

# Testing the lognormality of the galaxy and weak lensing convergence distributions from Dark Energy Survey maps

L. Clerkin,<sup>1\*</sup> D. Kirk,<sup>1</sup> M. Manera,<sup>1</sup> O. Lahav,<sup>1</sup> F. Abdalla,<sup>1,2</sup> A. Amara,<sup>3</sup> D. Bacon,<sup>4</sup> C. Chang,<sup>3</sup> E. Gaztañaga,<sup>5</sup> A. Hawken,<sup>6</sup> B. Jain,<sup>7</sup> B. Joachimi,<sup>1</sup> V. Vikram,<sup>7</sup> T. Abbott,<sup>8</sup> S. Allam,<sup>9</sup> R. Armstrong,<sup>10</sup> A. Benoit-Lévy,<sup>1,11,12</sup> G. M. Bernstein,<sup>7</sup> R. A. Bernstein,<sup>13</sup> E. Bertin,<sup>11,12</sup> D. Brooks,<sup>1</sup> D. L. Burke,<sup>14,15</sup> A. Carnero Rosell,<sup>16,17</sup> M. Carrasco Kind,<sup>18,19</sup> M. Crocce,<sup>5</sup> C. E. Cunha,<sup>14</sup> C. B. D’Andrea,<sup>4,20</sup> L. N. da Costa,<sup>16,17</sup> S. Desai,<sup>21,22</sup> H. T. Diehl,<sup>9</sup> J. P. Dietrich,<sup>21,22</sup> T. F. Eifler,<sup>7,23</sup> A. E. Evrard,<sup>24,25</sup> B. Flaugher,<sup>9</sup> P. Fosalba,<sup>5</sup> J. Frieman,<sup>9,26</sup> D. W. Gerdes,<sup>25</sup> D. Gruen,<sup>14,15</sup> R. A. Gruendl,<sup>18,19</sup> G. Gutierrez,<sup>9</sup> K. Honscheid,<sup>27,28</sup> D. J. James,<sup>8</sup> S. Kent,<sup>9</sup> K. Kuehn,<sup>29</sup> N. Kuropatkin,<sup>9</sup> M. Lima,<sup>16,30</sup> P. Melchior,<sup>10</sup> R. Miquel,<sup>31,32</sup> B. Nord,<sup>9</sup> A. A. Plazas,<sup>23</sup> A. K. Romer,<sup>33</sup> A. Roodman,<sup>15,26</sup> E. Sanchez,<sup>34</sup> M. Schubnell,<sup>24</sup> I. Sevilla-Noarbe,<sup>34</sup> R. C. Smith,<sup>8</sup> M. Soares-Santos,<sup>9</sup> F. Sobreira,<sup>16,35</sup> E. Suchyta,<sup>7</sup> M. E. C. Swanson,<sup>19</sup> G. Tarle<sup>25</sup> and A. R. Walker<sup>8</sup>

*Affiliations are listed at the end of the paper*

Accepted 2016 August 23. Received 2016 July 29; in original form 2016 May 5

## ABSTRACT

It is well known that the probability distribution function (PDF) of galaxy density contrast is approximately lognormal; whether the PDF of mass fluctuations derived from weak lensing convergence ( $\kappa_{\text{WL}}$ ) is lognormal is less well established. We derive PDFs of the galaxy and projected matter density distributions via the counts-in-cells (CiC) method. We use maps of galaxies and weak lensing convergence produced from the Dark Energy Survey Science Verification data over 139 deg<sup>2</sup>. We test whether the underlying density contrast is well described by a lognormal distribution for the galaxies, the convergence and their joint PDF. We confirm that the galaxy density contrast distribution is well modelled by a lognormal PDF convolved with Poisson noise at angular scales from 10 to 40 arcmin (corresponding to physical scales of 3–10 Mpc). We note that as  $\kappa_{\text{WL}}$  is a weighted sum of the mass fluctuations along the line of sight, its PDF is expected to be only approximately lognormal. We find that the  $\kappa_{\text{WL}}$  distribution is well modelled by a lognormal PDF convolved with Gaussian shape noise at scales between 10 and 20 arcmin, with a best-fitting  $\chi^2/\text{dof}$  of 1.11 compared to 1.84 for a Gaussian model, corresponding to  $p$ -values 0.35 and 0.07, respectively, at a scale of 10 arcmin. Above 20 arcmin a simple Gaussian model is sufficient. The joint PDF is also reasonably fitted by a bivariate lognormal. As a consistency check, we compare the variances derived from the lognormal modelling with those directly measured via CiC. Our methods are validated against maps from the MICE Grand Challenge  $N$ -body simulation.

**Key words:** gravitational lensing: weak – cosmology: observations – large-scale structure of Universe.

## 1 INTRODUCTION

It was first noted by Hubble that the distribution of galaxies in angular cells on the celestial sphere is well approximated by a lognormal (Hubble 1934). This has been confirmed observationally (Coles &

\*E-mail: lucyc@star.ucl.ac.uk

Jones 1991; Wild et al. 2005) as well as in  $N$ -body simulations (Bernardeau 1994a; Bernardeau & Kofman 1995; Kayo, Taruya & Suto 2001), which have shown that the underlying mass density field is expected to be lognormal.

Since the weak lensing convergence field along the line of sight is a weighted projection of the mass density contrast field, one might suspect that the lognormal distribution is a reasonable, if not exact, model of this too. This has been tested on numerical simulations and a lognormal probability distribution function (PDF) shown to be a reasonable model (Taruya et al. 2002; Hilbert, Hartlap & Schneider 2011). Even better fits to the convergence PDF, particularly in the tails of the distribution, have been obtained by generalizations of a lognormal PDF (Das & Ostriker 2006; Joachimi, Taylor & Kiessling 2011; Takahashi et al. 2011).

The Dark Energy Survey (DES; Dark Energy Survey Collaboration 2005, 2016a,b) presents an excellent opportunity to study both of these fields. DES was specifically conceived to produce cutting edge science from four different cosmological probes – large-scale structure (LSS), weak gravitational lensing, galaxy clusters and supernovae (SNe) – using the same instrument. The full survey involves five years of observations, currently in progress. In this paper, we focus on data produced during the pre-survey Science Verification (SV) series of observations.

This early data from DES allowed for the construction of two types of density fields. One is from luminous matter, i.e. galaxies of various types,  $\delta_g$ , which are biased tracers of the underlying dark matter field,  $\delta_m$ . The other uses the weak lensing of galaxy shapes to construct a convergence, or  $\kappa$  map (Chang et al. 2015; Vikram et al. 2015) that is directly sensitive to the integrated dark matter field out to the lensed galaxies.

Both maps trace the underlying density distribution in the Universe. Galaxies are biased tracers of matter density, preferentially clustering in overdense regions. Galaxy density contrast can then be considered a biased local tracer of the density field.

Weak lensing convergence on the other hand responds directly to the underlying density field and is therefore unbiased. However, gravitational deflection of light is a cumulative effect, sensitive to the integrated matter density along the line of sight from source galaxy to observer. The convergence field for a given galaxy source distribution therefore gives us information about the cumulative density field between observer and source, with the exact contribution of matter at different distances along the line of sight governed by the lensing efficiency function.

The purpose of the present study is to analyse the galaxy and mass maps from DES SV simultaneously, testing the two maps separately for lognormality, as well as analysing the joint distribution. To our knowledge, this is the first time that the lognormality of the weak lensing convergence field alone has been tested using data rather than numerical simulation (Taruya et al. 2002), and the first time the joint distribution has been tested for lognormality.

The *counts-in-cells* (CiC) method (e.g. Hubble 1934; White 1979; Gaztanaga 1992; Szapudi 1997; Bernardeau et al. 2002) is a natural way to measure the individual and joint PDFs. The CiC technique splits up a particular data set into spatial cells, in two or three dimensions, and takes an aggregate of the available information inside each cell. Statistical variation between cells then provides information on the properties of that cosmological field.

DES observations are ideally suited to this sort of analysis. The fact that DES provides a joint galaxy survey and convergence map data set produced from the same observations makes it easier to ensure consistency between data and to control for systematics. The SV data we use were taken before the start of the full five year

survey, covers  $139 \text{ deg}^2$  to full survey depth and forms a test bed for the kind of analyses planned on the final DES data. All of the analyses in this work are done first on galaxy and convergence maps from MICE simulations (Carretero et al. 2014; Fosalba et al. 2014, 2015; Crocce et al. 2015a; Hoffmann et al. 2015) in order to test our methodology.

The outline of the paper is as follows. In Section 2, we review the theory and formalism used. We describe the galaxy and weak lensing convergence maps from MICE simulations and DES used in Section 3, and summarize our CiC method in Section 4. In Section 5, we validate our CiC method on MICE Grand Challenge  $N$ -body simulations, checking that we see the expected lognormality in MICE  $\delta_g$  and the noise-free convergence. In Section 6, we present lognormal fits to the individual DES galaxy and convergence field distributions as well as their joint distribution. We check the validity of the lognormal model by measuring the variance of the fields and comparing this to the variance derived under the assumption of lognormality. We discuss the results in Section 7, and in the Appendices we give the formalism used to calculate moments from CiC, test the impact of systematic effects and confirm that assumptions we make in our method do not significantly affect our results.

## 2 LOGNORMAL MODELLING OF COSMIC FIELDS

Lognormal distributions are very common in nature, from the sizes of clouds, pebbles on a beach or crystals in ice cream; the length of sentences or words in a conversation; to populations of bacteria (see Limpert, Stahel & Abbt 2001; Gaskell 2004 and references therein). Many of these examples involve multiplicative processes, of either merging or fragmentation. Any process that can be written as a product of terms will, if there are many terms, tend to a lognormal distribution. This is because if a process  $X$  can be written as a product of independent terms, then  $\ln(X)$  will be a sum of independent terms, and via the central limit theorem these will be normally distributed. So  $\ln(X)$  is normally distributed, or,  $X$  is lognormally distributed.

There are many examples of the hierarchical merging or fragmenting of structure leading to lognormal distributions, such as the initial mass function of field stars, explained in terms of cloud fragmentation (Zinnecker 1984); X-ray flux variations, suggesting lognormal distribution of emitting regions (Gaskell 2004); luminosity functions of central galaxies, explained in terms of BCGs being formed by steps of mass adding/stripping (e.g. Taghizadeh-Popp, Heinis & Szalay 2012); and the angular momentum of disc galaxies (Marr 2015).

In this paper, we test the lognormality of the distribution of matter in the Universe via the galaxy density contrast field,  $\delta_g$ , and via the weak lensing convergence field,  $\kappa_{\text{WL}}$ . Each approach has particular observational and astrophysical noise associated with it, which we discuss and propose models for in this section.

### 2.1 Galaxies

In the standard picture of gravitational instability, the primordial density fluctuations that are the precursor of all structure in the universe are assumed to be a random Gaussian field. Once they enter the non-linear regime, with finite rms fluctuations, their PDF must deviate from a Gaussian to avoid non-zero probabilities being attached to regions with negative densities (Fry 1985). The exact form of the PDFs in this regime is not known but there are various phenomenological models that are fully specified statistically and satisfy the common sense requirement that the matter density

$\rho \geq 0$  (e.g. Saslaw & Hamilton 1984; Suto, Itoh & Inagaki 1990; Gaztanaga & Yokoyama 1993; Lahav et al. 1993; Suto 1993; Ueda & Yokoyama 1996).

One such model that is widely used is the lognormal. As well as being completely specified statistically and always having  $\rho \geq 0$ , it becomes arbitrarily close to Gaussian statistics at early times and has the advantage that it can be handled analytically. The merits of this model in a cosmological context are discussed extensively in Coles & Jones (1991). They explain possible motivations for using the lognormal model as empirical; kinematic; an application of the central limit theorem (as described above); and importantly, simplicity. It is one of the few random fields for which interesting quantities such as its moments can be calculated analytically.

It should be noted that despite these compelling motivations to use a lognormal in the statistical treatment of density perturbations, it does have shortcomings. In particular, it is not uniquely specified by its moments; many distributions can lead to the same set of moments. It must then be the case that information is lost in going from a lognormal field to its moments, an effect quantified in Carron (2011). However, it remains a popular and useful tool in analysis of the mass density contrast field.

Galaxies are biased tracers of the mass density contrast field. The 1D lognormal distribution of galaxy density contrast  $\delta_g = (\rho - \bar{\rho})/\bar{\rho}$  is given by

$$f(\delta_g)d\delta_g = \frac{1}{\omega\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\omega^2}\right) dx, \quad (1)$$

where  $x = \ln(1 + \delta_g) + \omega^2/2$  and  $\omega^2$  is the variance of the corresponding normal distribution  $f[\ln(1 + \delta_g)]$ . The offset  $\omega^2/2$  ensures that  $\langle \delta_g \rangle = 0$ . The width  $\omega$  is then the single free parameter of the 1D lognormal distribution.

If the lognormal distribution correctly describes the data, the variance of the overdensities will be related to the width,  $\omega$ , of the underlying Gaussian distribution by

$$\sigma_g^2 \equiv \langle \delta_g^2 \rangle = e^{\omega^2} - 1. \quad (2)$$

Due to the discrete nature of galaxies, shot noise is present. Assuming Poisson sampling of galaxies, the shot noise in the measurement of the distribution of galaxy overdensities can be accounted for by convolving the lognormal model with a Poisson distribution. The PDF of the galaxy counts  $N$  in a cell of given size is then given by

$$P(N) = \int_{-1}^{\infty} \frac{\bar{N}^N (1 + \delta_g)^N}{N!} e^{-N(1 + \delta_g)} f(\delta_g) d\delta_g. \quad (3)$$

This Poisson-sampled lognormal distribution has been shown to be a good fit to different galaxy populations in Coles & Jones (1991), Blanton (2000) and Wild et al. (2005). In this work, the median number of DES galaxies in a cell at the smallest scale considered (5 arcmin) is around 500. This corresponds to a relative uncertainty in the galaxy count of  $\sim 5$  per cent, so the shot noise term is important.

## 2.2 Weak lensing convergence

Various expressions for the convergence PDF have been proposed (Munshi & Jain 2000; Valageas 2000; Kainulainen & Marra 2011). The lognormal model has the advantage – as in the case of the matter density contrast – of mathematical convenience, while offering the chance to extract more information than assuming a purely Gaussian model for the convergence field. Following a study that showed

that a lognormal transformation of the matter density contrast increases the signal-to-noise (Neyrinck, Szapudi & Szalay 2009), Seo et al. (2011) performed an analogous study of the weak lensing convergence. They found that such a transform, when applied to the positively offset convergence, decorrelated angular frequencies and increased the signal-to-noise in the transformed power spectrum.

