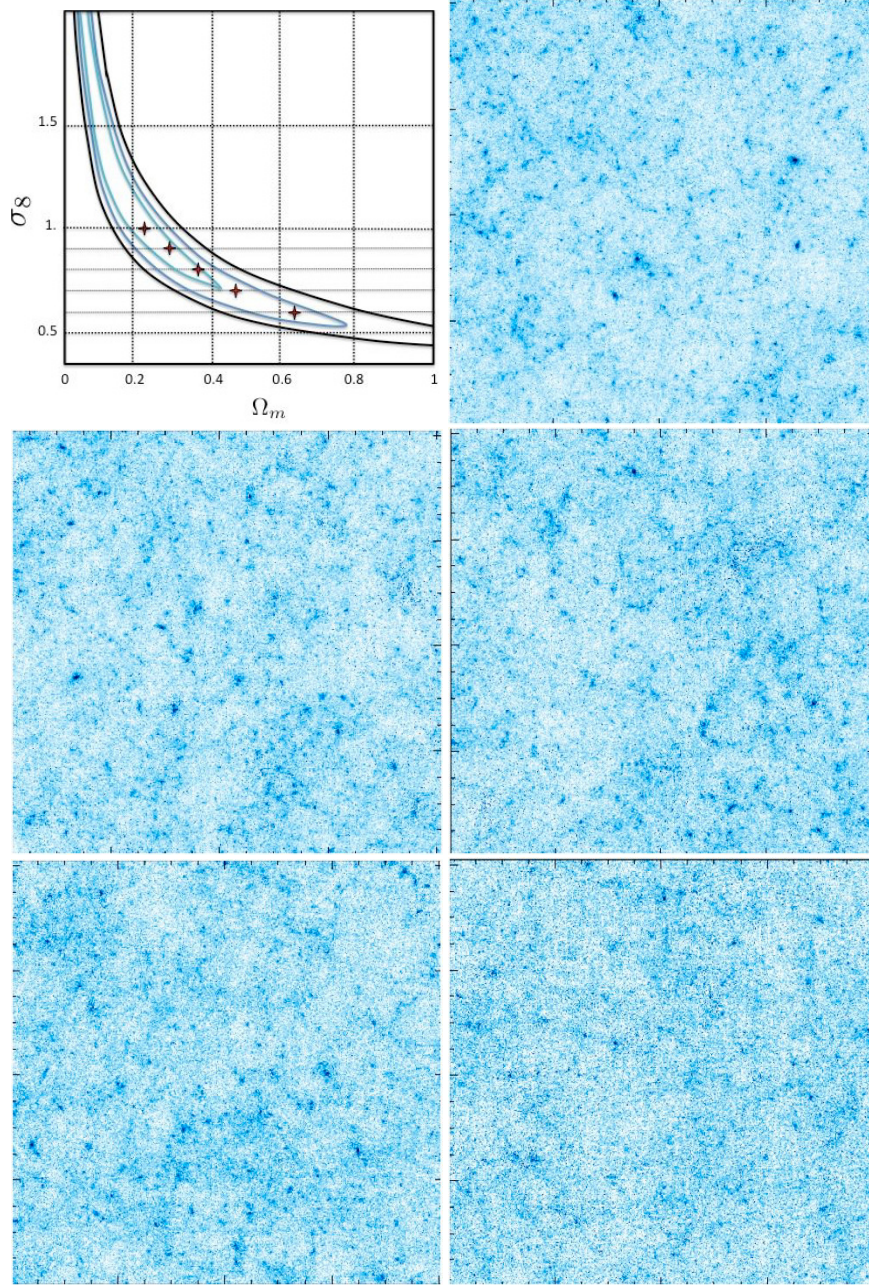


Fig. 1. *Upper left*, the 5 cosmological models along the (σ_8, Ω_m) -degeneracy. *Upper right*, one realization of the convergence κ for model 1 ($\sigma_8 = 1$ and $\Omega_m = 0.23$); *middle left*, for model 2 ($\sigma_8 = 0.9$ and $\Omega_m = 0.3$); *middle right*, for model 3 ($\sigma_8 = 0.8$ and $\Omega_m = 0.36$); *bottom left*, for model 4 ($\sigma_8 = 0.7$ and $\Omega_m = 0.47$) and *bottom right* for model 5 ($\sigma_8 = 0.6$ and $\Omega_m = 0.64$). Each map is 1.975×1.975 degrees down-sampled to 512×512 pixels.

– a threshold \mathcal{T}_2 : if the estimation of a given statistic in a map κ_i is above \mathcal{T}_2 , the map κ_i is classified to belong to the model m_2 , and not if it is below. We have used a statistical tool called FDR (false discovery rate) introduced by [Benjamini & Hochberg \(1995\)](#) to set these two thresholds (\mathcal{T}_1 and \mathcal{T}_2) correctly (see Appendix A).

This FDR method is a competitive tool that sets a threshold in an adaptive way without any assumption, given a false discovery rate (α). The false discovery rate is given by the proportion of false detections over the total number of detections. The threshold being estimated, we can derive a discrimination efficiency for each statistic. The discrimination efficiency measures the ability of a statistic to differentiate a model and another by calculating the ratio of detections (true or false) to the total number of

samples. It corresponds basically to the part of the distribution that does not overlap. The more the distributions overlap, the lower the discrimination efficiency will be.

Figure 2 represents the dispersion in the values of a given statistic estimated for 100 realizations of the model 1 (on the left) and 100 realizations of the model 2 (on the right). The two distributions barely overlap, it indicates a good discrimination that is to say the two models can easily be separated with this statistic. To be more quantitative, a threshold must be applied to each distribution to estimate a discrimination efficiency that corresponds to the part of the distribution delimited by the hatched area.

The formalism of the FDR method ensures that the yellow area delimited by corresponding to false detections will be small.

