MINKOWSKI FUNCTIONALS IN JOINT GALAXY CLUSTERING & WEAK LENSING ANALYSES

Nisha Grewal1*, Joe Zuntz2, Tilman Tröster1, and Alexandra Amon2
1Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh, EH9 3HJ, UK
2Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge, CB3 0HA, UK

Version August 23, 2022

ABSTRACT

We investigate the inclusion of clustering maps in a weak lensing Minkowski functional (MF) analysis of DES-like and LSST-like simulations to constrain cosmological parameters. The standard 3x2pt approach to lensing and clustering data uses two-point correlations as its primary statistic; MFs, morphological statistics describing the shape of matter fields, provide additional information for non-Gaussian fields. Previous analyses have studied MFs of lensing convergence maps; in this project we explore their simultaneous application to clustering maps. We employ a simplified linear galaxy bias model, and using a lognormal curved sky measurement and Monte Carlo Markov Chain (MCMC) sampling process for parameter inference, we find that MFs do not yield any information in the $\Omega_m$-$\sigma_8$ plane not already generated by a 3x2pt analysis. However, we expect that MFs should improve constraining power when nonlinear baryonic and other small-scale effects are taken into account. As with a 3x2pt analysis, we find a significant improvement to constraints when adding clustering data to MF-only and MF+$C_T$ shear measurements, and strongly recommend future higher order statistics be measured from both convergence and clustering maps.

1. INTRODUCTION

Statistics measured from gravitational lensing have been effective in probing the growth of large-scale structure and the behaviour of gravity on cosmological scales. Ongoing surveys like the Dark Energy Survey (DES; Flaugher 2005), Hyper Supreme-Cam2 (HSC; Aihara et al. 2017) survey, and Kilo-Degree Survey3 (KiDS; Kuijken et al. 2015) observe tens to hundreds of millions of galaxies. Covering 5000 deg$^2$, 1400 deg$^2$, and 1500 deg$^2$ respectively, these surveys construct detailed galaxy catalogs (Gatti et al. 2021b; Giblin et al. 2021; Aihara et al. 2022) and high resolution weak lensing maps (e.g. Jeffrey et al. 2021) and have extracted significant cosmological information from their data (Heymans et al. 2021; Miyatake et al. 2021; Abbott et al. 2022).

Cosmic shear is independent of galaxy bias and sensitive to the geometry and evolution of the Universe, making it particularly informative in the study of large scale structure (Hikage et al. 2019; Hamana et al. 2020; Asgari et al. 2021; Amon et al. 2022; Secco et al. 2022a). Using the $\Lambda$CDM model, which takes the assumption that the Universe comprises baryonic matter, cold dark matter, and dark energy, two-point statistics of weak lensing and galaxy clustering have been used successfully to constrain cosmological parameters that describe the model.

Measuring both convergence and clustering information using two-point statistics, such as correlation functions or the angular power spectra $C_\ell$, has proven particularly effective, yielding high precision constraints on cosmological parameters $\Omega_m$ and $\sigma_8$ (Heymans et al. 2021; Miyatake et al. 2021; Abbott et al. 2022).

The two-point measurements described above encode all the information present in purely Gaussian random fields. In the non-Gaussian fields generated by nonlinear structure formation, however, other (higher-order) statistics can contain additional information. One example is peak statistics, which, when applied to weak lensing maps, can break degeneracies between cosmological parameters while remaining robust against systematic effects (e.g. Dietrich and Hartlap 2010; Kratochvil et al. 2010; Liu et al. 2015; Kacprzak et al. 2016; Shan et al. 2017; Martinet et al. 2018; Peel et al. 2018b; Ajani et al. 2020; Zürcher et al. 2022). Other higher order statistics of galaxy density, galaxy shear, and cosmic microwave background (CMB) lensing like three point correlation functions and bi-spectra (e.g. Takada and Jain 2003; Vafaei et al. 2010; Sembolini et al. 2011; Petri et al. 2013; Fu et al. 2014; Secco et al. 2022b) as well as moments (e.g. Van Waerbeke et al. 2013; Petri et al. 2015; Vicinanza et al. 2016; Chang et al. 2018; Peel et al. 2018a; Vicinanza et al. 2018; Gatti et al. 2020; Gatti et al. 2021a) are able to achieve even higher precision.

Many statistics, including these, are measured from maps generated from galaxy catalogues. Convergence maps, which are projected measures of mass along the line of sight, are derived from shear catalogues of source galaxies. Clustering maps, which show the number density of galaxies, are built from lens galaxy position catalogues and are complementary tracers of cosmic structure (Elvin-Poole et al. 2018). Clustering maps have a higher signal to noise ratio, enabling more detailed results (Jeffrey et al. 2021). However, clustering maps are only able to probe the distribution of galaxies, so galaxy bias must be modelled to relate this to the density of dark matter, which dominates the overall matter density (Abbott et al. 2022).