The convergence field along a line of sight can be expressed as a weighted projection of the mass density contrast field:

$$\kappa(\theta) = \int_0^{\chi_{\text{hor}}} d\chi w(\chi) \delta[r(\chi)\theta, \chi], \quad (4)$$

where  $\chi$  is the comoving distance,  $\chi_{\text{hor}}$  is the angular diameter distance to the horizon and  $\delta[r(\chi)\theta, \chi]$  is the underlying matter density contrast field.  $w(\chi)$  is a geometrical weight function that depends on the relative separations of sources, lens and the observer (see e.g. Mellier 1999; Bartelmann & Schneider 2001 for reviews). It takes the form

$$w(\chi) = \frac{3}{2} \left(\frac{H_0}{c}\right)^2 \frac{\chi \Omega_0}{a(\chi)} \int_{\chi}^{\chi_{\text{hor}}} d\chi' n(\chi') \frac{\chi' - \chi}{\chi'}, \quad (5)$$

where  $n(\chi)$  is the source galaxy distribution.

The distribution of  $\kappa$  is not expected to be exactly lognormal, even if  $\delta$  is, since  $\kappa$  is a weighted projection of the mass density contrast field along line of sight. However, simulations have shown (Taruya et al. 2002) that the convergence field is well approximated by a lognormal outside the regime of extremely high  $\kappa$ . Hence, we choose in this work to model the noise-free  $\kappa$  field distribution with a shifted lognormal,

$$P(\kappa) = \begin{cases} \frac{\exp\left[-\frac{(\ln(\kappa_0 + \kappa) - \mu)^2}{2\omega^2}\right]}{(\kappa_0 + \kappa)\sqrt{2\pi\omega}} & \text{for } \kappa > -\kappa_0, \\ 0 & \text{otherwise,} \end{cases} \quad (6)$$

where the shift  $\kappa_0 = -\kappa_{\text{min}}$  and is called the ‘minimum convergence parameter’ (Hilbert et al. 2011). The lowest possible value of  $\kappa$  is given by  $-\kappa_0$  since the lognormal is only defined for a positive range. Parameters  $\mu$  and  $\omega$  are the mean and standard deviation of normally distributed  $\ln(\kappa_0 + \kappa)$ . If the convergence distribution is correctly described by the lognormal model, its moments about the origin are given by

$$\langle \kappa^n \rangle = \int_{-\infty}^{\infty} \kappa^n P(\kappa) d\kappa, \quad (7)$$

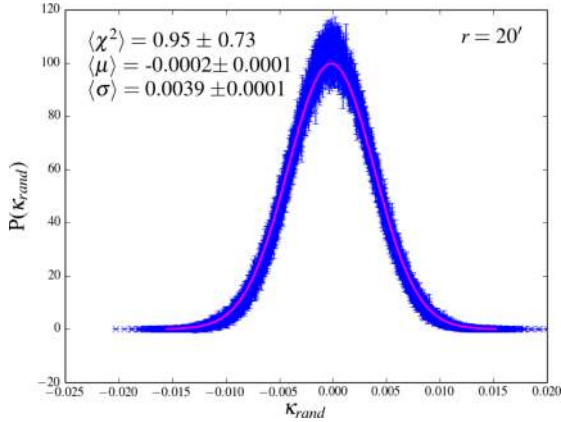
hence the mean convergence distribution is

$$\langle \kappa \rangle = e^{\mu + \omega^2/2} - \kappa_0 \quad (8)$$

and its variance is

$$\sigma_{\kappa}^2 \equiv \langle \kappa^2 \rangle - \langle \kappa \rangle^2 = e^{2\mu + \omega^2} (e^{\omega^2} - 1). \quad (9)$$

The value assigned to  $\kappa_0$  is a modelling choice that can be approached in several ways. The minimum measured value of  $\kappa$  could be used, but this is a noisy quantity and should not be used unless one has access to many realizations of  $\kappa$ . Or, treating  $\kappa$  analogously to  $\delta_g$ , we could consider a theoretical minimum corresponding to the convergence we would see, for a given source distribution, if the mass distribution was a pure void along the entire line of sight. For the MICE source distribution used in this work, this value is  $-0.050$ , and for the DES source distribution it is  $-0.053$ . Alternatively,  $\kappa_0$  can be derived by matching moments of the distribution (Xavier, Abdalla & Joachimi 2016), or it can simply be fitted as a free parameter (e.g. Hilbert et al. 2011). It is shown in Xavier



**Figure 1.** Demonstration of the Gaussianity of the noise in DES weak lensing convergence,  $\kappa$ , at a smoothing scale of 20 arcmin. Probability density distributions of the 100 realizations of the  $\kappa$  map, in which shears were randomized, are shown in blue, with jackknife errors. A Gaussian PDF was fitted to each, with the mean of the best-fitting PDFs shown in magenta. A Gaussian model is an excellent fit to the noise, with average goodness of fit  $\chi^2/\text{dof}=0.95 \pm 0.73$ .

et al. (2016) that for a simple lognormal distribution these different methods will give roughly the same value  $\kappa_0$ , but for a sum along the line of sight of correlated lognormals, they give different values of  $\kappa_0$ . If a simple lognormal model is assumed, then it cannot exactly reproduce every aspect (such as both the shape and moments) of the true convergence CiC PDF so we choose, where possible, to treat  $\kappa_0$  as a free parameter and fit it jointly with the lognormal width.

As for galaxies, we need to modify the lognormal to account for noise. The DES  $\kappa$  map is constructed (see Section 3.2) from measurements of shear, which is the change in the ellipticity of galaxies resulting from weak gravitational lensing. Since galaxies are intrinsically elliptical (i.e. in the absence of lensing), the observed shear is the sum of this intrinsic ellipticity and the shear caused by lensing. The variance of the intrinsic ellipticity, called shape noise, is the dominant source of noise in shear measurements, typically by a factor of more than an order of magnitude. An estimate of the shape noise in the DES  $\kappa$  map is provided by the 100 noise realizations described in Section 3.2.

To analyse the shape of the noise distribution, we construct PDFs via CiC (as described in Section 4) on each of the 100 maps. The resulting distributions appear Gaussian, as shown in Fig. 1, where the thick blue curve is made up of 100 superimposed noise distributions with jackknife error bars, and the magenta line shows the mean best-fitting Gaussian PDF. A Gaussian model provides an excellent fit, with average goodness of fit over the 100 maps  $\chi^2/\text{dof} = 0.95 \pm 0.73$ .

We therefore propose that the 1D probability distribution for the weak lensing convergence field is then given by a convolution of a lognormal distribution with a noise contribution modelled as Gaussian:

$$f(\kappa) = \frac{1}{\sqrt{2\pi}\sigma_n} \int_{-\infty}^{\infty} \exp\left[-\frac{(\kappa' - \kappa)^2}{2\sigma_n^2}\right] P(\kappa') d\kappa', \quad (10)$$

where  $P(\kappa)$  is the noise-free lognormal distribution given in equation (6), and  $\sigma_n$  is the Gaussian width of the shape noise. In this work,  $\sigma_n$  is determined from the 100 noise realizations.

### 2.3 Joint galaxy and weak lensing convergence field

Extending the approach taken with the 1D distributions, we can also try to model the joint distribution of galaxy density contrast  $\delta_g$  and weak lensing convergence  $\kappa_{\text{WL}}$  as a bivariate lognormal with PDF  $f(\delta_g, \kappa_{\text{WL}})$ . It is argued in Xavier et al. (2016) that the bivariate lognormal is limited in modelling the joint density and convergence distribution since the lognormal distribution's shape can prevent certain correlations or covariance matrices from being attainable. However, as in the case of the 1D distributions, the bivariate lognormal provides a relatively easy to implement, analytic model of the joint distribution, and it is interesting to see if this provides a reasonable fit to the data.

Following the notation used in Wild et al. (2005), the bivariate distribution is given by

$$f(\delta_g, \kappa) = \frac{|V|^{-1/2}}{2\pi} \exp\left[-\frac{(\tilde{g}_\delta^2 + \tilde{g}_\kappa^2 - 2r_{\text{LN}}\tilde{g}_\delta\tilde{g}_\kappa)}{2(1-r_{\text{LN}}^2)}\right], \quad (11)$$

where  $g_x = \ln(x) - \langle \ln(x) \rangle$  and  $\tilde{g}_x = g_x/\omega_x$ . Here  $\omega_x$  is the lognormal width of  $x$ , with  $x = (1 + \delta_g)$  and  $x = (1 + \kappa/\kappa_0)$  for the galaxy and convergence fields, respectively.

The lognormal correlation coefficient  $r_{\text{LN}}$  is given by

$$r_{\text{LN}} = \frac{\langle g_\delta g_\kappa \rangle}{\omega_\delta \omega_\kappa} \equiv \frac{\omega_{\delta\kappa}^2}{\omega_\delta \omega_\kappa} \quad (12)$$

and  $|V|$  is the determinant of the covariance matrix

$$V = \begin{pmatrix} \omega_\delta^2 & \omega_{\delta\kappa}^2 \\ \omega_{\delta\kappa}^2 & \omega_\kappa^2 \end{pmatrix}. \quad (13)$$

Note that  $r_{\text{LN}}$  and  $V$  are defined in log-density space, and so  $r_{\text{LN}}$  is not the same as the (linear) Pearson correlation coefficient  $\rho$ . The conditional probability

$$\begin{aligned} f(\delta_g|\kappa) &= \frac{f(\delta_g, \kappa)}{f(\kappa)} \\ &= \frac{\omega_\kappa}{(2\pi|V|)^{1/2}} \exp\left[-\frac{(\tilde{g}_\delta - r_{\text{LN}}\tilde{g}_\kappa)^2}{2(1-r_{\text{LN}}^2)}\right]. \end{aligned} \quad (14)$$

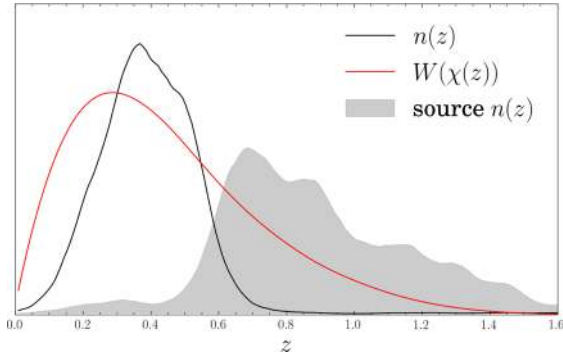
Since  $f(\delta_g, \kappa) = f(\kappa)f(\delta_g|\kappa)$ , we can combine equations (10) and (14) to give an expression for the joint PDF to which the appropriate Poisson and Gaussian convolutions can be applied, to account for shot and shape noise, respectively. The joint probability distribution is then

$$\begin{aligned} P(N, \kappa) &= \int_{-1}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left[-\frac{(\kappa' - \kappa)^2}{2\sigma_n^2}\right] f(\kappa) \\ &\quad \times \frac{\bar{N}^N (1 + \delta_g)^N}{N!} e^{-\bar{N}(1 + \delta_g)} f(\delta_g|\kappa) d\delta_g d\kappa', \end{aligned} \quad (15)$$

where  $N$  is the expected number of galaxies in a given cell and  $\kappa$  is the average convergence expected in a cell.

### 3 THE DATA

This paper uses the DES SV galaxy and shape catalogues. The SV data were gathered between 2012 November and 2013 February, shortly after DECam (Flaugher et al. 2015) commissioning, and before the beginning of the (five year) DES survey proper in 2013 August. The operation of the camera, survey planning, data analysis and reduction were all tested in preparation for starting year one of DES itself. The SV goal was to reproduce the properties of the full five year DES survey over a much smaller sky area.



**Figure 2.** Redshift kernels of the observables considered in this paper: the galaxy redshift distribution of the DES Benchmark galaxy sample using the best-fitting Skynet photo- $z$  estimation (black line), and the lensing efficiency function of the sources used to make the DES  $\kappa$  map (red line). Also shown is the redshift distribution of the source galaxies (shaded region). Each is shown with an arbitrary normalization to make comparison easier.

Five optical filters (*grizY*) are used, with exposure times of 90 s for *griz* and 45 s for *Y*. The final median depth in our region of interest, per band, was  $g \sim 24.0$ ,  $r \sim 23.9$ ,  $i \sim 23.0$  and  $z \sim 22.3$ .

In total, the SV data covered  $\sim 250$  deg<sup>2</sup> at close to the nominal depth of the full DES survey. The observing footprint was divided into regions to maximize overlap with other surveys and with several small fields used for SN searches.

In this paper, we concentrate on a large contiguous region of  $\sim 139$  deg<sup>2</sup> called the SPT-E field due to its overlap with the South Pole Telescope cosmic microwave background survey. This amount of contiguous data makes the SV SPT-E field a powerful data set in its own right, particularly for weak gravitational lensing where it rivals the full CFHTLenS (Erben et al. 2013) and is only slightly shallower.

### 3.1 DES galaxy sample

We use a particular subset of the DES SV galaxy catalogue known as the ‘Benchmark’ sample (Crocco et al. 2015b). First, a catalogue of galaxies suitable for LSS analysis was constructed from the SV data and dubbed the ‘Gold’ sample (Rykoff et al. 2016). Objects were included if detected in all five of the DES photometric bands. This covered  $\sim 244$  deg<sup>2</sup>, restricted to  $Dec. > -61$  to avoid the Large Magellanic Cloud and R Doradus regions. In addition, the Gold catalogue included masking of satellite trails and other artefacts, removal of regions where colours are severely affected by stray light and the application of additional stellar locus correction (Kelly et al. 2014).

From this Gold sample, the Benchmark sample was selected for cosmological analysis by imposing the additional conditions:

- (i)  $18.0 < i < 22.5$ ;
- (ii)  $0 < g - r < 3$ ,  $0 < r - i < 2$  and  $0 < i - z < 3$ ;
- (iii)  $wavg\_spread\_model > 0.003$  (star–galaxy separation);
- (iv)  $60 < RA < 95$  and  $-62 < Dec. < -40$  (SPT-E);

where  $g$ ,  $r$ ,  $i$  and  $z$  refer to SE extractor’s MAG\_AUTO quantity. The cuts on position restrict our analysis to the SPT-E region. The redshifts used in this paper come from the Skynet photo- $z$  pipeline (Graff & Feroz 2013; Bonnett et al. 2016).

The galaxy redshift distribution is shown in Fig. 2. The redshift range we use throughout this paper is  $0.1 < z < 1.5$ , chosen as in this region the galaxy redshift distribution overlaps with the lensing

efficiency function used to make the DES  $\kappa_{WL}$  map (see the next section).

### 3.2 DES $\kappa$ map

Shear measurement on DES SV galaxy images was performed with two independent pipelines: IM3SHAPE<sup>1</sup> (Zuntz et al. 2013) and NGMIX<sup>2</sup> (Sheldon 2014).

Extensive testing of both codes was carried out by the DES collaboration (see Jarvis et al. 2016 for details), and both pipelines passed all requirement tests for measurement of cosmic shear with the SV data set. A number of cuts were applied to both catalogues to remove stars, spurious detections, poor measurements and other effects that could bias shear measurement; these are also described in Jarvis et al. (2016).