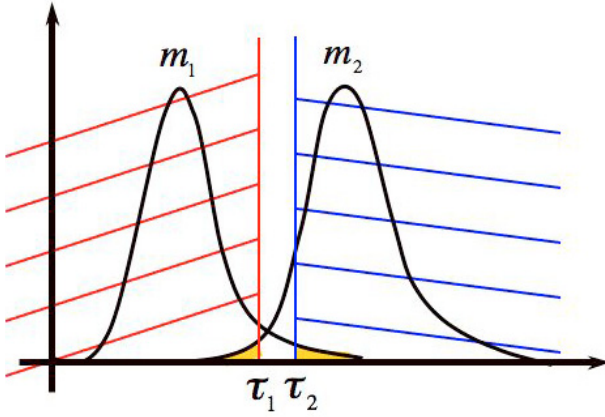


Fig. 2. The following two distributions correspond to the histogram of the values of a given statistic estimated on the 100 realizations of model 1 (m_1) and on the 100 realizations of model 2 (m_2). The discrimination achieved with this statistic between m_1 and m_2 is rather good: the two distributions barely overlap. To characterize the discrimination more quantitatively, the FDR method has been used to estimate the thresholds \mathcal{T}_1 and \mathcal{T}_2 . A false discovery rate (α) equal to 0.05 has been chosen. Then a discrimination efficiency can be derived.

4.2. A set of statistical tools

The first objective of this study is to compare different statistics to identify the one that places tighter constraints on the cosmological parameters. By adopting two-point statistics that contain all the information about a Gaussian signal leading to the (σ_8, Ω_m) -degeneracy, we have opted for statistics currently used to detect non-Gaussianity to probe the non-linear process of gravitational clustering. The statistics that we selected are the following:

- Skewness (S_κ):
The skewness is the third-order moment of the convergence κ and is a measure of the asymmetry of a distribution. The convergence skewness is related primarily to rare and massive dark-matter halos. The distribution will be more or less skewed positively depending on the abundance of rare and massive halos.
- Kurtosis (K_κ):
The kurtosis is the fourth-order moment of the convergence κ and is a measure of the peakedness of a distribution. A high kurtosis distribution has a sharper “peak” and flatter “tails”, while a low kurtosis distribution has a more rounded peak with wider “shoulders”.
- Bispectrum (B_κ):
There have been many theoretical studies of the three-point correlation function to constrain the cosmological parameters. However, the direct computation of three-point correlation function would take too long for our large maps. We therefore used its Fourier analog: the bispectrum that has been introduced Sect. 2.2. And we consider the equilateral configuration.
- Higher Criticism (HC):
The HC statistic was developed by Donoho & Jin (2004). It measures non-Gaussianity by identifying the maximum deviation after comparing the sorted p -values of a distribution with the sorted p -values of a normal distribution. A large HC value implies non-Gaussianity. We consider two different forms of HC (see Appendix B): HC^* and HC^+ .
- Peak counting (P_c):
We now investigate the possibility of using peak counting to

differentiate between cosmological models. By peak counting (or cluster count), we mean the number of halos that we can detect per unit area of sky (identified as peak above a mass threshold in mass maps). This cluster count enables us to constrain the matter power spectrum normalization σ_8 for a given Ω_m (see e.g., Bahcall & Fan 1998) and the formalism exist to predict the peak count to a given cosmological model (see e.g., Hamana et al. 2004).

– Wavelet peak counting (WPC):

We introduce a new statistic that we call wavelet peak counting (WPC). This statistic consists of estimating a cluster count per scale of a wavelet transform i.e., it is an approximate cluster count that depends on the size of the clusters. In the following, we show how WPC is superior to peak counting in characterizing the non-linear structure formation process.

All statistics are estimated in terms of the 2D convergence κ map. In practice, the complex geometry of surveys gives the convergence κ , which is in turn inferred from the shear field γ_i of the same complex geometry. However, it becomes possible with the inpainting method developed in Pires et al. (2008) to reconstruct a full convergence κ map. The solution proposed enables judiciously masked regions to be filled-in so as to reduce the impact of missing data on the estimation of both the power spectrum and higher-order statistical measures, while requiring $O(N \log N)$ operations.

4.3. Representations

The second objective of the present study was to compare different transforms and find the sparsest representation of weak lensing data that makes the discrimination easier. Some studies of CMB data used multiscale methods to detect non-Gaussianity (Aghanim & Forni 1999; Starck et al. 2004). Weak lensing maps exhibit both isotropic and anisotropic features. These kind of features can be represented more successfully by some basis functions. A transform is, indeed, optimal for detecting structures which have the same shape of its basis elements. We thus tested different representations:

- the Fourier transform;
- the anisotropic bi-orthogonal wavelet transform which we expect to be optimal for detecting mildly anisotropic features.
- The isotropic “à trous” wavelet transform which is well adapted to the detection of isotropic features such as the clumpy structures (clusters) in the weak lensing data;
- the ridgelet transform developed to process images including ridge elements which therefore provides a good representation of perfectly straight edges;
- the curvelet transform which allows us to approximate curved singularities with few coefficients and then provides a good representation of curved structures.

A description of these transforms is available in Appendix C. We note that the estimation of a given statistic at each resolution level of a representation can be considered to be individual statistic in itself.

5. Analysis and results

5.1. Treatment of the noise

As explained previously, the weak lensing mass maps are measured from the distortions of a finite number of background galaxies and therefore suffer from shot noise. Furthermore, each

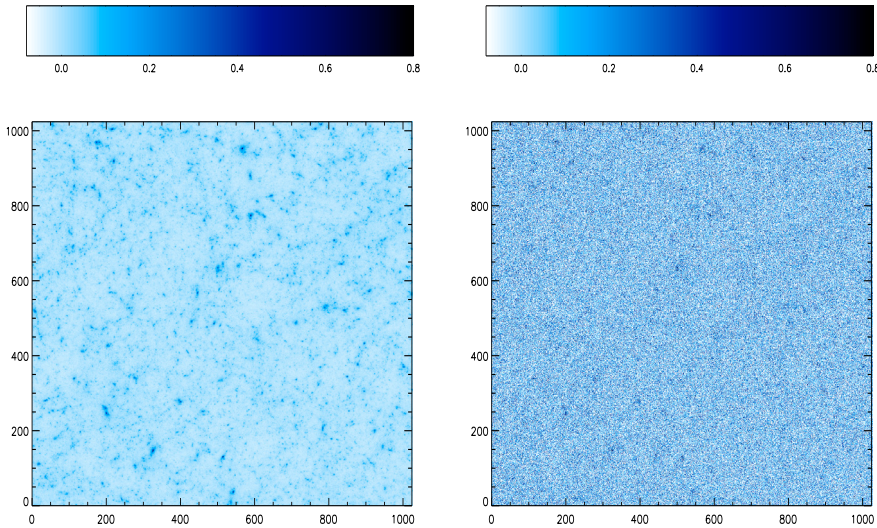


Fig. 3. *Left:* noiseless simulated mass map and *right:* simulated noisy mass map that we should obtain in space observations. The field is $1.975^\circ \times 1.975^\circ$.