The morphological descriptors Minkowski functionals (MFs) are another higher-order statistic applicable in this context (Minkowski 1903). MFs are unbiased functional integrals over fields, have low variance, and have been applied previously to non-Gaussian convergence maps (Mecke et al. 1993; Kratochvil et al. 2012; Petri et al. 2013; Farroni et al. 2020), CMB data (e.g. Schmalzing and Gorski 1998, and references thereto), and 3D density fields from spectroscopic data (Hikage et al. 2003; Wiegand and Eisenstein 2017; Sullivan et al. 2019; Appleby et al. 2022). For a purely Gaussian field within a $\Lambda$CDM model, MFs and $C_T$ contain the same information for $\Omega_m$ and $\sigma_8$. Adding MFs to a single lensing field analysis has been shown to greatly improve constraints on cosmological parameters, and their angular-
scale dependence is particularly effective in separat-
ing primordial non-Gaussianities from gravity-induced
non-Gaussianities (Munshi et al. 2011; Shirasaki et al.
2012; Vicinanza et al. 2019). Because they are sensitive
to small-scale structure, MFs have strong constraining
power. They have the potential to have additional re-
silience to some systematic uncertainties (Zürcher et al.
2021). Multiple MFs at different smoothing scales probe
more information than a single set of MFs, and more an-
alytical treatments like Hermite polynomials (Appleby
et al. 2022) could extract more information as well.

In this paper, we investigate the effectiveness of the
application of the first three MFs to multi-field simula-
tions of both convergence and clustering maps, with a
focus on forecasting potential constraining power using
upcoming datasets. Future surveys will be even wider
and deeper than current surveys, have the capabilities
to build more informed weak lensing maps, and provide
even more cosmological information. The Legacy Survey
Space and Time (LSST) at the Vera C. Rubin Observ-
atory (LSST Science Collaboration et al. 2009; LSST
Dark Energy Science Collaboration 2012) and the space
telescopes Euclid (Scaramella et al. 2021) and Nancy
Grace Roman (Spergel et al. 2015) will map nearly half
the sky, and so enable dramatically strong constraints
on the dark sector. The statistical approach we develop
here can eventually be applied to such datasets.

In Section 2 we cover the theory of MFs and power
spectra $C_\ell$, in Section 3 we go over the process by which
we use an MCMC to generate cosmological parameter
posterior values, in Section 4 we show the constraints on
cosmological parameters from various scenarios, and we
conclude in Section 5.

2. FORMALISM/THEORY

2.1. Minkowski Functionals

Minkowski functionals are mathematical descriptors
of the topology of continuous fields (Minkowski 1903;
Zürcher et al. 2021). For 2D random fields, there are
three functionals, quantifying area, perimeter, and mean
curvature of an excursion set (the region of a field above
a given threshold; Parroni et al. 2020). The first three
MFs are defined as

$$V_0(\nu) = \frac{1}{A} \int_A \Theta(\alpha(x) - \nu) d\phi d\theta, \quad (1)$$

$$V_1(\nu) = \frac{1}{4A} \int_A \delta(\alpha(x) - \nu) \sqrt{\alpha^2_\phi + \alpha^2_\theta} d\phi d\theta, \quad (2)$$

$$V_2(\nu) = \frac{1}{2\pi A} \int_A \delta(\alpha(x) - \nu) \left( \frac{2\alpha_\phi \alpha_\theta \alpha_\phi \theta - \alpha^2_\phi \alpha_\theta - \alpha^2_\phi \phi}{\alpha^2_\phi + \alpha^2_\theta} \right) d\phi d\theta, \quad (3)$$

where $x$ is the location in the field, $\nu$ is a chosen threshold, $A$ is the total area of the map, $\phi$ and $\theta$ are polar
coordinates, $\alpha(x)$ is the field value in two dimensions,
and $\alpha_\phi$, $\alpha_\theta$, $\alpha_{\phi\phi}$, $\alpha_{\phi\theta}$, and $\alpha_{\theta\theta}$ are derivatives of the field (Minkowski 1903; Petri et al. 2013). In Eq. (1) the Heaviside function $\Theta$ selects the field region above the threshold, whose area is calculated; Eq. (2) employs the Dirac delta function to select the perimeter of that region whose heights are at the same level as the threshold; similarly, Eq. (3) finds the curvature of the boundary, which also describes the connectivity of the field (Minkowski 1903). Since the fields we use are not normalised to have unit variance, the sensitivity of $\sigma_8$ comes from the overall amplitudes of $V_1$ and $V_2$.

Figure 1 and 2 illustrate the three MFs. The excursion
sets in Figure 1 give rise to the three MFs plotted as a
function of the threshold value in 2. If we visualize
the original 2D map as a topographic height map, the
excursion sets are the regions above a given altitude.

