

Increasing the lensing figure of merit through higher order convergence moments

Martina Vicinanza,^{1,2,3,4,*} Vincenzo F. Cardone,^{3,†} Roberto Maoli,^{2,‡} Roberto Scaramella,^{3,5,§} and Xinzhong Er^{3,||}

¹*Dipartimento di Fisica, Università di Roma “Tor Vergata”,
via della Ricerca Scientifica 1, 00133 Roma, Italy*

²*Dipartimento di Fisica, Università di Roma “La Sapienza”, Piazzale Aldo Moro, 00185 Roma, Italy*

³*I.N.A.F.-Osservatorio Astronomico di Roma, via Frascati 33, 00040 Monte Porzio Catone (Roma), Italy*

⁴*Instituto de Astrofísica e Ciências do Espaço, Universidade de Lisboa, Faculdade de Ciências,
Campo Grande, PT1749-016 Lisbon, Portugal*

⁵*I.N.F.N. - Sezione di Roma 1, Piazzale Aldo Moro, 00185 - Roma, Italy*



(Received 24 May 2017; published 22 January 2018)

The unprecedented quality, the increased data set, and the wide area of ongoing and near future weak lensing surveys allows one to move beyond the standard two points statistics, thus making it worthwhile to investigate higher order probes. As an interesting step toward this direction, we explore the use of higher order moments (HOM) of the convergence field as a way to increase the lensing figure of merit (FoM). To this end, we rely on simulated convergence to first show that HOM can be measured and calibrated so that it is indeed possible to predict them for a given cosmological model provided suitable nuisance parameters are introduced and then marginalized over. We then forecast the accuracy on cosmological parameters from the use of HOM alone and in combination with standard shear power spectra tomography. It turns out that HOM allow one to break some common degeneracies, thus significantly boosting the overall FoM. We also qualitatively discuss possible systematics and how they can be dealt with.

DOI: [10.1103/PhysRevD.97.023519](https://doi.org/10.1103/PhysRevD.97.023519)

I. INTRODUCTION

That the Universe is spatially flat and undergoing a phase of accelerated expansion is nowadays a well-known fact confirmed by an overwhelming flood of hard to confute evidence (see, e.g., [1] and references therein for a comprehensive although not updated review). What is responsible for this speed-up is, on the contrary, still frustratingly unknown notwithstanding the considerable numbers of papers proposing theoretical solutions to this problem. On the one hand, the concordance Λ CDM model [2], composed by a cold dark matter (CDM) component causing clustering and a cosmological constant Λ causing acceleration, makes excellent work of fitting the data on cosmological scales [3–7], but it is plagued by both theoretical shortcomings and unsatisfactory agreement with data on galactic scales [8]. On the other hand, dynamical models, collectively referred to as dark energy [9–11], including, e.g., an evolving scalar field rolling down its self-interaction potential, perform well at reproducing the observed data, but lack a solid foundation somewhat appearing as a way to shift the problem from what

is driving speed up to what is sourcing the field and its evolution. A radically different approach is possible if one takes cosmic speed-up as a first signature of a failure in our understanding of the laws of gravity. Rather than being evidence for something missing in the cosmic pie, acceleration then becomes a way nature is pointing us to a more general formulation of the theory of gravity with general relativity. This opens up the way to a full set of modified gravity theories [12,13] able to closely follow the same background evolution of the Λ CDM model, hence being in agreement with the same data.

Discriminating among these two rival approaches is the aim of modern cosmology with ongoing and planned surveys trying to achieve this goal by both reducing statistical and systematic errors (hence constraining the dark energy equation of state) and looking for observational probes able to test the growth of perturbations that evolves in different ways in the two competing scenarios. Thanks to the possibility offered to probe both the kinematics and the dynamics of the Universe, gravitational lensing has long been considered as an ideal tool to investigate the nature and the nurture of dark energy and to look for signatures of modified gravity. It is therefore not surprising that cosmic shear tomography has been recommended as the most promising technique to solve the problem of cosmic acceleration. After the pioneering small area surveys such as the CFHTLenS [14] and

*mvicinanza@fc.ul.pt

†winnyenodrac@gmail.com

‡roberto.maoli@roma1.infn.it

§roberto.scaramella@oa-roma.inaf.it

||er.xinzhong@oa-roma.inaf.it

DeepLens [15] programs that have shown the feasibility of the proposal, large area surveys have started to fully exploit the lensing potentialities with the first interesting results already coming from surveys in progress, such as DES [16], KIDS [17], and HSC [18]. Near future surveys, both ground (LSST [19]) and space based (Euclid [20], WFIRST [21]), will make it possible to make the error budget systematics dominated so that, provided these latter are taken under control, the hunt to the responsible of cosmic speed-up could likely come to a successful conclusion.

From a practical point of view, all the above quoted surveys aim at mapping cosmic shear from the ellipticity of large galaxy samples. Shear catalogs are then used to determine the two points correlation function splitting sources in redshift bins to perform tomography. This is then contrasted against theoretical expectations, thus constraining the cosmological parameters (see, e.g., [22]). Alternatively, one can rely on a three-dimensional (3D) analysis based on spherical Fourier-Bessel decomposition to constrain both dark energy models [23] and modified gravity theories [24]. It is worth noting that both these techniques mainly probe the growth of structures in the linear and quasilinear regimes so that the constraints on cosmological parameters are affected by severe degeneracies such as the well-known Ω_M - σ_8 one. To break this degeneracy, a possible way out consists in moving to a different tracer. Indeed, cosmic shear measurements can also be used to reconstruct the convergence field that quantifies the projection along the line of sight of the density contrast $\delta(\mathbf{x}) = [\rho(\mathbf{x}) - \bar{\rho}]/\bar{\rho}$. What makes convergence so attractive is the different scales, which is sensible, too. Indeed, the convergence field mainly probes the nonlinear scales [25–30], thus offering complementary information to the shear field. Moreover, on these scales, the collapse of structures introduce deviations from the Gaussianity which are strongly related to the underlying cosmological model and theory of gravity. The non-Gaussianity of the field can be quantified through higher than second order moments, which can then be used to break the Ω_M - σ_8 degeneracy [31,32]. Actually, higher order moments depend on the full set of cosmological parameters so that it is worth investigating whether they can represent a valuable help to narrow down the uncertainties on the dark energy equation of state, too. In other words, one can wonder whether convergence moments can help to increase the figure of merit (FoM) if used alone and/or in combination with the standard second order cosmic shear tomography.

The plan of the paper is as follows. In Sec. II, we derive a phenomenological yet well motivated relation connecting the theoretical moments with the observed ones taking care of issues related to noise and map reconstruction. How to predict moments from a given cosmological model is detailed in Sec. III, while Sec. IV presents the simulated maps we use to show that moments can be measured from wide enough surveys. Armed with these mock data, we can then validate the relation between theoretical and observed

moments that is carried out in Sec. V, thus making us confident that moments can indeed be used as cosmological probes. What can be gained by such a probe is investigated in Sec. VI through a Fisher matrix based forecast analysis. A discussion of possible systematics and how they could impact the results and be taken into account is presented in Sec. VII, while Sect. VIII is devoted to conclusions. Some further material is given in Appendixes A and B for the interested reader.

II. CONVERGENCE MOMENTS

The lensing effect of the large scale structure is responsible for the generation of the convergence field κ . However, we do not actually measure the convergence κ , but rather the ellipticity of the galaxies that are then used as input to reconstruct the shear field γ . A suitable algorithm is then used to get the convergence map. Provided the code is correctly working, one can relate the input field κ to the observed one κ_{obs} through the following linear relation:

$$\kappa_{\text{obs}} = (1 + m)\kappa + c + \mathcal{N}, \quad (1)$$

where (m, c) are the multiplicative and additive biases and \mathcal{N} is a noise term. Since both the signal and the noise vanish when averaged over a sufficiently large area, i.e., $\langle \kappa \rangle = \langle \mathcal{N} \rangle = 0$, one usually deals with $\tilde{\kappa}_{\text{obs}} = \kappa_{\text{obs}} - \langle \kappa_{\text{obs}} \rangle$ so that $\langle \tilde{\kappa}_{\text{obs}} \rangle = 0$. From Eq. (1), one naively gets

$$\tilde{\kappa}_{\text{obs}} = (1 + m)\kappa + \mathcal{N}, \quad (2)$$

which is the starting point for the derivation of the moments. Hereafter, to shorten the notation, we will drop the tilde sign from $\tilde{\kappa}_{\text{obs}}$ and denote with κ_{obs} the observed convergence after subtracting the mean value.

Let us now compute powers of κ_{obs} using Eq. (2) to get

$$\kappa_{\text{obs}}^2 = (1 + m)^2 \kappa^2 + 2(1 + m)\kappa\mathcal{N} + \mathcal{N}^2,$$

$$\kappa_{\text{obs}}^3 = (1 + m)^3 \kappa^3 + 3(1 + m)^2 \kappa^2 \mathcal{N} + 3(1 + m)\kappa \mathcal{N}^2 + \mathcal{N}^3,$$

$$\begin{aligned} \kappa_{\text{obs}}^4 &= (1 + m)^4 \kappa^4 + 4(1 + m)^3 \kappa^3 \mathcal{N} + 6(1 + m)^2 \kappa^2 \mathcal{N}^2 \\ &\quad + 4(1 + m)\kappa \mathcal{N}^3 + \mathcal{N}^4. \end{aligned}$$

Since κ and \mathcal{N} are uncorrelated, the expectation values of products such as $\kappa^n \mathcal{N}^m$ will simply be the product of the corresponding expectation values. Since both κ and \mathcal{N} have zero mean, it is then easy to get the following expressions for the moments of the observed convergence field:

$$\langle \kappa_{\text{obs}}^2 \rangle = (1 + m)^2 \langle \kappa^2 \rangle + \langle \mathcal{N}^2 \rangle,$$

$$\langle \kappa_{\text{obs}}^3 \rangle = (1 + m)^3 \langle \kappa^3 \rangle + \langle \mathcal{N}^3 \rangle,$$

$$\langle \kappa_{\text{obs}}^4 \rangle = (1 + m)^4 \langle \kappa^4 \rangle + 6(1 + m)^2 \langle \kappa^2 \rangle \langle \mathcal{N}^2 \rangle + \langle \mathcal{N}^4 \rangle.$$

The moments of the input convergence field κ should be theoretically predicted for a given cosmological model. However, since theory is imperfect (as we will see later), one can postulate the following ansatz:

$$\langle \kappa^n \rangle = (1 + \mu_n) \langle \kappa^n \rangle_{\text{th}} + \gamma_n, \quad (3)$$

where $\langle \kappa^n \rangle_{\text{th}}$ is the moment of order n as predicted from the theory, and (μ_n, γ_n) are calibration parameters to match theory with numerical simulations. Inserting these relations into the above ones, we finally end up with the following calibration relations:

$$\langle \kappa_{\text{obs}}^2 \rangle = (1 + m_2) \langle \kappa^2 \rangle + c_2 + \langle \mathcal{N}^2 \rangle, \quad (4)$$

$$\langle \kappa_{\text{obs}}^3 \rangle = (1 + m_3) \langle \kappa^3 \rangle + c_3 + \langle \mathcal{N}^3 \rangle, \quad (5)$$

$$\begin{aligned} \langle \kappa_{\text{obs}}^4 \rangle &= (1 + m_4) \langle \kappa^4 \rangle + c_4 + \langle \mathcal{N}^4 \rangle \\ &+ 6[(1 + m_2) \langle \kappa^2 \rangle + c_2] \langle \mathcal{N}^2 \rangle, \end{aligned} \quad (6)$$

where we have dropped the label ‘‘th’’ to shorten the notation and defined

$$1 + m_n = (1 + m)^n (1 + \mu_n), \quad c_n = (1 + m)^n \gamma_n.$$

It is worth noting how, although the calibration parameters are in total seven, namely $(\mu_2, \gamma_2, \mu_3, \gamma_3, \mu_4, \gamma_4, m)$, they are actually degenerate since they enter Eqs. (4)–(6) only through their combinations defining the six nuisance parameters $(m_2, c_2, m_3, c_3, m_4, c_4)$. One could naively think that the noise moments $(\langle \mathcal{N}^2 \rangle, \langle \mathcal{N}^3 \rangle, \langle \mathcal{N}^4 \rangle)$ are fully degenerate with (c_2, c_3, c_4) since both enter as additive terms. However, that this is not the case can be understood, noting that what moments of the observed convergence field are not directly measured on the map as it is, but only after smoothing it with a filter of aperture θ . As such, in the above relations, while (m_n, c_n) are constants, both the observed and theoretical moments $(\langle \kappa_{\text{obs}}^n \rangle, \langle \kappa^n \rangle)$ and the noise moments $\langle \mathcal{N}^n \rangle$ must be considered as functions of the smoothing angle θ . While the dependence of $\langle \kappa^n \rangle$ on θ will be derived later from a theoretical motivated approach, we can postulate here the following scaling for the noise moments:

$$\langle \mathcal{N}^n \rangle = \nu_n (\theta / \theta_0)^{-n} \quad (7)$$

with θ_0 an arbitrary reference scale (which we will set to 1 arcsec) and ν_n the value of the noise moment of order n on the smoothed map. Note that Eq. (7) is correct for a white noise, but can be taken as an intuitive general result no matter the noise power spectrum. In the particular case of a Gaussian noise, it is

$$\nu_3 = 0, \quad \nu_4 = 3\nu_2^2$$

so that the number of unknown noise quantities reduces to one. However, as a conservative choice, we will not rely on this assumption in the following. We therefore end up with a total of nine nuisance parameters to match the theoretically expected moments to the observed one as measured on a convergence map reconstructed from noisy shear data and smoothed with a filter of aperture θ .

III. MOMENTS FROM THEORY

Two main actors enter the scene of Eqs. (4)–(6). We investigate here how one of them can be estimated for a given cosmological model. To this end, we first remind the reader that the weak lensing convergence field is the result of inhomogeneities in the large scale matter distribution along the line of sight to a distant source. The two-dimensional (2D) convergence field κ is then related to the 3D density contrast δ as

$$\kappa = \int_0^{\chi_s} d\chi W(\chi) \delta(\chi) = \frac{c}{H_0} \int_0^{z_h} \frac{W(z) \delta(z)}{E(z)} dz, \quad (8)$$

where

$$\chi(z) = \frac{c}{H_0} \int_0^z \frac{dz'}{E(z')} \quad (9)$$

is the comoving distance to a source at redshift z (having assumed a spatially flat universe), and $E(z) = H(z)/H_0$ with $H(z)$ the cosmology dependent Hubble parameter and a subscript 0 denotes present day values. In Eq. (8), the integral extends up to the redshift of the last scattering surface z_h to take into account the cumulative effect of the full matter distribution. However, the sum is actually weighted by the lensing kernel $W(z)$, which takes into account the normalized source redshift distribution $n(z)$ through

$$W(z) = \frac{3\Omega_M H_0^2}{2c^2} (1+z) \chi(z) \int_z^{z_h} \frac{\chi(z') - \chi(z)}{\chi(z')} n(z') dz'. \quad (10)$$

Using a Fourier decomposition and the Limber flat sky approximation, it is then possible to compute the two points projected correlation function (see, e.g., [33] and references therein). Its average smoothed over an angle θ with a filter $\mathcal{W}(\theta)$ then gives the second order moment of the convergence field. A similar procedure can then be used to infer higher order moments starting from multipoints correlation functions and their corresponding multispectra. Under the assumption of validity of the hierarchical ansatz (which is the case in a highly nonlinear regime and in the quasilinear one in the limit of vanishing variance), all higher order multispectra can be written as weighted sums of the products of the matter power spectrum. It is then

possible to write the moments up to the fourth order of the convergence field as [33]

$$\langle \kappa^2 \rangle(\theta) = \mathcal{C}_2(\kappa_\theta), \quad (11)$$

$$\langle \kappa^3 \rangle(\theta) = 3\mathcal{Q}_3\mathcal{C}_3(\kappa_\theta^2), \quad (12)$$

$$\langle \kappa^4 \rangle(\theta) = (12\mathcal{R}_a + 4\mathcal{R}_b)\mathcal{C}_4(\kappa_\theta^3), \quad (13)$$

where we have defined

$$\kappa_\theta = 2\pi \int P[\ell/\chi(z), z] \mathcal{W}(\ell\theta_0) \ell d\ell, \quad (14)$$

$$\mathcal{C}_i(\kappa_\theta^n) = \frac{c}{H_0} \int_0^{z_h} \frac{W^i(z) \kappa_\theta^n(z)}{\chi^{2(i-1)}(z) E(z)} dz. \quad (15)$$

In Eq. (14), $P(k, z)$ is the matter power spectrum evaluated in $k = \ell/\chi(z)$ because of the flat sky approximation, while the coefficients ($\mathcal{Q}_3, \mathcal{R}_a, \mathcal{R}_b$) depend on the amplitude of the different topologies contributing to the multipoint correlation functions. Their values should be tailored against suitably designed N -body simulations that are not always available. We will therefore set their values to the following fiducial values:

$$\mathcal{Q}_3^{\text{fid}} = 1.00, \quad \mathcal{R}_a^{\text{fid}} = 7.29, \quad \mathcal{R}_b^{\text{fid}} = 16.23.$$

Equations (11)–(13) can be used as input in the calibration relations (4)–(6). However, these have been derived postulating that the actual theoretical quantities are linearly related to the ones we have derived here as in Eq. (3). Ideally, the theory summarized here should lead to a perfect match with the (unknown) true moments so that $\mu_n = \gamma_n = 0$ for all n . However, there are different arguments advising one to be conservative and allowing for (μ_n, γ_n) to deviate from null values. We sketch them qualitatively below.

- (i) Equations (12) and (13) define the third and fourth order moments up to a multiplicative term given by \mathcal{Q}_3 and $\mathcal{Q}_4 = 12\mathcal{R}_a + 4\mathcal{R}_b$. We have set these quantities to fiducial values, but they should actually be set matching multipoint correlation functions to those measured from high resolution N -body simulations. We can easily parametrize the systematic uncertainty on them through the multiplicative bias terms

$$1 + \mu_3 = \mathcal{Q}_3/\mathcal{Q}_3^{\text{fid}}, \quad 1 + \mu_4 = \mathcal{Q}_4/\mathcal{Q}_4^{\text{fid}}.$$

Note that, according to this argument, we should set $\mu_2 = 0$. Although we do not force this constraint for reasons explained below, we nevertheless note that we will indeed find that μ_2 is quite small. On the contrary, nothing prevents (μ_3, μ_4) from significantly deviating from the null value.

- (ii) The convergence moments are estimated from functionals of the smoothed mean convergence κ_θ defined in Eq. (14). Here, one should integrate over all the scales k , but what one actually does is to truncate the integration range to avoid extrapolating the nonlinear model for the matter power spectrum to scales where it is not tested. Although one expects that the scales outside the integration range give a little contribution, this term can be boosted by the functional defining $\langle \kappa^n \rangle$. It is therefore worth allowing for this possibility to introduce the additive bias term γ_n .