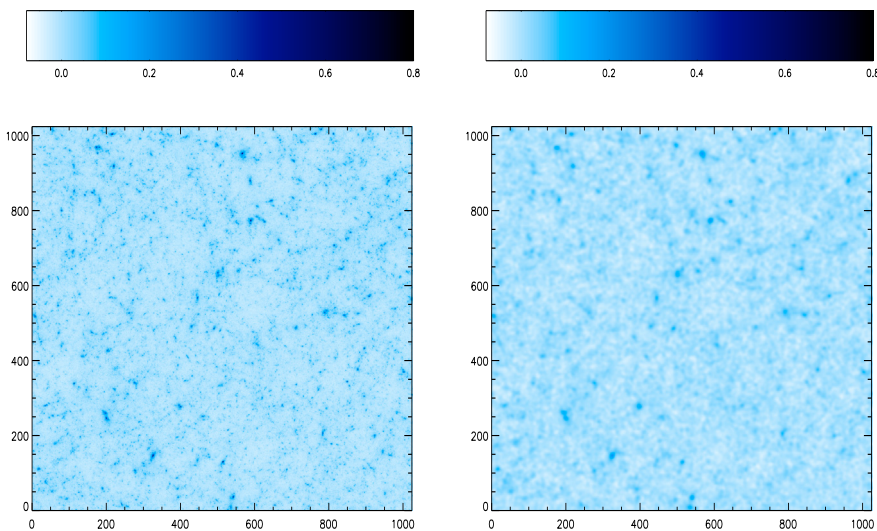


Fig. 4. *Left,* noiseless simulated mass map, and *right,* filtered mass map by convolution with a Gaussian kernel. The field is $1.975^\circ \times 1.975^\circ$.

galaxy provides only a noisy estimator of the distortion field. We added the expected level of Gaussian noise (see Sect. 3.3) to simulations of weak lensing mass maps to obtain simulated noisy mass maps corresponding to space observations. Figure 3 shows a noiseless simulated mass map (left) and a noisy simulated mass map (right) corresponding to space observations.

The noise has an impact on the estimated statistics and therefore needs to be considered.

5.1.1. No filtering

We started by applying the different transformations directly to the noisy data and, for each representation estimating the statistics described in the previous section, with the exception of cluster count and WPC, which required filtering.

We expect that the noise will make the third and fourth order statistics tend to zero. The more noisy the data are, the more the distribution will indeed look like a Gaussian.

5.1.2. Gaussian filtering

As a second step, we performed Gaussian filtering obtained by convolving the noisy mass maps κ_n with a Gaussian window G

of standard deviation σ_G :

$$\kappa_G = G * \kappa_n. \quad (11)$$

On the left, Fig. 4 shows the original map without noise, and on the right, the result obtained by Gaussian filtering of the noisy mass map displayed in Fig. 3 (right). The quality of the filtering depends strongly on the value of σ_G . For the simulations, an optimal value is about 0.9 arcmin. We then computed all our statistics for this new set of filtered mass maps.

5.1.3. MRLens filtering

We finally used a method of non-linear filtering based on the wavelet representation, i.e., the MRLens filtering proposed by Starck et al. (2006b). The MRLens¹ filtering is based on Bayesian methods that incorporate prior knowledge into data analysis. Choosing the prior is one of the most critical aspects of Bayesian analysis. Here a multiscale entropy prior is used. A full description of the method is given in the Appendix D. The MRLens

¹ The MRLens software is available at the following address: <http://www-irfu.cea.fr/Ast/878.html>

software that we used is available at the following address: <http://www-irfu.cea.fr/Ast/878.html>.

In Starck et al. (2006b), it was shown that this method outperforms several standard techniques such as Gaussian filtering, Wiener filtering, and MEM filtering). On the left, Fig. 5 shows the original map without noise, and on the right, the result of FDR filtering of the noisy mass map displayed Fig. 3 (right). The visual aspect indicates that many clusters are reconstructed and the intensity of the peaks are well recovered.

As before, all the statistics were estimated for these MRLens filtered mass maps. By essentially reconstructing the clusters, we anticipate that this MRLens method will improve statistical methods such as peak counting or WPC more than statistics that focus on the background.

5.2. Results

5.2.1. The discrimination methodology

As explained in Sect. 4.1, for each statistic described in the previous section, we can derive a discrimination efficiency between each two models out of the full set of 5 models. These values are given in Tables 3, 5, and 7 for three different statistics.

A mean discrimination efficiency for each individual statistic can be estimated by averaging the discrimination efficiency across all pairs of models. For statistics estimated in multiscale representations, the mean discrimination efficiency is calculated for each scale and we adopt the most reliable statistic.

Tables 2, 4, and 6 show the mean discrimination efficiency reached by a given statistic and a given transform, for respectively, (i) unfiltered mass maps; (ii) Gaussian filtered mass maps; and (iii) MRLens mass maps.

The mean discrimination efficiency of the Table 3 is about 40% and corresponds to the value at position (1,3) in Table 2 (i.e., the skewness of wavelet coefficients). The mean discrimination efficiency of Table 7 is about 97% and corresponds to the value at position (3,5) of Table 6.

Reliable differentiation between the five models is achievable if the mean discrimination efficiency is about 100%. If between 80% and 100%, the discrimination is possible except for adjacent models. At a level of below 40% no discrimination is possible even for distant models.