Shear measurements for a given galaxy are headless vectors and the cosmic shear field is therefore a spin-2 quantity. To allow us to perform our CiC analysis on a scalar quantity, we work with maps of weak lensing convergence,  $\kappa$ , a spin-0 field. This  $\kappa$ -reconstruction was performed using the Kaiser–Squires method (Kaiser & Squires 1993), and the production and initial analysis of these  $\kappa$  maps is described in detail in Vikram et al. (2015).

The Kaiser–Squires reconstruction method uses the relation of the Fourier transform of the observed shear,  $\hat{\gamma}_\ell$ , to that of the convergence,  $\hat{\kappa}_\ell$ ,

$$\hat{\kappa}_\ell = D_\ell^* \hat{\gamma}_\ell, \quad (16)$$

$$D_\ell = \frac{\ell_1^2 - \ell_2^2 + 2i\ell_1\ell_2}{|\ell|^2}, \quad (17)$$

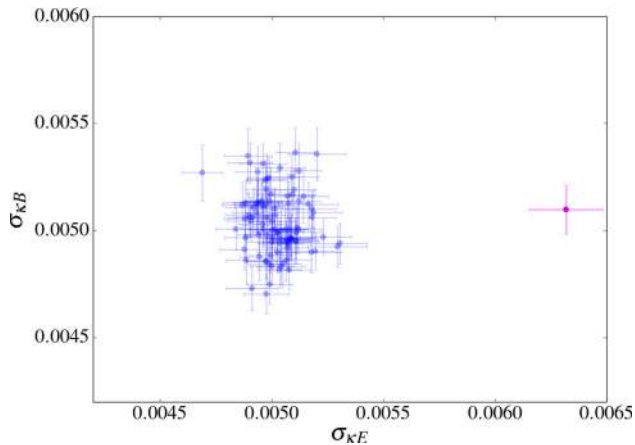
where  $\ell_i$  are the Fourier counterparts of the angular coordinates,  $\theta_i$ ,  $i = 1, 2$ . The inverse Fourier transform of equation (16) gives the convergence for the observed field in real space. In the absence of noise, systematics and masking, the convergence will be a real (spin-0) quantity. In reality, these effects produce a non-zero imaginary component. It is most convenient to express the real part of the convergence map as a map of curl-free E-modes, and the imaginary part as divergence-free B-modes. The  $\kappa$  maps have pixels of size 2 arcmin. For use in this analysis, the original flat sky  $\kappa$  maps are transformed into HEALPIX (Gorski et al. 2005) maps at resolution  $N_{\text{side}} = 4096$ . This is done by dividing each pixel of the flat sky maps into 25 sub-pixels, and creating a HEALPIX map by combining these sub-pixels. This procedure reduces inaccuracies in changing from one mapping system to another, and in tests gives the same angular power spectrum measurements as the flat sky map to well within the errors.

The source galaxy selection used to construct the  $\kappa$  map used in this paper took galaxies with redshifts in the range  $0.6 < z < 1.3$ . The resulting redshift efficiency function is shown in Fig. 2. The lensing efficiency function peaks at  $z \sim 0.3$ , and our selection of galaxies at  $0.1 < z < 0.5$  overlaps significantly with the range of redshifts to which the  $\kappa$  map is sensitive.

In addition to the E-mode  $\kappa$  map, we make use of a number of other products made in the course of the DES mass-mapping analysis. A B-mode map was constructed by rotating the measured galaxy

<sup>1</sup>The open source code can be downloaded at <https://bitbucket.org/joezuntz/im3shape/>

<sup>2</sup>The open source code can be downloaded at <https://github.com/esheldon/ngmix>



**Figure 3.** Standard deviations of the  $\kappa_E$ ,  $\kappa_B$  signal (magenta) and 100 realizations in which the shears have been randomized (blue) at a cell size of 20 arcmin. The random realizations of  $\kappa_E$  give an estimate of the shape noise contribution to  $\kappa_E$ ; this accounts for 80 per cent of the  $\kappa_E$  signal. The  $\kappa_B$  signal is also a good estimate of the shape noise, with the standard deviation of the  $\kappa_B$  signal agreeing with the rms standard deviation of  $\kappa_E$  random realizations within 2 per cent. These standard deviations are calculated via CiC and errors are from jackknife sampling.

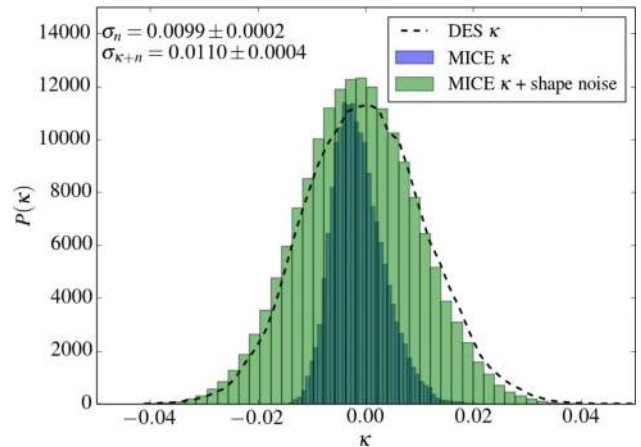
ellipticities by 45 deg. The physical process of weak gravitational lensing does not induce B-modes in the convergence field so the B-mode map is a test of systematic effects in our observations, shear measurement and  $\kappa$ -reconstruction; it should be consistent with zero within our reconstruction noise. We will refer to the B-mode reconstructed map as  $\kappa_B$ .

In addition to the E- and B-mode maps, we also make use of a series of noise-only realizations, made by taking the galaxy shape catalogue and rotating the measured shape of each galaxy by some random angle.  $\kappa$  maps were then constructed from each randomized catalogue in the usual way. This has the effect of destroying all cosmological information in the resulting maps, while retaining the same noise properties as the data (because the distribution of galaxies on the sky and in redshift remains the same, as does the overall ellipticity distribution across the sample). 100 of these noise realizations were made, and we use them to estimate the noise contribution in our measurement, as described in more detail in Section 2.2.

Fig. 3 shows the standard deviations of the  $\kappa_E$ ,  $\kappa_B$  signal (magenta) and 100 noise realizations (blue) for a cell size of 20 arcmin. This shows that the shape noise (given by the random realizations of  $\kappa_E$ ) accounts for 80 per cent of the  $\kappa_E$  signal, underlining the importance of accounting for shape noise in our modelling (as described in Section 3.2). The shape noise dominates the signal most at small scales, accounting for 89 per cent of the signal at 10 arcmin and dropping to 64 per cent at 40 arcmin. We can see that the  $\kappa_B$  signal is also a good estimate of the shape noise, with the standard deviation of the  $\kappa_B$  signal agreeing with the rms standard deviation of  $\kappa_E$  random realizations within 2 per cent. These standard deviations are calculated via CiC (see Appendix C for a prescription for calculating moments from CiC) and errors are from jackknife sampling (see Section 4).

### 3.3 MICE simulations

We validate our measurement of CiC from DES SV data using a special set of mock catalogues produced from  $N$ -body simulations for the DES collaboration. These come from the Marenstrum



**Figure 4.** Distribution of MICE  $\kappa_{WL}$  at an angular scale of 10 arcmin when DES-like shape noise is added. An estimate of the width of the shape noise distribution is obtained by fitting a Gaussian to the 100 random realizations of DES  $\kappa_{WL}$ . A noise contribution drawn from a Gaussian of this width is added to MICE  $\kappa_{WL}$  at the level of the cells used to construct the CiC distribution. The darker, narrow histogram is that of the shape noise-free MICE  $\kappa_{WL}$ ; the lighter histogram shows the distribution once the Gaussian shape noise is added; the black dashed line shows the observed distribution of DES  $\kappa_{WL}$ . The Gaussian width of the DES shape noise estimate,  $\sigma_n$ , is 0.0099 at this scale, which is 89 per cent of the width of the resulting noisy MICE  $\kappa_{WL}$  distribution.

Institut de Ciències de l’Espai Grand Challenges (MICE-GC hereafter) lightcone  $N$ -body simulation and associated halo catalogue.

These simulations have been used to produce mock galaxy catalogues for  $\sim 200$  million galaxies over  $5000 \text{ deg}^2$  up to a redshift of  $z = 1.4$ . There are also shear estimates for each galaxy made by ray-tracing through the  $N$ -body simulations. Every galaxy has a  $\kappa_{WL}$  value assigned from the integrated dark matter field.

The simulations are made with  $4096^3$  particles of mass  $2927 M_\odot h^{-1}$  in a box of side  $3072 h^{-1}$  Mpc. The MICE-GC has an assumed flat  $\Lambda$  cold dark matter cosmology:  $\Omega_m = 0.25$ ,  $\Omega_b = 0.044$ ,  $\Omega_\Lambda = 0.75$ ,  $\sigma_8 = 0.8$ ,  $h = 0.7$ ,  $n_s = 0.95$ . The MICE-GC DES mocks approximately reproduce the magnitude limits of the DES survey and are complete down to apparent magnitude  $i < 22.0$  at  $z = 0.5$ .

For use in this paper, we have reduced the effective number densities in the mock galaxy and shear catalogues to reflect the statistics of the DES SV samples as well as normalizing the redshift to reflect the distribution shown in Fig. 2. Each mock catalogue is projected on to a HEALPIX map of  $N_{\text{side}} = 8192$ , which is then degraded to match the resolution of our data maps where appropriate.

In order to be able to compare the distribution of DES  $\kappa_{WL}$  (which we know has a significant shape noise contribution) with simulations, we create a second MICE  $\kappa_{WL}$  sample that has DES-like shape noise added. An estimate of the width of the shape noise distribution is obtained by fitting a Gaussian to the 100 random realizations of DES  $\kappa_{WL}$ . A noise contribution drawn from a Gaussian of this width is added to MICE  $\kappa_{WL}$  at the level of the cells used to construct the CiC distribution.

Fig. 4 shows the effect of adding shape noise to MICE  $\kappa_{WL}$  in this way at an angular scale of 10 arcmin. The darker, narrow histogram is that of the shape noise-free MICE  $\kappa_{WL}$ , the lighter histogram shows the distribution once the Gaussian shape noise is added and the black dashed line shows the distribution of DES  $\kappa_{WL}$ . At a smoothing scale of 10 arcmin, the Gaussian width of the DES shape noise estimate is 0.0099, which is 89 per cent of the width of

the resulting noisy MICE  $\kappa_{\text{WL}}$  distribution; this falls to 63 per cent at 40 arcmin.

## 4 METHOD

### 4.1 Constructing PDFs via CiC

The CiC approach is a relatively simple way to measure the distribution of galaxies in a survey, but it is a surprisingly powerful tool. A general CiC distribution for galaxies can be denoted by  $f(N, V)$ , the probability of finding  $N$  galaxies in a volume of space  $V$ . This can be a 3D volume or, as is the case in this paper, a 2D area on the sky where we count over a population projected along the line of sight. Repeating this procedure with cells of varying radii,  $r$ , gives us the distribution  $f_r(N)$ , where the moments of  $f_r(N)$  are related to the volume integrals of the correlation functions of our underlying observable (Peebles 1980; Fry 1985; Fry & Gaztanaga 1994; Saslaw 2000).

We perform our CiC analysis on the galaxy density contrast and weak lensing convergence maps with HEALPIX pixelization of resolution  $N_{\text{side}} = 4096$ , which corresponds to an average pixel size of 0.9 arcmin. For galaxies, to construct the PDF, we sum the galaxy counts,  $N$ , inside 2D circular cells of fixed radius  $r$  in the range 10–40 arcmin. At the median redshift,  $z = 0.3$ , of the sources considered, this corresponds to physical scales of 3–10 Mpc. The smallest cells used are 10 times larger than the HEALPIX pixels in order to minimize edge effects, and this also avoids any difference in counts across our survey area due to the changing geometry of the HEALPIX pixels (see Appendix A for a discussion of this assumption). We chose to use randomly positioned circular cells rather than using the HEALPIX pixels themselves as this allows us to repeat the analysis straightforwardly at any smoothing scale, rather than using only the fixed scales of HEALPIX pixels. The criterion for accepting a cell is that 80 per cent of its area should fall in unmasked regions (again see Appendix A for discussion of this choice). We want to use enough cells that all pixels in the map are covered at least once, and find that this is achieved when the total area of the cells is 20 times that of the survey. We use a coverage of 100 times the total area.

Histograms of the counts give us the distribution  $f(N)$ , and this procedure is repeated with cells of different radii to obtain the distribution  $f_r(N)$ . Double counting of pixels is accounted for by jackknife errors on the height of each bin in the resulting histogram of counts. We divide the survey area into 152 approximately equal-area ( $1 \text{ deg}^2$ ) jackknife patches. For a fixed set of randomly generated cells, and removing one patch at a time, we re-make the galaxy and convergence PDFs and re-calculate the statistics of interest in order to produce covariances.

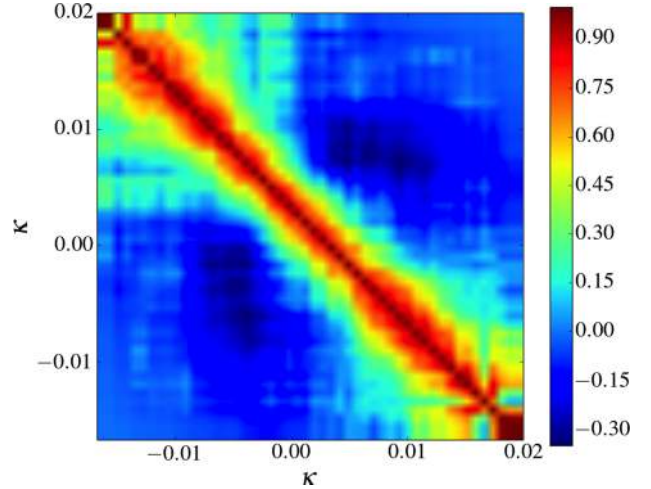
We repeat our CiC analysis on the DES reconstructed  $\kappa$  maps. The ‘count’ in each cell is now the average of the weak lensing convergence  $\kappa$  in pixels contained in that cell.

In Appendix B, we test the impact of spatially varying systematic effects on the DES  $\delta_{\text{g}}$  and  $\kappa_{\text{WL}}$  CiC distributions.

It is straightforward to generalize our CiC method to more than one observable. We simply throw the same circles on to each map (using the same mask for each), allowing us to compare counts at the same position for different observables.

### 4.2 Fitting the PDFs

We fit the lognormal models described in Section 2 to these distributions. For the MICE and DES galaxy density contrast distributions, we fit a Poisson-sampled lognormal using equation (3). For MICE



**Figure 5.** Correlation matrix of bin heights for a histogram of DES  $\kappa_{\text{WL}}$ , at a smoothing scale of 10 arcmin. Derived from jackknife sampling of the DES  $\kappa_{\text{WL}}$  map.

$\kappa_{\text{WL}}$ , which has no shape noise, we fit a plain lognormal model (equation 6). For the  $\kappa_{\text{WL}}$  distributions which include shape noise (i.e. DES  $\kappa_{\text{WL}}$  and the MICE  $\kappa_{\text{WL}}$  to which we add shape noise), we use equation (10).