- (iii) Notwithstanding the great efforts in modeling nonlinearities in the matter power spectrum, the matching between the theoretical one $P(k, z)$ and that estimated from N -body simulations is still not perfect on all scales and redshift of interest. If we denote with $P_N(k, z)$ this latter and take that this is the best representation of the unknown actual one, we can assume that the following relation holds true:

$$P_N(k, z) = (1 + \pi_1)P(k, z) + \pi_0 P(k_0, z_0),$$

where (π_0, π_1) are constant quantities, $P(k_0, z_0)$ is the value of the power spectrum in an arbitrary set reference point (k_0, z_0) , and the linear approximation is motivated by the assumption that the deviations are small enough. It is then only a matter of algebra to show that the moments evaluated using $P_N(k, z)$ and those from $P(k, z)$ are related by a linear relation as far as one neglects higher order terms in (π_0, π_1) .

These considerations provide a valid support to Eq. (3) and hence to the calibration relations derived from it. We will nevertheless validate the linear relation (3) later through a comparison with moments evaluated on a convergence map inferred from numerical simulations.

IV. MOMENTS FROM SIMULATED MAPS

To validate the theoretical approach worked out above and find the calibration parameters (m_n, c_n) , we need to contrast them against moments measured on a convergence map reconstructed by shear noisy data. Although such a kind of data set is starting to be available (see, e.g., [34] from CFHTLenS, and [35] from DES Science Verification data), the survey area is still too small to allow for reducing the statistical uncertainty to an acceptable level and prevent cosmic variance effects. Moreover, noiseless convergence maps are also needed in order to validate Eq. (3), which can obviously not be inferred from actual data. We are therefore forced to rely on a simulated data set.

A. The MICE lensing catalog

N -body light cone simulations are an ideal tool to build all-sky lensing maps. We rely, in particular, on the MICE

Grand Challenge (MICE-GC) containing about 70×10^9 dark matter particles in a $(3h^{-1} \text{ Gpc})^3$ volume [36,37]. The parent simulation is run in a flat Λ CDM with cosmological parameters set

$$(\Omega_M, \Omega_b, h, n_s, \sigma_8) = (0.25, 0.044, 0.70, 0.95, 0.80),$$

with $(\Omega_M, \Omega_b, h, n_s, \sigma_8)$ the matter and baryon density parameters, the Hubble constant in units of 100 km/s/Mpc, the spectral index and the variance of perturbations on the scale $R = 8h^{-1}$ Mpc, respectively. Galaxies are associated with dark matter haloes using a halo occupation distribution and a halo abundance matching technique [38] whose parameters are set to match local observational constraints, such as the local luminosity function [39,40], the galaxy clustering as a function of luminosity, and color [41] and color-magnitude diagrams [42].

Given its large volume and fine mass resolution, the MICE-GC simulation also allows an accurate modeling of the lensing observables from upcoming wide and deep galaxy surveys. Following the *onion universe* approach [43], all-sky lensing maps are constructed with subarcminute scale resolution. These lensing maps allow one to model galaxy lensing properties, such as the convergence, shear, and lensed magnitudes and positions. Tests have been performed to show that the galaxy lensing mocks can be used to accurately model lensing observables down to arcminute scales. We use the MICECAT v2.0 (kindly available to us from P. Fosalba), which updates the public release MICECAT v1.0 catalog to include less massive and hence lower luminosity galaxies. The galaxy catalog is complete up to $i \sim 24$ for $0 \leq z < 1.4$ so that we will hereafter focus our attention on this redshift range only.

B. Convergence maps and moments estimate

To build the convergence and shear maps we need for validation and calibration, we first cut $5 \times 5 \text{ deg}^2$ approximately square patches taking care that they are well separated so that they can be considered independent. This reduces the usable area to $\Omega = 3500 \text{ deg}^2$ divided into 140 subfields. Each subfield has an area large enough to ensure good statistics, but small enough to make deviations from flat sky approximation (used in the theoretical derivation) negligible. The MICECAT v2.0 catalog reports the redshift z and the Right Ascension and Declination coordinates (hereafter RA, Dec) on the sky of each galaxy. We use the redshift z to select galaxies only in the redshift range (0.1,1.4), and use a sinusoidal projection to convert the (RA, Dec) coordinates defined on the curved sky into the Cartesian ones (x, y) defined on the tangential plane. Following [34], we then arrange galaxies in approximately square pixels with side length 0.85 arcmin. Given that the number density of galaxies is about $n_g \approx 27 \text{ gal/arcmin}^2$, this size ensures that there is a large number of objects in each pixel without compromising the map resolution.

We construct two kind of maps. First, we use the convergence values reported in the MICECAT v2.0 catalog without any further modification. This can be considered the actual convergence map that can be used to infer the actual moments the theory should match. We will therefore use it to validate the linear relation (3) and estimate the calibration parameters (μ_n, γ_n) . Note that, strictly speaking, we do not need (μ_n, γ_n) since they are fully degenerate with (m, c) so that only the combined parameters (m_n, c_n) finally enter the calibration relations (4)–(6). We therefore use this ideal convergence map only as a way to show that the theory reproduces the high order convergence moments up to a linear transformation.

Actually, one does not observe the convergence itself, but rather the shear components¹ (γ_1, γ_2) . These quantities are available in the MICECAT v2.0 catalog, but we perturb them by adding random values of the intrinsic ellipticity on both the shear components to mimic the noise due to intrinsic ellipticity. We then smooth the map and apply the KS93 method [44] to get what we can consider the observed convergence map κ_{obs} . Moments estimated from this map are those that enter the left-hand side of Eqs. (4)–(6). Note that the smoothing procedure, which is typically used to reduce noise, critically depends on which filter is adopted. To investigate the impact of this choice on the results, we consider two popular cases, namely a Gaussian and a top hat filter. In the Fourier space, their window functions respectively read

$$\tilde{W}(\ell, \theta) = \begin{cases} \exp(-\ell^2 \theta^2 / 2) \\ 2J_1(\ell\theta) / (\ell\theta) \end{cases}, \quad (16)$$

where J_ν is the Bessel function of order ν and θ the filter aperture that we vary over the range (2, 20) arcmin.

The catalogs thus obtained are used as input for the estimate of high order convergence moments, which are simply given by

$$\langle \kappa^n \rangle = \mathcal{N}_{\text{pix}}^{-1} \sum \tilde{\kappa}_i^n (n = 2, 3, 4),$$

where the sum is over the \mathcal{N}_{pix} pixels in the map and $\tilde{\kappa}_i$ is the value of the convergence in the i th pixel after smoothing the map with a given filter. Note that we first subtract a constant offset in order to have $\langle \kappa \rangle = 0$ to be consistent with our derivation of the calibration relations.

This procedure is repeated for each of the 140 subfields, thus getting a list of $\langle \kappa^n \rangle$ values. We finally estimate the moments and the covariance matrix as

$$\langle \kappa^n \rangle_{\text{obs}} = \frac{\sum \langle \kappa^n \rangle(k)}{\mathcal{N}_f} (n = 2, 3, 4), \quad (17)$$

¹Actually, what is observed are the two components (e_1, e_2) of the ellipticity, but with no noise and intrinsic alignment there is no difference with the shear components.

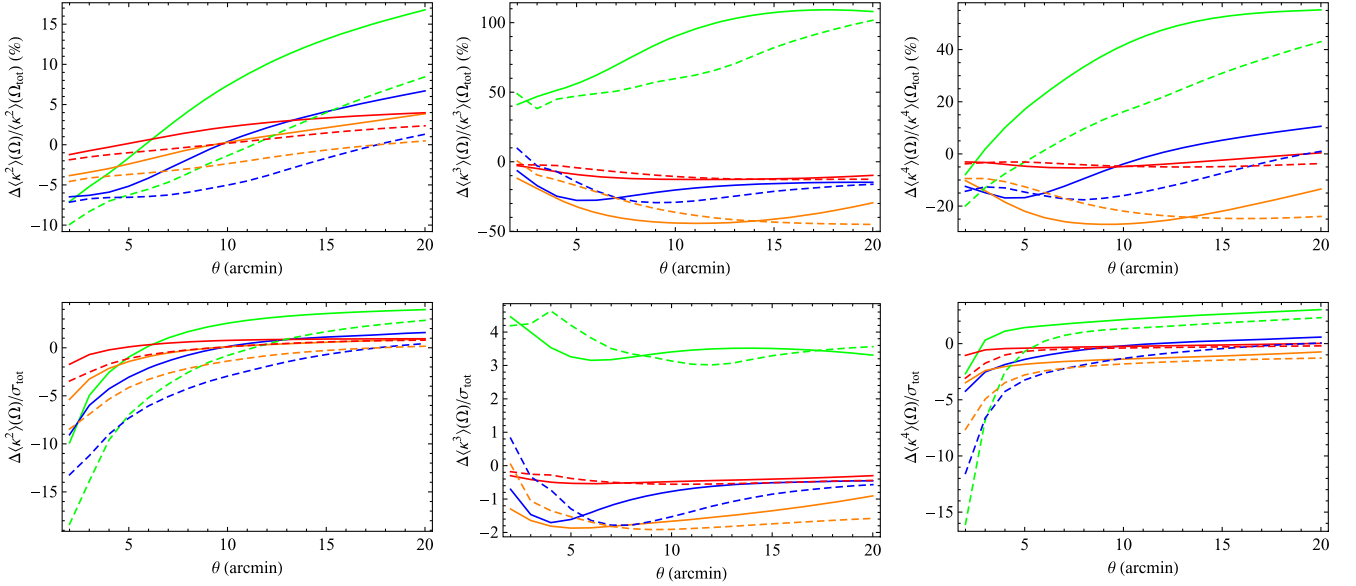


FIG. 1. Convergence of the HOM for different values of the survey area Ω . *Top*: Percentage deviation $\Delta\langle\kappa^n\rangle(\Omega)/\langle\kappa^n\rangle(\Omega_{\text{tot}})$ as a function of the smoothing angle θ for $\Omega = \{100, 500, 1500, 2500\}$ deg² (green, blue, orange, red lines) and $\Omega_{\text{tot}} = 3500$ deg². *Bottom*: Same as before but with the difference normalized with respect to the statistical error. Solid (dashed) lines refer to the results for the Gaussian (top hat) smoothing filter.

where $\langle\kappa^n\rangle(k)$ is the n th order moment estimated from the k th subfield, and the sum is over the \mathcal{N}_f subfields. For later applications, we remind the reader that it is $\mathcal{N}_f = \Omega/25$ with Ω in the used survey area ($\Omega = 3500$ deg²) giving $\mathcal{N}_f = 140$ for our reference MICECAT based case.

In practical applications, one will fit moments as a function of the smoothing scale θ so that the data vector is

$$\mathbf{D}_{\text{obs}} = \{ \langle\kappa^2\rangle_{\text{obs}}(\theta_1), \langle\kappa^2\rangle_{\text{obs}}(\theta_2), \dots, \langle\kappa^2\rangle_{\text{obs}}(\theta_N), \\ \langle\kappa^3\rangle_{\text{obs}}(\theta_1), \langle\kappa^3\rangle_{\text{obs}}(\theta_2), \dots, \langle\kappa^3\rangle_{\text{obs}}(\theta_N), \\ \langle\kappa^4\rangle_{\text{obs}}(\theta_1), \langle\kappa^4\rangle_{\text{obs}}(\theta_2), \dots, \langle\kappa^4\rangle_{\text{obs}}(\theta_N) \}, \quad (18)$$

where we will explore different choices for both the smoothing angle range ($\theta_{\text{min}}, \theta_{\text{max}}$) and the sampling $d\theta$ taking equispaced values giving $N = (\theta_{\text{max}} - \theta_{\text{min}})/d\theta$.

To estimate the data covariance matrix, one should run different simulations of the same survey, which is actually not possible. However, mimicking what is done with actual surveys (see, e.g., [34]) and in previous works based on simulated maps [45], we can use the \mathcal{N}_f subfields and estimate the covariance matrix elements as

$$\text{Cov}_{ij}^{\text{obs}} = \frac{1}{\mathcal{N}_f} \frac{\sum [D_{\text{obs}}^{i,k} - D_{\text{obs}}^i][D_{\text{obs}}^{j,k} - D_{\text{obs}}^j]}{\mathcal{N}_f - 1}, \quad (19)$$

where $D_{\text{obs}}^{i,k}$ is the i th component of the data vector \mathbf{D}_{obs} estimated on the convergence map of the k th subfield, D_{obs}^i is the value of the same component from the final vector (18), and the sum is over the \mathcal{N}_f subfields. It is worthwhile to

spend more time on Eq. (19). This is based on the implicit underlying assumption that all the subfields are a statistically faithful realization of the same underlying properties. As such, the covariance matrix for each subfield can be evaluated as given by the second multiplicative term in Eq. (19). This will give us the covariance for a field of area equal to the subfield one Ω_f so that a further $\Omega_f/\Omega = 1/\mathcal{N}_f$ factor is needed to scale the covariance to the full survey area Ω .

C. Convergence of moments

Higher order moments (HOM) probe the shape of the convergence probability distribution function (PDF). One can qualitatively expect that the higher the order, the better the PDF is characterized, but the larger must the statistics be to catch the details of its shape. This translates in a requirement on the survey area; i.e., one could wonder how large must the total area be so that HOM are reliably recovered from the convergence maps. To this end, we estimate HOM as a function of the survey area Ω which, in our case, actually means estimating moments using only a number of subfields $\mathcal{N}_f \leq 140$. Should the moments be reliably estimated, the functions² $\langle\kappa^n\rangle(\theta)$ must be independent of the survey area Ω as far as $\Omega \geq \Omega_{\text{min}}$ with Ω_{min} the final requirement on the survey area. How to set Ω_{min} actually depends on what one is interested in. To

²Hereafter, unless confusion is possible, we drop the label “obs” to shorten the notation. Moreover, we will set $(\theta_{\text{min}}, \theta_{\text{max}}, d\theta) = (2, 20, 1)$ arcmin. Unless otherwise stated, all the results are qualitatively valid also for other choices.

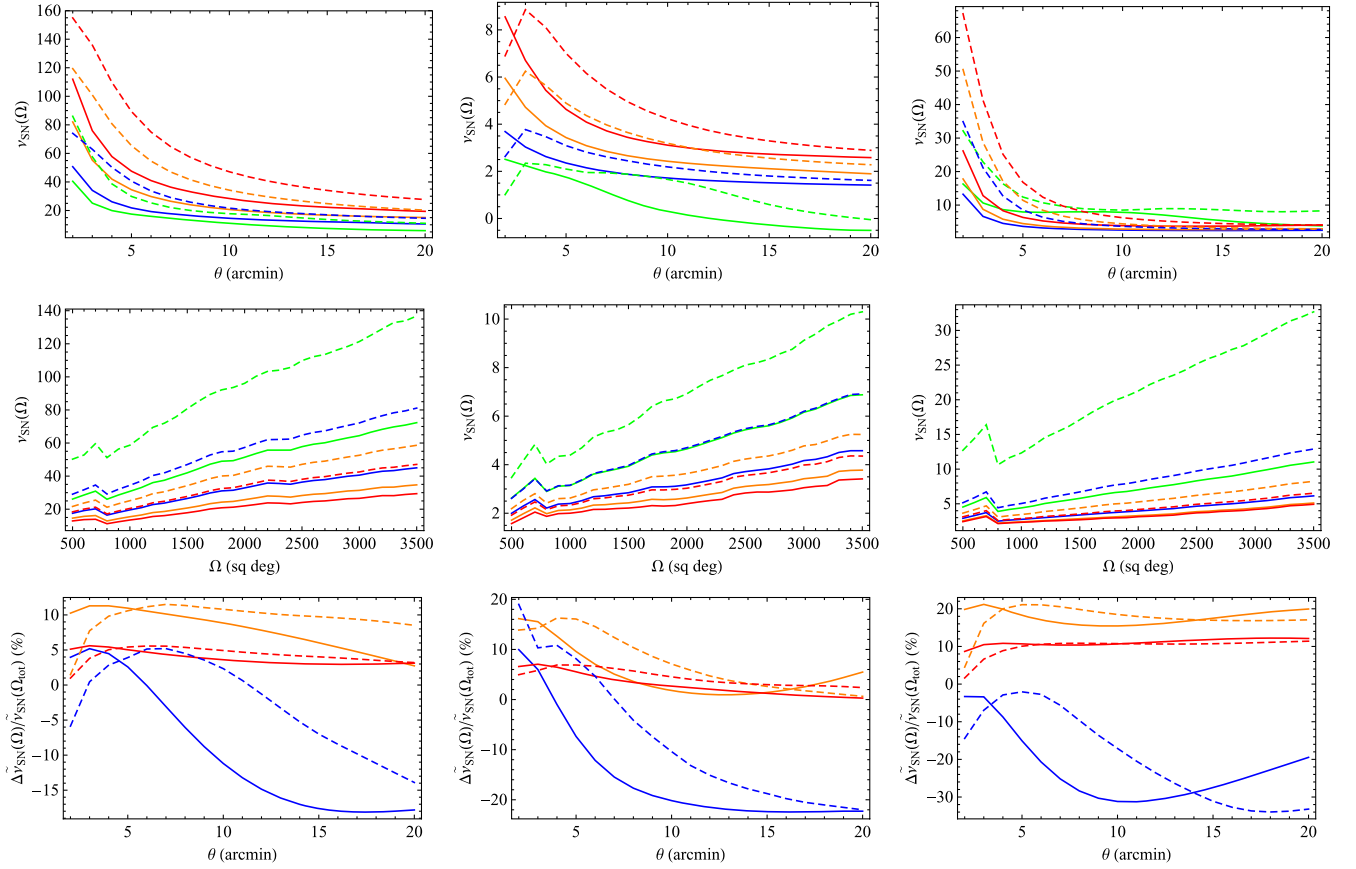


FIG. 2. Scaling of the HOM $\langle \kappa^2 \rangle$ (left), $\langle \kappa^3 \rangle$ (center), $\langle \kappa^4 \rangle$ (right) signal-to-noise (S/N) ratio ν_{SN} with smoothing angle θ and survey area Ω . *Top*: S/N ratio as a function of θ for $\Omega = \{100, 500, 1500, 2500\} \text{ deg}^2$ (green, blue, orange, red lines). *Center*: S/N ratio as a function of Ω for $\theta = \{4, 7, 10, 13\} \text{ arcmin}$ (green, blue, orange, red lines). *Bottom*: Percentage deviation of the rescaled S/N ratio (see text) from the value estimated from the full survey area. In all panels, solid (dashed) lines refer to the result for the Gaussian (top hat) smoothing filter.

understand this point, let us look at Fig. 1. Top panels show that the deviation of HOM from their fiducial value (i.e., the one estimated from the full area) can be quite significant with the difference being particularly dramatic for the third order moment (which probes the asymmetry of the convergence PDF). If we demand that $\Delta \langle \kappa^n \rangle / \langle \kappa^n \rangle$ is less than 10% for all orders, we get $\Omega_{\text{min}} \sim 2500 \text{ deg}^2$. However, this is a quite conservative limit that does not take into account the statistical errors. Indeed, one can relax the constraint asking that the systematic error on the HOM estimate is smaller than the statistical uncertainties σ . Bottom panels show that such a requirement is easier to fulfill depending on both the survey area (with smaller areas leading to larger σ hence less demanding constraints) and the smoothing scale θ . For instance, if a Gaussian filter is used to smooth the convergence maps, $\Omega_{\text{min}} = 500 \text{ deg}^2$ are enough to measure moments with a systematic bias smaller than the 2σ error as soon as $\theta > 10 \text{ arcmin}$. This area increases to $\Omega_{\text{min}} = 1500 \text{ deg}^2$ if one asks for all moments to be measured with a bias smaller than 2σ for smoothing scales $\theta > 5 \text{ arcmin}$. As a general result, we also find that the requirements on the minimum area are stronger for a top hat

filter, which turns out to be less efficient in canceling features introduced by the reconstruction from noisy data.

