**Astronomy
&
Astrophysics**

The non-Gaussianity of the cosmic shear likelihood
or how odd is the *Chandra* Deep Field South?

J. Hartlap[1], T. Schrabback[2,1], P. Simon[3], and P. Schneider[1]

[1] Argelander-Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, 53121 Bonn, Germany
    e-mail: `hartlap@astro.uni-bonn.de`
[2] Leiden Observatory, Universiteit Leiden, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands
[3] The Scottish Universities Physics Alliance (SUPA), Institute for Astronomy, School of Physics, University of Edinburgh,
    Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK

ABSTRACT

Aims. We study the validity of the approximation of a Gaussian cosmic shear likelihood. We estimate the true likelihood for a fiducial cosmological model from a large set of ray-tracing simulations and investigate the impact of non-Gaussianity on cosmological parameter estimation. We investigate how odd the recently reported very low value of σ_8 really is as derived from the *Chandra* Deep Field South (CDFS) using cosmic shear by taking the non-Gaussianity of the likelihood into account, as well as the possibility of biases coming from the way the CDFS was selected.
Methods. A brute force approach to estimating the likelihood from simulations must fail because of the high dimensionality of the problem. We therefore use independent component analysis to transform the cosmic shear correlation functions to a new basis, in which the likelihood approximately factorises into a product of one-dimensional distributions.
Results. We find that the cosmic shear likelihood is significantly non-Gaussian. This leads to both a shift of the maximum of the posterior distribution and a significantly smaller credible region compared to the Gaussian case. We re-analyse the CDFS cosmic shear data using the non-Gaussian likelihood in combination with conservative galaxy selection criteria that minimise calibration uncertainties. Assuming that the CDFS is a random pointing, we find $\sigma_8 = 0.68^{+0.09}_{-0.16}$ for fixed $\Omega_m = 0.25$. In a WMAP5-like cosmology, a value equal to or lower than this would be expected in ≈5% of the times. Taking biases into account arising from the way the CDFS was selected, which we model as being dependent on the number of haloes in the CDFS, we obtain $\sigma_8 = 0.71^{+0.10}_{-0.15}$. Combining the CDFS data with the parameter constraints from WMAP5 yields $\Omega_m = 0.26^{+0.03}_{-0.02}$ and $\sigma_8 = 0.79^{+0.04}_{-0.03}$ for a flat universe.

Key words. gravitational lensing – cosmology: cosmological parameters – methods: statistical – methods: numerical

1. Introduction

Weak gravitational lensing by the large-scale structure in the Universe, or cosmic shear, is becoming a more and more important tool to constrain cosmological parameters. It is largely complementary to other cosmological probes like the cosmic microwave background or the clustering of galaxies, and particularly sensitive to the matter density Ω_m and the normalisation of the matter power spectrum σ_8. Important constraints have already been obtained by Benjamin et al. (2007), who compiled a set of five weak lensing surveys, and from the CFHT Legacy Survey (Hoekstra et al. 2006; Semboloni et al. 2006; Fu et al. 2008). In subsequent years, a new generation of surveys like KIDS or Pan-STARRS (Kaiser & Pan-STARRS Collaboration 2005) will allow cosmic shear to be measured with statistical uncertainties that are much smaller than the systematic errors both on the observational and the theoretical sides. Strong efforts are now being made to find sources of systematics in the process of shape measurement and shear estimation (e.g. Massey et al. 2007a). In addition, new methods of shape measurement are being explored, such as the shapelet formalism (Refregier & Bacon 2003; Kuijken 2006) or the methods proposed in Bernstein & Jarvis (2002) and Miller et al. (2007).

It is equally important to have accurate theoretical model predictions that can be fit to the expected high-quality measurements. Currently, these models are all based on fitting formulae for the three-dimensional matter power spectrum derived from *N*-body simulations as given by Peacock & Dodds (1996) and more recently by Smith et al. (2003). However, these are only accurate at best to the percent level on the scales relevant to this and similar works when compared to ray-tracing simulations based on state-of-the-art *N*-body simulations (Hilbert et al. 2009), such as the Millennium Run (Springel et al. 2005). Therefore, there is a strong need for a large ray-tracing effort to obtain accurate semi-numerical predictions for a range of cosmological parameters.

While a tremendous effort is currently being directed to the solution of these problems, the actual process of parameter estimation has so far received relatively little attention. Obviously, the statistical data analysis has to achieve the same accuracy as the data acquisition if the aforementioned efforts are not to be wasted.

The standard procedure for converting measurements of second-order cosmic shear statistics into constraints on cosmological parameters is to write down a likelihood function and to determine the location of its maximum for obtaining estimates of the cosmological parameters of interest. To make this feasible, several approximations are commonly made. Despite the shear field being non-Gaussian due to nonlinear structure growth, lacking an analytical description the likelihood is most often approximated by a multivariate Gaussian distribution. The covariance matrix for the Gaussian likelihood then remains to be determined, which is an intricate issue by itself.

In most previous studies, the dependence of the covariance matrix on cosmological parameters has been ignored when writing down the likelihood function. Instead, it was kept fixed to some fiducial cosmological model. The dependence of the covariance matrix on the cosmological parameters has been investigated in Eifler et al. (2009) for the case of Gaussian shear fields. The authors find that this has a significant effect on the constraints on cosmological parameters (reducing the size of the credible regions) and will be particularly important for future large-area surveys.

There are several approaches to determine the covariance for the fiducial set of parameters: Hoekstra et al. (2006) use the covariance matrix derived for a Gaussian shear field. Although this is rather easy to compute (Joachimi et al. 2008), the errors are strongly underestimated particularly on small scales. Another option is to estimate the covariance from the data itself (e.g. Massey et al. 2007b). This will become sensible and feasible mostly for the upcoming large surveys, which can be safely split into smaller subfields without severely underestimating cosmic variance. A third possibility, which currently seems to be the most accurate, is to measure the covariance matrix from a large sample of ray-tracing simulations. Semboloni et al. (2007) have provided a fitting formula which allows one to transform covariances computed for Gaussian shear fields into covariances including non-Gaussianity. Another promising way, which would also easily allow one to take into account the dependence on cosmological parameters, is the semi-analytical computation using the halo model (Scoccimarro et al. 1999; Cooray & Hu 2001; Takada & Jain 2009).

However, all these works are based on the assumption that the likelihood is well approximated by a Gaussian. In this paper, we study the impact of this assumption on the shape of the posterior probability distribution of the matter density parameter Ω_m and the power spectrum normalisation σ_8. Furthermore, we compute Fisher matrix constraints for the four-dimensional parameter space spanned by Ω_m, σ_8, h_{100} and Ω_Λ. We propose a method to numerically compute the likelihood function from a large set of ray-tracing simulations based on the technique of *independent component analysis* (ICA, e.g. Jutten & Hérault 1991; Comon et al. 1991). ICA is a technique for the separation of independent source signals underlying a set of observed random variables, a statistical method related to factor analysis and principal component analysis (PCA). An approach similar to ours, called *projection pursuit density estimation*, which we use to verify our results, was proposed by Friedman et al. (1984).

In their cosmic shear analysis of the combined HST GEMS and GOODS data of the CDFS, Schrabback et al. (2007) (S07, from hereon) have found a very low value of $\sigma_8(\Omega_m = 0.3) = 0.52^{+0.11}_{-0.15}$. In the second part of this paper, we present a reanalysis of the cosmic shear data of S07. Using our estimate of the non-Gaussian likelihood, we investigate whether cosmic variance alone is responsible for producing the low σ_8-estimate or whether the criteria applied by Giacconi et al. (2001) to select a field suitable for deep X-ray observations have a share in this.

The outline of our paper is as follows: in Sect. 2, we describe our sample of ray-tracing simulations which we use for the likelihood estimation. In Sect. 3, we briefly review the lensing quantities relevant for this paper and Bayesian parameter estimation. We introduce our method of estimating the "true" likelihood and illustrate the impact of non-Gaussianity on parameter estimation using the example of a CDFS-like survey. In Sect. 4, we present the improved cosmic shear analysis of the CDFS and investigate possible reasons for the low power spectrum normalisation found in S07.