The histogram bins in  $\delta_{\text{g}}$  or  $\kappa_{\text{WL}}$  are correlated. This is demonstrated in Fig. 5, which shows the correlation matrix of bin heights of DES  $\kappa_{\text{WL}}$  at a smoothing scale of 10 arcmin. In fitting the lognormal model, we take into account these correlations by minimizing

$$\chi^2 = (\mathbf{f} - \mathbf{d})\mathbf{C}^{-1}(\mathbf{f} - \mathbf{d})'. \quad (18)$$

Here  $\mathbf{f}$  is the data vector of the lognormal fit at the bin centres,  $\mathbf{d}$  is the data vector of bin heights and  $\mathbf{C}$  is the covariance matrix. We remove weak eigenvectors of the covariance matrix via singular value decomposition.

## 5 VALIDATING METHODS ON MICE

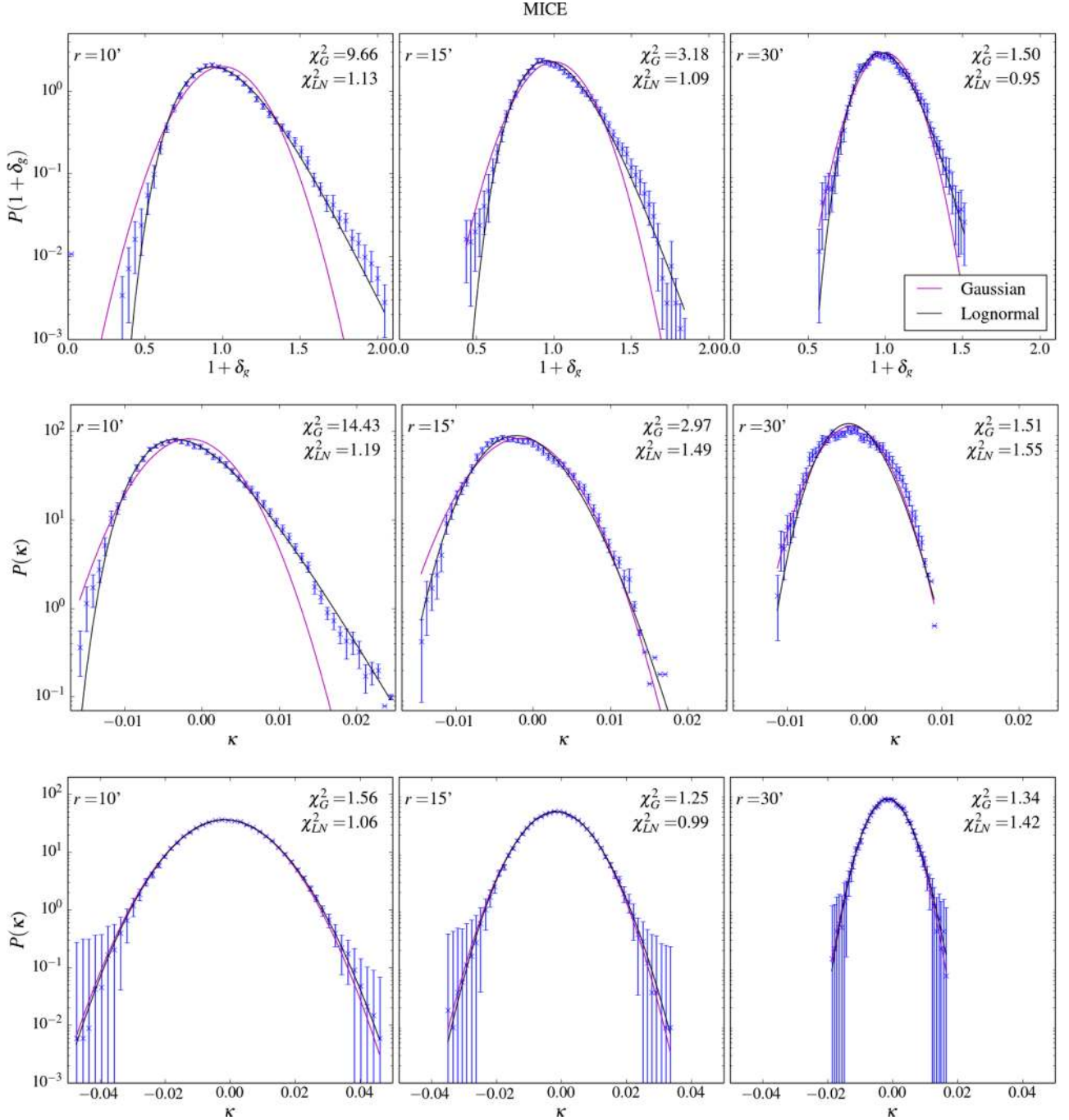
In this section, we verify the methods used to test the lognormality of DES  $\delta_{\text{g}}$  and  $\kappa_{\text{WL}}$  fields. After checking that the MICE  $\delta_{\text{g}}$  field is lognormal as we would expect, we see if this is true of the noise-less convergence field.

To enable easier comparison with the DES  $\kappa_{\text{WL}}$  results, we also look at the distribution of the simulation  $\kappa_{\text{WL}}$  for the MICE sample with number of galaxies and  $n(z)$  matched to our DES sample, and with DES-like shape noise added. We then look at the joint distribution of  $\delta_{\text{g}}$  and  $\kappa_{\text{WL}}$ , for the cases with and without shape noise.

As an additional check of the validity of the lognormal model, we compare the second moments of the distributions as calculated via CiC with those derived under the assumption of lognormality.

### 5.1 One-dimensional PDFs and lognormal fits

We first construct a simple histogram of  $\delta_{\text{g}}$  from the CiC to estimate the 1D PDF of  $\delta_{\text{g}}$ . The histogram uses 50 bins and we calculate jackknife errors on the bin heights as described in the previous section. The result for cells of radius of 10, 15 and 30 arcmin is shown in the upper panel of Fig. 6. We fit a Poisson-sampled lognormal distribution as described in equation (1) with  $w$  as the single free parameter. The best-fitting lognormal, which minimizes  $\chi^2$ , is shown as a black solid line and the best-fitting Gaussian



**Figure 6.** Upper row: measured 1D PDF of MICE galaxies at a smoothing scales of 10, 15 and 30 arcmin. The Poisson-sampled lognormal fit (black) provides a better fit to the galaxy CiC distribution than the Gaussian (magenta) at a scale of 10 arcmin. The distribution becomes increasingly Gaussian at larger scales. Middle row: same as above but for the MICE  $\kappa_{\text{WL}}$  PDF. Again the lognormal provides a good fit at the smallest scale, with the  $\kappa_{\text{WL}}$  distribution becoming more Gaussian at larger scales. Bottom row: fits to  $\kappa_{\text{WL}}$  using the subsample of MICE with DES-like galaxy density and  $n(z)$ , and to which DES-like shape noise has been added. This shape noise makes the distribution of  $\kappa_{\text{WL}}$  more Gaussian at all scales. All  $\chi^2$  are per degree of freedom.

(magenta) is shown for comparison. At a cell size of 10 arcmin (corresponding to about 3 Mpc at the median redshift  $z = 0.3$ ), it is clear that the lognormal model fits the data better, reflecting the non-linear clustering at this scale. The counting of information inside a cell can be thought of as a form of smoothing where the cells form a top-hat filter with a fixed size. As the size of our cells increases, we average information on increasingly large scales and lose sensitivity to the effects of non-linear clustering on small scales.

The lognormal distribution is designed to capture some of the information present as a result of non-linear evolution, so we would expect it to become less pronounced as the effective smoothing scale increases. This is indeed the case: at a cell radius of 10 arcmin, the lognormal model is highly favoured, with a  $\chi^2/\text{dof} = 1.13$ , compared to 9.66 for the Gaussian. At a cell size of 30 arcmin (corresponding to a physical scale of 8 Mpc at the median redshift), the distribution has become much more Gaussian with best-fitting



$\chi^2/\text{dof}$  for the Gaussian model now 1.50. The lognormal model is still favoured at this scale, with best-fitting  $\chi^2/\text{dof}=0.95$ .

The result for the MICE  $\kappa_{\text{WL}}$  PDF is shown in the middle panel of Fig. 6. Since there is no shape noise in the simulation, we fit a plain lognormal, shown by the black line. As discussed in Section 2.2, in order to fit a lognormal model to  $\kappa_{\text{WL}}$ , one must assign a value to  $\kappa_0$ , the minimum convergence parameter in equation (6). At 10 arcmin, we jointly fit  $\kappa_0$  and the lognormal width in equation (6), finding best-fitting  $\kappa_0 = 0.021$ . For larger scales, we find that it is not possible to jointly constrain  $\kappa_0$  and the width of the lognormal as they are degenerate. We therefore use the theoretically derived  $\kappa_0 = 0.050$ , described in Section 2.2.

Since the convergence is the weighted sum of the mass fluctuations along the line of sight, we expect it to be only approximately lognormal. At a smoothing scale of 10 arcmin, the lognormal is a good fit, with  $\chi^2/\text{dof}=1.19$ , and it is significantly preferred to the Gaussian model, which has a best-fitting  $\chi^2/\text{dof}=14.43$ . This lognormality of  $\kappa_{\text{WL}}$  at small scales is in line with Taruya et al. (2002) who found that a lognormal model was a good fit to simulated  $\kappa_{\text{WL}}$  at angular scales of 2–4 arcmin. Increasing the cell radius above 10 arcmin removes the clear preference for the lognormal, and the lognormal and Gaussian models fit the data equally well at cell radii of 30 arcmin. The fixed, theoretically derived  $\kappa_0 = 0.050$  allows the lognormal model with a single free parameter to fit the distribution well at 15 arcmin, but at larger cell radii this model does very slightly worse than the Gaussian model. This suggests that this value of  $\kappa_0$  may not be a good estimate for the minimum  $\kappa$  in the CiC PDF for larger cells. This makes sense as this  $\kappa_0$  corresponds to a pure void along the line of sight, which is a decreasingly likely observation as the cell radius increases.

The final row of Fig. 6 shows the distribution of  $\kappa$  using the subsample of MICE with DES-like galaxy density and  $n(z)$ , and to which DES-like shape noise has been added, as described in Section 3.3. The shape noise dominates the resulting distribution, particularly at smaller scales. The width of the distribution of shape noise is 74 per cent of the width of the noisy  $\kappa$  distribution at 40 arcmin, and at 10 arcmin it accounts for 89 per cent. We model the noisy  $\kappa$  distribution with a lognormal convolved with Gaussian noise as described in Section 2.2, using equation (10). Again we find that it is not possible to jointly constrain  $\kappa_0$  and the width of the lognormal at scales above 10 arcmin as they are degenerate. We therefore use the theoretically derived  $\kappa_0 = 0.049$ . It can be seen from Fig. 6 that at all scales the shape noise makes the noisy  $\kappa$  distribution much more Gaussian.

Despite the low signal-to-noise, at 10 arcmin the lognormal convolved with Gaussian noise provides a better fit than the simple Gaussian, with  $\chi^2/\text{dof}=1.06$  and 1.56, respectively. At scales larger than this, the Gaussian model performs as well as the lognormal. As with the noise-free convergence distribution, the theoretically derived  $\kappa_0$  seems to be a less suitable choice at larger scales as the Gaussian model provides a better fit for scales above 30 arcmin.

## 5.2 Joint galaxy–convergence distribution

In this subsection, we study the joint distribution of galaxy overdensities and weak lensing convergence and determine to what extent it can be described as a bivariate lognormal distribution. We look at joint distributions using both the full MICE sample and the subsample with DES-like galaxy density and  $n(z)$  and the addition of DES-like shape noise. As in the 1D case, the full sample with higher galaxy density allows us to better capture any lognormal behaviour,

and the DES-like sample allows us to compare the results for DES data given in the next section with simulations.

We can make a simple quantitative estimate of the relative correlation of  $\delta_{\text{g}}$  and  $\kappa_{\text{WL}}$  by calculating the Pearson product-moment correlation coefficient,  $r$ , for the joint PDF, where

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\langle (X - \bar{X})(Y - \bar{Y}) \rangle}{\sigma_X \sigma_Y}. \quad (19)$$

We begin with the joint distribution of  $\delta_{\text{g}}$  and  $\kappa_{\text{WL}}$  with no shape noise, which is shown in the upper panel of Fig. 7 for a smoothing scale of 15 arcmin. For the convergence, we plot  $1 + \hat{\kappa}$ , where  $\hat{\kappa} = (1 + \kappa_{\text{WL}}/\kappa_0)$ . Note that this results in a distribution that is narrower once we add shape noise; this is due to the lognormal shape of the noise-free  $\kappa_{\text{WL}}$  distribution versus the Gaussian shape of the noisy  $\kappa_{\text{WL}}$  distribution. The blue contours in the top-right section of this plot show the joint PDF, and the magenta dashed contours show the bivariate lognormal fit. Since there is no shape noise in this case, the bivariate fit is given by equation (15) but omitting the Gaussian convolution.

We expect the correlation coefficient  $\rho$  to be high since the galaxies considered are responsible for the lensing. This is indeed what we see:  $\rho$  is 0.81 at a smoothing scale 10 arcmin and 0.89 at 40 arcmin. We do not see full correlation because the relevant window functions – the lensing efficiency function of the source sample and the galaxy redshift distribution of the galaxy sample – do not overlap precisely.

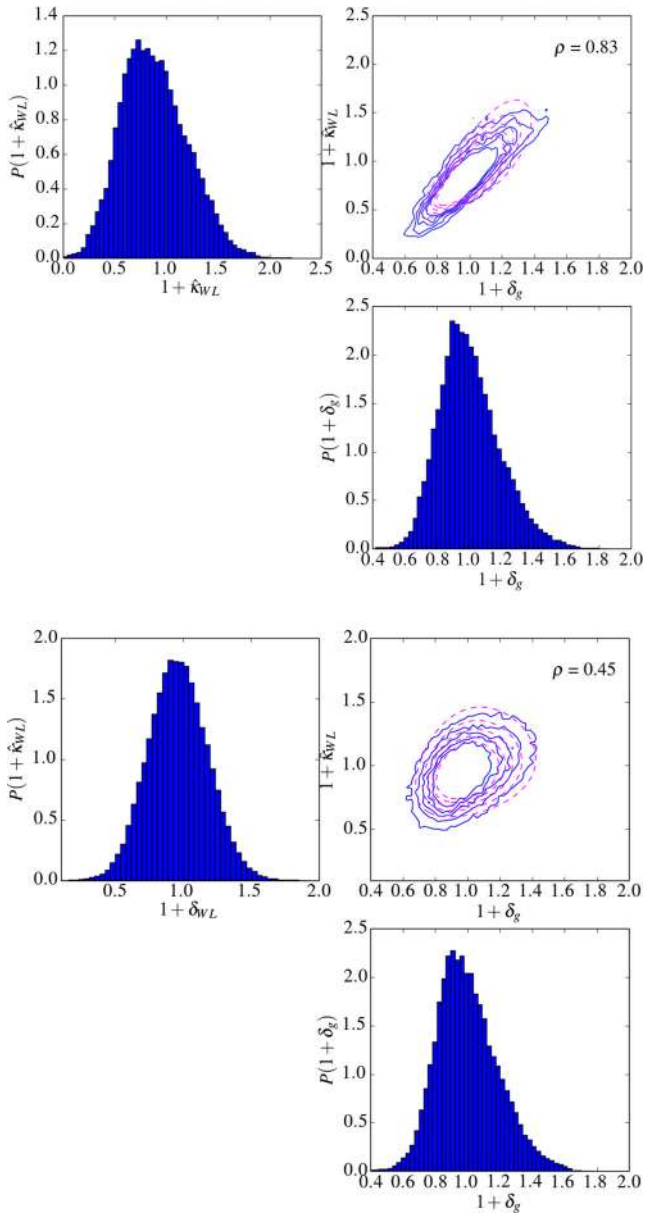
The lower panel of this figure shows the case where MICE  $\kappa_{\text{WL}}$  has had shape noise added. This noise reduces the correlation of the  $\kappa_{\text{WL}}$  with  $\delta_{\text{g}}$ , smearing out the joint distribution (shown on the top right of the figure) versus the shape noise-free case. The Pearson correlation coefficient is reduced to 0.45.