2.2. Power Spectra

We will compare the constraining power of MFs to the
standard 2D power spectrum $C_\ell$. The angular power
spectrum $C_{\ell}^{xy}$ measures the scale-dependent structure
of two fields $x, y$, which in this case are either weak lensing
convergence or galaxy density. Assuming an isotropic
system,

$$\langle a_{lm}^x a_{l'm'}^y \rangle = \delta_{ll'} \delta_{mm'} C_{\ell}^{xy}, \quad (4)$$

where $a_{lm}$ is the spherical harmonic transform of a field
(Kaplinghat et al. 2002). We project a 3D field $X(\chi)$ via
a weighting function $w_{\chi}(\chi)$ into the 2D field $x$ (Bartel-
mann and Maturi 2016):

$$x(\bar{\chi}) = \int_0^{\chi_{\min}} d\chi w_{\chi}(\chi) X(\bar{\chi}, \chi). \quad (5)$$

Using Limber’s approximation in a flat Universe, we
convert $P_{XY}$, the 3D (cross-)power spectrum of $X,Y$, into
the 2D angular power spectrum of $x,y$:

$$C_{\ell}^{xy} = \int_0^{\chi_{\min}} d\chi w_{\chi}(\chi) w_{\chi}(\chi)^2 P_{XY} \left( \ell + \frac{1}{2}, z(\chi) \right), \quad (6)$$

where $\ell$ is the scalar multipole, $\chi_{\min}$ is the comoving
distance to the source plane, $\chi$ is the comoving angular
diameter distance, and $w$ is a radial kernel function (Lim-
ber 1954; Bartelmann and Schneider 2001; Bartelmann
and Maturi 2016; Abbott et al. 2022). The kernel func-
ction for convergence is given by

$$w_{\chi}(\chi) = \frac{3 \Omega_m H_0^2}{2} \int_\chi^{\chi_{\max}} d\chi' n_g(\chi') \frac{\chi'}{a(\chi')} \frac{\chi' - \chi}{\chi}, \quad (7)$$

where $\Omega_m$ is the present-day matter density parameter,
$H_0$ is the Hubble parameter, $\chi_{\max}$ is the comoving
distance to the horizon, $n(\chi)$ is the source galaxy number density
distribution, and $a(\chi)$ is the scale factor (Bartelmann
and Schneider 2001). For clustering, the kernel function
is simpler:

$$n(\chi) = n(z(\chi)) \frac{dz}{d\chi}, \quad (8)$$

(Elvin-Poole et al. 2018). We do not use cross-
correlations of maps in our simulations (see Section
3.1.1), so the statistical structure of our power spectrum
analysis does not fully replicate 3x2pt analysis.
3. METHODOLOGY

In this paper we introduce the application of MFs to clustering maps. We compare the constraints on cosmological parameters from convergence maps, clustering maps, and a combination of the two. To measure these constraints, we explore the space of cosmological parameters with an MCMC process.

3.1. Simulations

In this work we generate curved sky lognormal simulations, which let us compute exact derivatives of the field. The lognormal sky simulations are relatively quick to evaluate and provide some degree of non-Gaussian signal. Our analysis has the flexibility to be measured from more sophisticated nonlinear, non-Gaussian simulations in the future.

3.1.1. Redshift

To build these simulated maps we first start from DES Y6-like, LSST Y1-like, and LSST Y10-like tomographic number densities based on those in Zhang et al. (2021), which are shown in Figure 3 and Table 1. These \( n(z) \) values are generated from an underlying true redshift distribution of the form \( n_{\text{true}}(z) \propto z^2 \exp(z/z_0)^\alpha \), divided into equal density tomographic bins and then convolved with a Gaussian error distribution with width \( \sigma(z) = (1 + z)\sigma_z \). Different values of the parameters \( \alpha \) and \( \sigma_z \) and the number of bins and total density are used for our three different scenarios, DES Y6-like, LSST Y1-like, and LSST Y10-like, using the values given in Table 1 of Zhang et al. (2021).

3.1.2. Map Simulation

Next we use the Dark Energy Science Collaboration’s (DESC; Abolfath et al. 2021) Core Cosmology Library (CCL; Chisari et al. 2019) to predict theory values for
Fig. 3.— $n_z$ redshift bins for DES Y6-like, LSST Y1-like, and LSST Y10-like (Hoyle et al. 2018; Zhang et al. 2021). Convergence distributions are displayed on the left, and clustering distributions are on the right.

|                  | Bin 1 | Bin 2 | Bin 3 | Bin 4 | Bin 5 | Bin 6 | Bin 7 | Bin 8 | Bin 9 | Bin 10 |
|------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| DES Y6-like      | 0.20  | 0.43  | 0.63  | 0.90  | 0.90-1.30 |     |       |       |       |        |
| LSST Y1-like     | 0.20  | 0.43  | 0.63  | 0.90  | 0.90-1.30 |     |       |       |       |        |
| LSST Y10-like    | 0.20  | 0.43  | 0.63  | 0.90  | 0.90-1.30 |     |       |       |       |        |
| DES Y6-like      | 1.7   | 1.28  | 1.41  | 1.55  | 1.69   |     |       |       |       |        |
| LSST Y1-like     | 1.7   | 1.28  | 1.41  | 1.55  | 1.69   |     |       |       |       |        |
| LSST Y10-like    | 1.13  | 1.19  | 1.25  | 1.32  | 1.38   | 1.45| 1.51  | 1.58  | 1.65  | 1.72   |

TABLE 1
Convergence map redshift bins in the top three rows and clustering map galaxy bias values in the bottom three rows (Hoyle et al. 2018; Zhang et al. 2021).