2. Ray-tracing simulations

We have performed a set of 10 N-body simulations using the publically available code GADGET-2 (Springel 2005), all of which are realisations of the same WMAP-5-like cosmology ($\Omega_m = 0.25$, $\Omega_\Lambda = 0.75$, $\Omega_b = 0.04$, $n_s = 1.0$, $\sigma_8 = 0.78$, $h_{100} = 0.73$). The simulation boxes are $L_{box} = 150\,h_{100}^{-1}$ Mpc on a side, populated by $N_p = 256^3$ dark matter particles with masses of $m_p = 1.2 \times 10^{10}\,h_{100}^{-1}\,M_\odot$. We have started the simulations at $z = 50$ and obtained snapshots from $z = 0$ to $z = 4.5$ in intervals of Δz corresponding to the box size, so that a suitable snapshot is available for each lens plane.

In the following, we only give a brief description of our ray-tracing algorithm and refer the reader to, for example, Jain et al. (2000) or Hilbert et al. (2009) for a more detailed introduction.

The ray-tracing is performed by dividing the dark matter distribution into redshift slices and projecting each slice onto a lens plane. Starting at the observer, light rays are shot through this array of lens planes. We assume that deflections only take place at the planes themselves, and that the rays propagate on straight lines in the space between two planes. In our case, each redshift slice corresponds to one output box of the N-body simulation and was projected as a whole onto a lens plane, preserving the periodic boundary conditions of the simulation box. To avoid repetition of structure along the line of sight, the planes were randomly shifted and rotated. The light rays are shot from the observer through the set of lens planes, forming a regular grid on the first plane. We then use FFT methods to compute the lensing potential on each lens plane, from which we obtain the deflection angle and its partial derivatives on a grid. The ray position and the Jacobian of the lens mapping for each ray are obtained by recursion: given the ray position on the current lens plane, its propagation direction (known from the position on the previous plane), and the deflection angle on the current plane interpolated onto the ray, we immediately obtain the ray position on the next plane. Differentiation of this recursion formula with respect to the image plane coordinates yields a similar relation for the Jacobian of the lens mapping, which takes into account the previously computed tidal deflection field (for a detailed description of the formalism used, see Hilbert et al. 2009). The recursion is performed until we reach the redshift cut-off at $z = 4.5$.

We obtain the final Jacobian for a given source redshift distribution by performing a weighted average over the Jacobians for the light paths to each lens plane. Since our aim is to create mock catalogues comparable to those of the CDFS field, we use the redshift distribution found for our revised galaxy catalogues (Smail et al. 1995, see Sect. 4.1):

$$p(z_s) = A \left(\frac{z_s}{z_0}\right)^\alpha \exp\left[-\left(\frac{z_s}{z_0}\right)^\beta\right],$$

where $z_0 = 1.55$, $\alpha = 0.59$, $\beta = 1.35$ and A is a normalisation constant. This corresponds to a mean source redshift of $\bar{z}_s = 1.54$. We then create the mock source catalogue by randomly sampling the resulting shear maps with $N_s = n_s\Omega^2$ galaxies, where $n_s = 68$ arcmin^{-2} is the number density of sources and $\Omega = 0\overset{\circ}{.}5$ is the side length of the simulated field. In total, we have produced 9600 quasi-independent realisations of the CDFS field, based on different random shifts and rotations of the lens planes and the various N-body simulations.

3. The non-Gaussianity of the cosmic shear likelihood

3.1. Cosmic shear

Perhaps the most common way to extract the lensing information from the measured shapes of distant galaxies is to estimate the two-point correlation functions of the distortion field. One defines two shear correlation functions (for more details, see e.g. Schneider 2006)

$$\xi_{\pm}(\theta) = \langle \epsilon_t(\boldsymbol{\vartheta})\epsilon_t(\boldsymbol{\theta} + \boldsymbol{\vartheta}) \rangle \pm \langle \epsilon_{\times}(\boldsymbol{\vartheta})\epsilon_{\times}(\boldsymbol{\theta} + \boldsymbol{\vartheta}) \rangle, \tag{1}$$

where $\epsilon_{t,\times}$ are the tangential and cross components of the measured ellipticity relative to the line connecting the two galaxies, and θ is the angular separation. An unbiased estimator for the shear correlation functions for a random distribution of galaxies is given in Schneider et al. (2002):

$$\hat{\xi}_{\pm}(\theta) = \frac{1}{N_p(\theta)} \sum_{i,j} \left(\epsilon_{it}\epsilon_{jt} \pm \epsilon_{i\times}\epsilon_{j\times} \right) \Delta_{\theta}(|\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_j|). \tag{2}$$

Here, i and j label galaxies at angular positions $\boldsymbol{\vartheta}_i$ and $\boldsymbol{\vartheta}_j$, respectively. The function $\Delta_{\theta}(\phi)$ is 1 if ϕ falls into the angular separation bin centred on θ, and is zero otherwise. Finally, N_p is the number of pairs of galaxies in the bin under consideration.

3.2. Parameter estimation

Let us assume that we have measured the shear correlation functions $\xi_{\pm}(\theta_i)$ on $p/2$ angular separation bins θ_i and now wish to infer some parameters $\boldsymbol{\pi}$ of our model $\boldsymbol{m}(\boldsymbol{\pi})$ for $\xi_{\pm}(\theta_i)$. For what follows, we define the joint data vector $\boldsymbol{\xi} = (\boldsymbol{\xi}_+, \ \boldsymbol{\xi}_-)^t$, which in total is supposed to have p entries.

Adopting a Bayesian point of view, our aim is to compute the posterior likelihood, i.e. the probability distribution of a parameter vector $\boldsymbol{\pi}$ given the information provided by the data $\boldsymbol{\xi}$:

$$p(\boldsymbol{\pi}|\boldsymbol{\xi}) = \frac{p(\boldsymbol{\pi})}{p(\boldsymbol{\xi})} p(\boldsymbol{\xi}|\boldsymbol{\pi}). \tag{3}$$

Here, $p(\boldsymbol{\pi})$ is the prior distribution of the parameters, which incorporates our knowledge about $\boldsymbol{\pi}$ prior to looking at the data; such can originate from previous measurements or theoretical arguments. The evidence $p(\boldsymbol{\xi})$ in this context simply serves as a normalisation factor. Hitherto, it has been assumed that the likelihood $p(\boldsymbol{\xi}|\boldsymbol{\pi})$ is a Gaussian distribution:

$$p(\boldsymbol{\xi}|\boldsymbol{\pi}) = \frac{1}{(2\pi)^{p/2} \det \mathsf{C}(\boldsymbol{\pi})^{1/2}} \\ \times \exp\left\{ -\frac{1}{2} [\boldsymbol{\xi} - \boldsymbol{m}(\boldsymbol{\pi})]^t \ \mathsf{C}^{-1}(\boldsymbol{\pi}) \ [\boldsymbol{\xi} - \boldsymbol{m}(\boldsymbol{\pi})] \right\}, \tag{4}$$

where $\mathsf{C}(\boldsymbol{\pi})$ is the covariance matrix of $\boldsymbol{\xi}$ as predicted by the underlying model. Usually, however, the dependence of the covariance matrix upon cosmological parameters is not taken into account. Rather, the covariance that is computed for a fixed fiducial set of parameters $\boldsymbol{\pi}_0$ is used in Eq. (4). Under this approximation, the likelihood is a function of the difference $\Delta(\boldsymbol{\pi}) = \boldsymbol{\xi} - \boldsymbol{m}(\boldsymbol{\pi})$ only:

$$p(\boldsymbol{\xi}|\boldsymbol{\pi}) = L_{\boldsymbol{\pi}_0} [\Delta(\boldsymbol{\pi})]. \tag{5}$$

3.3. Estimating the likelihood

The choice of the functional form of the likelihood as given by Eq. (4) is only approximate. Since the underlying shear field in the correlation function measurement becomes non-Gaussian in particular on small scales due to nonlinear structure formation, there is no good reason to expect the distribution of the shear correlation function to be Gaussian. Our aim therefore is to use a very large sample of ray-tracing simulations to estimate the likelihood and explore the effects of the deviations from a Gaussian shape on cosmological parameter constraints.

In this work, we have to sustain the approximation that the functional form of the likelihood does not depend on cosmology in order to keep computation time manageable. Our ray-tracing simulations were all done for identical cosmological parameters, which is our fiducial parameter vector $\boldsymbol{\pi}_0$. Thus, as in Eq. (5) the likelihood depends on cosmology only through the difference $\Delta(\boldsymbol{\pi}) = \boldsymbol{\xi} - \boldsymbol{m}(\boldsymbol{\pi})$.