When fitting a model to the data, one should be confident that not only the HOM central values but also the errors have been estimated in order to avoid unrealistically tight constraints on model parameters because of underestimating the uncertainties. We therefore investigate how the signal-to-noise ratio ν_{SN} scales with the survey area using the square root of the diagonal elements³ of the data covariance matrix to define ν_{SN} . Top and central panels in Fig. 2 show the expected behavior of $\nu_{\text{SN}}(\theta)$ with both the smoothing scale and the survey area. On the one hand, the smaller

³Actually, one should take into account also the nondiagonal elements and define the noise taking care of the correlations between moments of the same order evaluated at different smoothing scales or different orders at the same smoothing. We prefer to rely on diagonal elements only to mimic the case when one is using a single order only. For instance, one could be interested in using only $\langle \kappa^2 \rangle(\theta_0)$ with θ_0 a particular smoothing scale. In this case, the S/N ratio would be defined using only the error on this quantity so that one should check how this scales with the survey area, which is what we are investigating here.

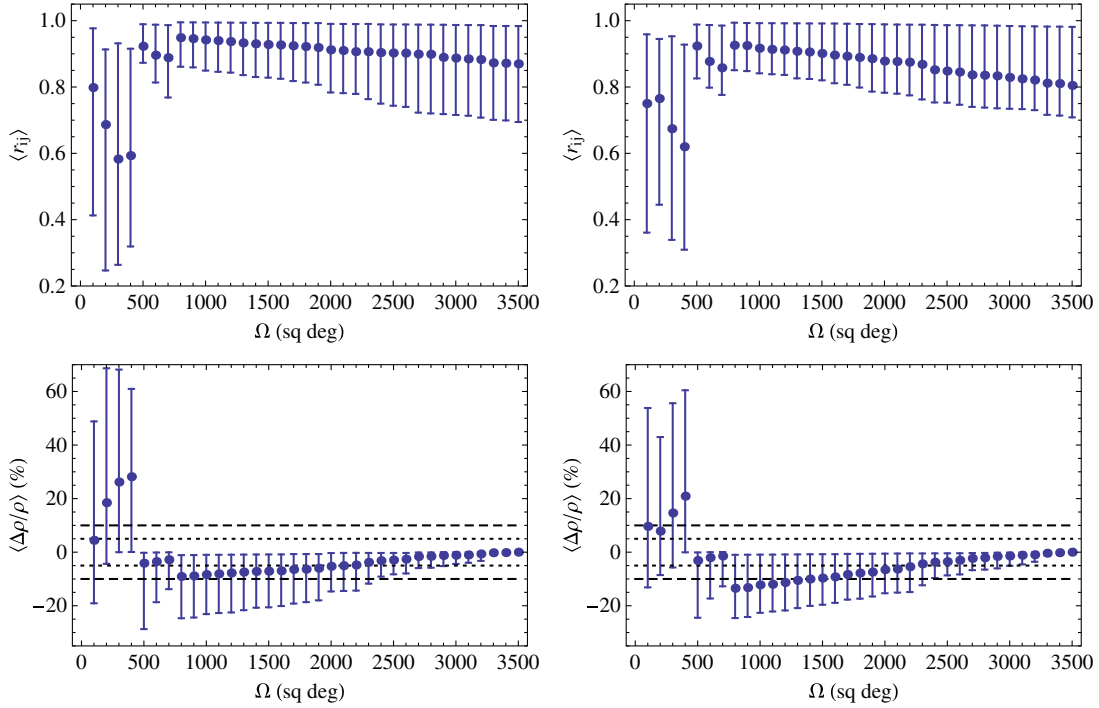


FIG. 3. Testing the convergence of the correlation matrix \mathbf{R} . *Top*: Average value and 68% confidence range of the independent element of \mathbf{R} for the Gaussian (left) and top hat (right) filters. *Bottom*: Average value and 68% confidence range of the percentage deviation of the \mathbf{R} elements from the value they get for $\Omega = 3500 \text{ deg}^2$. Dashed (dotted) lines mark the $\pm 10\% (\pm 5\%)$ limits.

the smoothing scales, the larger are the moments as one can naively understand considering the case of a Gaussian distribution. On the other hand, Eq. (19) suggests a scaling of the noise with $\sqrt{1/\Omega}$ so that we expect $\nu_{\text{SN}}(\theta) \propto \Omega^\alpha$ with $\alpha > 0$ and deviating from the naive $\alpha = 1/2$ value because the signal also depends on the area as inferred from Fig. 1. It is worth wondering which is the minimal area to make the S/N ratio estimate reliable. Note that for this test we have first to rescale the S/N for the different area, which is why we consider $\tilde{\nu}_{\text{SN}}(\theta) = \sqrt{\Omega}\nu_{\text{SN}}(\theta)$, and look for the minimal area needed to make this quantity deviate from the final value less than 20%. Bottom panels in Fig. 2 show a conservative cut $\Omega > 1500 \text{ deg}^2$ ensures reliable estimates of both HOM and their uncertainties.

It is worth noting that our definition of S/N, although quite intuitive, is somewhat optimistic since it neglects the strong correlation among the components of the data vector. We should therefore be confident that the correlation matrix \mathbf{R} too is stable. With its entries defined as

$$r_{ij} = \frac{\text{Cov}_{ij}^{\text{obs}}}{\sqrt{\text{Cov}_{ii}^{\text{obs}}\text{Cov}_{jj}^{\text{obs}}}},$$

there are $3N(3N+1)/2$ independent elements so that it is not immediate to understand whether the area is large enough to consider reliable the estimate of \mathbf{R} . As a qualitative test, we therefore look at $\langle r_{ij} \rangle$, i.e., the average value of its off diagonal elements which is plotted as a

function of the survey area Ω in the top panels of Fig. 3 where the error bars denote the 68% confidence range. Taken at face value, these results show that there is no convergence at all with $\langle r_{ij} \rangle$ slowly but constantly decreasing with Ω . However, the 68% confidence ranges well overlap for $\Omega \geq 1000 \text{ deg}^2$ suggesting that the variation can indeed be neglected. On the other hand, bottom panels show that $\langle \Delta\rho/\rho \rangle = \langle 100 \times [1 - r_{ij}(\Omega)/r_{ij}(\Omega_{\text{tot}})] \rangle$ becomes smaller than 10% as far as $\Omega \geq 1500 \text{ deg}^2$. Considering that this is only a qualitative test, we can conclude that an area as large as $\Omega \sim 1500 \text{ deg}^2$ allows a reliable estimate of the HOM, the on diagonal elements of the covariance matrix, and the full data correlation matrix.

As a final remark, we remind the reader that all the results in this section have been obtained relying on the MICECAT catalog, which has a definite redshift range and a galaxy number density $n_g \simeq 27 \text{ gal/arcmin}^2$. Should one or both of these characteristics be different, one might redo the analysis, which is, however, outside our aim here.

V. VALIDATING THE HOM CALIBRATION

To make moments a viable tool to constrain cosmological parameters, two preliminary steps are mandatory. First, one has to show that HOM can be reliably estimated from convergence maps reconstructed from noisy shear data. Second, one has to find a way to match theoretical predictions with observations through a phenomenological recipe taking care of what cannot be

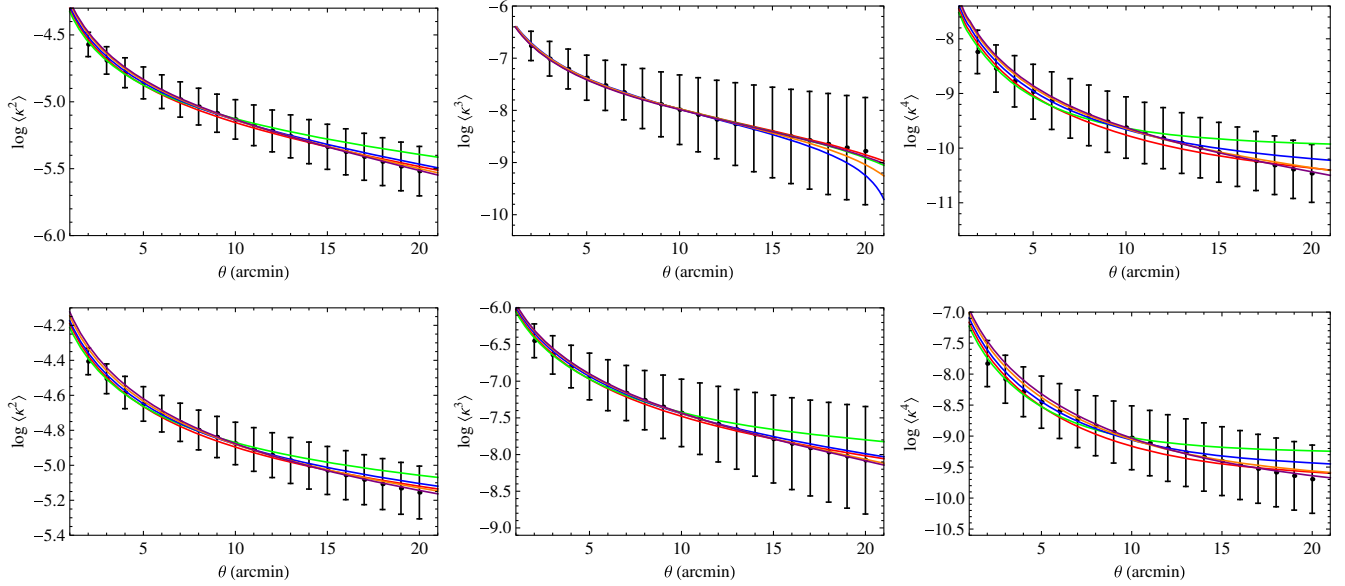


FIG. 4. HOM as estimated on the noiseless convergence map for Gaussian (top) and top hat (bottom) filters. Best fit curves are superimposed to the data with red, green, blue, orange, purple lines referring to the best fit parameters to the smoothing angle range (2,20), (2,12), (4,14), (6,16), (8,18) arcmin, respectively.

described within the theory. While the first step has been successfully completed in the above section, the second one is encapsulated in the calibration relations (4)–(6). Our aim here is to both validate their derivation and demonstrate that they can indeed fit the measured moments.

A. Idealized convergence moments

To this end, let us first consider the HOM as estimated from the convergence map provided by the MICECAT catalog itself. Since there is no noise and no reconstruction from shear is performed, HOM measured on this map may be considered the true ones that theory has to match. As explained in Sec. III, our theoretical predictions are given by Eq. (3) with $\langle \kappa^n \rangle_{\text{th}}$ computed as in Eqs. (11)–(13). Validating these relations therefore means finding the parameters (μ_n, γ_n) from a fit of the $\langle \kappa^n \rangle(\theta)$ data. The results are shown in Fig. 4 where we plot the best fit curves⁴ superimposed to the estimated HOM.⁵ As is apparent, the fit is quite good with the linear model passing through the points well within the statistical uncertainty. Moreover, the rms of the percentage residuals, defined as $\rho_{\text{rms}}(n) = 100 \times (1 - \langle \kappa^n \rangle_{\text{fit}} / \langle \kappa^n \rangle)$, turns out to be

$$\rho_{\text{rms}}(n) = \begin{cases} (4.88, 9.34, 21.5)\% \\ (4.55, 9.22, 21.2)\% \end{cases}$$

with the three values referring to moments of order $n = 2, 3, 4$, and the first (second) row for a Gaussian (top hat) filter. These small numbers definitely validate our assumption that the theoretical approach outlined in Sec. III is indeed able to reproduce the N -body derived HOM, thus providing a first validation of Eqs. (4)–(6).

Although the exact values of (μ_n, γ_n) are not of any use in the following, it is nevertheless instructive looking at the best fit values. For the Gaussian filter, we get

$$\begin{aligned} \mu_2 &= -0.002, & \gamma_2 &= -1.45 \times 10^{-6}, \\ \mu_3 &= 0.460, & \gamma_3 &= -2.82 \times 10^{-9}, \\ \mu_4 &= -0.191, & \gamma_4 &= 1.50 \times 10^{-11}. \end{aligned}$$

As expected, the multiplicative bias for the second order moment is negligibly small, while the larger values of (μ_3, μ_4) corrects for deviations of the (Q_3, Q_4) constant factors from the assumed fiducial values. On the other hand, the additive biases $(\gamma_2, \gamma_3, \gamma_4)$ are smaller than the lowest data values, confirming our expectation that cutting the integration range does not have a dramatic impact on the estimate of theoretical moments. Similar considerations also apply to the results for the top hat filter that read

$$\begin{aligned} \mu_2 &= -0.161, & \gamma_2 &= 0.74 \times 10^{-6}, \\ \mu_3 &= 0.298, & \gamma_3 &= -1.45 \times 10^{-9}, \\ \mu_4 &= -0.483, & \gamma_4 &= 1.72 \times 10^{-10}. \end{aligned}$$

⁴Best fit values and 68% confidence ranges from all the fits described in this section are given in Appendix A.

⁵Note that these plots and the following ones in Figs. 5 and 6 refer to the data as estimated from a single subfield so that the error bars are larger by a factor $\mathcal{N}_f^{1/2} \approx 12$. We make this choice in order to show that the fit to the full data set reproduces the data of each subfield too so that one can rely on the found calibration parameters even if only a smaller area is available.

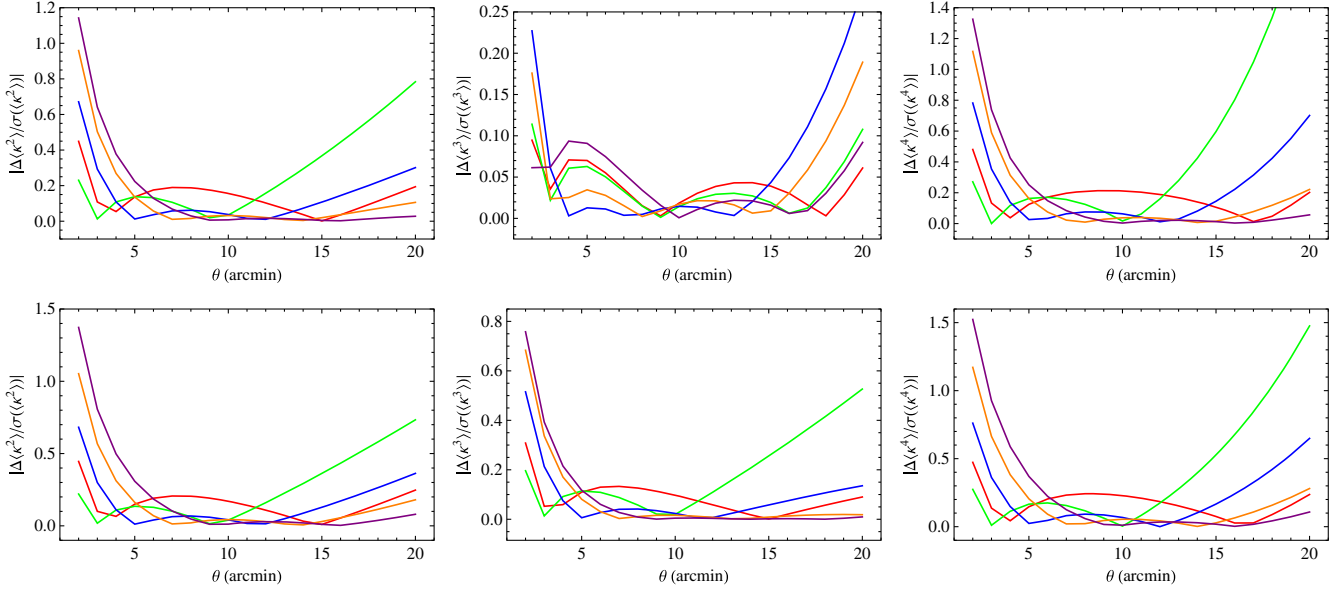


FIG. 5. Same as Fig. 4 but now we show the error normalized fit residuals (in absolute value).

The only remarkable difference is the multiplicative bias for the second order moment, which turns out to be definitely larger than for the Gaussian case suggesting that a top hat filter is less efficient in smoothing nonlinearities not correctly taken into account in the theoretical derivation.

Although the percentage residuals are already acceptably small, one can nevertheless wonder whether it is possible to further reduce them by cutting the smoothing angle range. Such an approach would be motivated by the consideration that the theory is inevitably based on assumptions that can break down for small θ , i.e., when small scale features in the convergence map are not smoothed out. We therefore repeat the fitting procedure limiting the smoothing angle range setting

$$(\theta_{\min}, \theta_{\max}) = (2, 12), (4, 14), (6, 16), (8, 18) \text{ arcmin},$$

thus getting the green, blue, orange, purple curves in Fig. 4. To better illustrate the results, in Fig. 5, we plot the error normalized best fit residuals (in absolute value) showing that the fit definitely improves; i.e., the green, blue, orange, purple curves stay below the red one over the corresponding fitting range. Moreover, the rms of the percentage residuals decreases up to an order of magnitude with the increase of θ_{\min} . This result can easily be explained noting that the larger the smoothing angle, the fewer nonlinearities contribute to the moments. Since these are the main sources of uncertainties in the theoretical modeling, it is not surprising that decreasing their impact makes the linear approximation better, thus reducing $\rho_{\text{rms}}(n)$.

B. Observed moments

The above results validate Eq. (3) showing that the theoretical modeling may be reconciled with actual

measurements on an idealized convergence map through a linear transformation. However, this is only half of the story. We indeed still have to demonstrate that the full calibration relations work at reproducing the observed HOM. These are estimated from the convergence map as reconstructed applying the KS93 method to noisy shear data. We therefore fit Eqs. (4)–(6) to the data⁶ measured on our simulated reconstructed maps getting the results in Fig. 6.

As a reassuring result, we find that the fit is again quite good even if the rms of percentage residuals is definitely larger than before. Indeed, for the Gaussian and top hat filters (first and second rows below), we find

$$\rho_{\text{rms}}(n) = \begin{cases} (7.84, 6.10, 53.3)\% \\ (8.57, 21.5, 42.2)\% \end{cases}$$

for $n = (2, 3, 4)$. Although the rms for the fourth order moment is quite large, it is nevertheless worth stressing that the best fit curve stays well within the error bars so that the fit is still acceptable. A look at the $\langle \kappa^4 \rangle(\theta)$ plot, however, shows that most of the deviations take place at large θ so that one could reduce the rms by cutting this range. We therefore repeat the fit for the four smaller ranges used before and actually find a somewhat contradictory result. Although $\rho_{\text{rms}}(n = 4)$ indeed reduces when fitting to the (2,12) and (4,14) ranges only, the best results are obtained for $\theta_{\min} \geq 6$ arcmin, suggesting that the theory has problems on small angular scales. This is again consistent with

⁶Note that we actually perform fits separately for the second and fourth order moments and the third order one since the calibration parameters (m_3, c_3, ν_3) only enter in the $\langle \kappa^3 \rangle_{\text{obs}}$ relation. Moreover, we do not assume the noise is Gaussian so that the three parameters (ν_2, ν_3, ν_4) are all free to vary.