The convergence and clustering power spectra $C_{\ell}$. CCL takes the redshift distributions $n(z)$ and cosmological parameters and computes $C_{\ell}$, which we pass with the $n(z)$ to the Full-sky Lognormal Astro-fields Simulation Kit (FLASK; Xavier et al. 2016) to simulate the convergence and clustering maps. FLASK generates continuous lognormal fields using spherical geometry (Xavier et al. 2016).

FLASK generates noisy clustering maps directly using the noise level we supply, and we manually add noise to the convergence maps it generates. In both cases we used Gaussian noise levels corresponding to the number densities in our scenarios. FLASK’s mock fields, as with real fields, are more non-Gaussian for the clustering than for the lensing, since they are integrated over a narrower redshift kernel, and the code has specific corrections to allow a sensible joint analysis in this intermediate non-linear and non-Gaussian regime (Xavier et al. 2016).

We originally used more realistic Poisson noise in our clustering maps, but this caused problems when trying to fix the random seed (see below) in our simulations. Changing the power spectra that FLASK takes as input changes the cosmic variance and noise generated in the system. But while with Gaussian noise the same number
of random variates must be generated even as the mean levels change, this is not the case for Poisson noise. This means that for Poisson noise, a small change to input power spectra no longer creates a small change to the resulting maps, since the sequence of random noise values changes significantly\(^7\).

### 3.1.3. Simulation Inputs

For this project, we use DES Y6-like, LSST Y1-like, and LSST Y10-like scenarios to find constraints on cosmological parameters. The cosmological parameter inputs we vary in the FLASK simulation are \(\Omega_m\) (total matter), \(\sigma_8\) (amplitude of the matter power spectrum), \(\Omega_b\) (baryonic matter), \(H_0\) (present-day rate of expansion of the Universe), and \(n_s\) (spectral index). We also vary linear galaxy bias values for the different lens bins. The fiducial values can be seen in Table 2.

| Parameter | Fiducial | Min | Max |
|-----------|----------|-----|-----|
| \(\Omega_m\) | 0.3      | 0.1 | 0.6 |
| \(\sigma_8\) | 0.8      | 0.3 | 1.2 |
| \(\Omega_b\) | 0.048    | 0.047 | 0.049 |
| \(H_0\) | 0.7      | 0.5 | 0.9 |
| \(n_s\) | 0.96     | 0.9 | 1.1 |

**TABLE 2**

Fiducial values of cosmological parameters with the minimum and maximum values given to the MCMC generator.

To test the constraining power of adding clustering maps to the model, we use three sets of tomographic maps: convergence maps, clustering maps, and a combination of the two. Generating these simulations is slow, so we optimised a model that was informative (relatively) efficient. We use \(N_{\text{side}} = 1024\) for the resolution parameter throughout, and we follow Petri et al. (2013)’s choices for Gaussian smoothing \(\theta_G\) and MF threshold count \(N_{\text{bins}}\), which can be seen in Table 3. While smoothing of 1 arcminute may be the most informative level, this is computationally taxing and essentially unsmoothed in our fields, thus subject to more noise. Going to smaller scales comes with other challenges, such as modulating uncertainties due to nonlinear galaxy bias and baryon feedback. Instead we use 5 arcminutes as the fiducial value, which is still informative at the \(N_{\text{side}} = 1024\) level.

We replicate DES Y6-like, LSST Y1-like, and LSST Y10-like scenarios in our simulations by applying the corresponding redshift bins, galaxy bias values, and sky fraction. The various scenarios are summarised in Table 3, and sample simulated convergence and clustering maps are shown in Figure 4. The “joint” case refers to both lensing and clustering.

### 3.2. Observables

To test the constraining power of MFs, we compare them to the standard \(C_\ell\) analysis in cosmology. We compare the constraining power of the following three estimators: \(C_\ell\), MF, and the combination of \(C_\ell\) and MF. We use the NaMaster\(^8\) (Alonso et al. 2019) library to calculate the full-sky auto-correlation convergence and density angular power spectra \(C_\ell\) of the fields. We use a bandpower window function with 50 multipoles per bandpower. The maximum \(\ell\) value we use is \(\frac{4N_{\text{side}}}{\pi}\), which is 1536 for \(N_{\text{side}} = 1024\), which has a pixel size of 3.4 arcminutes. This means information below 7 arcminutes in the maps is smoothed out.