5.2.2. Differentiation between the noisy mass maps

The mean discrimination efficiency obtained for unfiltered mass maps is displayed in Table 2. The peak counting and WPC differ from the others. They cannot be computed on unfiltered mass maps because the clusters cannot be extracted from noisy mass maps. Another point is that the bispectrum by definition can only be estimated in the Fourier domain.

Without filtering the results are poor and no discrimination can be achieved in direct space. The signal-to-noise ratio is indeed, weak as can be seen in Fig. 3 (right). The non-Gaussian signal is hidden by the Gaussian noise.

The different transforms appear to be unable to bring out efficiently the non-Gaussianity characteristics except for the isotropic wavelet transform (see Table 2), which performs rather well, whatever the statistic. This is probably because it is an optimal transform for detecting clusters. Indeed, the clusters are a direct probe of non-Gaussianity and by concentrating on the cluster information, the isotropic wavelet transform seems to be a better representation of non-Gaussianity.

The skewness in the wavelet transform representation appears to be the most successful statistic in unfiltered mass maps. Table 3 shows the discrimination efficiency obtained for the skewness on the second scale of an isotropic wavelet transform which is the scale for which optimal differentiation between models is achieved. We can see that the differentiation is only achieved between the farthest models, which is quite poor. It is illustrated in Fig. 6, where you can see that the 5 distributions overlap. Some groups have already used the skewness aperture mass statistic to try to break the σ_8 - Ω_m degeneracy, (see e.g., Kilbinger & Schneider 2005; Jarvis et al. 2004). This processing consists of convolving the noisy signal with Gaussian windows of different scales and is quite similar to an isotropic wavelet transform. They showed that by combining second and third-order statistics, the degeneracy can be diminished but not broken.

5.2.3. Discrimination in Gaussian filtered mass maps

To increase the signal-to-noise ratio, we applied a Gaussian filter to the noisy simulated mass maps. Table 4 shows the results.

After Gaussian filtering, the noise is removed but the structures are over-smoothed. Except for the direct space where the results are clearly improved by the noise removal, the results after Gaussian filtering are relatively unchanged. Some statistics are a bit improved by the noise removal, while others become worse.

In contrast, the peak counting and WPC that can now be estimated for these filtered mass maps, perform well. Table 5 shows the discrimination efficiency obtained with peak counting estimated on Gaussian-filtered mass maps. We can see that except for adjacent models, discrimination is now possible. We can verify these results by considering the 5 distributions displayed Fig. 7. The distributions, indeed, barely overlap for models that are not adjacent.

The ability of the weak-lensing cluster count (peak counting) to discriminate between the five different models chosen along the degeneracy, can be explained. If we assume that the dark matter lies at $z \simeq 0.5$, which would correspond to the maximum lensing efficiency for background galaxies at $z = 1$, and assuming a constant dark energy model, the number density of massive clusters in the weak lensing mass maps is sensitive to both the amplitude of the mass fluctuations σ_8 and the matter density parameter Ω_m , both at $z = 0$. If instead of σ_8 , we consider $\sigma_8^{z \simeq 0.5}$, which represents the amplitude of the fluctuations of the projected weak lensing mass map, the $\sigma_8^{z \simeq 0.5}$ is now a constant for the five models because the five corresponding weak lensing power spectrum are indistinguishable. This leaves the Ω_m parameter which controls the structure formation (see e.g., Bahcall & Fan 1998). A low value of Ω_m ensures that the structures form earlier and a high Ω_m ensures that the structures form later. Then, for a low value of Ω_m , the abundance of massive clusters at $z \simeq 0.5$ become more significant (see Fig. 1 upper right) than for a high Ω_m . The cluster count can then be used to differentiate between cosmological models. The massive cluster abundance has already been used to probe Ω_m (see e.g., Bahcall & Fan 1998).

5.2.4. Discrimination in MRLens filtered mass maps

Table 6 shows the results obtained when the MRLens filtering scheme is applied to the noisy simulated mass maps.

After MRLens filtering, the sensitivity of all transforms is greatly improved. However, the isotropic wavelet transform

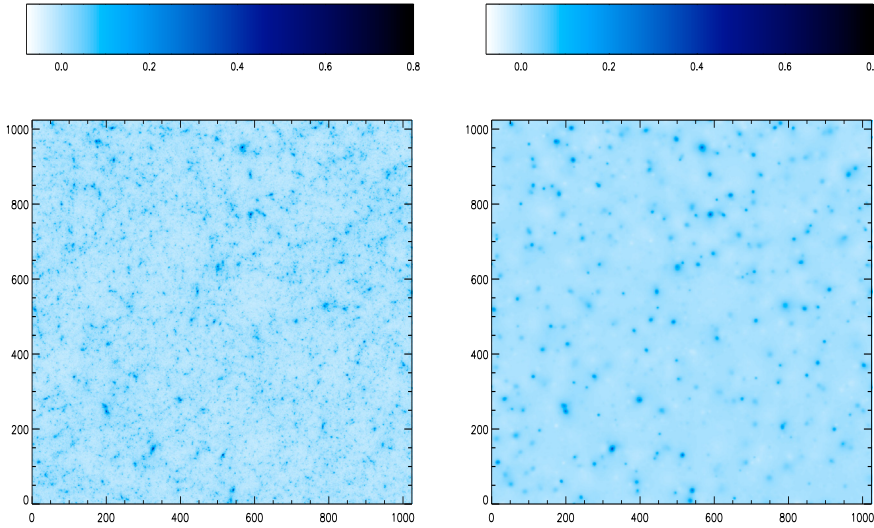


Fig. 5. *Left*, noiseless simulated mass map, and *right*, filtered mass map by the FDR multiscale entropy filtering. The field is $1.975^\circ \times 1.975^\circ$.