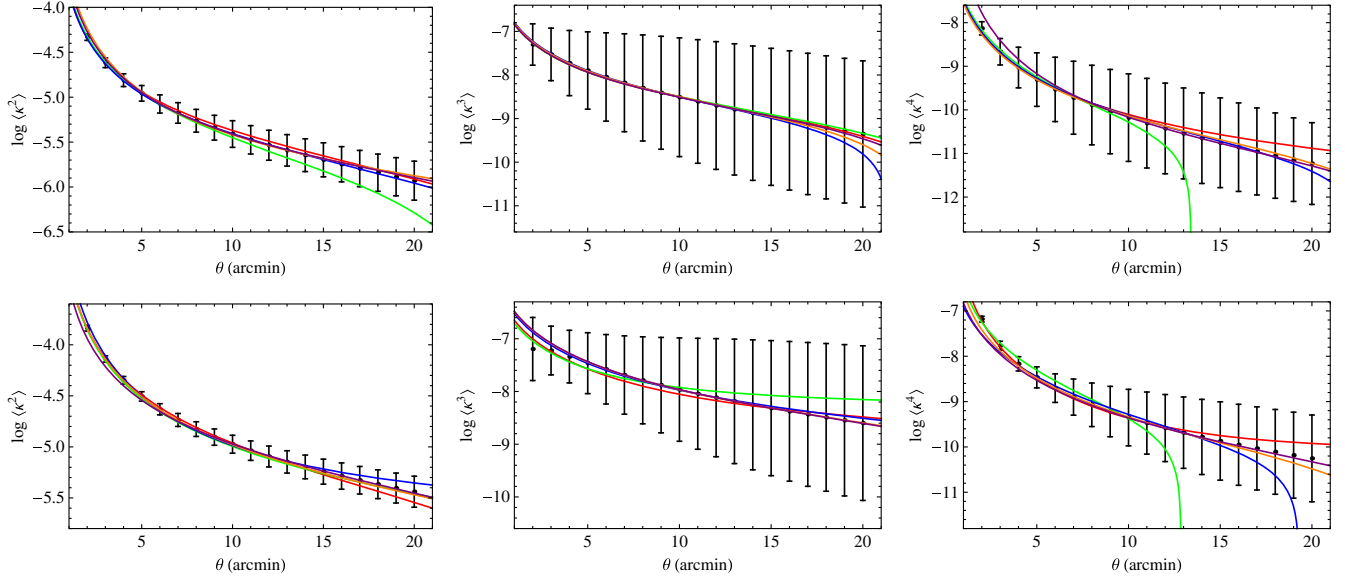


FIG. 6. Same as Fig. 4 but for moments from the map reconstructed from noisy shear data.

what we get in the previous paragraph. However, when fitting to the full range, the algorithm tries to improve the agreement on small θ since this is the range where the S/N is larger. We should therefore advocate against the use of small smoothing scales to get fourth order moment data, but this comes at the cost of giving away half the data set. We will investigate later which strategy (i.e., improving the rms of percentage residuals or retaining as much data as possible) is more convenient.

It is interesting to look at the best fit values of the calibration parameters. Considering in the rest of this paragraph only the Gaussian filter (since the results are qualitatively the same for the top hat case) and the fit to the full range,⁷ we get

$$\begin{aligned} m_2 &= -0.19_{-0.25}^{+0.28}, & c_2 &= -0.69_{-1.27}^{+1.35} \times 10^{-6}, \\ m_3 &= -0.53_{0.10}^{+0.38}, & c_3 &= -0.61_{-0.52}^{+0.71} \times 10^{-9}, \\ m_4 &= -0.11_{-0.33}^{+0.34}, & c_4 &= -0.10_{-3.15}^{+3.06} \times 10^{-11}, \end{aligned}$$

while the noise reference moments are

$$\begin{aligned} \nu_2 &= 0.21_{-0.21}^{+1.13} \times 10^{-4}, \\ \nu_3 &= -0.56_{-0.07}^{+0.27} \times 10^{-10}, \\ \nu_4 &= 2.24_{-2.24}^{+1.28} \times 10^{-10}. \end{aligned}$$

One could wonder whether these numbers are consistent with expectations. Using the expressions derived before, we can estimate the multiplicative bias m as

$$m = \left(\frac{1 + m_n}{1 + \mu_n} \right)^{1/n} - 1.$$

Ideally, we should find the same result no matter which order we choose for setting the calibration parameters. The error on m can then be qualitatively estimated propagating the errors on (m_n, μ_n) . Since the confidence ranges are asymmetric, we first follow [46] and correct for asymmetric errors with the following replacement rules:

$$m_n \rightarrow m_n + (\sigma_n^+ - \sigma_n^-), \quad \sigma(m_n) \rightarrow (\sigma_n^+ + \sigma_n^-)/2$$

with σ_n^\pm the positive and negative errors on m_n . We thus get

$$m(m_n, \mu_n) = \begin{cases} -0.08 \pm 0.15 & n = 2 \\ -0.20 \pm 0.09 & n = 3. \\ 0.03 \pm 0.10 & n = 4 \end{cases}$$

These numbers are consistent with each other within less than 2σ (taken as the sum in quadrature of the errors on the terms compared), which is a quite reassuring result if one also remembers that the fitting algorithm does not enforce the consistency of m aiming only at finding the values of the calibration parameters (m_n, c_n) that maximize the likelihood.

Somewhat more difficult is to judge whether the values of (ν_2, ν_3, ν_4) are reasonable or not making this question actually ill defined. On the one hand, we note that our simulated maps contain a Gaussian noise originating from the dispersion of galaxies intrinsic ellipticity. According to [47], we should expect a noise second order moment on the convergence map given by $\sigma_\kappa^2 = \sigma_\epsilon^2 / 2\theta_{pix}^2 n_g$. Setting $\sigma_\epsilon = 0.28$, $\theta_{pix} = 0.85$ arcmin, $n_g = 27$ gal/arcmin², we

⁷See Appendix A for the full list of calibration parameters.

get $\sigma_\kappa^2 = 2.0 \times 10^{-3}$, which is 2 orders of magnitude larger than our estimated ν_2 . However, σ_κ^2 does not take into account either the smoothing or the reconstruction so that a disagreement is far from being worrisome. On the other hand, we should find $(\nu_3, \nu_4) = (0, 3\nu_2^2)$ since we know that the noise is Gaussian. While ν_3 may indeed be considered vanishing since its value is 2 orders of magnitude smaller than the lowest $\langle \kappa^3 \rangle(\theta)$, ν_4 is definitely different from its expected Gaussian counterpart ($\nu_4 = 1.32 \times 10^{-9}$). The difference may come from both the reconstruction procedure and the fitting algorithm, which does not enforce any relation between ν_2 and ν_4 , thus leaving a degeneracy between these latter and c_4 . One could have changed the fitting code by forcing (ν_3, ν_4) to their Gaussian values (which also reduces the number of parameters). However, in practical applications, one cannot assume *a priori* that the noise is Gaussian so that we prefer to be conservative and not introduce any prior on the noise parameters.

Motivated by these considerations, we deem as successfully passed the validation test of Eqs. (4)–(6).

VI. HOM AS COSMOLOGICAL TOOLS

Having shown that it is possible to correctly match theoretical and observed moments, we can now investigate whether HOM are actually of any help in constraining cosmological parameters \mathbf{p}_c . As a preliminary remark, it is worth stressing that \mathbf{p}_c are not the only quantities that enter theory predictions. Indeed, one cannot assume *a priori* that the calibration parameters (m_n, c_n) are the same for every cosmological model so that having determined them from a fit to simulated data mimicking the actual ones is far from being sufficient to deem them as known quantities. Moreover, the noise properties of the field are fully unknown so that one has to add the noise reference moments (ν_2, ν_3, ν_4) to the list of quantities to be fitted for. Summarizing, the full parameters vector \mathbf{p} will be the union of the cosmological one,

$$\mathbf{p}_c = \{\Omega_M, \Omega_b, w_0, w_a, h, n_s, \sigma_8\},$$

and the nuisance one

$$\mathbf{p}_n = \{m_2, c_2, m_3, c_3, m_4, c_4, \nu_2, \nu_3, \nu_4\}.$$

Given the large number of parameters, we do not expect HOM alone to be able to put severe constraints on all of them. We will therefore set some parameters when considering HOM only as a data set. On the other hand, it has been suggested for a long time that HOM can help breaking degeneracies affecting second order statistics such as the cosmic shear tomography power spectra. We will therefore also investigate how constraints on the full set of parameters change when HOM are added to shear tomography.

A. Fisher matrix formalism

To answer the above questions, we rely on the Fisher matrix formalism [48]. We therefore compute the Fisher matrix elements given by

$$F_{ij} = \left\langle -\frac{\partial^2 \ln \mathcal{L}(\mathbf{p})}{\partial p_i \partial p_j} \right\rangle, \quad (20)$$

where the average is approximated by the likelihood value estimated in the fiducial model and nuisance parameters \mathbf{p}_{fid} , and, assuming a multivariate Gaussian, the likelihood function reads

$$-2 \ln \mathcal{L}(\mathbf{p}) \propto (\mathbf{D}_{\text{obs}} - \mathbf{D}_{\text{th}})^T \mathbf{Cov}^{-1} (\mathbf{D}_{\text{obs}} - \mathbf{D}_{\text{th}}), \quad (21)$$

where \mathbf{D}_{obs} has been defined in Eq. (18), and \mathbf{D}_{th} is its theoretical counterpart evaluated through Eqs. (4)–(6). The Fisher matrix elements therefore read

$$F_{ij} = \frac{\partial \mathbf{D}_{\text{th}}}{\partial p_i} \mathbf{Cov}^{-1} \frac{\partial \mathbf{D}_{\text{th}}}{\partial p_j}, \quad (22)$$

showing that a key role is played by the data covariance matrix, which we split up as

$$\mathbf{Cov} = \mathbf{Cov}^{\text{obs}} + \mathbf{Cov}^{\text{sys}} \quad (23)$$

with the first term giving the statistical errors as computed from Eq. (19). The second term is introduced to take into account the inaccuracies in the calibration procedures, i.e., the fact that we can predict HOM from theory using Eqs. (4)–(6) only up to a scatter quantified by the rms of percentage residuals. This systematics covariance matrix is diagonal⁸ so that its elements read

$$\text{Cov}_{ij}^{\text{sys}} = \rho_{\text{rms}}(i) \rho_{\text{rms}}(j) D_{\text{obs}}(i) D_{\text{obs}}(j) \delta_{ij}^K \quad (24)$$

with δ_{ij}^K the Kronecker δ and $\rho_{\text{rms}}(i)$ the rms of the percentage residuals for the order of the moment corresponding to the i th element of the data vector \mathbf{D}_{obs} . What actually enters the likelihood is the precision matrix, i.e., the inverse of the covariance matrix. However, it is known that this quantity may be biased if estimated from a not large enough number of independent realizations. Following [49], we therefore compute it as

⁸The systematics covariance matrix could actually be nondiagonal. Indeed, the expected values of the calibration parameters (m_n, c_n) are correlated being all related to the same multiplicative bias m . However, we have determined the calibration parameters from separate fits without forcing (m_2, m_3, m_4) to obey theoretical relations among them. This is expected to cancel any correlation among them and among the residuals of the calibration relations.

$$\mathbf{Cov}^{-1} = \frac{\mathcal{N}_f - \mathcal{N}_d - 2}{\mathcal{N}_f - 1} (\mathbf{Cov}^{\text{obs}} + \mathbf{Cov}^{\text{sys}})^{-1}, \quad (25)$$

where the multiplicative term corrects for the finite number of realizations with \mathcal{N}_d the length of the data vector. An important caveat is in order here. Since we have divided the full survey area Ω in subfields of 25 deg^2 each, it is $\mathcal{N}_f = \Omega/25$. On the other hand, the data vector length is $\mathcal{N}_d = \eta(\theta_{\text{max}} - \theta_{\text{min}})/d\theta = \eta\Delta\theta/d\theta$, with η the number of moments used (e.g., $\eta = 1$ if only moments of order 2 are used, or $\eta = 3$ if one uses all orders). The choice of $(\Omega, \Delta\theta, d\theta)$ must therefore be done taking care that the condition $\mathcal{N}_f > \mathcal{N}_d + 2$ is fulfilled in order to avoid having a physically meaningless negative definite inverse covariance matrix. For our fiducial choice, it is

$$(\Omega, \theta_{\text{min}}, \theta_{\text{max}}, d\theta, \eta) = (3500, 2, 20, 1, 3)$$

giving $\mathcal{N}_f = 140$ and $\mathcal{N}_d = 57$ so that the multiplicative factor takes the quite small value $(\mathcal{N}_f - \mathcal{N}_d - 2)/(\mathcal{N}_f - 1) \approx 0.58$, which reduces the Fisher matrix elements, and hence the forecasted constraints on the cosmological parameters. Since the total area is fixed, a possible way out could be to reduce the subfields area so that \mathcal{N}_f is larger, but we should then check whether the estimated HOM are still reliable. As an alternative, one can reduce the length of the data vector \mathcal{N}_d using a larger sampling $d\theta$ or only a subset of the data (so that η is smaller). While both these choices are possible, they both ask for reducing the number of constraints that can have the opposite effect of weakening overall the HOM constraining power.

Another effect also enters the game driving the choice of $(\theta_{\text{min}}, \theta_{\text{max}}, d\theta)$. As we have seen, $\rho_{\text{rms}}(n)$ is a function of $(\theta_{\text{min}}, \theta_{\text{max}})$ so that the systematics covariance matrix will be different depending on which smoothing angle range is used. The decrease of $\rho_{\text{rms}}(n)$ with θ_{min} would advocate in favor of larger values, but it is possible to show that the derivative terms into Eq. (22) are larger at small θ consistent with the naive expectation that a large smoothing angle cancels both the noise and the signal. The best range to use is, therefore, a compromise between these two opposite behaviors. A similar discussion also applies to the choice of the sampling $d\theta$. As shown before, moments of the same or different orders evaluated at two different smoothing angles (θ_1, θ_2) may be significantly correlated if $\theta_2 - \theta_1$ is smaller than a few times $d\theta$. As a consequence, the effective number of constraints is actually smaller than the length of the data vector. Moreover, should $d\theta$ be too small, the observed covariance matrix $\mathbf{Cov}^{\text{obs}}$ could become close to degenerate, thus making unstable its numerical inversion and hence unreliable the Fisher matrix itself. Whether this happens also depends on the θ range. Indeed, the larger is θ , the more the details of the convergence field are smoothed out and the moments reduce to those of a

Gaussian distribution. Should we therefore consider only large θ , the moments will be all correlated being simply a description of the smoothed noise field so that which $d\theta$ value does not matter at all.

Motivated by these considerations, we will therefore explore different combinations of $(\theta_{\text{min}}, \theta_{\text{max}}, d\theta)$. In particular, we consider five different ranges, namely

$$(\theta_{\text{min}}, \theta_{\text{max}}) = (2, 20), (2, 12), (4, 14), (6, 16), (8, 18) \text{ arcmin},$$

and sample them with $d\theta = 1, 2, 3, 4$ arcmin. Moreover, we will also investigate how the results depend on the smoothing filter used, and whether it is more convenient to use moments of all orders or only a subset of them.

As a final preliminary remark, a somewhat technical note is in order. As already said at the beginning of this section, one has to add the set of nuisance parameters \mathbf{p}_n to the list of quantities to be determined by the fit to the data. As a consequence, they also enter the Fisher matrix so that their fiducial values have to be set. It is, however, convenient to introduce a new nuisance parameter set $\tilde{\mathbf{p}}_n = \mathbf{p}_n/\mathbf{p}_n^{\text{fid}}$ with $\mathbf{p}_n^{\text{fid}}$ the values found from the fit to the simulated data set for a given fiducial cosmology. Should the calibration be independent of cosmology, we should expect $\tilde{p}_{in} = 1$ for all the components i of the scaled vector. We therefore use $\tilde{\mathbf{p}}_n$ as the nuisance parameter set and fix its fiducial value to unity. To break the degeneracies with cosmological parameters, we will add a prior matrix to the Fisher matrix one; i.e., we redefine the Fisher matrix elements as

$$F_{ij} = \frac{\partial \mathbf{D}_{\text{th}}}{\partial p_i} \mathbf{Cov}^{-1} \frac{\partial \mathbf{D}_{\text{th}}}{\partial p_j} + \pi_{ij}, \quad (26)$$

where π_{ij} are the elements of the prior matrix. This is a diagonal matrix whose first seven elements are set to zero (since we do not want to put any prior on the cosmological parameters), while the remaining elements are set to $1/\sigma_i^2$ with σ_i quantifying our belief in the prior knowledge of the nuisance parameters. Setting, e.g., $\sigma_i = 0.1$ means we are assuming that the true value of the i th component of the nuisance parameters vector lies within 10% of the corresponding fiducial one. To reduce the quantities to change in the analysis, we will set $\sigma_i = \sigma_p$ for all i and explore how constraints change as a function of σ_p .

B. Constraints on (Ω_M, σ_8) from HOM only

In the first application, one will probably be interested in using moments only as constraints. However, because of the large number of parameters (cosmological and nuisance) and the reduced numbers of effective degrees of freedom because of the large correlations among moments at different smoothing scales, one expects too weak constraints if all parameters are left free. One would therefore focus only on the two parameters where lensing is most

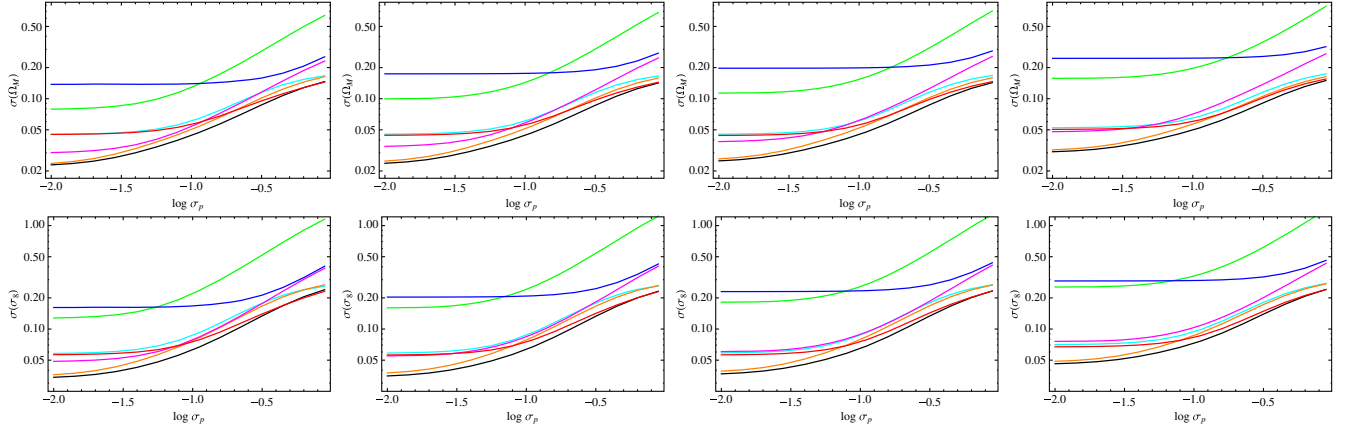


FIG. 7. Fisher matrix forecast of the 1σ uncertainty on Ω_M (top) and σ_8 (bottom) as a function of the prior parameter σ_p from the fit to data covering the smoothing angle range $(\theta_{\min}, \theta_{\max}) = (2, 20)$ arcmin sampled with step $d\theta = 1, 2, 3, 4$ arcmin (from left to right). A Gaussian filter has been used for smoothing, while black, green, cyan, blue, orange, magenta, red lines refer to the results using all moments, second order only, third order only, fourth order only, second and third, second and fourth, third and fourth, respectively.

sensible, too, i.e., the matter density Ω_M and the variance of linear perturbations σ_8 . We therefore compute the Fisher matrix, set all the cosmological parameters but (Ω_M, σ_8) , and marginalize over the nuisance parameters for any given choice of the prior σ_p , the fitting range $(\theta_{\min}, \theta_{\max})$, and the sampling $d\theta$. Although limited to two parameters only, this analysis will be helpful to elucidate how HOM actually works, hence selecting the best strategy.