Since $L_{\boldsymbol{\pi}_0}$ is the probability of obtaining the data $\boldsymbol{\xi}$ given the parameters $\boldsymbol{\pi}_0$, we in principle have to estimate the p-dimensional distribution of $\boldsymbol{\xi}$ from our sample of N ray-tracing simulations. However, due to the high dimensionality of the problem, a brute force approach to estimate the full joint distribution is hopeless. The problem would simplify considerably if we could find a transformation

$$\boldsymbol{s} = \boldsymbol{f}[\Delta(\boldsymbol{\pi})], \tag{6}$$

such that

$$p_s(\boldsymbol{s}|\boldsymbol{\pi}_0) = \prod_{i=1}^{n_{IC}} p_{s_i}(s_i|\boldsymbol{\pi}_0). \tag{7}$$

Here, $\boldsymbol{f}$ is in general a mapping from $\mathbb{R}^p$ to $\mathbb{R}^{n_{IC}}$ ($n_{IC} \leq p$) and $\boldsymbol{s} \in \mathbb{R}^{n_{IC}}$ is our new data vector. This would reduce the problem to estimating n_{IC} one-dimensional probability distributions instead of a single p-dimensional one. Equation (7) is equivalent to the statement that we are looking for a new set of basis vectors of $\mathbb{R}^{n_{IC}}$ in which the components s_i of the shear correlation function are statistically independent. It is virtually impossible to find the (in general nonlinear) mapping $\boldsymbol{f}$. However, it is possible to make progress if we make the ansatz that $\boldsymbol{f}$ is linear:

$$\boldsymbol{s} = \mathsf{A}\Delta(\boldsymbol{\pi}), \tag{8}$$

where $\mathsf{A} \in \mathbb{R}^{n_{IC} \times p}$ is the transformation or "un-mixing" matrix.

Our likelihood estimation procedure is as follows: the first step is to remove first-order correlations from the data vector by performing a PCA (e.g. Press et al. 1992). This yields a basis in which the components of $\boldsymbol{\xi}$ are uncorrelated. If we knew that the distribution of $\boldsymbol{\xi}$ were Gaussian, this would be sufficient, because in this case uncorrelatedness is equivalent to statistical independence. However, for a general distribution, uncorrelatedness is only a necessary condition for independence. Since we suspect that the likelihood is non-Gaussian, a second change of basis, determined by the ICA technique (described in detail in the next section), is carried out which then results in the desired independence. We then use a kernel density method (see e.g. Hastie et al. 2001; Venables & Ripley 2002, and references therein) to estimate and tabulate the one-dimensional distributions $p_{s_i}(s_i|\boldsymbol{\pi}_0)$ in this new basis. The density estimate is constructed by smoothing the empirical distribution function of the observations of s_i,

$$p_{s_i}^{emp}(x) = \frac{1}{N} \sum_{j=1}^{N} \delta_D\left(x - s_i^{(j)}\right), \tag{9}$$

where $s_i^{(j)}$ is the jth of N observations of s_i and δ_D is the Dirac delta-function, with a smooth kernel K. The estimate $\hat{p}_{s_i}$ of the desired density p_{s_i} then is given by

$$\hat{p}_{s_i}(x) = \frac{1}{Nb} \sum_{j=1}^{N} K\left(\frac{x - s_i^{(j)}}{b}\right), \tag{10}$$

where $s_i^{(j)}$ is the jth of N observations of s_i and b is the bandwidth. For the kernel K we use a Gaussian distribution. It has been shown that the shape of the Kernel K is of secondary importance for the quality of the density estimate; much more important is the choice of the bandwidth b. If b is too small, $\hat{p}_{s_i}$ is essentially unbiased, but tends to have a high variance because the noise is not properly smoothed out. On the other hand, choosing a bandwidth that is too large results in a smooth estimate with low variance, but a higher bias, because real small scale features of the probability density are smeared out. Our choice of the bandwidth is based on the "rule of thumb" (e.g. Silverman 1986; Scott 1992; Davison 2003): $b = 0.9 \min(\hat{\sigma}, R/1.34) N^{-1/5}$. Here, $\hat{\sigma}$ is the sample standard deviation and R is the inter-quartile range of the sample.

Constraints on cosmological parameters can now be derived as follows: we transform our set of model vectors and the measured correlation function to the new ICA basis:

$$\check{m}(\boldsymbol{\pi}) = A\, m(\boldsymbol{\pi}), \tag{11}$$

$$\check{\boldsymbol{\xi}} = A\, \boldsymbol{\xi}, \tag{12}$$

so that $s = \check{\boldsymbol{\xi}} - \check{m}(\boldsymbol{\pi})$. The ICA posterior distribution is then given by

$$p(\boldsymbol{\pi}|\boldsymbol{\xi}) \propto p(\boldsymbol{\pi}) \prod_{i=1}^{n_{IC}} p_{s_i}(\check{\xi}_i - \check{m}_i(\boldsymbol{\pi})|\boldsymbol{\pi}_0). \tag{13}$$

3.4. Independent component analysis

We now briefly outline the ICA method (Hyvärinen et al. 2001; Hyvärinen & Oja 2000), which we use to find the new basis in $\mathbb{R}^{n_{IC}}$ in which the components of $\boldsymbol{\Delta}$ are (approximately) statistically independent. ICA is best introduced by assuming that the data at hand were generated by the following linear model:

$$\boldsymbol{\Delta} = M\boldsymbol{s}, \tag{14}$$

where $\boldsymbol{s}$ is a vector of statistically independent source signals with non-Gaussian probability distributions and M is the mixing matrix. For simplicity, we will from now on only consider the case $n_{IC} = p$, in which case the mixing matrix M is simply the inverse of the un-mixing matrix A in Eq. (8). The goal of ICA is to estimate both M and $\boldsymbol{s}$ from the data.

An intuitive, though slightly hand-waving way to understand how ICA works is to note that a set of linear combinations Y_i of independent, non-Gaussian random variables X_j will usually have distributions that are more Gaussian than the original distributions of the X_j (central limit theorem). Reversing this argument, this suggests that the X_j could be recovered from a sample of the Y_i by looking for linear combinations of the Y_i that have the least Gaussian distributions. These linear combinations will also be close to statistically independent. A more rigorous justification of the method can be found in Hyvärinen et al. (2001).

The ICA algorithm consists of two parts, the first of which is a preprocessing step: after subtracting the mean $\bar{\boldsymbol{\Delta}} = \langle\boldsymbol{\Delta}\rangle$ from $\boldsymbol{\Delta}$, the data is whitened, i.e. a linear transformation $\tilde{\boldsymbol{\Delta}} = L(\boldsymbol{\Delta} - \bar{\boldsymbol{\Delta}})$ is

introduced such that $\langle\tilde{\boldsymbol{\Delta}}\tilde{\boldsymbol{\Delta}}^t\rangle = E$, where E is the unit matrix. This can be achieved by the eigen-decomposition of the covariance matrix $C = UDU^t$ of $\boldsymbol{\Delta}$, where $D = \text{diag}(d_1, \ldots, d_p)$, by choosing $L = D^{-1/2}U^t$. Note that U is orthonormal and that $d_i \geq 0$ for all i. As will be discussed below, each source signal s_i can only be determined up to a multiplicative constant using ICA. We choose these factors such that $\langle ss^t\rangle = E$. The effect of the whitening is that the new mixing matrix $\tilde{M} = LM$ between $\tilde{\boldsymbol{\Delta}}$ and $\boldsymbol{s}$ is orthogonal. This can be seen as follows: $E = \langle\tilde{\boldsymbol{\Delta}}\tilde{\boldsymbol{\Delta}}^t\rangle = \tilde{M}\langle ss^t\rangle\tilde{M}^t$. Since we have chosen $\langle ss^t\rangle = E$, the claim follows.

After the preprocessing, the components of $\tilde{\boldsymbol{\Delta}}$ are uncorrelated. This would be equivalent to statistical independence if their distributions were Gaussian. However, as this is not the case here, a further step is needed. It consists of finding a new set of orthogonal vectors $\boldsymbol{w}_i$ (the row vectors of $\tilde{M}$) such that the distributions $p_{z_i}(z_i)$ of

$$z_i = \tilde{\boldsymbol{\Delta}}_i \cdot \boldsymbol{w}_i \tag{15}$$

maximise a suitable measure of non-Gaussianity. A common method to achieve this is to minimise the entropy (or approximations thereof) of the z_i, which is defined by

$$H_{z_i} = -\int dy \, p_{z_i}(y) \log p_{z_i}(y). \tag{16}$$

Since it can be shown that the Gaussian distribution has the largest entropy of all distributions of equal variance, this can be rewritten as maximising the so-called negentropy of the z_i, defined by

$$J_{z_i} = H_{z_i^{\text{Gauss}}} - H_{z_i}. \tag{17}$$

Here, z_i^{Gauss} is a Gaussian random variable with the same variance as z_i and $J(z_i) \geq 0$. Starting from randomly chosen initial directions $\boldsymbol{w}_i$, the algorithm tries to maximise $J(z_i)$ iteratively (in practice, it is sufficient to use a simple approximation to the negentropy). For more details, the reader is again referred to Hyvärinen et al. (2001).