## 5.3 Comparison of moments

As well as seeing if the lognormal model gives a good least-squares fit to the galaxy and convergence CiC PDFs, as in the previous sections, we can also compare moments to check the validity of the lognormal modelling. Here we compare the variances derived directly from the CiC with those derived by fitting a lognormal model to the CiC PDF. The method for deriving moments of the galaxy and convergence fields from the CiC is described in Appendix C, and the variances derived under lognormal modelling are given by equations (2) and (9).

The variance of the MICE galaxy distribution,  $\sigma_{\text{g}}^2$ , is shown in the first panel of Fig. 8. Blue data points show the ratio of the variance derived by fitting a lognormal model to the CiC PDF and that calculated directly from the CiC. Magenta data points show the equivalent ratio for a Gaussian fit. Errors on the  $\sigma_{\text{g}}^2$  directly from CiC are produced by jackknife sampling; errors on the  $\sigma_{\text{g}}^2$  derived from lognormal fits are from the  $1\sigma$  width of the likelihood of the lognormal width. In all cases, shot noise is accounted for, via the Poisson convolution with the lognormal or Gaussian model for the fits (see Section 2.1), and as described in Appendix C for the CiC. The bottom panel of this figure shows the equivalent for the convergence distribution.

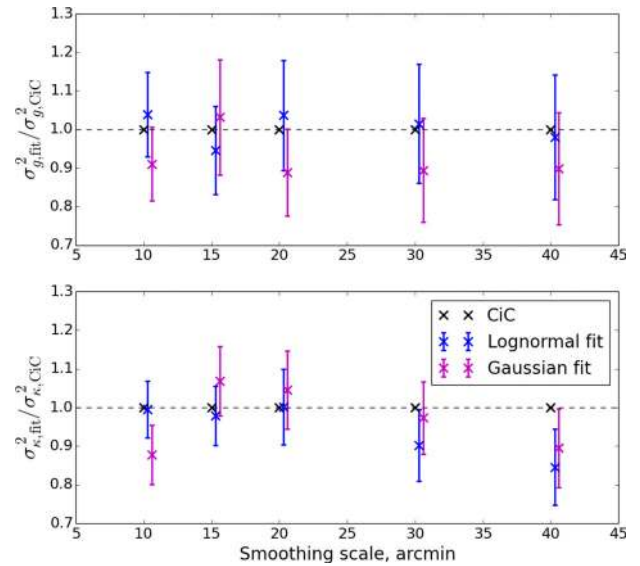
The lognormal model gives a good estimate of the variance of the MICE galaxy density contrast distribution at all scales. It gives a better estimate of the variance than a Gaussian model at all scales, and particularly at 10 arcmin. The lognormal model also gives a good estimate of the variance of the weak lensing convergence distribution at scales up to 20 arcmin. The poorer estimates at 30



**Figure 7.** Upper panel: joint CiC distribution of weak lensing convergence and galaxy density contrast for MICE simulation at a smoothing scale of 15 arcmin. The top-right plot shows the bivariate lognormal fit to MICE simulations. Contours for the simulation are given by the solid blue lines, with magenta dashed contours for the fit, both at 10, 30, 50, 70 and 90 per cent of the height of the joint distribution. Also shown are the 1D PDFs for  $1 + \delta_g$  and  $1 + \hat{\kappa}_{WL}$  individually. PDFs are calculated via the CiC method with cells of radius 15 arcmin. As in the rest of this paper, galaxies are selected over the redshift range  $0.1 < z < 0.5$  and weak lensing sources are restricted to the range  $0.6 < z < 1.3$ . This joint distribution has a Pearson correlation coefficient of  $r = 0.83$ . Lower panel: same but with DES-like shape noise added to  $\kappa_{WL}$ . The Pearson correlation coefficient drops to 0.45 with the addition of this shape noise.

and 40 arcmin are due to the fact that we fix  $\kappa_{\min}$  to the theory value at these scales.

These results suggest that within the ranges of scales discussed, the lognormal model can be used to estimate the two-point statistics of both the galaxy density contrast and weak lensing convergence distributions to reasonable accuracy in these simulations.



**Figure 8.** Upper panel: comparison of variances of MICE galaxy density contrast directly measured via CiC,  $\sigma_{g,CiC}^2$ , and from lognormal and Gaussian fits to the CiC PDF,  $\sigma_{g,fit}^2$ , as a function of smoothing scale. Blue data points show this ratio for lognormal fits, and magenta data points show the same but for Gaussian fits. Shot noise is accounted for in all variances. Data points are offset slightly in scale for clarity. Lower panel: same but for shape noise-free MICE weak lensing convergence.

## 6 TESTING LOGNORMALITY OF DES DENSITY AND CONVERGENCE FIELDS

Here we repeat the analysis of the previous section with DES galaxy and convergence maps, looking first distributions individually and then at their joint distribution.

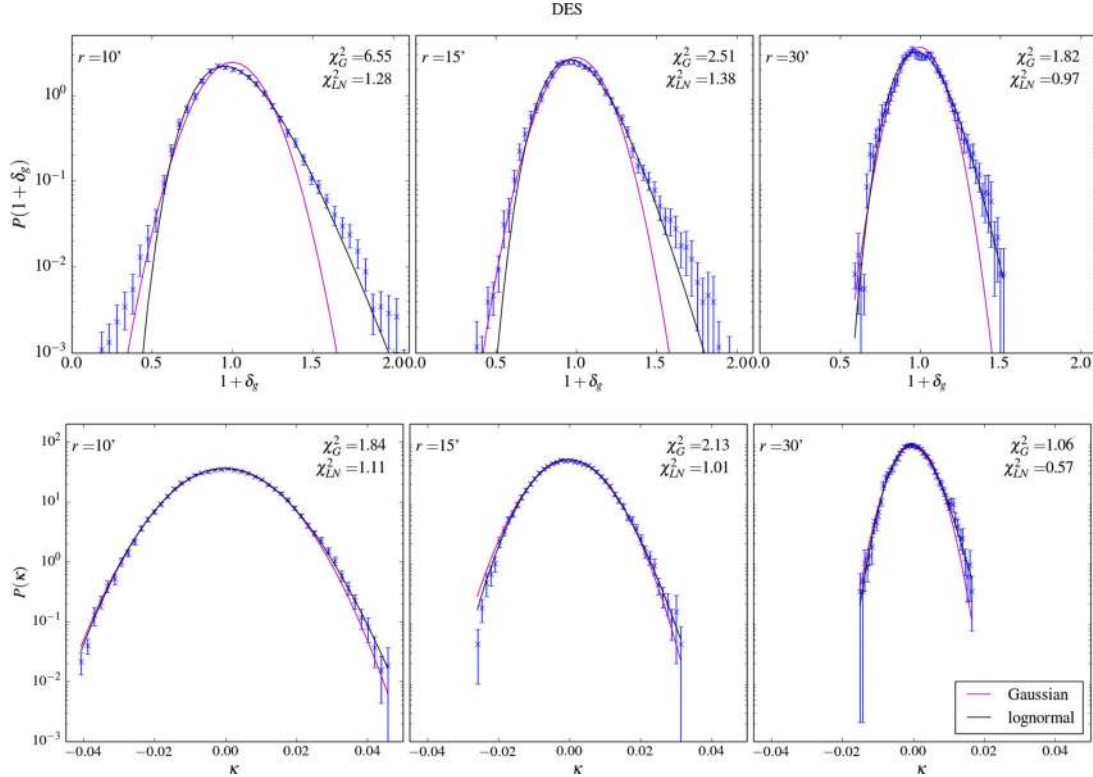
### 6.1 One-dimensional PDFs and lognormal fits

Fig. 9 shows 1D CiC PDFs for the DES galaxy density contrast (top row) and  $\kappa_{WL}$  (second row) fields for different cell radii. The lognormal fit is again shown in black, and for comparison a Gaussian fit is shown in magenta.

For the  $\delta_g$  PDF at 10 arcmin, the lognormal model is clearly favoured, with  $\chi^2/\text{dof} = 1.28$  compared to 6.55 for the Gaussian model. This confirms the expected lognormal behaviour at non-linear scales, indicating that our CiC procedure is capturing non-linear clustering information beyond the Gaussian assumption at smaller radii. As in the simulations, the  $\delta_g$  PDFs clearly appear more Gaussian at larger cell radii, although the lognormal model still provides a better fit than the Gaussian even at 30 arcmin, with  $\chi^2/\text{dof}$  of 0.97 and 1.82, respectively.

The best-fitting values of the free parameters of the lognormal fits to the DES galaxy density contrast distribution, the  $\chi^2$ , the number of degrees of freedom (dof) and the second moment of the best-fitting lognormal PDF are given in Table 1, for smoothing scales of 10–40 arcmin.

The second row of Fig. 9 shows the DES  $\kappa_{WL}$  distribution. We find that it is possible to jointly constrain  $\kappa_0$ , the minimum convergence parameter in equation (6), and the width of the lognormal at all smoothing scales. The best-fitting values of  $\kappa_0$  as well as the lognormal width  $\omega$ , best-fitting  $\chi^2$  and the variance of the best-fitting lognormal PDF are given in Table 2. The best-fitting  $\kappa_0 = 0.021$  and  $\omega = 0.235$  at cell radius 10 arcmin are in good agreement with the results from the MICE simulation at this scale, which are 0.023



**Figure 9.** Upper row: measured 1D PDF of DES galaxies at a smoothing scales of 10, 15 and 30 arcmin. At 10 arcmin, the Poisson-sampled lognormal fit (black) provides a much better fit than the Gaussian (magenta), demonstrating the lognormality of the galaxy CiC distribution at this scale. At larger scales, the distribution becomes more Gaussian. Lower row: same but for  $\kappa_{\text{WL}}$ . Here the lognormal model includes Gaussian shape noise, which provides a good fit at all scales. Error bars on the counts PDFs are jackknife errors. All  $\chi^2$  are per degree of freedom.

**Table 1.** Best-fitting parameters and derived statistics from lognormal fits to CiC PDFs of DES galaxy density contrast, for varying cell radii. First column gives the cell radius, and the second column is the width of the best-fitting Poisson-sampled lognormal. The following columns are the minimum  $\chi^2$  for Gaussian and Poisson-sampled lognormal fits, and the number of degrees of freedom. The final column is the variance of the best-fitting lognormal PDF, derived from the lognormal width, with  $1\sigma$  errors given by the likelihood of the lognormal width.

$r$ (arcmin)	$\omega$	$\chi_G^2$	$\chi_{\text{LN}}^2$	dof	$\sigma_g^2 \times 10^{-2}$
10	0.184	72.05	14.08	11	$3.44 \pm 0.30$
15	0.156	27.61	15.18	11	$2.46 \pm 0.28$
20	0.146	21.48	12.36	12	$2.15 \pm 0.29$
30	0.126	30.94	16.49	17	$1.60 \pm 0.16$
40	0.112	22.88	18.48	22	$1.26 \pm 0.17$

and 0.228, respectively. Note that for larger scales we fix  $\kappa_0$  at the theory value of 0.05 in the simulations, so we would not expect close agreement of the best-fitting lognormal width with that of the data at these scales.

The  $\kappa_{\text{WL}}$  distribution appears quite Gaussian at all scales due to the Gaussian shape noise, the distribution of which has a width of 70–90 per cent of the width of the  $\kappa_{\text{WL}}$  distribution. Despite this low signal-to-noise, as in the case of simulated  $\kappa_{\text{WL}}$ , we find that the lognormal model with Gaussian shape noise (black line) provides a better fit than the simple Gaussian model (magenta line) at small scales. At 10 arcmin the lognormal model has  $\chi^2 = 1.11$  and the Gaussian 1.84, corresponding to  $p$ -values of 0.35 for the lognormal model (i.e. within  $1\sigma$ ) and 0.07 for the Gaussian model. At 15 arcmin the advantage of the lognormal model over the Gaussian

**Table 2.** Same as Table 1 but for DES weak lensing convergence. The lognormal fit accounts for shape noise, so the statistics quoted are for the de-noised  $\kappa_{\text{WL}}$  distribution. The additional information given versus Table 1, in the second column, is the best-fitting minimum convergence parameter  $\kappa_0 = -\kappa_{\text{min}}$ .

$r$ (arcmin)	$\kappa_0$	$\omega$	$\chi_G^2$	$\chi_{\text{LN}}^2$	dof	$\sigma_\kappa^2 \times 10^{-5}$
10	0.021	0.235	18.41	11.10	10	$2.44 \pm 0.45$
15	0.017	0.248	19.17	9.09	9	$1.59 \pm 0.34$
20	0.016	0.238	10.92	4.63	10	$1.46 \pm 0.29$
30	0.009	0.314	11.66	6.27	11	$0.84 \pm 0.21$
40	0.008	0.300	14.82	8.58	13	$0.71 \pm 0.19$

is clear, with best-fitting  $\chi^2/\text{dof}$  of 1.01 and 2.13, respectively. At scales larger than this, the Gaussian model provides a good fit with best-fitting  $\chi^2/\text{dof}$  of 1.09, 1.06 and 1.14 at 20, 30 and 40 arcmin. The lognormal model is overfitting the data at these scales, with  $\chi^2/\text{dof}$  of 0.46, 0.57 and 0.66 at the same scales, so the Gaussian model is sufficient in this regime.