Let us start by investigating whether all moments have to be used or comparable results can be obtained using only a given combination of second, third, and fourth order moments.⁹ Figure 7 shows the forecasted 1σ constraints on the parameters (Ω_M, σ_8) setting $(\theta_{\min}, \theta_{\max}) = (2, 20)$ arcmin and four values of the sampling step $d\theta$. As expected, the constraints on (Ω_M, σ_8) are quite weak and strongly dependent on the value of the prior σ_p . A saturation roughly takes place for $\sigma_p < 0.03$ suggesting that, for smaller values, the results are limited by the HOM statistical and systematic errors. For a 10% prior on the nuisance parameters and using all moments, we get $\sigma(p_\mu)/p_\mu \approx 18(6)\%$ for $p_\mu = \Omega_M(\sigma_8)$ which points at HOM as interesting tools to constrain these two cosmological parameters. However, the constraints are strongly dependent on priors with $\sigma(p_\mu)/p_\mu \approx 37(22)\%$ for $p_\mu = \Omega_M(\sigma_8)$ if no prior is set. These are definitely too large to deem HOM as actually useful which, however, does not come as a surprise having HOM being suggested as complementary probes to be used in combination with standard cosmic shear tomography to break degeneracies and not as a single tool. We will nevertheless keep

⁹Unless otherwise stated, all the results in this section refer to HOM estimated after smoothing the reconstructed convergence map with a Gaussian filter. The results are qualitatively unchanged for the top hat case, too, the only differences being highlighted at the end of this section if necessary.

investigating their use as stand-alone probes to better highlight which strategy is most suited to strengthen the constraints.

One can naively expect that the strongest constraints are obtained when all moments are combined. This is actually not so obvious. Indeed, using all moments maximizes \mathcal{N}_d in Eq. (25), thus reducing the inverse covariance matrix and hence the F_{ij} elements. Moreover, because of the strong correlation among moments of different orders estimated at the same θ and moments of the same order at close θ values, the effective number of degrees of freedom is smaller than the length of the data vector \mathbf{D}_{obs} so that it is not clear whether adding more orders automatically translates to more constraints. On the other hand, the number of nuisance parameters is smaller if only one order or only two out of three are used. That is why we have explored all seven possibilities that are plotted as different colored lines in Fig. 7. Notwithstanding the above cautionary considerations, it turns out that the best case scenario is indeed the most obvious one. The best constraints are obtained when the full HOM data set (black line) is used, but we nevertheless note that almost equivalent results are obtained for the case with only second and fourth order moments used (magenta line). Dropping off the fourth order moments to use only second and third order ones (orange line) leads to constraints on (Ω_M, σ_8) which are comparable but yet definitely weaker, while using only one set of moments (green, cyan, and blue lines) gives significantly weaker constraints. Such results suggest that most of the cosmological information is encapsulated in second and fourth order moments, while third order ones contain less information but can still be used to replace fourth order moments because of the smaller systematic uncertainty.

Let us now move to the question about which smoothing angle range $(\theta_{\min}, \theta_{\max})$ should be used to minimize the

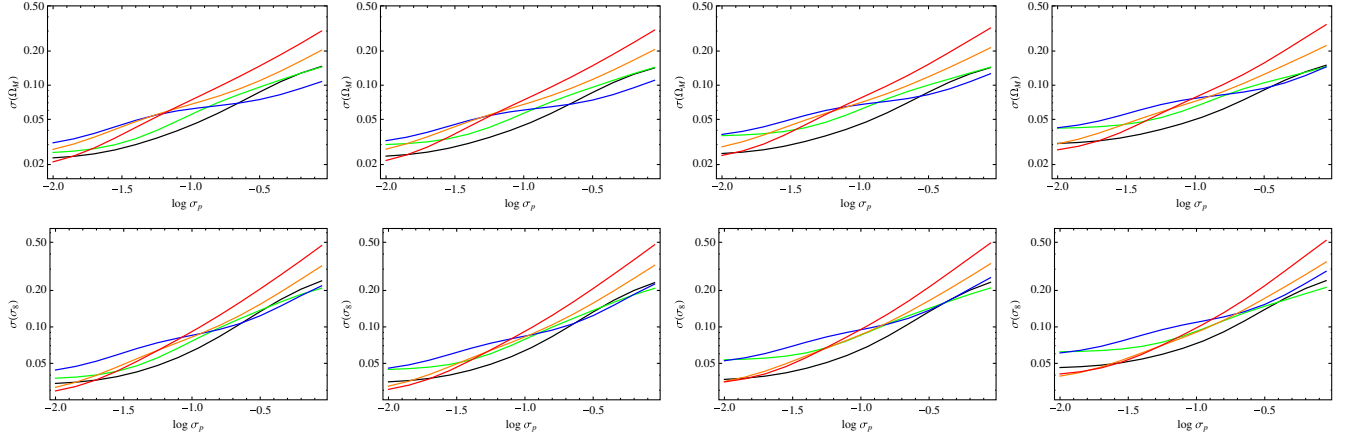


FIG. 8. Same as Fig. 7 but all moments are used, while black, green, blue, orange, red lines refer to using data over the ranges (2, 20), (2, 12), (4, 14), (6, 16), (8, 18) arcmin, respectively, sampled with step $d\theta = 1, 2, 3, 4$ arcmin from left to right.

forecasted errors on (Ω_M, σ_8) . On one hand, for a fixed sampling $d\theta$, the wider is $\Delta\theta = \theta_{\max} - \theta_{\min}$, the larger is the length N of the data vector. On the other hand, the larger is $\Delta\theta$, the larger is \mathcal{N}_d , thus making the inverse covariance matrix smaller. Moreover, moving to larger θ_{\min} values allows one to reduce (up to an order of magnitude) the systematics contribution to the data covariance matrix which should help in strengthening the constraints. To explore which effects dominate, we use all HOM and plot the constraints as a function of the prior parameter σ_p for the five different ranges we have introduced before and four different sampling step values. Results in Fig. 8 show that the strongest constraints are obtained when the full (2, 20) arcmin (black line) range is used, but comparable results can be reached if the upper limit is cut to 12 arcmin only (green line). It is instructive to note that the worst results are obtained for the intermediate range (4, 14) arcmin (blue line). Such a result can be qualitatively explained as follows. Most of the cosmological information is encapsulated in the dependence of HOM on cosmological parameters at small smoothing angles. In other words, the derivatives $\partial\mathbf{D}_{\text{th}}/\partial p_i$ entering the Fisher matrix elements F_{ij} are decreasing function of θ so that F_{ij} is larger when small θ ranges are used. On the other hand, the overall $\rho_{\text{rms}}(n)$ is smaller when the fit is limited to ranges with large θ_{\min} , thus increasing \mathbf{Cov}^{-1} and hence F_{ij} . These two different effects dominate in the extreme ranges, while they compensate for each other in the intermediate one, which therefore provides the weakest constraints.

Figure 8 can also be used to infer how the results change with the sampling step $d\theta$. Again, two competing effects are at work here. On the one hand, the larger is $d\theta$, the smaller are the elements of the correlation matrix \mathbf{R} and the larger is the multiplicative term into Eq. (25), thus leading to an overall increase of F_{ij} . On the contrary, the larger is $d\theta$, the smaller is the length of the data vector \mathbf{D}_{obs} so that the number of degrees of freedom in the fit is smaller, hence

weakening the constraints. Figure 8 shows that the two effects approximately compensate for each other if the full (2, 20) arcmin range is used, while it is safer to set $d\theta \leq 3$ arcmin if one wants to use only a subset of the data.

We finally conclude this analysis with a note on the results when the top hat filter is used. All the results are qualitatively unchanged pointing at a preference for using all HOM data over the full smoothing angle range with the finest sampling. The only remarkable difference is the magnitude of the constraints on (Ω_M, σ_8) which turns out to be dramatically weaker. We indeed find that $\sigma(\Omega_M)$ and $\sigma(\sigma_8)$ can be larger up to a factor of 2 for the same value of the prior parameter σ_p . This is likely related to the top hat filter being too aggressive in performing noise reduction with the smoothing procedure removing also part of the signal where cosmological information would have been usable. However, we warn the reader that what is of most interest is not the use of HOM alone, but the combination with shear tomography so that we postpone to the next paragraph the decision whether to dismiss the top hat filter overall.

C. Joint use of shear tomography and HOM

It has long been suggested that higher than second order statistics can help break degeneracies among cosmological parameters that limit the efficiency of cosmic shear tomography. It is therefore our aim here to investigate whether this is indeed the case and to which extent the combination of shear tomography and HOM improves the constraints on underlying cosmology. To this end, we assume that shear tomography and HOM are independent probes so that the total Fisher matrix simply reads

$$\mathbf{F} = \mathbf{F}_{\text{HOM}} + \mathbf{F}_{\text{WL}} + \mathbf{\Pi}$$

with \mathbf{F}_{HOM} and \mathbf{F}_{WL} the HOM and shear tomography Fisher matrices, and $\mathbf{\Pi}$ the priors matrix defined as explained before. We refer the reader to Appendix B for

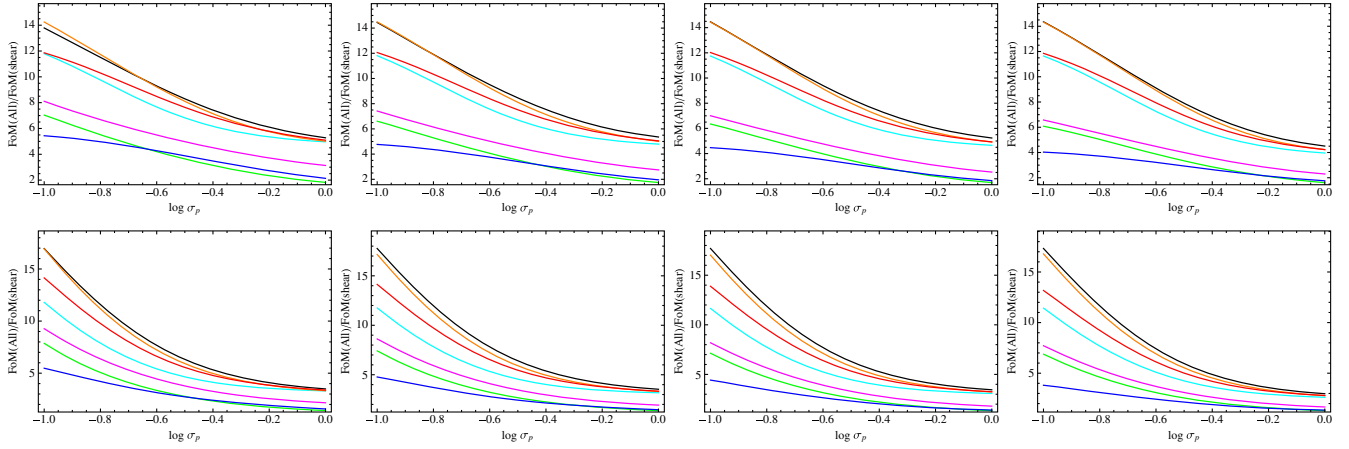


FIG. 9. FoM ratio as a function of the prior parameter σ_p . Black, green, cyan, blue, orange, magenta, red lines refer to using all HOM, second order only, third only, fourth only, second and third, second and fourth, third and fourth in combination with cosmic shear tomography based on three (top) and six (bottom) redshift bins. A Gaussian filter is used to smooth the convergence map with smoothing angles spanning the range (2, 20) arcmin with sampling step $d\theta = 1, 2, 3, 4$ arcmin from left to right.

details on the estimate of \mathbf{F}_{WL} , while we only remark here that it is based on an idealized survey having the same area, number density, and redshift distribution of the MICECAT catalogue we have used as input for the HOM Fisher matrix. We will consider two different options for the tomography splitting sources in three or six redshift bins.

Since we are interested in investigating how much the use of HOM improves the constraints with respect to the case when shear tomography only is used, we will always discuss ratios of quantities so that it is immediately clear to grasp any improvement. Although the details of an actual survey may be different from the idealized one we are considering, we are confident that taking the ratios partially washes out these differences, making our results more general. However, we are well aware that this point should be addressed with care, which is nevertheless outside the aim of this preliminary investigation.

With these caveats in mind, let us first look at Fig. 9 where we plot the ratio between the FoM with and without using HOM, where we remind the reader that it is

$$\text{FoM} = \sqrt{\frac{1}{\det \mathbf{C}(w_0, w_a)}}$$

with $\mathbf{C}(w_0, w_a)$ the inverse of the Fisher matrix marginalized over all parameters but (w_0, w_a) . We hold fixed the smoothing angle range to the full one and look at how the results change when the full HOM data set or only a part of it is used. The most remarkable result we find is that the FoM increases by a factor ~ 2 even in the worst case scenario when only a very poor prior ($\sigma_p = 1$) is set and fourth order moments only (blue line) are used. Results become roughly independent on the prior as far as $\sigma_p \leq 0.03$ which is encouraging since it suggests that a realistic

prior on the HOM nuisance parameters is enough to get a valuable FoM increase. Which combinations of HOM is best suited depends on the sampling step adopted. However, as a general rule, lines clearly separate into two groups depending on whether second order moments are included in the combined shear + HOM data set. When $\langle \kappa^2 \rangle(\theta)$ data are included, the FoM increase is maximized and is roughly the same no matter which sampling step is used or which other data are added.

This result can be interpreted noting that both shear tomography and $\langle \kappa^2 \rangle(\theta)$ are second order quantities so that they are most sensible to the cosmological parameters than higher order ones. As such, the derivatives with respect to cosmological parameters entering the total Fisher matrix elements receive a further contribution, thus ameliorating the overall result. Higher than second order moments then help break the Ω_M - σ_8 degeneracy that allows one to better constrain these two quantities and, as a consequence, to also strengthen the constraints on the other parameters, hence leading to a larger FoM. The breaking of the Ω_M - σ_8 degeneracy takes place even when second order moments are not used, thus motivating the factor 2.5 boost of the FoM in these cases too. However, for all cases, the FoM ratio is maximized when the full HOM data set is jointly fitted with shear tomography so that we take this as our final recommendation.

Let us now look at Fig. 10 where we investigate the impact of changing the HOM data range assuming that all moments are used in combination with shear tomography. As soon as a prior with $\sigma_p \leq 0.3$ is adopted, the preferred range is always the full one (2, 20) arcmin or its narrower version (2, 12) arcmin. This is consistent with a picture where most of the cosmological information is encoded in the small smoothing angle regime, which is expected given that the more the map is smoothed, the more difficult it is to

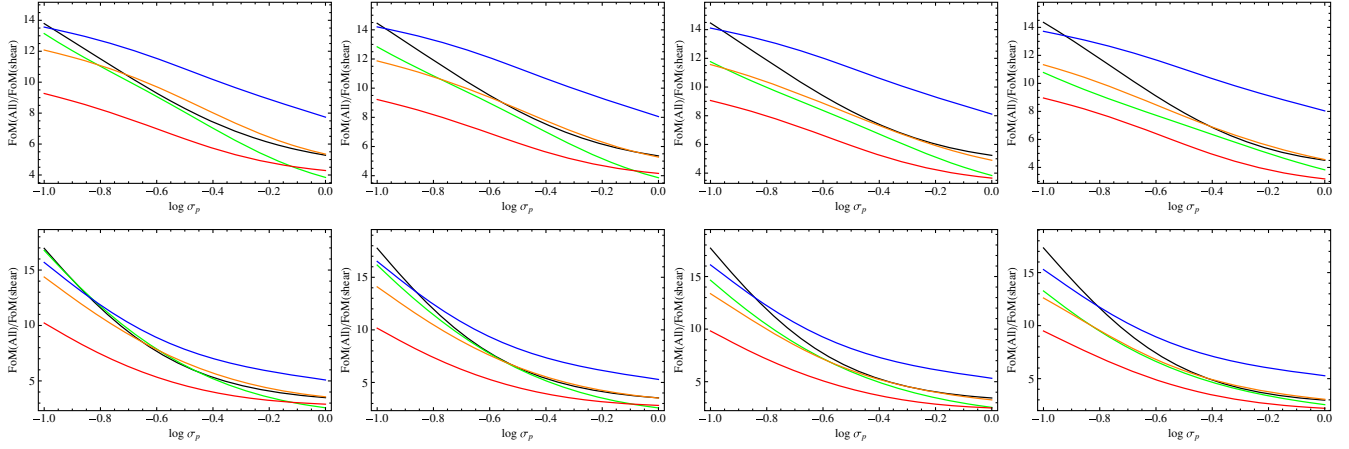


FIG. 10. Same as Fig. 9 but changing the smoothing angle range and using all HOM in combination with shear tomography (with three and six redshift bin results plotted in top and bottom panels). Black, green, blue, orange, red lines refer to the cases with $(\theta_{\min}, \theta_{\max}) = (2, 20), (2, 12), (4, 14), (6, 16); (8, 18)$ arcmin, respectively, with sampling $d\theta = 1, 2, 3, 4$ arcmin from left to right.

detect those deviations from non-Gaussianity, which bear the imprint of the underlying cosmological model. Which of the two ranges is most efficient in boosting the FoM then depends on the sampling step and the details (three or six redshift bins) of the shear tomography, but we can advocate the use of the full one since it is the preferred case in all the panels but in the bottom left one. On the contrary, there is no significant dependence of the results on the sampling step $d\theta$. This is likely a consequence of the total Fisher matrix being dominated by the shear contribution, which makes HOM important mainly to break degeneracies. Since this happens no matter which step is used, the differences become meaningless, thus deleting any preference for any particular value.

The FoM is a global measure of the efficiency of a given set of cosmological probes to constrain the dark energy equation of state (EoS). Yet, it is interesting to look at the

constraints on the single cosmological parameters and how they improve when HOM are added to shear tomography. Figure 11 is the answer to this question when all HOM are combined with the six bin shear tomography considering the five different smoothing angle ranges sampled with a step $d\theta = 1$ arcmin. As is apparent, HOM are particularly efficient at reducing the errors on all parameters with a particular emphasis on EoS quantities (w_0, w_a, w_p), the matter density Ω_M , and the variance σ_8 of the linear matter power spectrum. It is worth noting that the range more convenient to reduce the error on a given parameter depends on the parameter itself, but the full (2, 20) arcmin range is ranked first or second for all parameters.