ICA suffers from several ambiguities, none of which, however, is crucial for this work. First of all, the amplitudes of the source signals cannot be determined, since any prefactor λ to the signal s_i can be cancelled by multiplication of the corresponding column of the mixing matrix by $1/\lambda$. Secondly, the order of the independent components is not determined, since any permutation of the s_i can be accommodated by corresponding changes to M. Thirdly, ICA does not yield a unique answer if at least some of the s_i are Gaussian – the subset of Gaussian signals is only determined up to an orthogonal transformation. This is not an issue in our context, since the Gaussian signals will be uncorrelated thanks to the preprocessing steps, and uncorrelatedness implies statistical independence for Gaussian random variables.

Several interpretations of ICA and algorithms exist and are described in detail in Hyvärinen et al. (2001). In this work, we use an implementation of the fastICA algorithm (Hyvärinen & Oja 1997) for the R language (R Development Core Team 2007)[1].

3.5. Tests

In this section, we present the results of a number of tests we have performed to insure that our results are not affected by convergence issues or statistical biases of any kind.
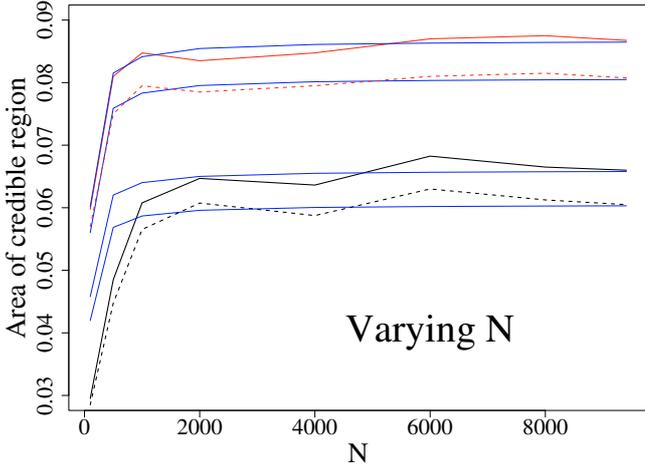
---

[1] http://www.r-project.org/

Fig. 1. Area of the 68% (dashed lines) and 99% (solid lines) credible regions in the Ω_m-σ_8-plane as function of the sample size N, for the Gaussian likelihood (red, upper curves) and the likelihood computed using the ICA algorithm (black, lower curves). Blue lines are the predicted areas based on Eq. (18).

The `fastICA` algorithm requires a set of randomly chosen directions w_i as initial conditions. It then iteratively computes corrections to these vectors in order to increase the negentropy of the projections of the data vectors onto these directions (Eq. (15)), followed by an orthonormalisation step. It is not clear a priori whether the algorithm will settle in the same negentropy maxima for different sets of initial vectors. This concern is backed by the fact that at least some of the $p_{s_i}(s_i|\pi_0)$ might be very close to Gaussian, which might hamper convergence even further. We have therefore tested whether we obtain the same set of basis vectors from a large number of different initial conditions. We find that this is indeed the case for those w_i for which the distribution of $p_{z_i}(z_i = \tilde{\Delta} \cdot w_i)$ departs significantly from a Gaussian. As expected, the directions leading to a rather Gaussian p_{z_i} are different for different starting values, reflecting the inability of ICA to distinguish between Gaussian source signals. However, the posterior distributions derived using our algorithm do not differ notably when using different initial conditions. This is even true if the `fastICA` algorithm does not formally converge (i.e. when the differences of some of the basis vectors between two iterations is not small): after a few hundred iterations, the non-Gaussian directions are determined and do not change anymore. The reason for not reaching convergence is that the algorithm still tries to find negentropy maxima in the subspace of Gaussian directions.

As has been noted in Hartlap et al. (2007), statistical biases can become significant already for the Gaussian approximation of the likelihood (Eq. (4)): care has to be taken if the covariance matrix of the correlation function (given on p bins) is estimated from a finite set of N simulations or observations. Inverting the estimated covariance yields a biased estimate of the inverse:

$$\left\langle \hat{C}^{-1} \right\rangle = \frac{N-1}{N-p-2} \, \Sigma^{-1} \text{ for } p < N-1, \tag{18}$$

where $\hat{C}$ is the estimated and Σ the true covariance matrix. This bias leads to an underestimation of the size of credible regions by a factor of $(N-p-2)/(N-1) \approx 1 - p/N$. We suspect that a similar bias occurs in our likelihood estimation procedure. In Fig. 1, we therefore plot the area of the 68% and 99% credible regions of the posterior distribution for Ω_m and σ_8 (keeping all other cosmological parameters fixed to their fiducial values)

as functions of the number N of observations of the correlation functions used to estimate the ICA transformation (black curves). To exclude noise effects from the analysis, we use the theoretical prediction of the correlation function for the fiducial cosmological parameters as data vector. We set $p = 30$ throughout. For comparison, we also show the areas computed using the Gaussian likelihood (red curves). In the latter case, the bias predicted by Eq. (18) is clearly visible as a decrease of the area when N becomes small. The ICA method suffers from a similar bias, although the behaviour at small N seems to be slightly different. More important, though, is the fact that this bias is unimportant for reasonably large sample sizes ($N \gtrsim 2000$). Since we always use the full sample ($N = 9600$) in the following, this bias is completely negligible.

Our method to estimate the likelihood crucially depends on the assumption that a linear transformation makes the components of the shear correlation vectors statistically independent. A necessary condition for mutual statistical independence of all s_i is pairwise independence. The components i and j are called pairwise statistical independent if $p(s_i, s_j) = p_{s_i}(s_i) \, p_{s_j}(s_j)$. We therefore compare the joint pairwise distributions $p(s_i, s_j)$ to the product distributions $p_{s_i}(s_i) \, p_{s_j}(s_j)$, where we estimate $p(s_i, s_j)$ using a two-dimensional extension (using a bi-variate Gaussian kernel) of the kernel density method given by Eq. (10). We give two examples in Fig. 2, where we compare the joint and product distributions of the two most-non-Gaussian components and two nearly Gaussian components. As expected, a simple PCA is not enough to achieve pairwise statistical independence in the non-Gaussian case. Only after performing the ICA, pairwise independence is achieved.

A more rigorous test for mutual statistical independence for the multivariate, continuous case was proposed by Chiu et al. (2003). It is based on the observation that if x is a continuous random variable and $P(x)$ is its cumulative distribution function (CDF), then $z = P(x)$ is uniformly distributed in [0, 1]. If we are given a set of statistically independent random variables s_i, this means that the joint distribution of $z_i = P_i(s_i)$, where again P_i is the CDF of s_i, is uniform in the multidimensional unit cube. On the other hand, if the assumption of statistical independence of the s_i is violated, the joint density p_z of the z_i is given by

$$p_z(z) = p_z \left[P_1(s_1), \ldots, P_n(s_n) \right]$$
$$= p_s(s_1, \ldots, s_n) \left| \frac{\partial z}{\partial s} \right|^{-1}$$
$$= \frac{p_s(s_1, \ldots, s_n)}{\prod_{i=1}^{n} p_i(s_i)}. \tag{19}$$

Here, $p_i(s_i)$ is the distribution function of s_i only and p_s is the joint distribution function of $s_1, \ldots, s_n$. This means that the joint distribution of the z_i is not uniform if the s_i are statistically dependent. Therefore, we can test if the s_i we obtain from the ICA procedure are indeed independent by computing their empirical cumulative distribution functions, carrying out the above transformation and finally testing for multivariate uniformity. Such a test was described in Liang et al. (2001), to which we refer the reader for more details.