Table 2. Mean discrimination efficiencies (in percent) achieved on noisy mass maps with a false discovery rate of $\alpha = 0.05$.

| Basis/statistics | S | K | HC_n^* | HC_n^+ | B_l | P_c | WPC |
|---------------------------------|------|------|----------|----------|-------|-------|-----|
| Direct space | 21.0 | 1.6 | 1.8 | 1.9 | × | × | × |
| Fourier space | 0.7 | 10.7 | 3.0 | 21.3 | 5.75 | × | × |
| Isotropic wavelet transform | 40.5 | 29.3 | 24.9 | 34.0 | × | × | × |
| Bi-orthogonal wavelet transform | 7.3 | 10.8 | 4.1 | 2.9 | × | × | × |
| Ridgelet transform (block = 16) | 6.2 | 3.7 | 7.2 | 5.6 | × | × | × |
| Ridgelet transform (block = 32) | 8.3 | 15.2 | 9.5 | 20.4 | × | × | × |
| Fast curvelet transform | 1.1 | 4.2 | 2.6 | 4.2 | × | × | × |

Table 3. Discrimination efficiencies (in percent) achieved in unfiltered mass maps with the skewness estimated on the second scale of an isotropic wavelet transform.

| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---------|---------|---------|---------|---------|---------|
| model 1 | × | 4 | 55 | 86 | 99 |
| model 2 | 6 | × | 8 | 52 | 88 |
| model 3 | 23 | 3 | × | 6 | 77 |
| model 4 | 81 | 7 | 2 | × | 13 |
| model 5 | 99 | 81 | 12 | 9 | × |

remains the most efficient transform certainly helped by the MRLens filtering that uses this transform.

The most effective statistic remains peak counting, which is also helped by the MRLens filtering that essentially reconstructs the clusters. However, the other statistics also achieve good results in these MRLens filtered mass maps compared to those with Gaussian filtered mass maps. This is probably because the MRLens filtering, by favoring clusters reconstruction, helps all statistics that search for non-Gaussianity.

The best result is obtained with WPC on the third scale of the isotropic wavelet transform (see Table 7). Figure 8 shows the 5 distributions that barely overlap. This statistic allows us to discriminate between different models even for adjacent models for which the discrimination is challenging.

A comparison of these results with those obtained for noisy mass maps by the skewness in a wavelet representation (Table 3), shows that the accuracy of the constraints on σ_8 and Ω_m is improved greatly by using WPC estimated for MRLens filtered mass maps.

6. Discussion

As stated earlier, the formalism of the halo model provides a prediction of the number of clusters contained in a given field for a given cosmological model (Press and Schechter 1974; Sheth and Tormen 1999; Hamana et al. 2004). However, we have to consider that just a fraction of the clusters, present in the sky, will be detected. It follows that we have to take into account the selection effects caused by the observation quality and the data processing method. The solution that is currently used consists of modeling the selection effects by estimating the selection function. An analytic model can be developed by considering all the selection effects. An alternative consists in using a Monte Carlo approach that enables us to take into account all the selection effects that could not be considered in the analytical approach. This study will be done in a future work. When, the selection function is specified, the connection between observations and theory is straightforward. The cosmological parameters can thus be estimated from WPC.

For a perfect differentiation between the 5 cosmological models with WPC upper limits error in the cosmological parameters can be given by considering the spacing between two adjacent models. For space observations covering 4 square degrees, the upper limit error in σ_8 is 8%, in the range of $\sigma_8 \in [0.6, 1]$, and the upper limit error in Ω_m is 12%, in the range of $\Omega_m \in [0.23, 0.64]$. In future work, an accurate estimation of the error should be done.

7. Conclusion

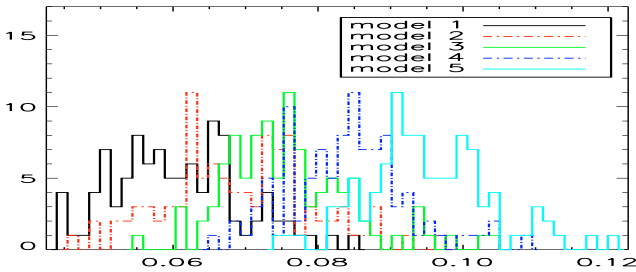
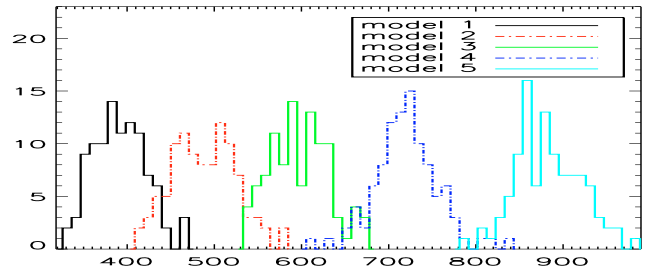
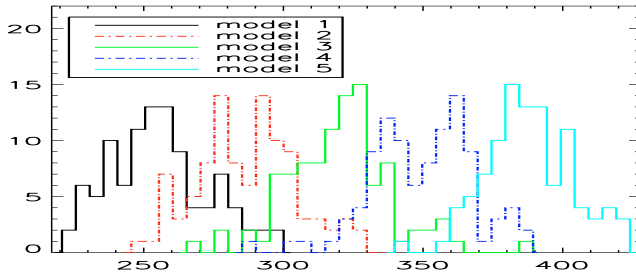
By using only two-point statistics to constrain the cosmological model, various degeneracies between cosmological parameters are possible, such as the $(\sigma_8-\Omega_m)$ -degeneracy. In this paper, we

Table 4. Mean discrimination efficiencies (in percent) achieved on Gaussian-filtered mass maps with a false discovery rate of $\alpha = 0.05$.

| Basis/statistics | S | K | HC_n^* | HC_n^+ | B_l | P_c | WPC |
|---------------------------------|------|------|----------|----------|-------|-------|------|
| Direct space | 39.2 | 34.3 | 28.8 | 40.2 | × | 79.3 | × |
| Fourier space | 2.2 | 2.2 | 3.8 | 2.9 | 6.75 | × | × |
| Isotropic wavelet transform | 31.9 | 29.5 | 24. | 38.4 | × | × | 79.2 |
| Bi-orthogonal wavelet transform | 12.2 | 10.5 | 5.1 | 6.8 | × | × | × |
| Ridgelet transform (block = 16) | 7.4 | 20.2 | 22.7 | 35.3 | × | × | × |
| Ridgelet transform (block = 32) | 2.5 | 7.3 | 4. | 9.1 | × | × | × |
| Fast curvelet transform | 1.3 | 4.7 | 4.8 | 8.3 | × | × | × |