The reduction of the errors indeed comes from the HOM breaking the Ω_M - σ_8 degeneracy. This can be efficiently seen from Fig. 12 where blue (red) lines mark the 68%, 95%, 99% confidence ranges in the (Ω_M, σ_8) plane for

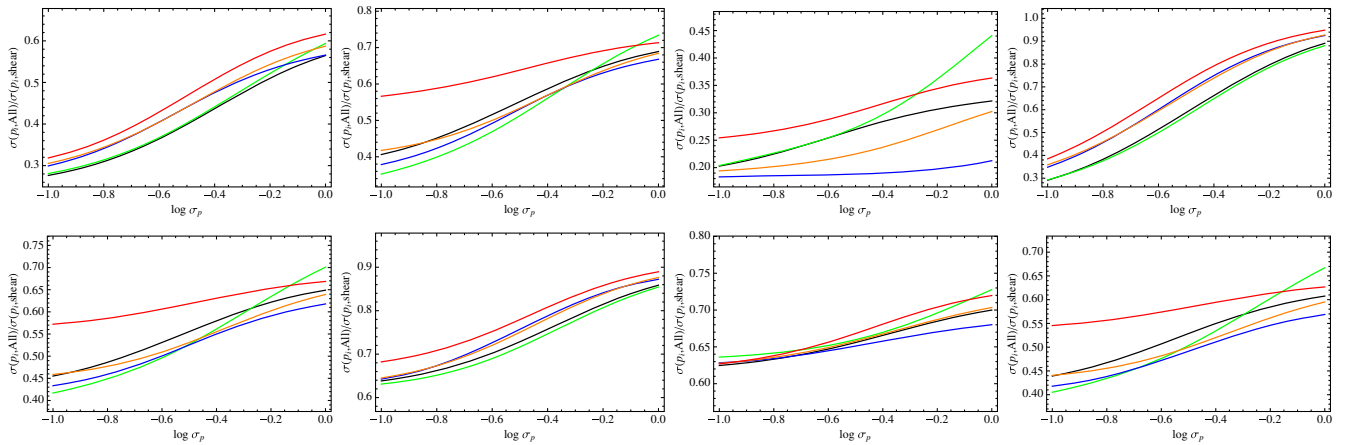


FIG. 11. Ratio between the marginalized errors on the cosmological parameters for the case with the full HOM data set is used and shear tomography is computed splitting galaxies into six redshift bins. Same color code as Fig. 10 is used to denote the smoothing angle range used, while the sampling step changes from left to right as in that same figure. Panels are for $(\Omega_M, w_0, w_a, w_p, \Omega_b, h, n_s, \sigma_8)$ from top left to bottom right, with w_p the dark energy EoS at the pivot redshift.

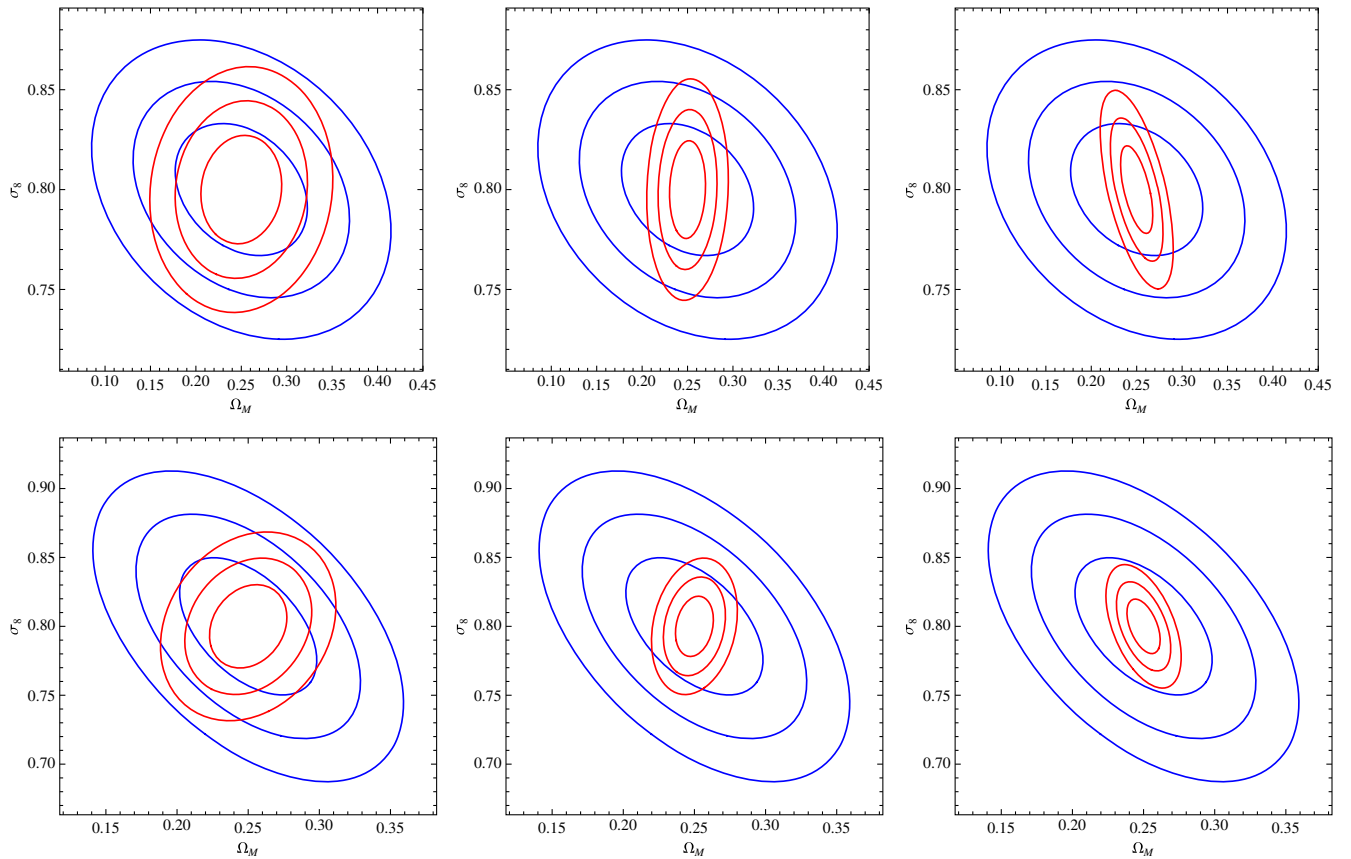


FIG. 12. The 68%, 95%, 99% confidence ranges in the (Ω_M, σ_8) plane as forecasted from the shear only (blue lines) and shear + HOM (red lines) Fisher matrix analysis. Top (bottom) panels refer to the case when three (six) redshift bins are used for tomography, while we use all HOM for the Gaussian case, the full smoothing angle range, a step $d\theta = 1$ arcmin and $\sigma_p = 1, 0.1, 0.01$ from left to right.

shear tomography with (without) the HOM contribution. It is clear that, no matter the details of the HOM configuration and whether three or six bins are used in shear tomography, the use of HOM definitely changes the orientation of the contours rotating them so that they become more parallel to the coordinate axes. This is indeed what one expects when the Ω_M - σ_8 degeneracy is broken. As a consequence, the constraints on (Ω_M, σ_8) improves, which then start a sort of *chain reaction* leading to an improvement of the constraints on the other cosmological parameters, too.

All the results discussed so far refer to the case when a Gaussian filter is used to smooth the map. Adopting a top hat filter does not change the main outcome of this analysis; i.e., HOM still increases the FoM by a factor of 2.5 in the worst case scenario. However, two main differences occur. First, the best strategy still claims the use of second, third, and fourth order moments, but now giving away the fourth order ones leaves the FoM ratio almost unchanged. Second, the preferred smoothing angle range is now the (8, 18) arcmin, and large θ_{\min} values are typically preferred. Such an opposite result with respect to the Gaussian case can be traced back to the systematics errors coming from the calibration of fourth order moments. Indeed, $\rho_{\text{rms}}(n=4)$ makes fourth order moment systematics comparable to

statistical errors on small smoothing scales so that one has to move to the large ones to reduce the data covariance matrix elements and, hence, the overall FoM. Contrary to what has been suggested by the use of HOM to constrain (Ω_M, σ_8) only, we therefore conclude that a top hat filter can still provide valid help to increase the cosmic shear tomography FoM.

VII. CONVERGENCE HOM SYSTEMATICS

The results presented up to now are implicitly based on a number of realistic yet simplifying assumptions. There are indeed some effects that we have neglected but that can potentially impact our analysis introducing uncorrected systematics. Some of them are discussed below trying to be as quantitative as possible.

A. Shear multiplicative and additive biases

The convergence map has been reconstructed from the noisy shear data implicitly assuming that these are a faithful reproduction of the true shear field. Actually, because of imperfections in the codes used to perform galaxy shape measurements and of other instrumental related systematics, the observed shear is a biased

representation of the actual one. It is common practice [50] to model the relation between the observed and the actual shear as linear, i.e., $\gamma_{\text{obs}} = (1 + m)\gamma + c$, with (m, c) the multiplicative and additive biases (to not be confused with the same quantities we have introduced in Sec. II). Considering that shear is a two components field, the i th component will read

$$\gamma_{\text{obs},i}(\vartheta_1, \vartheta_2) = [1 + m + m_i(\vartheta_1, \vartheta_2)]\gamma_1(\vartheta_1, \vartheta_2) + c + c_i(\vartheta_1, \vartheta_2),$$

where $(\vartheta_1, \vartheta_2)$ are coordinates on the sky, and, without loss of generality, we have split the additive and multiplicative biases as the sum of a constant part common to both components and a varying part characteristic of each component. Moving to the Fourier space, we get

$$\hat{\gamma}_{\text{obs},i}(\ell_1, \ell_2) = (1 + m)\hat{\gamma}_i(\ell_1, \ell_2) + c + \hat{\mathcal{G}}_i(\ell_1, \ell_2) + \hat{c}_i(\ell_1, \ell_2),$$

with (ℓ_1, ℓ_2) the component of the wave number vector, $\hat{f}(\ell_1, \ell_2) = \mathcal{F}[f(\vartheta_1, \vartheta_2)]$, and \mathcal{F} the Fourier transform operator whose properties we have used to get the above relation. In particular, the function $\hat{\mathcal{G}}_i(\ell_1, \ell_2)$ is the convolution of the Fourier transforms of the i th components of the shear and multiplicative bias. According to the KS93 algorithm, the Fourier transform of the convergence field will then be given by

$$\hat{\kappa}_{\text{obs}}(\ell_1, \ell_2) = \frac{\ell_1^2 - \ell_2^2 + 2i\ell_1\ell_2}{\ell_1^2 + \ell_2^2} \hat{\gamma}_{\text{obs}}.$$

Making some simple algebra, moving back to the real space, and adding noise, the observed convergence field will finally read

$$\begin{aligned} \kappa_{\text{obs}}(\vartheta_1, \vartheta_2) &= (1 + m)\kappa_E(\vartheta_1, \vartheta_2) + c + \mathcal{N}(\vartheta_1, \vartheta_2) \\ &+ \kappa_{E\text{mul}}(\vartheta_1, \vartheta_2) + \kappa_{E\text{add}}(\vartheta_1, \vartheta_2) \\ &+ i\kappa_{B,\text{tot}}(\vartheta_1, \vartheta_2), \end{aligned} \quad (27)$$

where the real E terms are given by

$$\kappa_X(\vartheta_1, \vartheta_2) = \mathcal{F}^{-1}[\hat{\kappa}_X(\ell_1, \ell_2)]$$

with $X = (E, E_{\text{mul}}, E_{\text{add}})$ and

$$\hat{\kappa}_X(\ell_1, \ell_2) = \frac{\ell_1^2 - \ell_2^2}{\ell_1^2 + \ell_2^2} \hat{f}_1(\ell_1, \ell_2) - \frac{2\ell_1\ell_2}{\ell_1^2 + \ell_2^2} \hat{f}_2(\ell_1, \ell_2)$$

and $(\hat{f}_1, \hat{f}_2) = (\hat{\gamma}_{\text{obs},1}, \hat{\gamma}_{\text{obs},2}), (\hat{\mathcal{G}}_1, \hat{\mathcal{G}}_2), (\hat{c}_1, \hat{c}_2)$ for $X = E, E_{\text{mul}}, E_{\text{add}}$, respectively. A similar expression holds for the imaginary B component, which will still be the sum of three terms, each one being the Fourier antitransform of a function having the following structure:

$$\hat{\kappa}_Y(\ell_1, \ell_2) = \frac{\ell_1^2 - \ell_2^2}{\ell_1^2 + \ell_2^2} \hat{f}_2(\ell_1, \ell_2) + \frac{2\ell_1\ell_2}{\ell_1^2 + \ell_2^2} \hat{f}_1(\ell_1, \ell_2)$$

with $Y = B, B_{\text{mul}}, B_{\text{add}}$ and (\hat{f}_1, \hat{f}_2) the same as before.

Equation (27) reduces to our starting relation (1) when one assumes that the B component vanishes and the multiplicative and additive biases are not present so that both $\kappa_{E\text{mul}}$ and $\kappa_{E\text{add}}$ are identically null. Shear systematics therefore have a double effect. First, they introduce a further B term so that their presence can be detected by the nonvanishing of this part. Second, they add two further terms to the real E part, thus making the observed convergence field a biased reproduction of the actual one even if one assumes $m = c = 0$. Dropping the E label for shortness, taking only the real part, and subtracting the mean value, we then have

$$\tilde{\kappa}_{\text{obs}} = (1 + m)\kappa + \mathcal{N} + \tilde{\kappa}_{\text{mul}} + \tilde{\kappa}_{\text{add}} \quad (28)$$

with $\tilde{\kappa}_{\text{mul}} = \kappa - \langle \kappa_{\text{mul}} \rangle$ and $\tilde{\kappa}_{\text{add}} = \kappa - \langle \kappa_{\text{add}} \rangle$. Note that we cannot assume *a priori* that the two systematic terms $(\kappa_{\text{mul}}, \kappa_{\text{add}})$ have a null average so that they still enter $\tilde{\kappa}$ contrary to the constant offset term c . Equation (28) is the same as Eq. (2) provided one redefines the noise term as

$$\tilde{\mathcal{N}} = \mathcal{N} + \tilde{\kappa}_{\text{mul}} + \tilde{\kappa}_{\text{add}}.$$

Under the reasonable assumption that the multiplicative and additive shear biases do not correlate with the signal and the noise, it is straightforward to show that one can repeat the full derivation of the calibration relations (4)–(6) carried out in Sec. II provided that noise moments are replaced with those of the effective noise $\tilde{\mathcal{N}}$, which are a combination of the moments of the actual noise field \mathcal{N} and the two systematics fields $(\tilde{\kappa}_{\text{mul}}, \tilde{\kappa}_{\text{add}})$.

We therefore safely conclude that the presence of shear systematics has the only impact to change the nuisance parameters of the HOM calibration but does not alter the main conclusion that HOM can be calibrated and hence used as cosmological tools.

B. Intrinsic alignment

Even under the idealized conditions of no systematics from shape measurement, the observed shear can still be biased because of intrinsic alignment [51]. In the weak lensing limit, the observed ellipticity e of a galaxy is the sum of the intrinsic ellipticity e_{int} and the shear γ , i.e., $e = e_{\text{int}} + \gamma$. Averaging over a large number of galaxies, one gets a shear estimator as $\gamma = \langle e \rangle$ under the assumption that galaxies are randomly oriented. Actually, because of intrinsic alignment (IA), one has $\langle e_{\text{int}} \rangle \neq 0$ so that the observed shear is $\gamma_{\text{obs}} = \gamma + \gamma_{\text{IA}}$. It is straightforward to understand that the γ_{IA} terms originate an additive contribution in the convergence that now reads

$$\kappa_{\text{obs}} = (1 + m)\kappa + \mathcal{N} + (\kappa_{\text{IA}} - \langle \kappa_{\text{IA}} \rangle),$$

where we have already subtracted the mean value over the map so that it is $\langle \kappa_{\text{obs}} \rangle = 0$. Note that one cannot assume *a priori* that $\langle \kappa_{\text{IA}} \rangle = 0$ so that we should carry this term along the derivation. For shortness of notation, we will, however, denote with κ_{IA} the last term in the above sum.

One can now follow the same steps carried out in Sec. II to derive the calibration relations (4)–(6), but there is now an important difference. While we can still assume that κ and \mathcal{N} are uncorrelated, this is not true for κ and κ_{IA} so that the expectation value of terms as $\kappa^n \kappa_{\text{IA}}^m$ has to be taken into account. Under the still realistic assumption that κ_{IA} and \mathcal{N} are uncorrelated, we finally end up with the following extended calibration relations:

$$\langle \kappa_{\text{obs}}^2 \rangle = (1 + m_2)\langle \kappa^2 \rangle + c_2 + \langle \mathcal{N}^2 \rangle + 2(1 + m)\langle \kappa \kappa_{\text{IA}} \rangle + \langle \kappa_{\text{IA}}^2 \rangle, \quad (29)$$

$$\begin{aligned} \langle \kappa_{\text{obs}}^3 \rangle &= (1 + m_3)\langle \kappa^3 \rangle + c_3 + \langle \mathcal{N}^3 \rangle \\ &+ 3(1 + m)^2 \langle \kappa^2 \kappa_{\text{IA}} \rangle + 3(1 + m)\langle \kappa \kappa_{\text{IA}}^2 \rangle \\ &+ \langle \kappa_{\text{IA}}^3 \rangle, \end{aligned} \quad (30)$$

$$\begin{aligned} \langle \kappa_{\text{obs}}^4 \rangle &= (1 + m_4)\langle \kappa^4 \rangle + c_4 + \langle \mathcal{N}^4 \rangle \\ &+ 6[(1 + m_2)\langle \kappa^2 \rangle + c_2]\langle \mathcal{N}^2 \rangle \\ &+ 4(1 + m)^3 \langle \kappa^3 \kappa_{\text{IA}} \rangle + 4(1 + m)\langle \kappa \kappa_{\text{IA}}^3 \rangle \\ &+ 6(1 + m)^2 \langle \kappa^2 \kappa_{\text{IA}}^2 \rangle + 6\langle \kappa_{\text{IA}}^2 \rangle \langle \mathcal{N}^2 \rangle \\ &+ \langle \kappa_{\text{IA}}^4 \rangle. \end{aligned} \quad (31)$$

For $\kappa_{\text{IA}} = 0$, Eqs. (29)–(31) reduce to the calibration relations (4)–(6) we have used throughout the paper. When IA is present, additional terms appear that have to be dealt with. Some possible strategies are outlined below.

- (i) As a brute force approach, one can consider the additional terms as further nuisance parameters to be determined from the fit itself. Although this is a legitimate procedure, it is likely quite hard to actually implement. Indeed, the dimension of the nuisance vector \mathbf{p}_n becomes dramatically close to that of the data vector, thus forcing one to use a smaller sampling step $d\theta$. Such a choice, however, will likely push the correlation matrix elements closer to 1, thus reducing the effective number of quasi-independent constraints. Second, the greater the number of fitting parameters, the greater the possibility of degeneracies among them. Both effects will operate to increase the forecasted errors on cosmological parameters, thus making HOM less and less useful even when combined with shear tomography. Moreover, in a Fisher matrix forecast analysis, one should also find out fiducial values for

the additional nuisance parameters that are actually hard to set.