We give inputs for cosmological parameters, linear galaxy biases, pixel count, fraction of sky, and the amount of Gaussian smoothing applied to the map simulations.

| \(\theta\) | \(N_{\text{side}}\) | \(N_{\text{bins}}\) | mask | \(n_s\) bins | MF/\(C_\ell\) | Map Type |
|----------|-----------------|-----------------|------|-------------|-------------|----------|
| \(\theta_G\) | \(N_{\text{side}}\) | \(N_{\text{bins}}\) | mask | \(n_s\) bins | MF/\(C_\ell\) | lens/clust |
| \(\theta_G\) | \(N_{\text{side}}\) | \(N_{\text{bins}}\) | mask | \(n_s\) bins | MF/\(C_\ell\) | lens/clust |

**TABLE 3**

Scenarios considered in this paper: comparing statistic type, map type, Gaussian smoothing \(\theta_G\) at 1, 5, and 15 arcminutes, and survey model: DES Y6-like, LSST Y1-like, and LSST Y10-like.

---

\(^7\) An alternative solution would be to fix the mean of the noise field even as the signal changes.

\(^8\) https://github.com/LSSTDESC/NaMaster

---

**FIG. 4.**— Small 5x5 degree flat sky piece of the curved sky clustering (left) and convergence (right) maps of redshift bin 4 at our fiducial cosmology, using \(N_{\text{side}} = 1024\) and 5 arcminutes of Gaussian smoothing. The convergence map colourbar shows convergence of each pixel, and the clustering map colourbar shows the galaxy over-density.
values by taking a sum over the number of pixels \( N \):

\[
V_0(\nu_j) = \frac{1}{N} \sum_i \Theta(\alpha(x_i) - \nu), \tag{9}
\]

\[
V_1(\nu_j) = \frac{1}{4N} \sum_i \Delta(\alpha(x_i) - \nu) \sqrt{\alpha_{\phi}^2 + \alpha_\theta^2}, \tag{10}
\]

\[
V_2(\nu_j) = \frac{1}{2\pi N} \sum_i \Delta(\alpha(x_i) - \nu) \left( \frac{2\alpha_{\phi}\alpha_\theta - \alpha_{\phi}^2 - \alpha_\theta^2}{\alpha_{\phi}^2 + \alpha_\theta^2} \right), \tag{11}
\]

where \( \Delta \) is 1 when \( \alpha(x_i) \) is between \( \nu_j \) and \( \nu_{j+1} \) and 0 otherwise. The thresholds are evenly spaced from \( \mu - 3\sigma \) to \( \mu + 3\sigma \), where \( \mu \) is the field value mean and \( \sigma \) is the field value standard deviation. We use HEALPix\(^9\) to calculate first and second derivatives of the fields using a series of spherical harmonic transforms (Górski et al. 2005; Zonca et al. 2019).

### 3.3. Likelihood

![Correlation coefficient matrix of all observables for LSST Y1-like lensing + clustering analysis: 300 MFs and 300 \( C_\ell \). There is some correlation between the MFs and some between the MFs and \( C_\ell \). The ticks represent the range of the observables for the following statistics: \( V_0 \) clustering, \( V_5 \) lensing, \( V_1 \) clustering, \( V_1 \) lensing, \( V_2 \) clustering, \( V_2 \) lensing, \( C_\ell \) clustering, \( C_\ell \) lensing.](image)

We follow a Bayesian procedure to measure our constraining power (Trotta 2017). To calculate likelihoods, we measure our observables on a suite of sky map realisations made using DES Y6-like fiducial values for cosmological parameters and biases. Once we have measured the \( C_\ell \) and/or MFs on each clustering and/or convergence map set, we calculate the likelihood at other cosmologies with a Gaussian likelihood.

\[
L \propto -0.5(\bar{x} - \bar{\mu})^4C^{-1}(\bar{x} - \bar{\mu}), \tag{12}
\]

where \( \bar{\mu} \) represents mean observables over the suite of fiducial realizations, and \( C^{-1} \) is the inverse covariance matrix of the same suite. The simulated data point \( \bar{x} \) is evaluated at other non-fiducial values of the input parameters. The calculations of \( \mu \) and \( C^{-1} \) vary a random seed in map simulations and require a large number of simulations at the fiducial cosmology. To find the covariance of the statistics, we look at the correlation of the concatenation of the observables: MFs \( V_0 \), \( V_1 \), \( V_2 \), and/or the \( C_\ell \). The correlation matrix derived from this covariance is shown in Figure 5.

The number of fiducial realisations must be greater than or equal to the number of data points for the covariance matrix to be invertible. In our MF analysis, the number of data points is found by multiplying the number of thresholds used by three (the number of functionals) and the number of tomographic bins used, and the length of the \( C_\ell \) is determined by the map resolution and binning. If there are too few fiducial simulations, the noise in the covariance overwhelms the signal and can cause errors in the constraining power calculation (Petri et al. 2013). One way to avoid this issue is to use more realisations, but this rapidly becomes time-consuming.

To make the correction for the finite number of simulations, we follow Petri et al. (2013) and make the Anderson (2003) adjustment to the inverse:

\[
\langle \hat{C}^{-1}_\ell \rangle = \frac{N}{N - p - 1} \Sigma^{-1}, \tag{13}
\]

where \( \Sigma^{-1} \) is the sample inverse covariance matrix, \( \hat{C}^{-1}_\ell \) is an estimator for the inverse covariance matrix, \( N \) is the number of realisations minus one (since we find the mean from the data), and \( p \) is the number of data points (Anderson 2003; Hartlap et al. 2006). With \( N_{\text{side}} = 1024 \), there are 300 observables from \( C_\ell \) and MFs each, so we have 600 data points total. We run the fiducial simulation up to 4000 times for each scenario.