Table 5. Discrimination efficiencies (in percent) achieved on Gaussian-filtered mass maps with the peak counting statistic on direct space given a False Discovery Rate $\alpha = 0.05$.

| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---------|---------|---------|---------|---------|---------|
| model 1 | × | 50 | 95 | 100 | 100 |
| model 2 | 46 | × | 43 | 96 | 100 |
| model 3 | 94 | 8. | × | 63. | 100 |
| model 4 | 100 | 97 | 50 | × | 78 |
| model 5 | 100 | 100 | 99 | 63 | × |

**Fig. 6.** Distribution of the skewness calculated from the second scale of an isotropic wavelet transform on the simulated realizations of the 5 models. It illustrates the results of the Table 3. No discrimination is possible except between the farthest models (i.e., between models 1 and 5).**Fig. 8.** Distribution of the wavelet peak counting estimated at the third scale of an isotropic wavelet transform on the simulated realizations of the 5 models. It illustrates the results of the Table 7. We obtain a good discrimination even for adjacent models.**Fig. 7.** Distribution of the peak counting estimated directly on the simulated realizations of the 5 models. It illustrates the results of the Table 5. The discrimination is possible except between adjacent models (that is to say between model 1 and model 2, model 2 and model 3, model 3 and model 4, model 4 and model 5).

have considered a range of non-Gaussian statistics to attempt to place tighter constraints on cosmological parameters. For this purpose, we have run N -body simulations of 5 models selected along the $(\sigma_8\text{-}\Omega_m)$ -degeneracy, and we have examined different non-Gaussian statistical tools in different representations, comparing their abilities to differentiate between models. Using non-Gaussian statistics, we have searched for non-Gaussian signal for small scales caused by gravitational clustering.

The main conclusions of our analysis are the following:

1. The isotropic wavelet transform has been found to be the best representation of the non-Gaussian structures in weak

lensing data because it can differentiate between the models the most reliably.

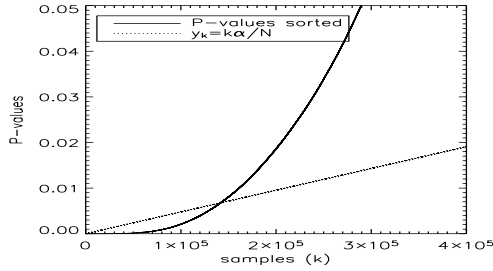
2. We have shown that a wavelet transform denoising method such as the MRLens filtering, which reconstructs essentially the non-Gaussian structures (the clusters), helps the statistics to characterize the non-Gaussianity more reliably.
3. We have introduced a new statistic called wavelet peak counting (WPC), which consists of estimating a cluster count per scale of an isotropic wavelet transform.
4. WPC has been found to be the best statistic compared to the others that we have tested (skewness, kurtosis, bispectrum, HC , P_c), and we have shown that this statistic estimated on MRLens filtered maps provide a strong discrimination between the 5 selected models.

In this paper, we have compared systematically a wide range of non-Gaussian statistics proposed in the literature, and WPC has been found to be the best statistic. By comparing models selected along the $(\sigma_8\text{-}\Omega_m)$ -degeneracy, this study shows that the power spectrum (or the two-point correlation function) and WPC should therefore be used simultaneously. However, these two statistics probe the same field, therefore the correlations introduced by the combined measurement need to be taken into account (see Takada & Bridle 2007). This will be investigated in the future.

Another issue, discussed in Sect. 6, is selection effects. We need to determinate the selection function to estimate the cosmological parameters and accurately estimate their errors. This study will be completed in a future work.

Table 6. Mean discrimination efficiencies (in percent) achieved on MRLens filtered mass maps with a false discovery rate of $\alpha = 0.05$.

| Basis/statistics | S | K | HC_n^* | HC_n^+ | B_l | P_c | WPC |
|---------------------------------|------|------|----------|----------|-------|-------|------|
| Direct space | 65.2 | 54.5 | 72.9 | 8.4 | × | 93.2 | × |
| Fourier space | 3.2 | 3.4 | 30.0 | 38.4 | 9.95 | × | × |
| Isotropic wavelet transform | 81.7 | 68.1 | 74.8 | 74.1 | × | × | 97.3 |
| Bi-orthogonal wavelet transform | 52.2 | 67.3 | 29.7 | 65.0 | × | × | × |
| Ridgelet transform (block = 16) | 45.1 | 54.5 | 77.1 | 37.2 | × | × | × |
| Ridgelet transform (block = 32) | 61.5 | 52.6 | 70.3 | 69.0 | × | × | × |
| Fast curvelet transform | 25. | 65.5 | 68.8 | 63.3 | × | × | × |

**Fig. 9.** Finding a threshold graphically using the FDR procedure.**Table 7.** Discrimination efficiencies (in percent) achieved on MRLens filtered mass maps with WPC on the third scale of an isotropic wavelet transform.

| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---------|---------|---------|---------|---------|---------|
| model 1 | × | 86 | 100 | 100 | 100 |
| model 2 | 87 | × | 94 | 100 | 100 |
| model 3 | 100 | 92 | × | 94 | 100 |
| model 4 | 100 | 100 | 93 | × | 99 |
| model 5 | 100 | 100 | 100 | 100 | × |

Finally, while peak counting and WPC provide a lot of information, further statistics such as the cluster count per mass and the spatial cluster correlation will provide further constraints. Future work will be needed to fully exploit these approaches.

Appendix A: The FDR method

The false discovery rate (FDR) is a statistical approach to the multiple testing problem, introduced by [Benjamini & Hochberg \(1995\)](#). The FDR method offers an effective way of selecting an adaptative threshold, without any assumption. The FDR threshold is determined from the observed p -value distribution, and hence depends on the amount of signal in the data.