At scales below 15 arcmin where the lognormal model is a good fit, the width of the lognormal distribution also specifies the higher moments of the distribution, such as skewness and kurtosis. Higher moments of the matter density distribution can be employed to probe cosmology (e.g. Bernardeau 1994b; Bernardeau, van Waerbeke & Mellier 1996). For example, in conventional analyses of weak lensing two-point statistics, a strong degeneracy between  $\sigma_8$  and  $\Omega_M$  exists (Simon et al. 2015; Kacprzak et al. 2016; Kwan et al. 2016). With the one-point statistic, this degeneracy is broken because  $\sigma_8$  is related to the skewness of the matter density contrast distribution,

with higher  $\sigma_8$  values associated with greater skewness. In fact, the skewness scales linearly with  $\sigma_8$  (see Appendix D). Since we have access to the integrated matter density via the convergence PDF both from simulations and DES data, we can use this scaling relation, that the skewness of the convergence PDF  $\gamma \propto \sigma_8$ , to estimate  $\sigma_8$  for the data:

$$\sigma_8^{\text{DES}} = \sigma_8^{\text{MICE}} \times \frac{\gamma^{\text{DES}}}{\gamma^{\text{MICE}}}. \quad (20)$$

Various definitions of skewness appear in the literature; here we use

$$\gamma \equiv \frac{\mu_3}{\sigma^3}, \quad (21)$$

where  $\sigma$  is the standard deviation of the convergence PDF, given by equation (9), and  $\mu_3$  is its third central moment,

$$\mu_3 = \int_{-\infty}^{\infty} (\kappa - \langle \kappa \rangle)^3 P(\kappa) d\kappa \quad (22)$$

with  $P(\kappa)$  given by equation (6). The skewness of the lognormal PDF is then

$$\gamma = (e^{\omega^2} + 2)\sqrt{e^{\omega^2} - 1}, \quad (23)$$

where  $\omega$  is the width of the lognormal fit to the convergence PDF.

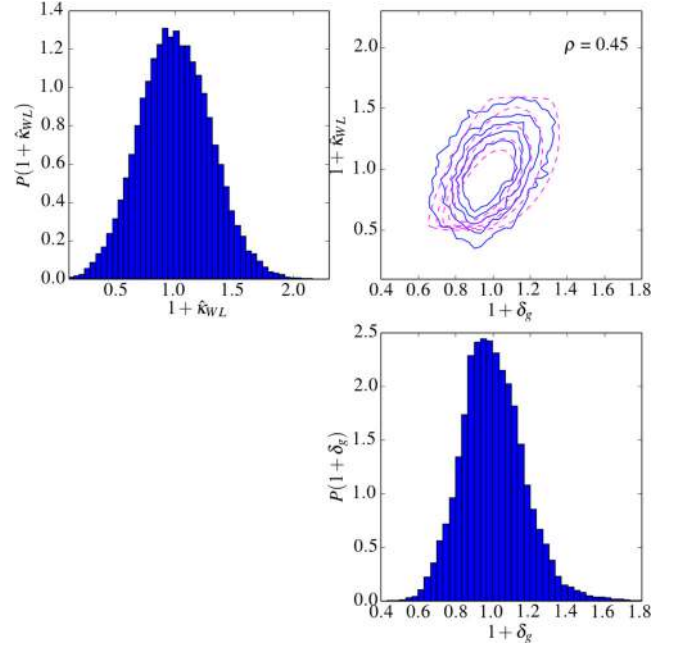
As a consistency check of the lognormal modelling of the data, we can see if this value of  $\sigma_8$  agrees with recently derived constraints in the literature. At a cell size of 10 arcmin, the scale at which the lognormal fit is best, the lognormal widths of the MICE and DES convergence PDFs are  $0.228 \pm 0.002$  and  $0.235 \pm 0.021$ , respectively, where errors are the  $1\sigma$  width of the likelihood. The value of  $\sigma_8$  used in the simulations is 0.8. This gives an estimate of  $\sigma_8$  for the data of  $0.826 \pm 0.022$ , which is consistent with the Planck Collaboration's recently revised value of  $0.8174 \pm 0.0081$  (Planck Collaboration 2016).

## 6.2 Joint galaxy–convergence distribution

The joint distribution of DES galaxy density contrast and weak lensing convergence data at an angular scale of 15 arcmin is shown in the top-right panel of Fig. 10. The data are shown by the blue contours, and the bivariate fit is shown by the magenta dashed contours. The individual 1D PDFs for  $1 + \delta_g$  and  $1 + \kappa/\kappa_0$  are also shown.

Before we account for shot noise in the galaxies and shape noise in the convergence, the galaxy counts and  $\kappa_{\text{WL}}$  have a Pearson correlation coefficient of 0.45. This is in line with what we see in the MICE simulations once DES-like shape noise is added (bottom row of Fig. 7).

Once we account for these sources of noise, the correlation coefficient is 0.82, again in line with the noise-free MICE simulations,



**Figure 10.** Joint distribution of weak lensing convergence and galaxy density contrast for DES at smoothing scale 15 arcmin. Upper-right panel: fit of bivariate lognormal to DES SV data. Contours for the data are given by the solid blue lines, with magenta dashed contours for the fit, both at 10, 30, 50, 70 and 90 per cent of the height of the joint distribution. Also shown are the individual 1D PDFs for  $1 + \delta_g$  and  $1 + \kappa_{\text{WL}}$ . DES Benchmark galaxies are used, selecting the redshift range  $0.1 < z < 0.5$  and weak lensing sources from the IM3SHAPE catalogue are used over the range  $0.6 < z < 1.3$ . All redshifts are best fits from the SkyNet pipeline. PDFs are calculated via the CiC method with cells of radius 15 arcmin. This joint distribution has a Pearson correlation coefficient of  $r = 0.45$ .

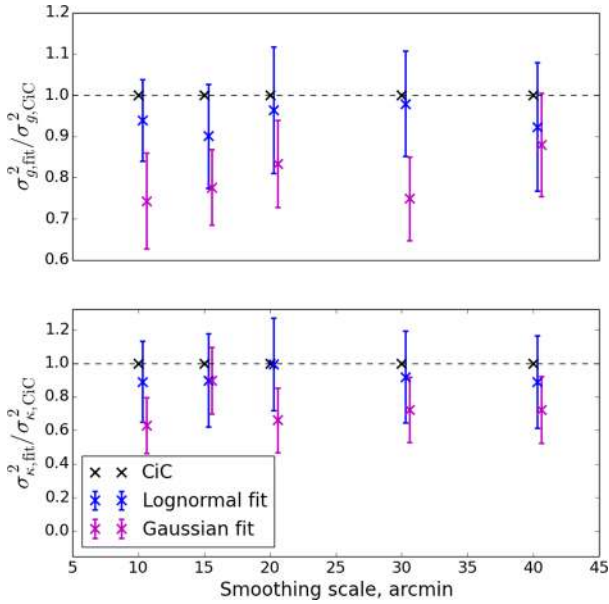
where the Pearson correlation coefficient was 0.83 (top row of Fig. 7).

## 6.3 Comparison of second moments

In this section, we repeat the analysis of Section 5.3, checking the validity of the lognormal model by comparing variances derived from the lognormal assumption with those measured directly from the CiC. The variance of the DES galaxy and convergence PDFs are shown in Fig. 11, which has the same form as Fig. 8. The variances derived directly from the CiC are given in Table 3. Shape noise in the DES convergence distribution has been accounted for via the Gaussian convolution given in equation (10) for the lognormal fit, and as per equation (C7) for the variance derived directly from the CiC.

**Table 3.** Second moments of DES galaxy density contrast and weak lensing convergence, as calculated by CiC, for different cell radii. Shot and shape noise have been accounted for, and these are the de-noised moments. The final column gives our estimate of the shape noise of the weak lensing convergence. This is derived from the 100 realizations of the  $\kappa_{\text{WL}}$  map with randomized shears, which we find to agree with the second moment of the  $\kappa_{\text{WL}}$  B-mode within 2 per cent.

$r$ (arcmin)	$\sigma_g^2$	$\sigma_\kappa^2$	$\sigma_n^2$
10	$(3.70 \pm 0.22) \times 10^{-2}$	$(2.52 \pm 0.41) \times 10^{-5}$	$(1.00 \pm 0.03) \times 10^{-4}$
15	$(2.76 \pm 0.20) \times 10^{-2}$	$(1.69 \pm 0.32) \times 10^{-5}$	$(4.68 \pm 0.18) \times 10^{-5}$
20	$(2.26 \pm 0.18) \times 10^{-2}$	$(1.39 \pm 0.25) \times 10^{-5}$	$(2.60 \pm 0.16) \times 10^{-5}$
30	$(1.65 \pm 0.14) \times 10^{-2}$	$(9.84 \pm 1.66) \times 10^{-6}$	$(1.18 \pm 0.06) \times 10^{-5}$
40	$(1.38 \pm 0.16) \times 10^{-2}$	$(8.40 \pm 1.30) \times 10^{-6}$	$(6.75 \pm 0.44) \times 10^{-6}$



**Figure 11.** Same as Fig. 8, but for DES galaxies (upper panel) and convergence (lower panel).

The lognormal model with appropriate noise contribution gives an estimate of the variance that is consistent with that calculated directly from the CiC, for both  $\delta_g$  and  $\kappa_{\text{WL}}$ , at all scales from 10 to 40 arcmin.

For both the galaxy density contrast and the convergence distributions, the Gaussian model provides a less accurate estimate of the variance calculated directly from the CiC at all scales. For both distributions, the Gaussian and lognormal approaches underestimate the variance as compared to measuring it directly from the CiC. This is because in constructing the CiC PDF to which we fit the lognormal model, we bin the cell counts. We account for noise via singular value decomposition, and one of the things this removes is contributions to the fit from the outermost bins, which have very few cell counts. This makes the effective distribution narrower, with a lower variance, than if these noisy data points were included. In calculating the variance directly from the CiC (as described in Appendix C), this binning is not necessary and all cells, including those with the most extreme values of  $\delta_g$  or  $\kappa_{\text{WL}}$ , are included in the calculation, resulting in a larger variance in  $\delta_g$  or  $\kappa_{\text{WL}}$ . This effect is less stark in the MICE simulations where there are a greater number of galaxies than in the DES data, so fewer bins are discarded due to low counts of cells. This underestimation of the variance, however, is not significant within the errors.

## 7 DISCUSSION

We have tested the lognormality of the DES galaxy density contrast and weak lensing convergence PDFs at angular scales of 10–40 arcmin (corresponding to physical scales of 3–10 Mpc at median redshift  $z = 0.3$ ). In the context of this work, estimating the CiC PDF is a way of quantifying the non-linear growth of mass and galaxy fluctuations, as well as the visual impression of comparing the  $\kappa_{\text{WL}}$  mass maps with the galaxy distribution on the same patch of the sky. It is also a test of systematics.

Our main findings are as follows.

(i) In agreement with many earlier papers, we find that the 1D DES galaxy PDF is well fitted by a lognormal model, taking into

account Poisson shot noise, with best-fitting  $\chi^2/\text{dof}=1.28$  versus 6.55 for a Gaussian model at a scale of 10 arcmin.

(ii) In modelling the weak lensing convergence distribution, it is important to account for shape noise since the width of this noise is a significant fraction (70–90 per cent) of the width of the  $\kappa_{\text{WL}}$  signal. We find that the shape noise estimate derived from the 100 realizations of DES  $\kappa_{\text{WL}}$  in which the shears have been randomized agrees with that of the  $\kappa_{\text{B}}$  mode within 2 per cent at all scales from 10 to 40 arcmin, and that the distribution of the shape noise can be well modelled by a Gaussian PDF. This allows us to model the  $\kappa_{\text{WL}}$  distribution with a lognormal convolved with Gaussian PDF. In future work, it would be interesting to investigate the spatial correlation of this noise.

(iii) The convergence field is not expected to be exactly lognormal even if the mass density contrast field is, as it is a weighted projection of the mass density field along the line of sight. We find however, in agreement with previous work on simulations, that the 1D  $\kappa_{\text{WL}}$  PDF is well fitted by a lognormal model, taking into account shape noise. This is the first such measurement on data. The best-fitting  $\chi^2/\text{dof}$  for the lognormal model is 1.11, compared to 1.84 for a Gaussian model, corresponding to  $p$ -values of 0.35 (i.e. within  $1\sigma$ ) and 0.07, respectively. At scales above 15 arcmin, the Gaussian model is a sufficient approximation.

(iv) The bivariate  $(\kappa_{\text{WL}}, \delta_g)$  PDF is also well fitted by a bivariate lognormal.

(v) De-noised second moments derived via the lognormal fit are consistent with variances derived directly from the data up to scales of 40 arcmin, for both the DES galaxy density contrast and weak lensing convergence distributions.

This pilot study could be extended to much larger areas with weak lensing surveys such as the full DES (5000 deg<sup>2</sup>) survey, LSST (20 000 deg<sup>2</sup>) and *Euclid* (15 000 deg<sup>2</sup>). In this work, we have tested the lognormality of the  $\kappa_{\text{WL}}$  PDF; with the higher signal/noise that future surveys will provide it might be possible to deduce from the observed  $\kappa_{\text{WL}}$  PDF whether or not the underlying matter density field is lognormal – essentially inverting equation (4).

On the scales explored in this work, the lognormal model is a good fit to both the galaxy and convergence PDFs, which means that all moments of these distributions can be derived from the fitted lognormal width. These moments could be used to probe cosmology, such as constraining  $\sigma_8$  as briefly explored in this work, as well the matter density. On scales outside of this range, if the lognormal model was found not to provide a good fit, all the moments of the distribution are still contained in the CiC. Pujol et al. (2016) have used the DES  $\kappa_{\text{WL}}$  map and re-weighted galaxy map to measure galaxy bias. The method used in this work, of constructing PDFs via CiC and cross-correlating them, could also be used to extract information on galaxy bias. Fitting a bivariate lognormal model to the joint distribution would also allow one to learn about non-linear and stochastic bias (Wild et al. 2005).

In the production of covariances and simulated realizations of the convergence field, the assumption that the underlying convergence field follows lognormal statistics is commonly made (Hilbert et al. 2011; Joachimi et al. 2011; Alsing, Heavens & Jaffe 2016). This assumption is based on simulation results (Taruya et al. 2002); our results showing that the DES one-point convergence PDF is well fitted by a lognormal model provide more firm evidence to support this approach.

Finally, it could also be interesting to repeat this analysis using manipulations of the shear field than other  $\kappa_{\text{WL}}$  that avoid the reconstruction noise due to the Kaiser–Squires method.

## ACKNOWLEDGEMENTS

We would like to thank Ludovic Van Waerbeke for extremely useful exchanges on the formalism for the calculation of the convergence second moment.

We thank the anonymous referee for suggesting that the skewness of the convergence PDF can be used to estimate the amplitude of fluctuations,  $\sigma_8$ , as presented in Section 6.1.

The MICE simulations have been developed at the MareNostrum supercomputer (BSC-CNS) thanks to grants AECT-2006-2-0011 through AECT-2015-1-0013. Data products have been stored at the Port d'Informació Científica (PIC), and distributed through the CosmoHub webportal ([cosmohub.pic.es](http://cosmohub.pic.es)). Funding for this project was partially provided by the Spanish Ministerio de Ciencia e Innovación (MICINN), projects 200850I176, AYA2009-13936, AYA2012-39620, AYA2013-44327, ESP2013-48274, ESP2014-58384, Consolider-Ingenio CSD2007-00060, research project 2009-SGR-1398 from Generalitat de Catalunya, and the Ramon y Cajal MICINN programme.