- (ii) To the opposite extreme, one can note that IA is a small scale effect so that its impact will be reduced when smoothing with a large enough filter aperture. One can therefore suppose that the smoothing procedure completely washes out the effect so that the terms related to IA in Eqs. (29)–(31) are negligible or can be compensated by a change in the remaining nuisance parameters. Ideally, one could also look for a suitably defined filter that makes sure this assumption comes true. Whether this is the case or not is, however, hard to say given the present day lack of any study of the impact of IA on map reconstruction.
- (iii) Rather than considering the IA related terms as quantities to be fitted for, one can reduce the dimensionality of the problem by estimating them theoretically. To this end, one can consider the effective IA convergence field κ_{IA} as an actual one and rely on the formalism developed in [33] to compute both the moments of κ_{IA} and the expectation value of the cross terms $\langle \kappa^n \kappa_{\text{IA}}^m \rangle$. One should then replace the matter power spectrum $P(k, z)$ in Eqs. (14) with the IA power spectrum taking one of the models available in the literature [52]. These are assigned by some characteristic parameters that then add to the rest of the nuisance parameters. Such an approach allows one not only to reduce the dimension of \mathbf{p}_n but also to make the IA terms dependent on the cosmological parameters, thus adding further information in the HOM. However, to the best of our knowledge, a theoretical estimate of the IA moments with the formalism of [33] has never been attempted nor validated so that it is still premature to judge whether such a theory based treatment of the impact of IA on HOM is viable or not.

Which of the above strategies is best suited to tackle the impact of IA on HOM is not possible to say at the moment of writing. Indeed, lensing light-cone simulations including IA in the shear field are needed to perform map reconstruction and HOM measurement and calibration and to test a theoretical approach to the IA moments estimate. We are nevertheless confident that IA systematics can be consistently included in the HOM framework as soon as a suitable simulated data set will be available.

C. Measurement noise and denoising

Equation (1) has been taken as our starting point, but we have actually not shown that it actually holds since its derivation is straightforward. We have then included a noise term in our simulated data by adding a random ellipticity to the value of the shear in each position of the map to take care of the nonvanishing dispersion of the

intrinsic ellipticity distribution. It is worth noting, however, that this is not the only source of noise. One should indeed add the component coming from measurement errors typically related to instrumental effects (e.g., shape measurement will be degraded close to the edge of the image). The details of how to take care of this further component depend on the survey at hand, but some general comments are nevertheless possible and will be sketched below.

To this end, we can note that the noise field may be considered as an additional field that enters the convergence map reconstruction. But this is exactly the same situation we have already dealt with in the previous paragraph for IA. We can therefore repeat the same derivation as above, but with the further simplification that this noise field can be considered uncorrelated from the convergence field. Equations (29)–(31) reduce to

$$\langle \kappa_{\text{obs}}^2 \rangle = (1 + m_2) \langle \kappa^2 \rangle + c_2 + \langle \mathcal{N}^2 \rangle + \langle \mathcal{M}^2 \rangle, \quad (32)$$

$$\langle \kappa_{\text{obs}}^3 \rangle = (1 + m_3) \langle \kappa^3 \rangle + c_3 + \langle \mathcal{N}^3 \rangle + \langle \mathcal{M}^3 \rangle, \quad (33)$$

$$\begin{aligned} \langle \kappa_{\text{obs}}^4 \rangle &= (1 + m_4) \langle \kappa^4 \rangle + c_4 + \langle \mathcal{N}^4 \rangle \\ &\quad + 6[(1 + m_2) \langle \kappa^2 \rangle + c_2] \langle \mathcal{N}^2 \rangle \\ &\quad + 6[(1 + m)^2 \langle \kappa^2 \rangle + c_2] \langle \mathcal{M}^2 \rangle \\ &\quad + 6 \langle \mathcal{M}^2 \rangle \langle \mathcal{N}^2 \rangle + \langle \mathcal{M}^4 \rangle, \end{aligned} \quad (34)$$

where we have denoted with \mathcal{M} the additional noise field after subtracting its average value. Equations (32)–(34) are actually the same as our standard calibration relations (4)–(6) provided define a total noise field $\mathcal{N}_{\text{tot}} = \mathcal{N} + \mathcal{M}$ so that all the results we have discussed in previous sections still hold. As the only difference, we get an increase in the magnitude of the noise moments. A second somewhat obvious difference will be the S/N ratio of the measured moments that will be smaller by an amount that depends on the ratio \mathcal{M}/\mathcal{N} and hence on the survey details.

One can also wonder whether it is possible to fully remove the contribution of the noise moments to improve the convergence HOM measurements. Such a procedure has been suggested in [34] where it was argued that a better estimate of the HOM is provided by the corrected moments

$$\langle \kappa_{\text{corr}}^2 \rangle = \langle \kappa_{\text{obs}}^2 \rangle - \overline{\langle \kappa_{\text{rnd}}^2 \rangle}, \quad (35)$$

$$\langle \kappa_{\text{corr}}^3 \rangle = \langle \kappa_{\text{obs}}^3 \rangle, \quad (36)$$

$$\langle \kappa_{\text{corr}}^4 \rangle = \langle \kappa_{\text{obs}}^4 \rangle - 6 \overline{\langle \kappa_{\text{obs}}^2 \rangle \langle \kappa_{\text{rnd}}^2 \rangle} - \overline{\langle \kappa_{\text{rnd}}^4 \rangle}, \quad (37)$$

where quantities with the label “rnd” are measured over noise only maps, and the overline denotes the average over a large number of noise only map realizations.

To test whether this denoising procedure works, we first build 100 noise only shear maps keeping fixed both the

position and the shape of galaxies on every 25 deg² patch, but we randomize their orientation. For a given filter and smoothing length, we then use these maps to estimate the terms entering Eqs. (35) and (37). Note that we are here considering only the intrinsic ellipticity noise (i.e., we set $\mathcal{N}_{\text{tot}} = \mathcal{N}$), but the results of this test should qualitatively hold whatever kind of noise is actually affecting the input shear map.

If denoising is actually working, we should expect that the corrected moments match those measured on the noiseless convergence map. We therefore also perform map reconstruction from an idealized shear map with no noise. We then compare the second and fourth order corrected moments¹⁰ with those measured on this map to finally get

$$\overline{\langle \kappa_{\text{corr}}^2 \rangle} / \overline{\langle \kappa_{\text{free}}^2 \rangle} = 0.88 \pm 0.01, \quad \overline{\langle \kappa_{\text{corr}}^4 \rangle} / \overline{\langle \kappa_{\text{free}}^4 \rangle} = 0.17 \pm 0.02$$

with quantities with the label “free” measured on the noise free map. These numbers clearly point to a failure of the denoising procedure since one should actually find values of order unity. On the contrary, such low numbers are evidence that denoising is removing both the noise and (part of) the signal. As a possible way out, one could try to empirically modify Eqs. (35)–(37) by adding some multiplicative factors to the correction terms and tailoring them so that the noiseless moments are exactly recovered. However, this would require a large number of both noise and signal simulations to be sure that these corrective factors do not depend on the noise properties and the cosmological parameters, but are rather universal. This is outside our aims here so that we conservatively argue against using denoising in moment estimates.

D. Masking

Camera CCD defects, bright stars saturating the image, satellite trails, and various spurious artifacts must be masked out before any weak lensing analysis. On the contrary, we have here assumed that no masking is needed so that the full subfield area can be used to convert the shear field into the convergence map through the KS93 method. Actually, this is far from the truth in realistic applications so that it is worth wondering whether and to which extent masking impacts the HOM estimate.

To this aim, one can first note that moments describe the shape of the convergence PDF so that they are global quantities. Unless masking preferentially selects areas with a given κ value, it is therefore expected that the κ distribution is overall unchanged so that HOM are unaffected by masking. Such a qualitative discussion would actually hold if masking were performed on the convergence map itself. On the contrary, masking is performed on

¹⁰We do not test third order moments since the denoising procedure does not apply any correction to them.

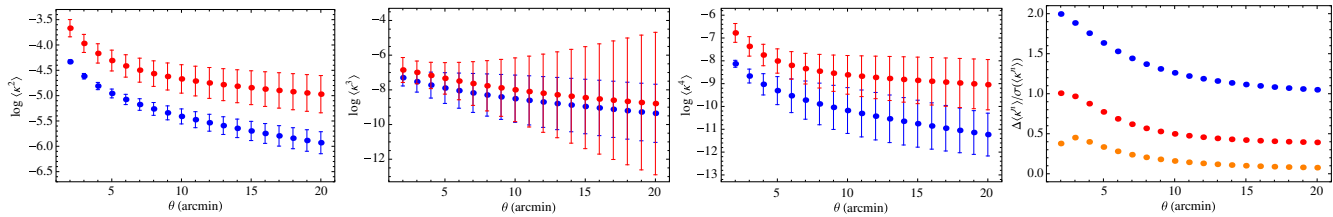


FIG. 13. Comparing convergence HOM measured on maps reconstructed from unmasked (blue) and masked (red) shear data. First three panels show the second, third, and fourth order moments, while the fourth one plots the difference between the two data sets normalized with respect to the statistical error (with blue, orange, red points referring to second, third, fourth order moments).

the shear field introducing holes that must be filled somewhat in order to run the KS93 algorithm. Different techniques [34,53] have been developed to tackle this problem with the final aim of making the reconstructed convergence map as similar as possible to the actual one. As far as this goal is reached, we can still assume that the global properties of the convergence field are unaffected, hence leading to no bias on the HOM.

To see whether this is indeed the case, we carry on a preliminary analysis using the filling form method already applied to reconstruct the convergence field from CFHTLenS data [34]. As a first step, we construct a simple mask to be applied on the catalog shear data by randomly positioning circles with radii between 0.18 arcsec to 1 arcmin in such a way that the total masked area is 20%. We generate this kind of mask for each of the 140 shear fields cut from the MICECAT catalog and input them to the map reconstruction code. Moments are then estimated from these reconstructed maps, getting the results shown in Fig. 13. Note that this figure refers to moments estimated from a single subfield since this is the area used to generate the mask.

It is apparent that HOM from maps reconstructed from masked data are only shifted upwards with respect to the unmasked case with the shift being more or less statistically meaningful depending on the moment order. Such results are not fully unexpected and can be qualitatively explained as follows. First, we note that the second order moment is increased as a result of the smaller number of cells used to compute the convergence PDF distributions, which reduces the statistics and thus increases the variance of the distribution because of the larger impact of the noise. This, however, does not significantly change the symmetry properties of the distribution so that the third order moment is the least affected. On the contrary, the fourth order moment is increased as a consequence of the increase of the second order one. Indeed, should the distribution be Gaussian, a factor of 2 boost in the second order moment translates into a factor of 4 boost in the fourth order one. Although the convergence PDF is not Gaussian, the combination with the Gaussian noise makes this effect still partially at work, thus motivating the increase we find.

What is important to stress, however, is that masking does not invalidate our calibration approach. It is indeed

still possible to rely on Eqs. (4)–(6) which is our fundamental assumption for the use of HOM as cosmological probes. Masking will change the values of the calibration parameters, but these quantities are actually of no interest for cosmology. Moreover, it is also worth noting that the use of inpainting techniques will likely reduce the difference among maps reconstructed from masked and unmasked data, thus making the difference in calibration parameters likely negligible. While this is a point we will address in a separate paper, we can nevertheless safely conclude that masking introduces systematics in the analysis which do not bias the estimate of the cosmological parameters.

VIII. CONCLUSIONS

More and better data does not only imply more and better statistics. Improved observational techniques and wider areas are two promises that ongoing and future lensing surveys will respect, thus making available a large data set of unprecedented quality data. Rather than limiting ourselves to do better what has already been done up to now, one can think to take the next step forward. From the point of view of lensing studies, this means moving away from second order statistics (two points correlation function and cosmic shear tomography) and consider higher order estimators. Motivated by this idea, we have here explored the use of higher order moments of the convergence field to increase the FoM of lensing surveys.

To make HOM viable cosmological tools, one has first to show that it is possible to match theory and observations, i.e., that reliable calibration relations exist so that the predictions for a given set of cosmological parameters are in agreement with the moments estimated from the corresponding convergence maps. Equations (4)–(6) are our answer to this question, making it possible to connect the theoretical predictions from Eqs. (11)–(13) to the HOM estimated from the MICECAT simulated catalog data. The successful comparison with simulations has validated our approach so that we can safely rely on it to use HOM as cosmological probes.

A Fisher matrix analysis has then shown that, while as stand-alone probe moments are unable to put meaningful constraints on (Ω_M, σ_8) , the joint use of shear tomography

and HOM can boost the FoM by up to a factor of 14 and no less than a factor of 2. Such an astonishing result is due to HOM breaking the $\Omega_M\text{-}\sigma_8$ degeneracy, which makes it possible to strengthen the constraints on all the cosmological parameters. This degeneracy breaking takes place whatever smoothing angle range and sampling step is used, but we advocate the use of the full (2, 20) arcmin range with sampling $d\theta = 1$ arcmin since this choice guarantees a better statistics and hence maximizes the FoM boost.

As encouraging as they are, these results have to be considered still preliminary and taken *cum grano salis*. First, we note that they refer to an idealized lensing survey having area, redshift distribution, and number density different from ongoing (DES, KIDS) and future (Euclid) surveys. In particular, the shear tomography only FoM is likely grossly underestimated because of the small number of redshift bins we have used. The impact of the number of bins can be appreciated by noting that the shear FoM increases by a factor of 2 when we move from three to six redshift bins. Since we have considered the FoM ratio, it is clear that, if the denominator is underestimated, the full ratio is overestimated. It is, however, reassuring that the FoM ratio in Figs. 9 and 10 is roughly the same between top and bottom panels, suggesting that the FoM increase does not depend on the number of bins. However, changes in the redshift distribution and number density affect both shear tomography and HOM, and it is not possible to say in advance if the changes go in the same direction, thus leaving the FoM ratio unchanged.

To test whether this is the case, one should repeat the analysis presented in this exploratory paper based on simulations mimicking a given survey in terms of redshift distribution, area, and number density. A possible example is the recently released multidark weak lensing light cone [54]. This will offer the possibility to not only extend the redshift range but also adjust the source redshift distribution, thus allowing one to explore the impact of this quantity on the validity of our linear calibration relation. Moreover, being multidark based on a different input cosmological model, we could also check whether the results in the present paper still hold when the background cosmology is changed. We stress that, in order for such a test to be passed with green lights, it is not important that the calibration coefficients (m, c) are the same since they will likely depend on both cosmology and the redshift range probed, but only that the relation between theoretical and observed moments can be approximated as a linear one with small scatter. Although such a confirmation test is needed, we are nevertheless confident that this is indeed the case since there is nothing in our procedure related to which cosmological model is the input for the simulation. The impact on the FoM ratio is, in contrast, hard to reliably forecast.

As a first step toward the use of HOM as supplementary cosmological probes, our analysis has been based on a

number of assumptions that could be violated, thus originating unaccounted for systematics. Some of them have been semiquantitatively addressed in Sec. VII where we have shown that the main problems may come from IA only. How troublesome IA is in lensing analysis, however, is not unexpected news because IA is the main systematics also in standard shear tomography. As discussed in Sec. VII B, taking care of IA requires specifically designed simulations that implement it in the shear field itself. Although unavailable at the moment of writing, work is in progress to make them available to the lensing community so we are confident that the time will come when this issue will correctly be taken into account. We note that a theoretical approach calibrated on simulations can turn IA from a problem to an opportunity since it should add further cosmology dependent terms in HOM, thus increasing their constraining power and hence the joint FoM.

As a conclusive remark, we want to stress that, as preliminary as they are, the results presented in this paper convincingly show that moving beyond second order statistics is an efficient way to deepen our understanding of the dark energy permeating the Universe. It is therefore entirely possible that the wisest use of future data is not doing more of the same, but rather some of the higher.

ACKNOWLEDGMENTS

The authors acknowledge E. Jullo, T. Kitching, and S. Pires for comments on a preliminary version of the manuscript. We also thank F. Bernardeau for discussion on higher order perturbation theory and C. Schimd for help with the effect of masking. We acknowledge the use of data from the MICE simulations, publicly available at <http://maia.ice.cat/mice/>; in particular, we thank the MICE team for allowing us to use the MICECAT v2.0 catalog prior to publication. V. F. C. and X. E. are funded by Italian Space Agency (ASI) through Contract Euclid-IC (I/031/10/0) and acknowledge financial contribution from the agreement ASI/INAF/I/023/12/0. We acknowledge the support from the grant MIUR PRIN 2015 ‘‘Cosmology and Fundamental Physics: illuminating the Dark Universe with Euclid’’. M. V. acknowledges support from Fundação para a Ciência e a Tecnologia (FCT) through the research grant IF/01518/2014.

APPENDIX A: CALIBRATION PARAMETERS

We summarize here the results of the HOM calibration analysis. Tables I and II report the median and the 68% confidence range of the calibration parameters, and the rms of the percentage residuals for the idealized and noisy moments, respectively. Results for fits to limited smoothing angle ranges are also given.

Concerning the results for noisy moments, we remind the reader that second and fourth order moments are jointly fit since the same parameters enter both calibration relations.

TABLE I. Calibration parameters, (m_n, c_n) , and rms percentage residuals, $\rho_{\text{rms}} = \text{rms}[100 \times (1 - \langle \kappa^n \rangle(\text{fit}) / \langle \kappa^n \rangle(\text{obs}))]$, for different moments. Note that c is in units of $(10^{-6}, 10^{-9}, 10^{-11})$ for $(\langle \kappa^2 \rangle, \langle \kappa^3 \rangle, \langle \kappa^4 \rangle)$, respectively.

Gaussian filter																
θ_0 range	(2-20)			(2-12)			(4-14)			(6-16)			8-18			
n	m_n	c_n	ρ_{rms}	m_n	c_n	ρ_{rms}	m_n	c_n	ρ_{rms}	m_n	c_n	ρ_{rms}	m_n	c_n	ρ_{rms}	
2	0.00 ± 0.05	-1.45 ± 0.38	4.88	-0.08 ± 0.06	-1.45 ± 0.66	3.30	0.04 ± 0.04	-1.45 ± 0.04	1.68	0.11 ± 0.03	-1.45 ± 0.20	0.93	0.16 ± 0.02	-1.45 ± 0.02	-1.45 ± 0.10	0.50
3	0.46 ± 0.05	-2.82 ± 0.29	6.34	0.48 ± 0.08	-2.82 ± 0.12	4.48	0.59 ± 0.03	-2.82 ± 0.03	1.75	0.54 ± 0.05	-2.82 ± 0.36	3.04	0.43 ± 0.07	-2.82 ± 0.34	3.48	
4	-0.19 ± 0.13	1.50 ± 1.20	21.5	-0.31 ± 0.12	1.50 ± 5.11	15.9	-0.04 ± 0.10	1.50 ± 0.10	8.07	0.14 ± 0.07	1.50 ± 0.80	4.45	0.25 ± 0.05	1.50 ± 0.35	2.49	
Top hat filter																
θ_0 range	(2-20)			(2-12)			(4-14)			(6-16)			8-18			
n	m_n	c_n	ρ_{rms}	m_n	c_n	ρ_{rms}	m_n	c_n	ρ_{rms}	m_n	c_n	ρ_{rms}	m_n	c_n	ρ_{rms}	
2	-0.16 ± 0.04	0.74 ± 0.65	4.55	-0.23 ± 0.04	0.74 ± 0.93	2.70	-0.13 ± 0.04	0.74 ± 0.57	1.54	-0.06 ± 0.03	0.74 ± 0.41	1.04	0.00 ± 0.03	0.74 ± 0.30	0.75	
3	0.30 ± 0.09	-1.46 ± 1.94	9.23	0.21 ± 0.11	-1.46 ± 5.36	7.42	0.42 ± 0.06	-1.46 ± 0.06	3.23	0.53 ± 0.03	-1.46 ± 0.77	1.37	0.57 ± 0.01	-1.46 ± 0.19	0.39	
4	-0.48 ± 0.09	17.2 ± 6.3	21.2	-0.55 ± 0.08	17.2 ± 19.1	14.5	-0.39 ± 0.07	17.2 ± 8.4	8.44	-0.26 ± 0.06	17.2 ± 4.3	5.47	-0.15 ± 0.05	17.2 ± 2.4	3.70	

TABLE II. Same as Table I but for noisy moments. For each moment, there is now one further parameter, namely the order n moment of the noise for a smoothing scale $\theta_0 = 1$ arcmin. Note that (c_2, c_3, c_4) are in units of $(10^{-6}, 10^{-9}, 10^{-11})$, while (ν_2, ν_3, ν_4) are in units of $(10^{-4}, 10^{-10}, 10^{-10})$, respectively.