Applying the test to the s_i that we have obtained from our ICA procedure, we have to reject statistical independence at 99% confidence. This means that the ICA does not remove all dependencies between the components of the shear correlation function. This result, however, does not give an indication of how these residual dependencies affect our likelihood estimate and the conclusions regarding constraints on cosmological parameters. We therefore compare the constraints derived from the ICA
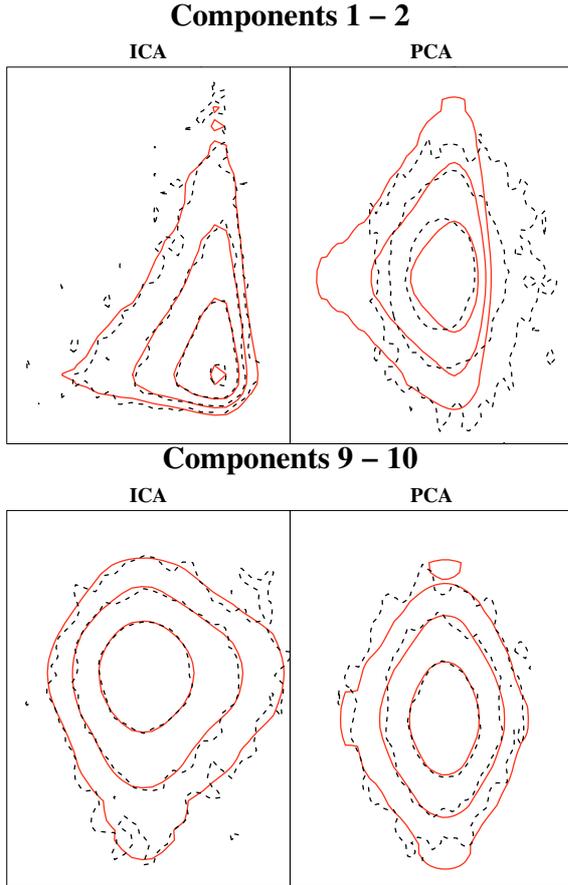
Components 1 – 2

ICA   PCA

Components 9 – 10

ICA   PCA

Fig. 2. Comparison of the joint distributions $p(s_i, s_j)$ (black dashed contours) and the product $p_{s_i}(s_i) \, p_{s_j}(s_j)$ (solid red contours) for the two most non-Gaussian components ($i = 1$, $j = 2$) and two rather Gaussian ones ($i = 9$, $j = 10$), after performing ICA (*left panels*) and PCA (*right panels*). The components have been ranked and labelled according to their non-Gaussianity; the ith PCA component is in general not the same as ith ICA-component. *In the right panel* of each plot, the distributions with respect to the PCA basis vectors are shown and in *the left panel*, the distributions in the ICA basis are displayed. Statistical independence is indicated by $p(s_i, s_j) = p_{s_i}(s_i) \, p_{s_j}(s_j)$.

likelihood with the constraints from the likelihood estimated using an alternative method, called *projection pursuit density estimation* (PPDE; Friedman et al. 1984), which we describe in detail in Appendix A. This method is free from any assumptions regarding statistical independence and therefore provides an an ideal cross-check for the ICA method. For the comparison, we have computed the shear correlation functions with $p = 10$, and we also use $n_{IC} = 10$ independent components. The resulting contours in the $\Omega_m - \sigma_8$-plane are shown in Fig. 3. Both posterior likelihoods are very similar, although the credible regions of the PPDE posterior have a slightly smaller area than the contours of the ICA posterior (which actually supports the findings presented in the next section). Given the good agreement of the two methods, we will henceforth only make use of the ICA procedure, which is considerably faster and numerically less contrived than PPDE.

3.6. Results on the posterior

The most interesting question is how much the posterior distribution computed from the non-Gaussian ICA likelihood will differ
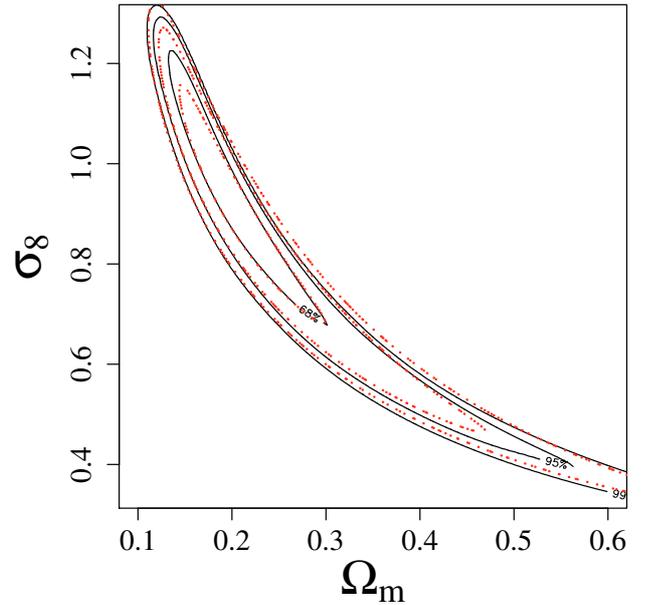

Fig. 3. Comparison of the posterior likelihoods for (Ω_m, σ_8), computed using the ICA likelihood (black contours) and the PPDE likelihood (red contours). Shown are the contours of the 68%, 95% and 99% credible regions.

from the Gaussian approximation. We have investigated this for the case of the CDFS and the parameter set (Ω_m, σ_8). Here and henceforth, we use 15 angular bins for ξ_+ and ξ_- in the range from $12''$ to $30'$, i.e. $p = 30$. For the data vector, we do not use the correlation functions from our simulations, but take the theoretical prediction for our fiducial parameter set instead. This allows us to study the shape of the posterior likelihood independent of noise in the data and biases due to the fact that the theoretical model does not quite match the mean correlation function from the simulations. In Fig. 4, we show the contours of the posterior computed in this way from the likelihood estimated using our ICA method (left panel) and from the Gaussian likelihood. We have assumed $\sigma_{|\epsilon|} = 0.45$ for the dispersion of the intrinsic galaxy ellipticities. The shape of the ICA posterior is different from that of the Gaussian approximation in three respects: it is steeper (leading to smaller credible regions), the maximum is shifted towards higher σ_8 and lower Ω_m, and the contours are slightly tilted. The first two differences can be traced back to the shape of the distributions of the individual ICA components: most of the distribution functions p_{s_i} are generally slightly steeper than a Gaussian and most of the non-Gaussian components are in addition strongly skewed, thus shifting the peak of the posterior. Generally, these differences are more pronounced in the direction of the Ω_m–σ_8-degeneracy and towards lower values of both parameters, where the posterior is shallower.

Of more practical relevance is how the parameter constraints change when the ICA likelihood is used for the analysis of large weak lensing surveys. Here, we consider surveys consisting of N_f CDFS-like fields. Bayesian theory states that if N_f is large enough, the posterior probability distribution of the parameters becomes Gaussian, centred on the true parameter values, with covariance matrix $(N_f \, \mathsf{F})^{-1}$ (e.g. Gelman et al. 2004). Here, F is the Fisher matrix (Kendall et al. 1987), which is defined by

$$\mathsf{F}_{\alpha\beta} = \left\langle \frac{\partial \log L}{\partial \pi_\alpha} \frac{\partial \log L}{\partial \pi_\beta} \right\rangle, \tag{20}$$
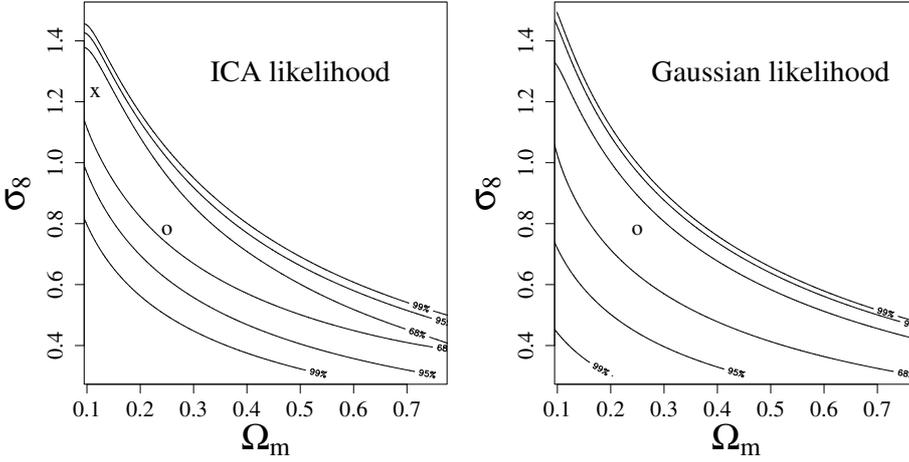
Fig. 4. Comparison of the posterior likelihoods for (Ω_m, σ_8), computed using the ICA likelihood (*left panel*) and the Gaussian approximation (*right panel*). Shown are the contours of the 68%, 95% and 99% credible regions. The maximum of the ICA posterior is denoted by "x"; the maximum of the posterior based on the Gaussian likelihood coincides with the fiducial parameter set and is marked by the symbol "o".