We are grateful for the extraordinary contributions of our CTIO colleagues and the DECam Construction, Commissioning and Science Verification teams in achieving the excellent instrument and telescope conditions that have made this work possible. The success of this project also relies critically on the expertise and dedication of the DES Data Management group.

Funding for the DES Projects has been provided by the US Department of Energy, the US National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey.

The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l'Espai (IEEC/CSIC), the Institut de Física d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, the National Optical Astronomy Observatory, the University of Nottingham, the Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, Texas A&M University and the OzDES Membership Consortium.

The DES data management system is supported by the National Science Foundation under Grant Number AST-1138766. The DES participants from Spanish institutions are partially supported by MINECO under grants AYA2012-39559, ESP2013-48274, FPA2013-47986, and Centro de Excelencia Severo Ochoa

SEV-2012-0234. Research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) including ERC grant agreements 240672, 291329 and 306478.

This paper has gone through internal review by the DES collaboration.

LC thanks the Perren Fund for a studentship.

OL, DK and MM acknowledge support from a European Research Council Advanced Grant FP7/291329.

## REFERENCES

- Alsing J., Heavens A. F., Jaffe A. H., 2016, MNRAS, preprint ([arXiv:1607.00008](https://arxiv.org/abs/1607.00008))
- Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
- Bernardeau F., 1994a, A&A, 291, 24
- Bernardeau F., 1994b, ApJ, 433, 1
- Bernardeau F., Kofman L., 1995, ApJ, 443, 479
- Bernardeau F., van Waerbeke L., Mellier Y., 1996, A&A, 322, 1
- Bernardeau F., Colombi S., Gaztañaga E., Scoccimarro R., 2002, Phys. Rep., 367, 1
- Blanton M., 2000, ApJ, 544, 63
- Bonnett C. et al., 2016, Phys. Rev. D, 94, 042005
- Carretero J., Castander F. J., Gaztanaga E., Crocce M., Fosalba P., 2014, MNRAS, 447, 646
- Carron J., 2011, ApJ, 738, 86
- Chang C. et al., 2015, Phys. Rev. Lett., 115, 051301
- Coles P., Jones B., 1991, MNRAS, 248, 1
- Crocce M., Castander F. J., Gaztañaga E., Fosalba P., Carretero J., 2015a, MNRAS, 453, 1513
- Crocce M. et al., 2015b, MNRAS, 455, 4301
- Dark Energy Survey Collaboration, 2005, preprint ([astro-ph/0510346](https://arxiv.org/abs/astro-ph/0510346))
- Dark Energy Survey Collaboration, 2016a, Phys. Rev. D, 94, 022001
- Dark Energy Survey Collaboration, 2016b, MNRAS, 460, 1270
- Das S., Ostriker J. P., 2006, ApJ, 645, 1
- Erben T. et al., 2013, MNRAS, 433, 2545
- Flaugher B. et al., 2015, AJ, 150, 150
- Fosalba P., Gaztanaga E., Castander F. J., Crocce M., 2014, MNRAS, 447, 1319
- Fosalba P., Crocce M., Gaztanaga E., Castander F. J., 2015, MNRAS, 448, 2987
- Fry J. N., 1984, ApJ, 279, 499
- Fry J. N., 1985, ApJ, 289, 10
- Fry J. N., Gaztanaga E., 1994, ApJ, 425, 1
- Gaskell C. M., 2004, ApJ, 612, L21
- Gaztanaga E., 1992, ApJ, 398, L17
- Gaztanaga E., 1994, MNRAS, 12, 12
- Gaztanaga E., Yokoyama J., 1993, ApJ, 403, 450
- Gorski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
- Graff P., Feroz F., 2013, Astrophysics Source Code Library, record ascl:1312.007
- Hilbert S., Hartlap J., Schneider P., 2011, A&A, 536, A85
- Hoffmann K., Bel J., Gaztanaga E., Crocce M., Fosalba P., Castander F. J., 2015, MNRAS, 447, 1724
- Hubble E., 1934, ApJ, 79, 8
- Jarvis M. et al., 2016, MNRAS, 460, 2245
- Joachimi B., Taylor A. N., Kiessling A., 2011, Spectrum, 22, 27
- Kacprzak T. et al., 2016, MNRAS, 463, 3653
- Kainulainen K., Marra V., 2011, Phys. Rev. D, 84, 063004
- Kaiser N., Squires G., 1993, ApJ, 404, 441
- Kayo I., Taruya A., Suto Y., 2001, ApJ, 561, 22
- Kelly P. L. et al., 2014, MNRAS, 439, 28
- Kwan J. et al., 2016, MNRAS, 465, 2567
- Lahav O., Itoh M., Inagaki S., Suto Y., 1993, ApJ, 402, 387
- Leistedt B. et al., 2016, ApJS, 226, 24
- Limpert E., Stahel W. A., Abbt M., 2001, BioScience, 51, 341

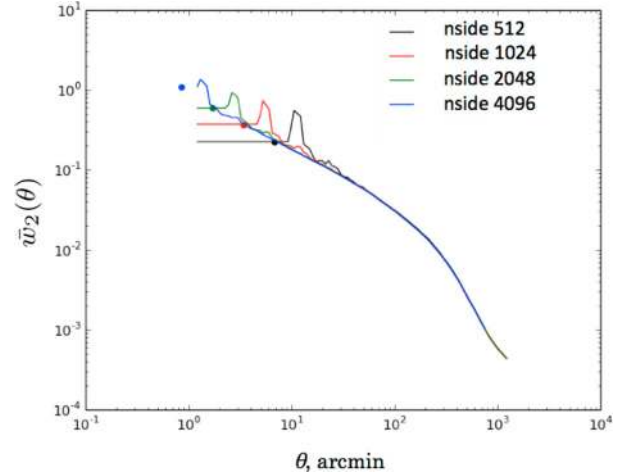
- Marr J. H., 2015, MNRAS, 453, 2214  
 Mellier Y., 1999, ARA&A, 37, 127  
 Munshi D., Jain B., 2000, MNRAS, 318, 109  
 Neyrinck M. C., Szapudi I., Szalay A. S., 2009, ApJ, 698, L90  
 Peebles P. J. E., 1980, Ann. New York Acad. Sci., 336, 161  
 Peebles P. J. E., Harrison E., 1994, Am. J. Phys., 62, 381  
 Planck Collaboration XLVI, 2016, A&A, preprint (arXiv:1605.02985)  
 Pujol A. et al., 2016, MNRAS, 462, 35  
 Rykoff E. S. et al., 2016, ApJS, 224, 1  
 Saslaw W. C., 2000, The Distribution of the Galaxies. Cambridge Univ. Press, Cambridge  
 Saslaw W. C., Hamilton A. J. S., 1984, ApJ, 276, 13  
 Seo H.-J., Sato M., Dodelson S., Jain B., Takada M., 2011, ApJ, 729, L11  
 Sheldon E. S., 2014, MNRAS, 444, L25  
 Simon P. et al., 2015, MNRAS, 449, 1505  
 Suto Y., 1993, Prog. Theor. Phys., 90, 1173  
 Suto Y., Itoh M., Inagaki S., 1990, ApJ, 350, 492  
 Szapudi I., 1997, ApJ, 497, 10  
 Szapudi I., Szalay A. S., 1993, ApJ, 408, 43  
 Taghizadeh-Popp M., Heinis S., Szalay A., 2012, ApJ, 755, 143  
 Takahashi R., Oguri M., Sato M., Hamana T., 2011, ApJ, 742, 15  
 Taruya A., Takada M., Hamana T., Kayo I., Futamase T., 2002, ApJ, 571, 638  
 Ueda H., Yokoyama J., 1996, MNRAS, 280, 754  
 Valageas P., 2000, A&A, 356, 771  
 Van Waerbeke L. et al., 2013, MNRAS, 433, 3373  
 Vikram V. et al., 2015, Phys. Rev. D, 92, 022006  
 White S. D. M., 1979, MNRAS, 186, 145  
 Wild V. et al., 2005, MNRAS, 356, 247  
 Xavier H. S., Abdalla F. B., Joachimi B., 2016, MNRAS, 459, 3693  
 Zinnecker H., 1984, MNRAS, 210, 43  
 Zuntz J., Kacprzak T., Voigt L., Hirsch M., Rowe B., Bridle S., 2013, MNRAS, 434, 1604

## APPENDIX A: TESTS OF SAMPLING METHODS

Our CiC analysis has made particular choices for cell size and distribution when accounting for the mask and creating the underlying HEALPIX maps. In this appendix, we test each of these assumptions and demonstrate that the conclusions of our analysis are robust to our methodological choices.

HEALPIX tessellations are made up of pixels with equal area, but not equal shape. Using circles that encompass several pixels will reduce the effect of the different pixel shapes, more so the larger the circles relative to the pixels. To check that the effect of varying shapes is effectively mitigated in this way, we measure the area averaged two-point correlation  $\bar{w}_2(\theta)$  for DES galaxies at different HEALPIX resolutions (512, 1024, 2048 and 4096).  $\bar{w}_2(\theta)$  is estimated by the shot noise corrected variance (Szapudi & Szalay 1993),  $\sigma_{\text{g}}^2$ , given in equation (C4). Fig. A1 shows that when the cell's size is close to the pixel size, the correlation function is not smooth due to the effects of pixel shape. Once the cells are several times larger than the average pixel separation, the correlation function becomes smooth, and the correlation functions based on the different HEALPIX resolutions converge. This confirms that the method of using circular cells several times larger than the pixel resolution does not suffer from the effects of pixel shape, and that the underlying pixel resolution is not important as long as the minimum cell size considered is sufficiently large.

The other sampling assumption we test is the threshold,  $\Delta$ , at which we decide to discard a randomly positioned cell because too much of it is masked. We only include cells that have a fraction of their area greater than  $\Delta$  unmasked, and the counts are up-weighted



**Figure A1.** Two-point correlation of MICE galaxy density contrast distribution as a function of scale as calculated via CiC, using underlying HEALPIX maps with different resolutions. The HEALPIX maps have  $N_{\text{side}}$  (see the main text) 512, 1024, 2048, 4096 corresponding to pixels of sizes shown by the filled circles. At scales approaching the pixel size, edge effects are visible.

to make it the equivalent of a whole cell. Here we explain the choice of  $\Delta = 0.8$  used in this work.

The first panel of Fig. A2 shows the PDF of MICE galaxy counts for which the  $\Delta$  takes different values. For values of  $\Delta > 0.5$ , there is not much difference in the histograms. The middle panel shows the variances of these different distributions, with errors produced by jackknife sampling. We find that the effect on the variance of changing  $\Delta$  is not significant within the errors.

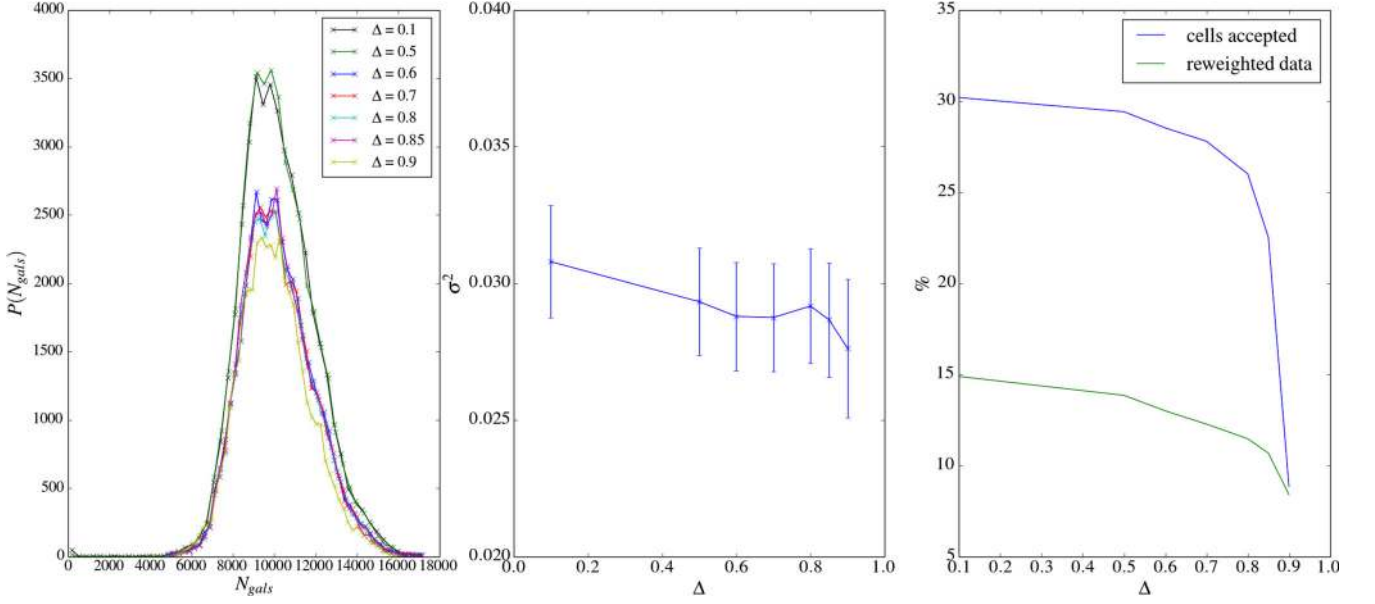
If we chose a very high threshold, such as requiring 90 per cent of a cell to be unmasked in order for it to be used, we would throw away a lot of cells landing near the edge of the survey and give greater statistical weight to areas away from the edges. If we set  $\Delta$  too low, so that cells with a large fraction of their area masked are kept, we will end up re-weighting a lot the data near the edges. So we would like to strike a balance between these two effects. In the right-hand panel of Fig. A2, the blue line shows the fraction of cells randomly thrown that are accepted. The green line shows fraction of the resulting counts that come from re-weighting these cells. We can see that  $\Delta = 0.8$  is the highest value that can be allowed before the number of cells we discard drops off significantly, and that at this value the fraction of data re-weighted is not too high (around 10 per cent). Hence,  $\Delta = 0.8$  seems to be a sensible choice.