This technique was described by [Miller et al. \(2001\)](#); [Hopkins et al. \(2002\)](#); [Starck et al. \(2006b\)](#); [Pires et al. \(2006\)](#) for several astrophysical applications. Instead of controlling the chance of any false positives, FDR controls the expected proportion of false positives. The FDR is given by the ratio of declared active pixels (12) that are false positives, i.e.,

$$\mathcal{FDR} = \frac{V_{ia}}{D_a}, \quad (12)$$

where V_{ia} is the number of pixels that are truly inactive but declared active, and D_a is the number of pixels declared active. The FDR formalism ensures that, *on average*, the false discovery rate is no higher than α , which lies between 0 and 1. This procedure guarantees control over the FDR in the sense that:

$$\mathcal{E}(\mathcal{FDR}) \leq \frac{T_i}{V} \cdot \alpha \leq \alpha, \quad (13)$$

Where, the unknown factor $\frac{T_i}{V}$ is the proportion of truly inactive pixels, where T_i is the number of inactive pixels, and V is the total number of pixels.

In the FDR procedure, we assume that P_1, \dots, P_n denote the p -values from the N tests, listed from lowest to highest and

$$d = \max \left\{ k : P_k < \frac{k \cdot \alpha}{c_N \cdot N} \right\}, \quad (14)$$

where $c_N = 1$, if p -values are statistically independent.

We now declare all pixels with p -values less than or equal to P_d to be active.

Graphically, this procedure corresponds to plotting P_k versus $\frac{k}{N}$, superposing the line through the origin of slope $\frac{\alpha}{c_N}$ (see [Fig. 9](#)), and finding the last point at which P_k falls below the line, defined to be P_d . From this p -value P_d , we can derive a threshold \mathcal{T} . All pixel values greater than \mathcal{T} have a p -value less than P_d and are declared active.

Appendix B: The higher criticism definition

To define HC, we first convert the individual κ_i into p -values. Let $p_i = P\{N(0, 1) > \kappa_i\}$ be the i th p -value, and $p_{(i)}$ denote the p -values sorted in increasing order. The higher criticism statistic is defined to be

$$HC_n^* = \max_i \left| \frac{\sqrt{n}[i/n - p_{(i)}]}{\sqrt{p_{(i)}(1 - p_{(i)})}} \right|, \quad (15)$$

or in a modified form :

$$HC_n^+ = \max_{i: 1/n \leq p_{(i)} \leq 1-1/n} \left| \frac{\sqrt{n}[i/n - p_{(i)}]}{\sqrt{p_{(i)}(1 - p_{(i)})}} \right|. \quad (16)$$

Appendix C: Description of the representations

The anisotropic bi-orthogonal wavelet transform

The most commonly used wavelet transform is the undecimated bi-orthogonal wavelet transform (OWT). Using the OWT, a 2D signal S can be decomposed as follows:

$$S(x, y) = \sum_{k_x, k_y} \phi_{J, k_x, k_y}(x, y) C_J(k_x, k_y) \quad (17)$$

$$+ \sum_d \sum_{k_x, k_y} \sum_{j=1}^J \psi_{j, k_x, k_y}^d(x, y) w_j^d(k_x, k_y), \quad (18)$$

where ϕ and ψ^d are, respectively, the scaling function and the wavelet functions that prioritize the horizontal, vertical, and diagonal directions. We indicate with J is the number of resolutions used in the decomposition, w_j^d the wavelet (or detail) coefficients of scale j and direction d , and C_J a smooth version of the original signal S .

The undecimated isotropic wavelet transform

The undecimated isotropic wavelet transform (UIWT) decomposes an $n \times n$ image κ into :

$$\kappa(x, y) = C_J(x, y) + \sum_{j=1}^J w_j(x, y),$$

where C_J is a coarse or smooth version of the original image κ , and w_j represents the details in κ on scale 2^j (see [Starck and Murtagh 2006](#), for details).

The ridgelet transform

Classical multiresolution methods address only a portion of the entire range of interesting phenomena which is roughly isotropic one on all scales and at all locations. The ridgelet transform was proposed as an alternative to the wavelet representation of image data.

Given a function $f(x_1, x_2)$, the ridgelet transform is the superposition of elements of the form $a^{-1/2}\psi((x_1 \cos \theta + \sin \theta - b)/a)$, ψ is the wavelet, $a > 0$ the scale parameter, b the location parameter, and θ the orientation parameter. The ridgelet is constant along lines $x_1 \cos \theta + x_2 \sin \theta = \text{const}$, and, transverse to these ridges, it is a wavelet.

The curvelet transform

Ridgelets essentially focus on straight lines rather than curves. However they can be adapted to representing objects with curved edges using an appropriate multiscale localization. If one uses a sufficiently fine scale to capture curved edges, such edges are almost straight. As a consequence, the curvelet transform has been introduced, in which ridgelets are used in a localized manner.

The idea of the curvelet transform ([Candès and Donoho 1999](#); [Starck et al. 2003](#)) is to first decompose the image into a set of wavelet planes, then decompose each plane into several blocks (the block size can change on each scale level) before finally analyzing each block with a ridgelet transform. The finer the scale, the more sensitive to the curvature the analysis is. As a consequence, curved singularities can be well represented with very few coefficients.

Appendix D: The MRLens filtering

The MRLens filtering ([Starck et al. 2006b](#)) is a non-linear filtering based on Bayesian theory that searches for a solution maximizing the a posteriori probability. Choosing the prior is one of the most critical aspects of Bayesian analysis. The MRLens filtering uses a multiscale entropy prior.

Assuming Gaussian noise, the MRLens filtering solves the minimization

$$J(\kappa) = \frac{\|k_n - \kappa\|^2}{2\sigma_n^2} + \beta \sum_{j=1}^J \sum_{k,l} h_n((W\kappa)_{j,k,l}), \quad (19)$$

where σ_n is the noise standard deviation, J is the number of scales, β is the regularization parameter, and W is the wavelet transform operator. Full details of the minimization algorithm can be found in [Starck et al. \(2001\)](#), in addition to a description of how to determine the regularization parameter β automatically.