### 3.4. Sampling

We use an MCMC to evaluate posterior values of cosmological parameters based on the output of the Gaussian log-likelihood function in Eq. (12). The cosmological parameter priors can be seen in Table 2 and the bias priors can be seen in Table 1. For this project we have wrapped the emcee ensemble sampler (Foreman-Mackey et al. 2013) via the cosmological parameter estimator framework CosmoSIS (Zuntz et al. 2015) to run and parallelise the likelihood calculation. We use 40 walkers to generate emcee chains with tens of thousands of samples. As the sampler takes a given number of iterations to explore the parameter space before it settles onto a stationary distribution, we truncate the first several thousand steps of the chains to eliminate the early ‘burn-in’ piece.

Since we are using a simulation to generate the maps and hence observables at each step of our chain, we would ideally marginalise over a large number of possible random seeds for the chain to find the mean observables for a given cosmology or else utilize a more sophisticated simulation-based inference framework (see, e.g. Cranmer et al. 2020). Instead, we use the same random seed for every point in the chain, relying on the fact that with our configuration choices, the FLASK simulated fields are a smooth function of the cosmological parameters. Though

---

\[^{9}\] http://healpix.sourceforge.net
this process generates a noisy estimate of the model for a point in parameter space, it does yield a correct unnormalised likelihood, and the sampling acceptance criterion is correct.

Our DES Y6-like model has four redshift bins of source galaxies and five redshift bins of lens galaxies, LSST Y1-like has five redshift bins of source galaxies and five redshift bins of lens galaxies, and LSST Y10-like has five redshift bins of source galaxies and ten redshift bins of lens galaxies. We use only a simple linear model for galaxy bias, with the fiducial bias values calculated using

\[ b_i = \frac{1}{D(\pi_i)} \]  

(14)

where \( D(\pi_i) \) is the growth factor as a function of redshifts at the fiducial cosmology.

3.5. Masking

Every survey comes with a sky mask, representing incomplete survey sky coverage, regions blocked by bright stars or other sources, and removals to keep systematic effects out of galaxy maps (Abbott et al. 2022). By limiting the available data, sky masks cause increased variance of fields and affect the estimated MFs values, since MFs are topological descriptors (Shirasaki et al. 2013). As such, MFs are only effective cosmological probes in simulations if masking is taken into account.

In this paper we use a simple polar cap sky fraction mask, in contrast to a more realistic analysis with complex small-scale mask features. For speed, we mask pixels after the simulations and derivatives are computed. However, taking derivatives of a field using harmonic transforms is only possible with a complete (unmasked) field. As such, in an analysis of real data, masked regions must be smoothed and convolved with field values around the mask. This must be carefully addressed in future analyses of real data.

4. RESULTS

Figure 6 shows a comparison of constraints on cosmological parameters from three sets of convergence and clustering statistics: MFs alone, \( C_\ell \) alone, and a combination of MFs and \( C_\ell \). The \( C_\ell \) contours are typical for a 3x2pt analysis; MFs alone show the same degeneracy direction at \( C_\ell \) and show significant constraining power with standard deviations displayed in Table 4. Adding MFs to the 3x2pt analysis does not lead to significantly tighter constraints than \( C_\ell \) alone, with only a 15% decrease in \( S_8 \) standard deviation, implying MFs and clustering \( C_\ell \) contain largely the same information. The parameters \( \Omega_8 \) and \( \sigma_8 \) separately are relatively poorly constrained.

Figure 7 compares constraints for the MF and \( C_\ell \) analysis of convergence, clustering, and a combination of the two. As with standard 3x2pt analyses, adding clustering data to convergence adds significant constraining power to an MF plus \( C_\ell \) analysis; the joint constraint is stronger than the sum of its parts, pointing to internal degeneracies being broken. The decreased error for the combined model can be compared with higher errors of the individual models in Table 4, demonstrating again the power of including clustering maps in the simulation.

Figure 8 shows constraints from the full convergence plus clustering, MF plus \( C_\ell \) analysis for 1, 5, and 15 arcminute Gaussian smoothing. The resulting uncertainties are quantified in Table 4, and are not statistically different. This is largely as expected for given our \( N_{\text{side}} \) of 1024, which corresponds to 3.4 arcminute pixels, but the noticeable decrease in the \( S_8 \) constraint size for 15 arcminute smoothing may reflect the different trade-off for smoothing in MF analyses where noise suppression leads to better recovery of the excursion set boundary. With the tomographic number density that we are using here the measurements are noisy even over relatively large smoothing scales - we find the noise contribution to the value dominates at all the three smoothing scales. Since MFs are higher order statistics, it is not possible to separate out the noise term, but the effect of this noise is included in both the theory and observed values in our MCMC, so does not matter too critically - it is the variation of the noise after that subtraction that affects the constraining power.