Gaussian filter																				
θ_0 range	(2-20)			(2-12)			(4-14)			(6-16)			8-18							
n	m_n	c_n	ν_n	ρ_{rms}	m_n	c_n	ν_n	ρ_{rms}	m_n	c_n	ν_n	ρ_{rms}	m_n	c_n	ν_n	ρ_{rms}				
2	$-0.19^{+0.28}_{-0.25}$	$-0.69^{+1.35}_{-1.27}$	$0.21^{+1.13}_{-0.21}$	7.83	$-0.12^{+0.27}_{-0.24}$	$-1.03^{+2.14}_{-1.75}$	$0.14^{+0.88}_{-0.14}$	6.72	$-0.25^{+0.22}_{-0.22}$	$-0.72^{+1.23}_{-1.11}$	$0.10^{+0.97}_{-0.10}$	4.44	$-0.32^{+0.34}_{-0.29}$	$-0.28^{+0.78}_{-0.57}$	$0.15^{+1.48}_{-0.15}$	2.19	$-0.42^{+0.31}_{-0.25}$	$-0.10^{+0.50}_{-0.35}$	$0.13^{+1.5}_{-0.13}$	3.03
3	$-0.53^{+0.38}_{-0.10}$	$-0.61^{+0.71}_{-0.52}$	$0.30^{+1.97}_{-2.16}$	6.10	$-0.56^{+0.27}_{-0.07}$	$-0.32^{+0.72}_{-0.66}$	$0.12^{+3.82}_{-4.97}$	5.85	$-0.53^{+0.16}_{-0.09}$	$-0.78^{+0.82}_{-0.50}$	$-0.35^{+1.72}_{-1.16}$	1.44	$-0.56^{+0.20}_{-0.11}$	$-0.48^{+0.63}_{-0.56}$	$0.41^{+1.07}_{-1.54}$	2.36	$-0.59^{+0.33}_{-0.12}$	$-0.41^{+0.54}_{-0.42}$	$-0.07^{+0.74}_{-0.67}$	3.01
4	$-0.11^{+0.34}_{-0.33}$	$-0.10^{+3.06}_{-3.15}$	$2.24^{+2.24}_{-2.24}$	53.3	$-0.14^{+0.35}_{-0.30}$	$-0.04^{+6.00}_{-6.14}$	$0.54^{+6.85}_{-0.54}$	24.6	$-0.21^{+0.37}_{-0.30}$	$0.33^{+2.11}_{-2.33}$	$2.86^{+2.05}_{-2.86}$	11.6	$-0.18^{+0.36}_{-0.34}$	$-0.24^{+1.07}_{-0.90}$	$1.07^{+2.82}_{-1.07}$	12.1	$-0.24^{+0.33}_{-0.33}$	$-0.07^{+0.66}_{-0.66}$	$5.46^{+4.31}_{-5.46}$	4.18
Top hat filter																				
θ_0 range	(2-20)			(2-12)			(4-14)			(6-16)			8-18							
n	m_n	c_n	ν_n	ρ_{rms}	m_n	c_n	ν_n	ρ_{rms}	m_n	c_n	ν_n	ρ_{rms}	m_n	c_n	ν_n	ρ_{rms}				
2	$-0.03^{+0.17}_{-0.24}$	$-1.67^{+3.98}_{-2.71}$	$1.17^{+3.25}_{-1.15}$	8.56	$-0.06^{+0.28}_{-0.23}$	$-1.87^{+5.50}_{-4.27}$	$1.42^{+3.12}_{-1.40}$	4.92	$0.03^{+0.13}_{-0.22}$	$-1.79^{+2.75}_{-2.41}$	$0.51^{+2.64}_{-0.50}$	2.66	$-0.09^{+0.15}_{-0.21}$	$-1.44^{+3.29}_{-1.79}$	$0.32^{+3.42}_{-0.32}$	2.21	$-0.16^{+0.15}_{-0.17}$	$-1.12^{+1.68}_{-1.52}$	$0.23^{+3.44}_{-0.23}$	1.22
3	$-0.58^{+0.39}_{-0.15}$	$-0.02^{+1.65}_{-1.62}$	$-2.67^{+6.68}_{-4.89}$	21.5	$-0.58^{+0.41}_{-0.16}$	$1.61^{+3.43}_{-4.22}$	$-2.04^{+8.14}_{-7.26}$	20.8	$-0.59^{+0.09}_{-0.04}$	$-0.12^{+0.96}_{-0.74}$	$-1.18^{+3.82}_{-4.18}$	3.86	$-0.55^{+0.11}_{-0.05}$	$-1.02^{+1.29}_{-0.57}$	$-0.81^{+4.13}_{-3.28}$	0.78	$-0.54^{+0.02}_{-0.05}$	$-1.25^{+0.96}_{-0.26}$	$-0.73^{+1.82}_{-2.06}$	0.19
4	$-0.13^{+0.30}_{-0.27}$	$-4.63^{+24.3}_{-33.7}$	$11.14^{+61.1}_{-11.14}$	42.2	$-0.07^{+0.22}_{-0.20}$	$-10.9^{+80.5}_{-66.2}$	$6.51^{+441}_{-6.51}$	35.4	$-0.19^{+0.41}_{-0.30}$	$-5.62^{+21.8}_{-13.5}$	$9.54^{+786}_{-9.54}$	13.0	$-0.21^{+0.31}_{-0.30}$	$-0.41^{+5.91}_{-6.16}$	$18.6^{+114}_{-18.6}$	5.90	$-0.15^{+0.35}_{-0.35}$	$-0.32^{+4.02}_{-4.10}$	$11.5^{+145}_{-11.5}$	4.79

As a side effect, the rms on third order moments will typically be smaller than the one for second and fourth order ones since in this case the fit has to adjust a smaller set of parameters and only minimize the residuals with respect to a single set of moments. A further remark concerns the ν_4 parameter that turns out to be practically unconstrained at the upper end (with the upper limit of the 68% confidence ranges being up to 2 orders of magnitude larger than the median value). Actually, we checked that this is a consequence of the very small contribution the noise gives to the overall fourth order moments so that even values of ν_4 as large as its upper limit are negligible with respect to the signal. A similar discussion also holds for ν_3 . Although such large errors could be reduced by shrinking the statistical errors on moments with wider area surveys, we stress that we do not actually worry about the uncertainties on the calibration parameters since these quantities are marginalized over in the present Fisher matrix analysis and in future application to real data.

APPENDIX B: SHEAR FISHER MATRIX

We briefly sketch below how we computed the cosmic shear tomography Fisher matrix. Its elements are

$$F_{\alpha\beta} = \sum_{\ell=\ell_{\min}}^{\ell_{\max}} \frac{(2\ell+1)f_{\text{sky}}\Delta\ell}{2} \times \text{Tr} \left[\frac{\partial \mathbf{C}(\ell)}{\partial p_\alpha} \mathbf{Cov}^{-1}(\ell) \frac{\partial \mathbf{C}(\ell)}{\partial p_\beta} \mathbf{Cov}^{-1}(\ell) \right], \quad (\text{B1})$$

where f_{sky} is the fraction of the sky covered by the survey, the sum is over the ℓ mode within the range $(\ell_{\min}, \ell_{\max})$, $\mathbf{C}(\ell)$ is the $\mathcal{N}_{\text{bin}} \times \mathcal{N}_{\text{bin}}$ lensing matrix whose elements $\mathcal{C}_{ij}(\ell)$ read

$$\mathcal{C}_{ij}(\ell) = \frac{c}{H_0} \int_0^{z_h} \frac{\mathcal{W}_i^y(z)\mathcal{W}_j^y(z)}{E(z)\chi^2(z)} P \left[\frac{\ell+1/2}{\chi(z)}, z \right] dz, \quad (\text{B2})$$

having assumed no contribution from IA. These also enter the covariance matrix $\mathbf{Cov}(\ell)$ which is assumed to be Gaussian so that its elements are simply given by

$$\mathbf{Cov}_{ij}(\ell) = \mathcal{C}_{ij}(\ell) + \frac{\gamma_{\text{int}}^2 \delta_{ij}^K}{f_i n_g \times 3600(180/\pi)^2} \quad (\text{B3})$$

with δ_{ij}^K the Kronecker symbol, and $f_i n_g$ the source number density for the i th redshift bin. We will use equipopulated bins, it is $f_i = 1/\mathcal{N}_{\text{bin}}$, while we set $\gamma_{\text{int}} = 0.22$ for the variance of the intrinsic ellipticity distribution.

In Eq. (B2), $P(k, z)$ is the same matter power spectrum we have used to compute the theoretical HOM, while $\mathcal{W}_i^y(z)$ is the same as $W(z)$ in Eq. (10) provided one replaces $n(z)$ with $n_i(z)$, this latter quantity being the redshift distribution for the i th bin. This is obtained from the full redshift distribution $n(z)$ convolving with the selection function of the i th bin, which also takes into account photometric redshift uncertainties. Finally, we set $(\ell_{\min}, \ell_{\max}) = (10, 5000)$ and use 50 equispaced logarithm bins in ℓ adjusting $\Delta\ell$ accordingly.

-
- [1] D. H. Weinberg, M. J. Mortonson, D. J. Eisenstein, C. Hirata, A. Riess, and E. Rozo, *Phys. Rep.* **530**, 87 (2013).
 [2] T. Padmanabhan, *Phys. Rep.* **380**, 235 (2003).
 [3] P. A. R. Ade *et al.* (Planck Collaboration), *Astron. Astrophys.* **594**, A13 (2016); **594**, A18 (2016).
 [4] F. Beutler, C. Blake, M. Colless, D. Heath Jones, L. Staveley-Smith, L. Campbell, Q. Parker, W. Saunders, and F. Watson, *Mon. Not. R. Astron. Soc.* **416**, 3017 (2011).
 [5] W. J. Percival *et al.*, *Mon. Not. R. Astron. Soc.* **401**, 2148 (2010).
 [6] F. Beutler, C. Blake, J. Koda, F. A. Marín, H.-J. Seo, A. J. Cuesta, and D. P. Schneider, *Mon. Not. R. Astron. Soc.* **455**, 3230 (2016).
 [7] M. Betoule *et al.*, *Astron. Astrophys.* **568**, A22 (2014).
 [8] C. P. Burgess, *arXiv:1309.4133*.
 [9] E. J. Copeland, M. Sami, and S. Tsujikawa, *Int. J. Mod. Phys. D* **15**, 1753 (2006).
 [10] L. Amendola and S. Tsujikawa, *Dark Energy: Theory and Observations* (Cambridge University Press, Cambridge, UK, 2010).
 [11] L. Amendola (The Euclid Theory Working Group), *Living Rev. Relativity* **16**, 6 (2013).
 [12] T. Clifton, P. G. Ferreira, A. Padilla, and C. Skordis, *Phys. Rep.* **513**, 1 (2012).
 [13] A. Joyce, B. Jain, J. Khoury, and M. Trodden, *Phys. Rep.* **568**, 1 (2015).
 [14] C. Heymans *et al.*, *Mon. Not. R. Astron. Soc.* **427**, 146 (2012).
 [15] M. J. Jee, J. A. Tyson, M. D. Schneider, D. Wittman, S. Schmidt, and S. Hilbert, *Astrophys. J.* **765**, 74 (2013).
 [16] Dark Energy Survey Collaboration, *Mon. Not. R. Astron. Soc.* **460**, 1270 (2016).
 [17] J. T. A. de Jong *et al.*, *Astron. Astrophys.* **582**, A62 (2015).
 [18] H. Aihara *et al.*, *arXiv:1704.05858*.
 [19] LSST Science Collaboration, *arXiv:0912.0201*.
 [20] R. Laureijs *et al.*, *arXiv:1110.3193*.
 [21] A. Green *et al.*, *arXiv:1208.4012*.
 [22] M. J. Jee, J. A. Tyson, M. D. Schneider, D. Wittman, S. Schmidt, and S. Hilbert, *Astrophys. J.* **765**, 74 (2013); M. Kilbinger *et al.*, *Mon. Not. R. Astron. Soc.* **430**, 2200 (2013); M. J. Jee, J. A. Tyson, S. Hilbert, M. D. Schneider,

- S. Schmidt, and D. Wittman, [arXiv:1510.03962](https://arxiv.org/abs/1510.03962); T. Abbott *et al.*, [arXiv:1507.05552](https://arxiv.org/abs/1507.05552).
- [23] T. D. Kitching *et al.*, *Mon. Not. R. Astron. Soc.* **442**, 1326 (2014).
- [24] G. Pratten, D. Munshi, P. Valageas, and P. Brax, [arXiv:1602.06711](https://arxiv.org/abs/1602.06711).
- [25] B. Jain, U. Seljak, and S. White, *Astrophys. J.* **530**, 547 (2000).
- [26] D. Munshi and B. Jain, *Mon. Not. R. Astron. Soc.* **318**, 109 (2000).
- [27] P. Valageas, *Astron. Astrophys.* **356**, 771 (2000).
- [28] M. Takada and B. Jain, *Mon. Not. R. Astron. Soc.* **348**, 897 (2004).
- [29] M. Takada and M. White, *Astrophys. J.* **601**, L1 (2004).
- [30] P. Valageas, D. Munshi, and A. J. Barber, *Mon. Not. R. Astron. Soc.* **356**, 386 (2005).
- [31] F. Bernardeau, L. van Waerbeke, and Y. Mellier, *Astron. Astrophys.* **322**, 1 (1997).
- [32] B. Jain and U. Seljak, *Astrophys. J.* **484**, 560 (1997).
- [33] D. Munshi and B. Jain, *Mon. Not. R. Astron. Soc.* **322**, 107 (2001).
- [34] L. van Waerbeke *et al.*, *Mon. Not. R. Astron. Soc.* **433**, 3373 (2013).
- [35] C. Chang *et al.*, *Phys. Rev. Lett.* **115**, 051301 (2015); V. Vikram *et al.*, *Phys. Rev. D* **92**, 022006 (2015).
- [36] P. Fosalba, M. Crocce, E. Gaztañaga, and F. J. Castander, *Mon. Not. R. Astron. Soc.* **448**, 2987 (2015).
- [37] M. Crocce, F. J. Castander, E. Gaztañaga, P. Fosalba, and J. Carretero, *Mon. Not. R. Astron. Soc.* **453**, 1513 (2015).
- [38] J. Carretero, F. J. Castander, E. Gaztañaga, M. Crocce, and P. Fosalba, *Mon. Not. R. Astron. Soc.* **447**, 646 (2015).
- [39] M. R. Blanton *et al.*, *Astrophys. J.* **592**, 819 (2003).
- [40] M. R. Blanton, R. H. Lupton, D. J. Schlegel, M. A. Strauss, J. Bronkman, M. Fukugita, and J. Loveday, *Astrophys. J.* **631**, 208 (2005).
- [41] I. Zehavi *et al.*, *Astrophys. J.* **736**, 59 (2011).
- [42] M. R. Blanton *et al.*, *Astron. J.* **129**, 2562 (2005).
- [43] P. Fosalba, E. Gaztañaga, F. J. Castander, and M. Crocce, *Mon. Not. R. Astron. Soc.* **447**, 1319 (2015).
- [44] N. Kaiser and G. Squires, *Astrophys. J.* **404**, 441 (1993).
- [45] A. Petri, M. May, Z. Haiman, and J. M. Kratochvil, *Phys. Rev. D* **90**, 123015 (2014); A. Petri, J. Liu, Z. Haiman, M. May, H. Lam, and J. M. Kratochvil, *Phys. Rev. D* **91**, 103511 (2015); A. Petri, M. May, and Z. Haiman, *Phys. Rev. D* **94**, 063534 (2016).
- [46] G. D'Agostini, [arXiv:physics/0403086](https://arxiv.org/abs/physics/0403086).
- [47] T. Hamana, M. Takada, and N. Yoshida, *Mon. Not. R. Astron. Soc.* **350**, 893 (2004).
- [48] M. Tegmark, A. N. Taylor, and A. F. Heavens, *Astrophys. J.* **480**, 22 (1997).
- [49] J. Hartlap, P. Simon, and P. Schneider, *Astron. Astrophys.* **464**, 399 (2007).
- [50] C. Heymans *et al.*, *Mon. Not. R. Astron. Soc.* **368**, 1323 (2006); R. Massey *et al.*, *Mon. Not. R. Astron. Soc.* **376**, 13 (2007); **429**, 661 (2013); M. Cropper *et al.*, *Mon. Not. R. Astron. Soc.* **431**, 3103 (2013); V. F. Cardone, M. Martinelli, E. Calabrese, S. Galli, Z. Huang, R. Maoli, A. Melchiorri, and R. Scaramella, *Mon. Not. R. Astron. Soc.* **439**, 202 (2014).
- [51] B. Joachimi *et al.*, *Space Sci. Rev.* **193**, 1 (2015); A. Kiessling *et al.*, *Space Sci. Rev.* **193**, 67 (2015); D. Kirk *et al.*, *Space Sci. Rev.* **193**, 139 (2015).
- [52] P. Catelan, M. Kamionkowski, and R. D. Blandford, *Mon. Not. R. Astron. Soc.* **320**, L7 (2001); C. M. Hirata and U. Seljak, *Phys. Rev. D* **70**, 063526 (2004); B. Joachimi, R. Mandelbaum, F. B. Abdalla, and S. L. Bridle, *Astron. Astrophys.* **527**, A26 (2011); N. Chisari, S. Codis, C. Laigle, Y. Dubois, C. Pichon, J. Devriendt, A. Slyz, L. Miller, R. Gavazzi, and K. Benabed, *Mon. Not. R. Astron. Soc.* **454**, 2736 (2015).
- [53] S. Pires, J. L. Starck, A. Amara, R. Teyssier, A. Refregier, and J. Fadili, *Mon. Not. R. Astron. Soc.* **395**, 1265 (2009).
- [54] C. Giocoli *et al.*, *Mon. Not. R. Astron. Soc.* **461**, 209 (2016).