## APPENDIX B: SYSTEMATIC EFFECTS

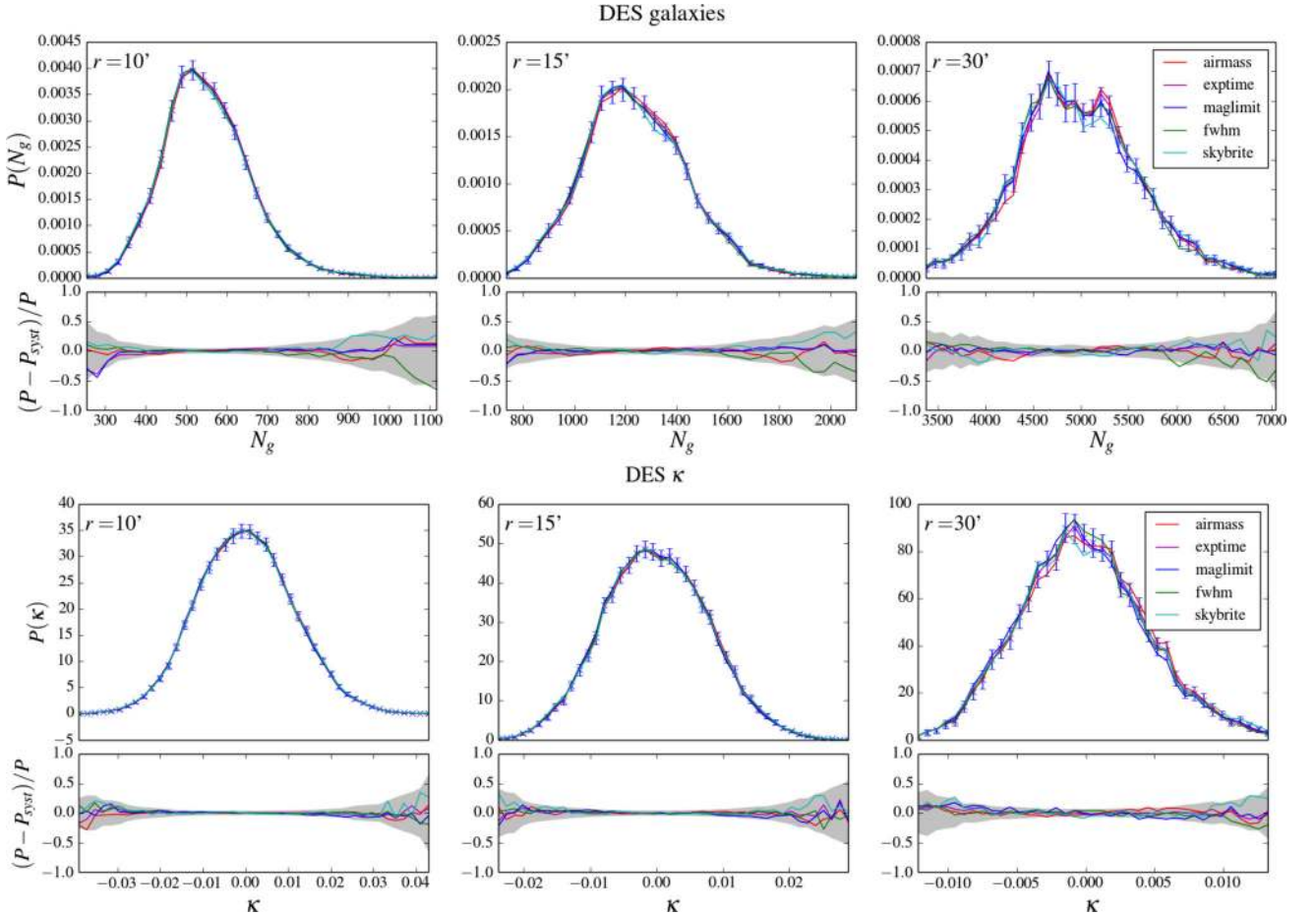
We investigate the potential impact on our results of spatially varying systematics on the probability distributions of DES galaxies and weak lensing convergence. The systematics we consider are varying the amount of air mass dependent on the distance of the observed field from the zenith, exposure time, magnitude limit, atmospheric seeing and sky brightness. The values of these properties were mapped across the DES SV area as described in Leistedt et al. (2016).

We compare PDFs of  $\delta_{\text{g}}$  and  $\kappa_{\text{WL}}$  for the full samples used in this work versus when the areas worst affected by these systematics are removed. We produce PDFs with the 20 per cent of worst affected pixels masked, for each systematic in turn.

Fig. B1 shows the resulting distributions. The top row shows DES galaxy number density at cell radii of 10, 15, 30 arcmin (blue



**Figure A2.** Left-hand panel: effect of the mask threshold  $\Delta$  (the fraction of a cell that must be unmasked in order to be included in the analysis) on the resulting probability distribution of MICE galaxies at 20 arcmin. Middle panel: effect of this threshold on the variance of the galaxy distribution. Right-hand panel: the fraction of the cells randomly thrown that are kept in the analysis (blue line) and, of those kept, the fraction of the data that come from re-weighting these cells (green line).



**Figure B1.** Upper row: PDF of DES galaxy number density at cell radii of 10, 15, 30 arcmin (blue data points) with jackknife errors. Coloured lines show PDFs with cuts for each systematic effect in turn. Lower panel shows the fractional difference between the full sample and those with systematics cuts, with errors shown by the grey shaded region. Lower row: same but for DES  $\kappa_{\text{WL}}$ .



data points) with jackknife errors. Coloured lines show PDFs with cuts for each systematic effect in turn. The lower panel shows the fractional difference between the full sample and those with systematics cuts, with errors shown by the grey shaded region. The same for DES  $\kappa_{\text{WL}}$  is shown in the bottom row.

Here we can see that the PDFs of the cut data are broadly consistent with that of the full data, given the jackknife errors. For DES galaxies, for each systematic effect at least 95 per cent of the bin heights after the cuts are made fall within the jackknife errors of the original distribution, and all are within  $1.5\sigma$  of the original distribution. For DES  $\kappa_{\text{WL}}$ , at least 93 per cent of the new bin heights fall within the jackknife errors of the original distribution. All are within  $1.9\sigma$  of the original data points. We can see that the effect of the systematics on the distribution increases with scale. Importantly, PDFs of the cut  $\kappa_{\text{WL}}$  data at scales below 20 arcmin, which is where we detect lognormality of  $\kappa_{\text{WL}}$ , are completely consistent with the original distributions, i.e. all of the new bin heights fall within the errors of the original distribution. This simple test is reassuring and indicates that our lognormal fits to the DES  $\delta_g$  and  $\kappa_{\text{WL}}$  distributions are not likely to be affected by these systematic effects.

### APPENDIX C: MEASUREMENT OF MOMENTS FROM CiC

In Section 5, we use the variance, as calculated via CiC, as a check that the lognormal model accurately recovers the characteristics of the galaxy and convergence distributions. In this appendix, we show how the moments of these underlying distributions are estimated from the CiC, including how noise is accounted for.

Let us first consider the distribution of galaxies. The  $p$ th central moment of the distribution of counts at angular scale  $\theta$  is given by

$$m_p = \frac{1}{N} \sum_{i=1}^N \left( \frac{n_i}{\Delta_i} - \bar{n} \right)^p, \quad (\text{C1})$$

where  $N$  is the number of cells,  $n_i$  is the number of galaxies, with mean  $\bar{n}$ , and  $\Delta_i$  is the fraction of the cell that is unmasked. We discard cells that are more than 80 per cent masked (see Appendix A). The overdensity of a cell is defined as  $\delta_i \equiv \frac{(n_i/\Delta_i - \bar{n})}{\bar{n}}$ . We then have

$$m_p = \bar{n}^p \langle \delta_i^p \rangle. \quad (\text{C2})$$

We want to estimate the variance of the underlying galaxy density field using the counts of galaxies, so we need to account for shot noise. If we assume that the galaxies Poisson sample the underlying density field, the shot noise-corrected second moment,  $k_2$ , is given by

$$k_2 = m_2 - \bar{n} \quad (\text{C3})$$

(Szapudi & Szalay 1993; Gaztanaga 1994). Our estimate of the de-noised variance of the galaxy density field is then

$$\sigma_g^2 = (m_2 - \bar{n})/\bar{n}^2. \quad (\text{C4})$$

For weak lensing convergence, the observable in each cell is the mean value of the convergence in the HEALPIX pixels contained in that cell,  $\kappa_i$ , so the moments of the CiC are given by

$$m'_p = \frac{1}{N} \sum_{i=1}^N (\kappa_i - \bar{\kappa})^p, \quad (\text{C5})$$

where  $\bar{\kappa}$  is the mean  $\kappa_i$  across all cells. Note that there is no need to include the unmasked fraction of the cell,  $\Delta$ , since we are using the average convergence in a cell rather than summing counts. We

do however discard cells that are more than 80 per cent masked, as with the galaxies.

Since the underlying convergence field is continuous, there is no need to correct for discreteness and our estimator for its moments is simply given by  $m'_p$ . The variance of the observed convergence field is then

$$\sigma_{\kappa, \text{obs}}^2 = m'_2. \quad (\text{C6})$$

For the MICE convergence equation (C6) is sufficient, but in the case of DES we need to account for shape noise. The shape noise in the DES  $\kappa$  map is estimated from the 100 noise realizations, as discussed in Section 3.2. Following Van Waerbeke et al. (2013), we assume that the de-noised variance of the DES convergence,  $\sigma_\kappa^2$ , is then given by

$$\sigma_\kappa^2 = \sigma_{\kappa, \text{obs}}^2 - \sigma_n^2, \quad (\text{C7})$$

where  $\sigma_n^2$  is the mean of the variances measured via CiC from each of the noise maps, which is our estimate of the shape noise contribution.

### APPENDIX D: RELATION OF SKEWNESS TO $\sigma_8$

In Section 6.1, we use a scaling relation between the amplitude of the matter power spectrum,  $\sigma_8$ , and the skewness of the matter density field. Here we show how this relation comes about.

The power spectrum,  $P(k)$ , and bispectrum,  $B(k_1, k_2, k_3)$ , which are the Fourier transforms of the two- and three-point correlation functions, are defined as

$$\langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2) \rangle = \delta_D(\mathbf{k}_1 + \mathbf{k}_2)P(k) \quad (\text{D1})$$

$$\langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3) \rangle = \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3)B(k_1, k_2, k_3), \quad (\text{D2})$$

where  $\delta(\mathbf{k})$  is the Fourier transform of the density contrast and  $\delta_D$  is the Dirac function, which results from the assumption of a homogeneous Universe. The additional assumption of isotropy implies that  $P(k_1)$  and  $B(k_1, k_2, k_3)$  only depend on the lengths  $|k_1|$  and  $|k_1, k_2, k_3|$ . The variance of the matter density field smoothed at comoving scale  $R$  is

$$\sigma^2(R) = \int_0^\infty \frac{dk}{k} k^3 \frac{P(k)}{2\pi^2} W^2(kR), \quad (\text{D3})$$

where  $W(kR)$  is the window function corresponding to the smoothing of the density field (e.g. Peebles & Harrison 1994). The conventional normalization of the amplitude of the matter power spectrum,  $\sigma_8 \equiv \sigma(8 h^{-1} \text{Mpc})$ . Hence

$$P(k) \propto \sigma_8^2. \quad (\text{D4})$$

According to perturbation theory, the first non-trivial contribution to the bispectrum, at tree-level (i.e. with no loops), comes from second-order terms in the expansion of the cosmic density fluctuation, and is given by (Fry 1984)

$$B(k_1, k_2, k_3) = P(k_1)P(k_2)F_2(\mathbf{k}_1, \mathbf{k}_2) + \text{cyc}, \quad (\text{D5})$$

where cyc refers to the two terms obtained by making cyclic permutations of the indices of the first term, and the form of kernel  $F_2$  is given in Fry (1984). Hence, we have the approximate relation

$$B(k_1, k_2, k_3) \propto P^2(k). \quad (\text{D6})$$

We define the skewness of the matter density  $\delta(k)$  as

$$\gamma(\delta) \equiv \frac{\langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3) \rangle}{(\sqrt{\langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2) \rangle})^3}, \quad (\text{D7})$$

therefore

$$\gamma(\delta) = \frac{B(k_1, k_2, k_3)}{P(k)^{3/2}}. \quad (\text{D8})$$

Combining this with equations (D4) and (D6) gives the result that the skewness of the matter density field scales linearly with  $\sigma_8$ :

$$\gamma(\delta) \propto \sigma_8. \quad (\text{D9})$$

<sup>1</sup>*Astrophysics Group, Department of Physics and Astronomy, University College London, 132 Hampstead Road, London NW1 2PS, UK*

<sup>2</sup>*Department of Physics and Electronics, Rhodes University, PO Box 94, Grahamstown 6140, South Africa*

<sup>3</sup>*Department of Physics, ETH Zurich, Wolfgang-Pauli-Strasse 16, CH-8093 Zurich, Switzerland*

<sup>4</sup>*Institute of Cosmology & Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK*

<sup>5</sup>*Institut de Ciències de l'Espai (ICE, IEEC/CSIC), E-08193 Bellaterra (Barcelona), Spain*

<sup>6</sup>*Istituto Nazionale di Astrofisica – Osservatorio Astronomico di Brera, Via E. Bianchi 46, I-23807 Merate, Italy*

<sup>7</sup>*Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA*

<sup>8</sup>*Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, Casilla 603, La Serena, Chile*

<sup>9</sup>*Fermi National Accelerator Laboratory, PO Box 500, Batavia, IL 60510, USA*

<sup>10</sup>*Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA*

<sup>11</sup>*CNRS, UMR 7095, Institut d'Astrophysique de Paris, F-75014 Paris, France*

<sup>12</sup>*Sorbonne Universités, UPMC Univ Paris 06, UMR 7095, Institut d'Astrophysique de Paris, F-75014 Paris, France*

<sup>13</sup>*Carnegie Observatories, 813 Santa Barbara St., Pasadena, CA 91101, USA*

<sup>14</sup>*Kavli Institute for Particle Astrophysics & Cosmology, Stanford University, PO Box 2450, Stanford, CA 94305, USA*

<sup>15</sup>*SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA*

<sup>16</sup>*Laboratório Interinstitucional de e-Astronomia – LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ 20921-400, Brazil*

<sup>17</sup>*Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ 20921-400, Brazil*

<sup>18</sup>*Department of Astronomy, University of Illinois, 1002 W. Green Street, Urbana, IL 61801, USA*

<sup>19</sup>*National Center for Supercomputing Applications, 1205 West Clark St., Urbana, IL 61801, USA*

<sup>20</sup>*School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK*

<sup>21</sup>*Excellence Cluster Universe, Boltzmannstr. 2, D-85748 Garching, Germany*

<sup>22</sup>*Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, D-81679 Munich, Germany*

<sup>23</sup>*Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109, USA*

<sup>24</sup>*Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA*

<sup>25</sup>*Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA*

<sup>26</sup>*Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA*

<sup>27</sup>*Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA*

<sup>28</sup>*Department of Physics, The Ohio State University, Columbus, OH 43210, USA*

<sup>29</sup>*Australian Astronomical Observatory, North Ryde, NSW 2113, Australia*

<sup>30</sup>*Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo, CP 66318, CEP 05314-970, São Paulo, SP, Brazil*

<sup>31</sup>*Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain*

<sup>32</sup>*Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, E-08193 Bellaterra (Barcelona), Spain*

<sup>33</sup>*Department of Physics and Astronomy, Pevensey Building, University of Sussex, Brighton BN1 9QH, UK*

<sup>34</sup>*Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), E-28040 Madrid, Spain*

<sup>35</sup>*ICTP South American Institute for Fundamental Research, Instituto de Física Teórica, Universidade Estadual Paulista, 01140-070 São Paulo, Brazil*

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