where $\langle \cdot \rangle$ denotes the expectation value with respect to the likelihood function. If the likelihood is Gaussian and if the covariance matrix C does not depend on cosmology, one can show that

$$\mathsf{F}_{\alpha\beta} = \sum_{i,j} \mathsf{C}_{ij}^{-1} \frac{\partial m_i(\pi)}{\partial \pi_\alpha} \frac{\partial m_j(\pi)}{\partial \pi_\alpha} . \tag{21}$$

Equation (20) provides us with a way to estimate the Fisher matrix for the non-Gaussian likelihood. For each ray-tracing realisation of the CDFS, we compute the logarithm of the posterior distribution $\log p(\pi|\xi)$ and its derivatives with respect to the cosmological parameters at the fiducial parameter values. Since we use uniform priors for all cosmological parameters, the derivatives of the log-posterior are identical to those of the log-likelihood. We can then compute the Fisher matrix by averaging over all realisations:

$$\hat{\mathsf{F}}_{\alpha\beta} = \frac{1}{N} \sum_{k=1}^{N} \frac{\partial \log p(\pi|\xi)}{\partial \pi_\alpha} \frac{\partial \log p(\pi|\xi)}{\partial \pi_\beta} . \tag{22}$$

In Appendix B, we show that the expression for the Fisher matrix of the ICA likelihood can be evaluated further to be

$$\mathsf{F}_{\alpha\beta} = \sum_i \frac{\partial \breve{m}_i}{\partial \pi_\alpha} \frac{\partial \breve{m}_i}{\partial \pi_\beta} \int \mathrm{d}s_i \, p_{s_i}(s_i) \left(\frac{\partial \log p_{s_i}(s_i)}{\partial s_i} \right)^2 . \tag{23}$$

This equation allows a simpler, alternative computation of F from the estimated $p_{s_i}(s_i)$, as discussed in Appendix B.

We have used Eqs. (21) and (23) to compute the Fisher matrices for a 1500-deg^2 survey ($N_f = 6000$). We fit for four cosmological parameters (Ω_m, σ_8, h_{100}, Ω_Λ), keeping all other parameters fixed to their true values. To visualise the posterior, we compute two-dimensional marginalised posterior distributions for each parameter pair as well as the one-dimensional marginals for each parameter. The results are shown in Fig. 5. A general feature of the ICA likelihood, which has already been apparent in the 2D-analysis (Fig. 4), is that the credible intervals are significantly smaller than the ones derived from the Gaussian likelihood. For the two-dimensional marginal distributions, the area of the 68% credible regions derived from the ICA likelihood are smaller by $\approx 30{-}40\%$. The one-dimensional constraints are tighter by $\approx 10{-}25\%$. In addition we find that the ICA Fisher ellipses in some cases are slightly tilted with respect to those computed using the Gaussian likelihood. This is particularly apparent for parameter combinations involving the Hubble parameter. Note that the shift of the maximum observed in the two-dimensional case for a single CDFS-like field is absent here because it was assumed for the Fisher analysis that the posterior is centred on the true parameter values.

4. How odd is the Chandra Deep Field South?

4.1. The CDFS cosmic shear data

The second part of this work is based on the cosmological weak lensing analysis of the combined HST GEMS and GOODS data of the CDFS (Rix et al. 2004; Giavalisco et al. 2004), which was presented in S07. The mosaic comprises 78 ACS/WFC tiles imaged in F606W, covering a total area of $\sim 28' \times 28'$. We refer the reader to the original publication for details on the data and weak lensing analysis, which applies the KSB+ formalism (Kaiser et al. 1995; Luppino & Kaiser 1997; Hoekstra et al. 1998).

In S07, the cosmic shear analysis was performed using two different signal-to-noise and magnitude cuts. The first one selects galaxies with $S/N > 4$ and has no magnitude cut, and the second one applies a more conservative selection with $S/N > 5$ and $m_{606} < 27.0$, where S/N is the shear measurement signal-to-noise ratio as defined in Erben et al. (2001). The drizzling process in the data reduction introduces correlated noise in adjacent pixels. While these correlations are ignored in the computation of S/N, an approximate correction factor (see S07) is taken into account for S/N^{true}, making the above cuts $S/N^{\mathrm{true}} \gtrsim 1.9$ and $S/N^{\mathrm{true}} \gtrsim 2.4$ respectively. The two selection criteria yielded moderately different σ_8-estimates of $0.52^{+0.11}_{-0.15}$ and $0.59^{+0.11}_{-0.14}$ for $\Omega_m = 0.3$ (median of the posterior), not assuming a flat Universe. The errors include the statistical and redshift uncertainties. This translates to $\sigma_8 = 0.57^{+0.12}_{-0.16}$ and $0.65^{+0.12}_{-0.15}$ for our fiducial cosmology with $\Omega_m = 0.25$. The difference of the two estimates was considered as a measure for the robustness and hence systematic accuracy of our shear measurement pipeline. While the analysis of the "Shear TEsting Programme 2" (STEP2) image simulations (Massey et al. 2007a) indicated no significant average shear calibration bias for our method, a detected dependence on galaxy magnitude and size could effectively bias a cosmic shear analysis through the redshift dependence of the shear signal (see also Semboloni et al. 2009). In order to better understand the difference between the two estimates found in S07, and to exclude any remaining calibration uncertainty in the current analysis, we further investigate the shear recovery accuracy as a function of the signal-to-noise ratio using the STEP2 simulations in Appendix C. Here we conclude that our KSB+ implementation under-estimates gravitational shear for very noisy galaxies with $S/N^{\mathrm{true}} \lesssim 2.5$, which likely explains the lower signal found in S07 when all galaxies with $S/N > 4$ ($S/N^{\mathrm{true}} \gtrsim 1.9$) were considered. For the more conservative selection criteria we find no significant mean shear calibration bias and a variation as a
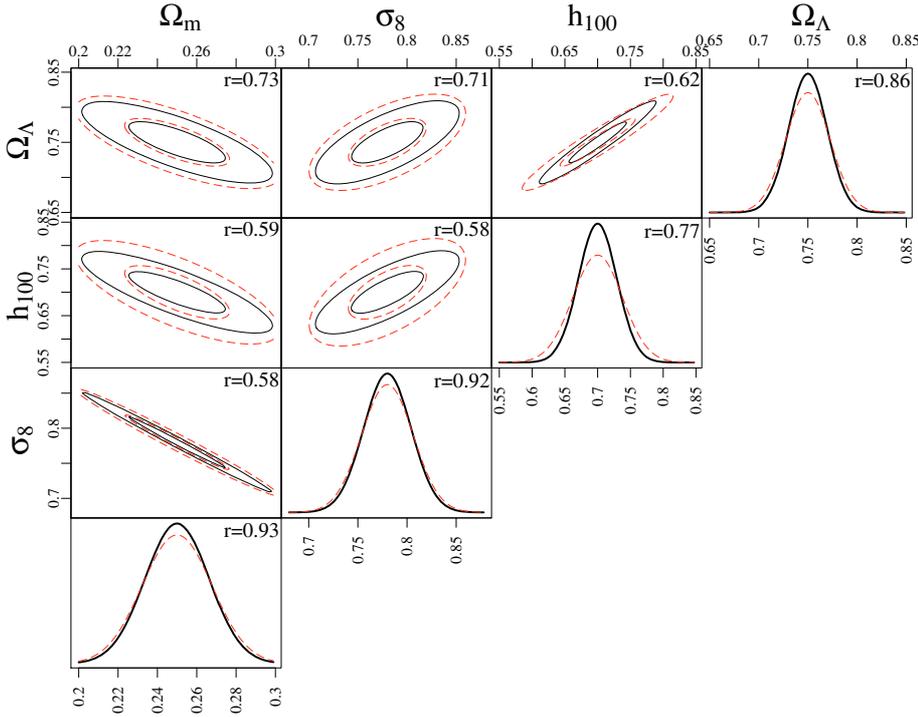
Fig. 5. Fisher matrix constraints for a hypothetical 1500-deg^2 survey, consisting of 6000 CDFS-like fields. The plots on the diagonal show the 1D marginals, the off-diagonal plots the 2D marginals derived from the full 4D posterior. The red dashed (black solid) lines/contours have been computed using the Fisher matrix of the Gaussian likelihood (the ICA likelihood).

function of magnitude and size of $\lesssim \pm 5\%$. Therefore we base our current analysis on the more robust galaxy sample with $S/N > 5$ ($S/N^{\text{true}} \gtrsim 2.4$) and $m_{606} < 27.0$, which yields a galaxy number density of 68 arcmin^{-2}. Based on the simulations, any remaining calibration uncertainty should be negligible compared to the statistical uncertainty.