In [Starck et al. \(2006b\)](#), it was shown that the MRLens filtering outperforms the existing methods. The MRLens² filtering has already been used in several applications to weak lensing data, and was in particular selected to filter the dark-matter mass map obtained by the Hubble Space Telescope in the COSMOS field.

References

- Aghanim, N., & Forni, O. 1999, *A&A* 347, 409
 Bacon, D. J., Massey, R. J., Refregier, A. R., et al. 2003, *MNRAS*, 344, 673
 Bahcall, N. A., & Fan, X. 1998, *ApJ*, 504, 1
 Benabed, K., & Scoccimarro, R. 2006, *A&A*, 456, 421
 Benjamin, J., Heymans, C., Semboloni, E., et al. 2007, *MNRAS*, 381, 702
 Benjamini, Y., & Hochberg, Y. 1995, *J. R. Stat. Soc. B.*, 57, 289
 Bernardeau, F., van Waerbeke, L., & Mellier, Y. 1997, *A&A*, 322, 1
 Bernardeau, F., van Waerbeke, L., & Mellier, Y. 2003, *A&A*, 397, 405
 Candès, E., & Donoho, D. 1999, in *Curve and Surface Fitting: Saint-Malo*, ed. A. Cohen, C. Rabut, & L. Schumaker (Nashville, TN: Vanderbilt University Press)
 Cooray, A., & Hu, W. 2001, *ApJ*, 548, 7
 Dahle, H. 2006, *Astrophys. J.*
 Donoho, D., & Jin, J. 2004, *Annals of Statistics*, 32, 962
 Fu, L., Semboloni, E., Hoekstra, H., et al. 2008, *A&A*, 479, 9
 Hamana, T., Takada, M., & Yoshida, N. 2004, *MNRAS*, 350, 893
 Hoekstra, H., Yee, H. K. C., & Gladders, M. D. 2002, *ApJ*, 577, 595
 Hoekstra, H., Mellier, Y., van Waerbeke, L., et al. 2006, *ApJ*, 647, 116
 Hopkins, A. M., Miller, C. J., Connolly, A. J., et al. 2002, *Astron. J.*, 123, 1086
 Jarvis, M., Bernstein, G., & Jain, B. 2004, *MNRAS*, 352, 338
 Jin, J., Starck, J.-L., Donoho, D. L., Aghanim, N., & Forni, O. 2005, *EURASIP Journal on Applied Signal Processing*, 2005, 15, 2470, 15, 2470
 Kilbinger, M., & Schneider, P. 2005, *A&A*, 442, 69
 Maoli, R., Van Waerbeke, L., Mellier, Y., et al. 2001, *A&A*, 368, 766
 Massey, R., Rhodes, J., Refregier, A., et al. 2004, *AJ*, 127, 3089
 Massey, R., Refregier, A., Bacon, D. J., Ellis, R., & Brown, M. L. 2005, *MNRAS*, 359, 1277
 Massey, R., Rhodes, J., Ellis, R., et al. 2007, *Nature*, 445, 286
 Miller, C. J., Genovese, C., Nichol, R. C., et al. 2001, *Astron. J.*, 122, 3492
 Pen, U.-L., Lu, T., van Waerbeke, L., et al. 2003, *MNRAS*, 346, 994
 Pires, S., Juin, J. B., Yvon, D., et al. 2006, *A&A*, 455, 741
 Pires, S., Starck, J., Amara, A., et al. 2008, *MNRAS*, Submitted
 Press, W. H., & Schechter, P. 1974, *ApJ*, 187, 425
 Refregier, A., Rhodes, J., & Groth, E. J. 2002, *API*, 572, L131
 Schneider, P. 1996, *MNRAS*, 283, 837
 Schneider, P., & Lombardi, M. 2003, *A&A*, 397, 809
 Schneider, P., van Waerbeke, L., Jain, B., et al. 1998, *MNRAS*, 296, 873
 Schneider, P., Kilbinger, M., & Lombardi, M. 2005, *A&A*, 431, 9
 Semboloni, E., Mellier, Y., van Waerbeke, L., et al. 2006, *A&A*, 452, 51
 Sheth, R. K., & Tormen, G. 1999, *MNRAS*, 308, 119
 Starck, J.-L., & Murtagh, F. 2006, *Astronomical Image and Data Analysis, Astronomical image and data analysis*, ed. J.-L. Starck, & F. Murtagh, *Astron. Astrophys. Library* (Berlin: Springer)
 Starck, J.-L., Murtagh, F., Querre, P., et al. 2001, *A&A*, 368, 730
 Starck, J.-L., Donoho, D. L., & Candès, E. J. 2003, *A&A*, 398, 785
 Starck, J.-L., Aghanim, N., & Forni, O. 2004, *A&A*, 416, 9
 Starck, J.-L., Abrial, P., Moudou, Y., et al. 2006a, *A&A*, 446, 1191
 Starck, J.-L., Pires, S., & Réfrégier, A. 2006b, *A&A*, 451, 1139
 Takada, M., & Bridle, S. 2007, *New J. Phys.*, 9, 446
 Takada, M., & Jain, B. 2003, *MNRAS*, 344, 857
 Takada, M., & Jain, B. 2004, *MNRAS*, 348, 897
 Teyssier, R. 2002, *A&A*, 385, 337
 Vale, C., & White, M. 2003, *ApJ*, 592, 699
 Van Waerbeke, L., Mellier, Y., Pelló, R., et al. 2002, *A&A*, 393, 369
 Vielva, P., Martínez-González, E., Barreiro, R. B., Sanz, J. L., & Cayón, L. 2004, *Astrophys. J.*, 609, 22
 Vielva, P., Wiaux, Y., Martínez-González, E., et al. 2006, *New Astron. Rev.*, 50, 880
 Wiaux, Y., Vielva, P., Barreiro, R. B., Martínez-González, E., & Vanderghyest, P. 2008, *MNRAS*, 385, 939

² The complete MRLens software package to perform weak lensing filtering can be downloaded from <http://www-irfu.cea.fr/Ast/878.html>