The trade-off between larger smoothing kernels to yield lower noise and smaller ones to retain small-scale power depends greatly on the details of the analysis and model. Understanding this relationship between scale and MF constraining power requires further study.

The constraints for the distributions of DES Y6-like, LSST Y1-like, and LSST Y10-like scenarios are shown in Table 4. As expected, surveys with more area, bins, and/or depth have stronger constraints.

| Analysis Type | \( \Omega_8 \) | \( \sigma_8 \) | \( S_8 \) |
|---------------|---------------|---------------|---------------|
| CI            | 0.0040        | 0.0049        | 0.0014        |
| MF            | 0.0132        | 0.0155        | 0.0019        |
| CI+MF         | 0.0045        | 0.0052        | 0.0012        |
| Map Type      | \( \Omega_8 \) | \( \sigma_8 \) | \( S_8 \) |
| Lensing       | 0.0090        | 0.0118        | 0.0029        |
| Clustering    | 0.0101        | 0.0075        | 0.0147        |
| Lensing+Clustering | 0.0045 | 0.0052 | 0.0012 |
| Smoothing (arcmin) | \( \Omega_8 \) | \( \sigma_8 \) | \( S_8 \) |
| 1            | 0.0038        | 0.0045        | 0.0012        |
| 5            | 0.0045        | 0.0052        | 0.0012        |
| 15           | 0.0034        | 0.0014        | 0.0009        |
| Survey (sky fraction) | \( \Omega_8 \) | \( \sigma_8 \) | \( S_8 \) |
| DES Y6-like (12.5%) | 0.0053 | 0.0072 | 0.0017 |
| LSST Y1-like (44%) | 0.0045 | 0.0053 | 0.0012 |
| LSST Y10-like (44%) | 0.0006 | 0.0004 | 0.0006 |

TABLE 4

Errors on cosmological parameters for different scenarios.

5. CONCLUSION

In this paper we investigate the impact of including clustering measurements in analyses of Minkowski functionals combined with power spectra on cosmological constraints. Using simulated convergence and clustering maps, we measure the constraining power of the two statistics for DES Y6-like, LSST Y1-like, and LSST Y10-like surveys. While MFs have been previously proven to be useful in convergence map analyses, here we explore their application to photometric clustering maps for the first time. We compare analyses of varying measurement
Fig. 6.— ΛCDM constraints for power spectra (blue), MFs (green), and joint analysis (pink) measured from convergence + clustering maps using LSST Y1-like redshift bins. Contours show 68% and 95% confidence levels. The left plot shows constraints on $\Omega_m$ and $\sigma_8$, and the right plot shows constraints on $\Omega_m$ and $S_8$. Adding MFs to the 3x2pt analysis does not lead to significantly tighter constraints than $C_\ell$ alone, implying MFs and clustering $C_\ell$ contain largely the same information.

Fig. 7.— ΛCDM constraints for joint MF and $C_\ell$ statistics measured from convergence (green), clustering (blue), convergence and clustering (shaded in pink) maps using LSST Y1-like redshift bins. Contours show 68% and 95% confidence levels. The left plot shows constraints on $\Omega_m$ and $\sigma_8$, and the right plot shows constraints on $\Omega_m$ and $S_8$. As with standard 3x2pt analyses, adding clustering data to convergence adds significant constraining power to an MF plus $C_\ell$ analysis.

statistics, survey data properties, statistical power, and smoothing levels and present constraints on $\Omega_m$, $\sigma_8$, and the better constrained $S_8$.

The MF measurements have the same $S_8$ degeneracy direction as power spectrum measurements, so we focus on this parameter when inspecting the impact of analysis type, map type, and smoothing amount, as it is generally more robust to analysis choices. We find that MFs probe similar cosmological information to clustering measurements in a 3x2pt analysis, and therefore the improvement on cosmological constraints for $\Omega_m$, $\sigma_8$, and $S_8$ is limited, at least in our simplified lognormal map simulation. An important limitation to our analysis, our use of only the auto-correlation $C_\ell$, strengthens this conclusion, since the constraining power of the full 3x2pt analysis is even stronger than that presented here. Other limitations of our analysis, such as our omission of photo-z and shear-related nuisance parameters, are unlikely to change this conclusion, since they affect both MF and $C_\ell$.

If there is value in including MFs in a 3x2pt analysis, it will perhaps become apparent only when we understand and incorporate small scale effects (i.e. baryons, intrinsic alignments, boost factors, mass reconstruction
on small scales, nonlinear galaxy bias) (e.g. Osato et al. 2021) combined with noise. That is, MFs are expected to have more potential on such nonlinear scales.

Similar to the standard 3x2pt case, in the $C_\ell$+MF analysis we find that the addition of clustering measurements has a significant improvement on the constraints. For the LSST Y1-like case, we find an improvement of over 50% for all three cosmological parameters. Therefore, we recommend future higher order statistics be measured from both convergence and clustering maps.