Note that Heymans et al. (2005) found a higher estimate of $\sigma_8 (\Omega_{\text{m}}/0.3)^{0.65} = 0.68 \pm 0.13$ from GEMS, where they extrapolated the redshift distribution from the relatively shallow COMBO-17 photometric redshifts (Wolf et al. 2004). Using deeper data from the GOODS-MUSIC sample (Grazian et al. 2006), S07 were able to show that the COMBO-17 extrapolation significantly underestimates the mean redshift for GEMS, leading to the difference in the results for σ_8.

In Fig. 6, we show the posterior distribution for σ_8 based on this sample of galaxies. For the fit, all other cosmological parameters were held fixed at the fiducial values chosen for our ray-tracing simulations. This avoids complications in the discussion of cosmic variance and field selection biases due to the effect of parameter degeneracies. We choose a flat prior for σ_8, with a lower boundary of $\sigma_{8,\text{min}} = 0.35$ to cut off the tail of the posterior distribution towards small values of the power spectrum normalisation, which is caused by the fact that the difference (and therefore the likelihood) between the data and the model vectors changes only very little when σ_8 (and therefore the shear correlation function) is very small. We have performed the fit for the ICA likelihood as well as for the Gaussian approximation to the likelihood. For the latter, the covariance matrix was in one case estimated from the full sample of our ray-tracing simulations, and in the other case computed analytically assuming that the shear field is a Gaussian random field (Joachimi et al. 2008). The striking similarity of the posterior densities derived from the ICA likelihood and using the Gaussian covariance matrix for this particular data vector is merely a coincidence and is in general not seen for our set of simulated correlation functions.

For estimates of σ_8, we use the maximum of the posterior (henceforth we write ICA-MAP for the maximum of the non-Gaussian likelihood, and Gauss-MAP if the Gaussian

Table 1. Estimates of σ_8 from the CDFS.

	ICA likel.	Gaussian likel. (ray-tracing cov.)	Gaussian likel. (Gaussian cov.)
MAP	$0.68^{+0.09}_{-0.16}$	$0.59^{+0.10}_{-0.19}$	$0.68^{+0.10}_{-0.14}$
Median	$0.62^{+0.11}_{-0.11}$	$0.57^{+0.15}_{-0.15}$	$0.64^{+0.10}_{-0.14}$

approximation is used), although we also quote the median (ICA median) for comparison with S07. In the first case, our credible intervals are highest posterior density intervals, whereas for the median we choose to report the interval for which the probability of σ_8 of being below the lower interval boundary is as high as being above the upper boundary. The results are summarised in Table 1.

4.2. Cosmic variance

The original estimates for σ_8 given in S07 and those found in the previous section for the Gaussian likelihood are rather low compared to the value reported by WMAP5 (Dunkley et al. 2009). This problem appears less severe when the full non-Gaussian likelihood is used, but the σ_8-estimate is still rather low. It is therefore interesting to know whether this can be fully attributed to cosmic variance or whether the way in which the CDFS was originally selected biases our estimates low.

To begin, we determine the probability of finding a low σ_8 in a CDFS-like field if the pointing is completely random. We estimate the sampling distribution of the σ_8-MAP estimators for Gaussian and ICA likelihoods from the full sample of our ray-tracing simulations. We compute the posterior likelihood for σ_8 using a uniform prior in the range $\sigma_8 \in [0.35; 1.8]$ and determine the MAP estimator $\hat{\sigma}_8$. As in the previous sections, we do this using both the Gaussian and the ICA likelihoods. To separate possible biases of the estimators from biases that might arise because the model prediction based on Smith et al. (2003) does not quite fit our simulations, we correct the simulated correlation
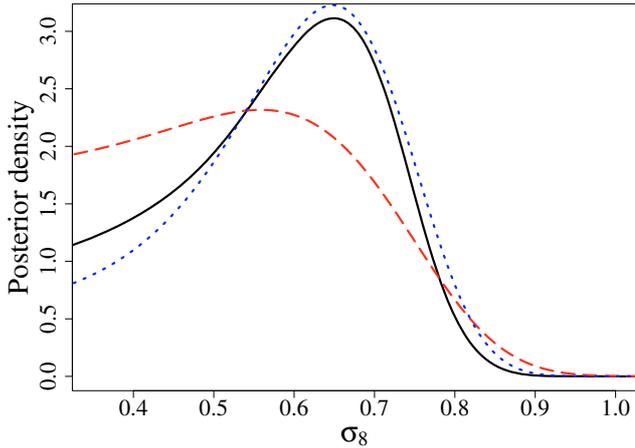
Fig. 6. Posterior distributions for σ_8 as computed from the CDFS data. The black solid line corresponds to the ICA likelihood, the red dashed line is from the Gaussian likelihood whose covariance matrix was estimated from the ray-tracing simulations. The blue dotted line was computed from the Gaussian likelihood with an analytically computed covariance matrix, assuming that the shear field is Gaussian. The similarity of the posterior densities derived from the ICA likelihood and using the Gaussian covariance matrix is purely coincidental, occurring only for this particular data vector.

functions for this: if $\boldsymbol{\xi}^{(i)}$ is the correlation function measured in the ith realisation, then

$$\boldsymbol{\xi}_{\mathrm{rc}}^{(i)} = \boldsymbol{\xi}^{(i)} - \langle\boldsymbol{\xi}\rangle + \boldsymbol{m}(\boldsymbol{\pi}_0), \tag{24}$$

is the "re-centred" shear correlation, where $\langle\boldsymbol{\xi}\rangle$ is the mean of all realisations and $\boldsymbol{m}(\boldsymbol{\pi}_0)$ is our fiducial model.

The resulting sampling distributions of $\hat\sigma_8$ are shown in Figs. 7 (original ξ) and 8 (re-centred ξ). All the distributions are well fit by a Gaussian. With the original correlation functions, we obtain estimates $\hat\sigma_8$ which are too high on average. This reflects the fact that the power spectrum fitting formula by Smith et al. (2003) underpredicts the small scale power in the simulations (see also Hilbert et al. 2009). If we correct for this, we see that the maximum of the ICA likelihood is a nearly un-biased estimator of σ_8 in the one-dimensional case considered here, and in addition has a lower variance than the maximum of the Gaussian likelihood.

We estimate the probability of obtaining a power spectrum normalisation as low as the one measured in the CDFS or lower, $\mathrm{Prob}(\hat\sigma_8 < \hat\sigma_8^{\mathrm{CDFS}})$, by the ratio of the number of realisations which fulfil this condition to the total number of simulations. These estimates agree very well with those computed from the best fitting Gaussian distribution. The results for the MAP and median estimators are summarised in Table 2. As expected from the above considerations, we find higher probabilities for the re-centred correlation functions. In this case, the ICA-MAP esti-mator yields 13% for the probability of obtaining an equally low or lower σ_8 than the CDFS. This reduces to $\approx 5\%$ when the uncorrected correlation functions are used, because the mis-fit of our theoretical correlation functions to the simulations bi-ases the σ_8-estimates high. If we assume that our simulations are a reasonable representation of the real Universe, we can ex-pect the same bias when we perform fits to real data. Therefore, $\mathrm{Prob}(\hat\sigma_8 < \hat\sigma_8^{\mathrm{CDFS}}) \approx 0.05$ as derived from the uncorrected cor-relation functions is most likely closest to reality. The proba-bilities computed from the Gauss-MAP estimates are generally smaller than the ICA-MAP values because of the lower value of
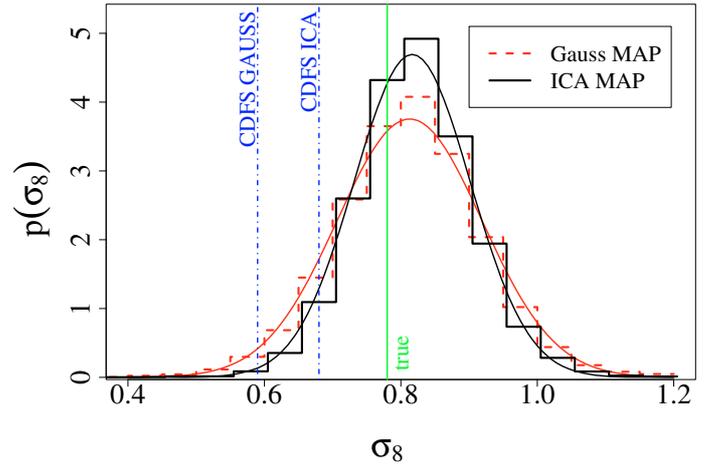