In this project we use curved sky lognormal maps at Gaussian smoothing scales used by Petri et al. (2013) in convergence mapping, but the analysis has the flexibility to be measured from more sophisticated nonlinear, non-Gaussian data in the future. Our conclusions are significantly dependent on the specific form the non-Gaussian field takes, and so that the realism of the lognormal form of the field matters a great deal. On the intermediate scales that we use, previous research has shown that lognormal fields are a good approximation to real lensing fields Clerkin et al. (2016), so the general conclusions drawn here should be reasonable for clustering fields too.

An MF analysis that included all possible smoothing scales simultaneously would incorporate all the information available in a $C_\ell$ measurement Schmalzing and Gorski 1998, see e.g.. While such an analysis is not possible in practice, using a small number of different MF smoothing scales at the same is feasible and could improve the power of MFs.

The maps we use require high computational power, but since the primary constraining power of MFs will come at small scales, using full sky maps may not have been necessary for forecasts. We emphasise, however, that real measurements must consider curved sky effects.

As such, we make simplifications to the maps for computational efficiency. To usefully apply these statistics to real data one must take into account observing conditions, complex masking, baryon effects, and correlations between redshift bins. This is already true for current surveys like DES, HSC, and KiDS, but will be of heightened importance for upcoming efforts like Rubin, Euclid, and Roman. Masks present a particular challenge for many higher order statistics like MFs, since they increase the complexity of derivatives and other calculations.

This investigation motivates the use of MFs combined with $C_\ell$ in future analyses of convergence and clustering. Other potential applications for MFs include measuring them from shear maps, instead of convergence maps, or on combinations of fields. For practical data analysis, it is also critical to speed up or avoid the slow likelihood calculations used here, such as by using a faster implementation, an emulator, fitting function, neural network, or use likelihood-free inference. Building on the MF analysis in this work will improve and inform the statistical model used to constrain cosmological parameters when applied to future data.

6. ACKNOWLEDGMENTS

NG thanks Alex Hall for the useful discussions and Charlie Mpetha for the motivation. TT acknowledges support from the Leverhulme Trust. AA received support from a Kavli Fellowship at Cambridge University. We thank the referee for productive discussion.

Results in this paper made use of many software packages, including Numpy, Scipy, FLASK, CCL, CosmoSIS, NaMaster, Emcee, ChainConsumer, and Healpy/HEALPix.

REFERENCES

T. M. C. Abbott, M. Aguena, A. Alarcon, S. Allam, O. Alves, A. Amon, F. Andrade-Oliveira, J. Annis, et al. Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and weak lensing. Phys. Rev. D, 105(2):023520, 2022. doi:10.1103/PhysRevD.105.023520.

B. Abolfathi, D. Alonso, R. Armstrong, E. Aubourg, H. Awan, Y. N. Babuji, F. E. Bauer, R. Bean, et al. The lst desc de2 simulated sky survey. The Astrophysical Journal Supplement Series, 253(1):31, Mar 2021. ISSN 1538-4365. doi:10.3847/1538-4365/abd62c.
A. Wiegand and D. J. Eisenstein. The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: higher order correlations revealed by germ–grain minkowski functionals. *Monthly Notices of the Royal Astronomical Society*, 467(3):3361–3378, Feb 2017. doi:10.1093/mnras/stx292.

H. S. Xavier, F. B. Abdalla, and B. Joachimi. Improving lognormal models for cosmological fields. *Monthly Notices of the Royal Astronomical Society*, 459(4):3693–3710, Apr 2016. ISSN 1365-2966. doi:10.1093/mnras/stw874.

Z. Zhang, C. Chang, P. Larsen, L. F. Secco, J. Zuntz, and the LSST Dark Energy Science Collaboration. Transitioning from Stage-III to Stage-IV: Cosmology from galaxy×CMB lensing and shear×CMB lensing. 2021. doi:10.48550/ARXIV.2111.04917

A. Zonca, L. Singer, D. Lenz, M. Reinecke, C. Rosset, E. Hivon, and K. Gorski. healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in python. *Journal of Open Source Software*, 4(35):1298, March 2019. doi:10.21105/joss.01298.

J. Zuntz et al. Cosmosis: Modular cosmological parameter estimation. *Astronomy and Computing*, 12:45–59, Sep 2015. ISSN 2213-1337. doi:10.1016/j.ascom.2015.05.005.

D. Zürcher, J. Fluri, R. Sgier, T. Kacprzak, M. Gatti, C. Doux, L. Whiteway, A. Refregier, et al. Dark energy survey year 3 results: Cosmology with peaks using an emulator approach. *Monthly Notices of the Royal Astronomical Society*, 511(2):2075–2104, Jan 2022. doi:10.1093/mnras/stac078.

D. Zürcher, J. Fluri, R. Sgier, T. Kacprzak, and A. Refregier. Cosmological forecast for non-gaussian statistics in large-scale weak lensing surveys. *Journal of Cosmology and Astroparticle Physics*, 2021(01):028–028, Jan 2021. doi:10.1088/1475-7516/2021/01/028.