Fig. 7. Sampling distributions of the MAP estimators of σ_8, derived from 9600 realisations of the CDFS. All other parameters were held fixed at their fiducial values for the fit. The histogram with red dashed lines has been obtained from the Gaussian likelihood, the one with solid lines from the ICA likelihood. Also shown are the best fitting Gaussian distributions. We indicate the fiducial value of σ_8 and our estimates from the CDFS with vertical lines.
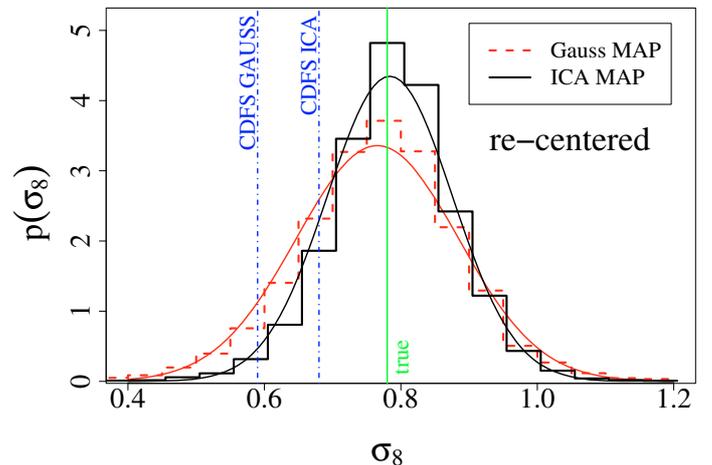

Fig. 8. Same as Fig. 7, but using re-centred correlation functions.

$\hat\sigma_8^{\mathrm{CDFS}}$ found using these estimators, even though the sampling distributions of the Gauss estimators are broader.

4.3. Influence of the CDFS selection criteria

We now investigate if and by how much the way in which the CDFS was selected can bias our estimates of the power spectrum normalisation low. Several local criteria had to be fulfilled by the future CDFS, such as a low galactic HI density, the absence of bright stars and observability from certain observatory sites. Since these conditions do not reach beyond our galaxy, we do not expect them to affect the lensing signal by the cosmological large-scale structure.

Furthermore, the field was chosen such that no extended X-ray sources from the ROSAT All-Sky Survey (RASS), in particular galaxy clusters, are in the field of view. This is po-tentially important, since it is known from halo-model calcula-tions that the cosmic shear power spectrum on intermediate and small scales is dominated by group- and cluster-sized haloes. Therefore, the exclusion of X-ray clusters might bias the selec-tion of a suitable line of sight towards under-dense fields. On the
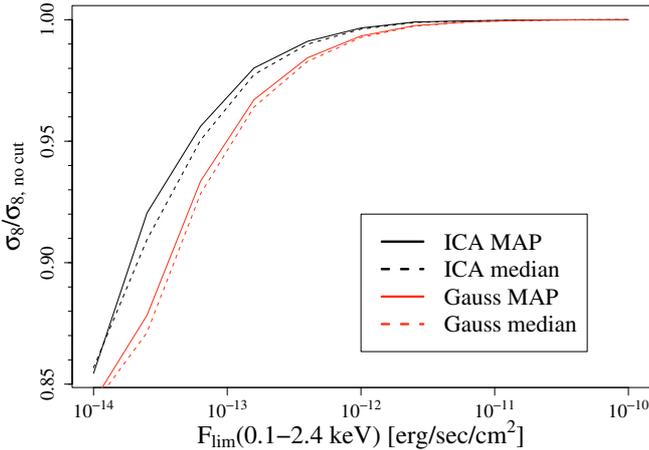
Fig. 9. The average values of the ICA-MAP (solid black line) and Gauss-MAP (solid red line) estimators computed from CDFS realisations that do not contain clusters with an X-ray flux larger than F_{lim}. For comparison, we also plot the averages of the corresponding median estimators (dashed lines).

Table 2. Prob($\sigma_8 < \hat{\sigma}_8^{\mathrm{CDFS}}$) for the CDFS.

	Gauss (MAP)	Gauss (median)	ICA (MAP)	ICA (median)
re-centred CF	6.8%	8.6%	12.9%	9.0%
original CF	1.8%	3.0%	5.4%	3.4%

other hand, the RASS is quite shallow and thus only contains very luminous or nearby clusters, which have a limited impact on the lensing signal due to their low number or low lensing efficiency. We quantify the importance of this criterion using the halo catalogues of our N-body-simulations. To each halo, we assign an X-ray luminosity in the energy range from 0.1 to 2.4 keV using the mass-luminosity relation given in Reiprich & Böhringer (2002) and convert this into X-ray flux using the halo redshift. We then compute the average of the σ_8 estimates from all fields which do not contain a cluster brighter than a certain flux limit. It is difficult to define an exact overall flux limit to describe the CDFS selection, because the RASS is rather heterogeneous. However, it is apparent from Fig. 9 that even a very conservative limit of 10^{-13} erg/s/cm^2 will change the average σ_8 estimate by at most 3−5%. This bias is therefore most likely not large enough to explain our CDFS result alone.

Finally, the CDFS candidate should not contain any "relevant NED source". This is very hard to translate into a quantitative criterion, in particular because our simulations contain only dark matter. We model the effect of imposing this requirement by demanding that there be less than n_{halo} group- or cluster-sized haloes ($M > 10^{13} \, M_\odot/h$) in the redshift range from $z = 0$ and $z = 0.5$ in a CDFS candidate. The impact of this criterion on the estimated value of $\hat{\sigma}_8$ using the ICA- and Gauss-MAP estimators is shown in Fig. 10. As expected, the median $\hat{\sigma}_8$ is a monotonically increasing function of n_{halo}. For fields with less than ≈ 12 massive haloes, the probability of obtaining a power spectrum normalisation as low as in the CDFS rises above $\approx 20\%$. Given that the average number of massive haloes in the specified redshift range is 18.5, it does not seem to be too unreasonable that fields with less than ≈ 12 such haloes could be obtained by selecting "empty" regions in the NED. This is also in qualitative agreement with Phleps et al. (2007), who find that the CDFS
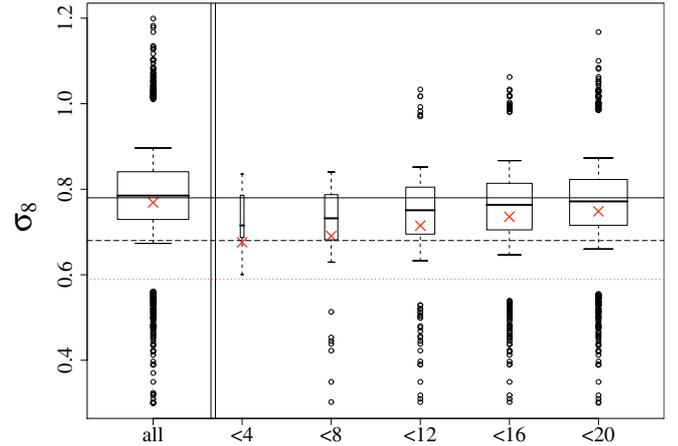
Fig. 10. Dependence of the ICA-MAP-estimator for σ_8 on the number of group- and cluster-sized haloes n_{halo} between $z = 0$ and $z = 0.5$. For each n_{halo}-bin, we summarise the distribution of the corresponding subsample of simulated CDFS-fields by giving a box plot: the thick horizontal line in each box denotes the median, the upper and lower box boundaries give the upper and lower quartiles of the distribution of the sample values. The error bars ("whiskers") extend to the 10% and 90% quantiles, respectively. To visualise the tails of the distributions, the most extreme values are given as points. The width of each box is proportional to the square root of the sample size. For comparison, we also show for each subsample the median of the Gauss MAP estimators as red crosses. The solid black horizontal line indicates the true value of σ_8, the black dashed line the ICA-MAP estimate for the CDFS and the red dotted line the Gauss-MAP estimate. The average number of haloes with $M > 10^{13} \, M_\odot$ and $z \leq 0.5$ in a CDFS-like field is $\bar{n}_{\mathrm{halo}} = 18.5$.

is underdense by a factor of ≈ 2 in the redshift range from $z \approx 0.2$ to $z \approx 0.4$.

We estimate the impact of this selection criterion on the estimates of cosmological parameters by treating the number of haloes in the CDFS as a nuisance parameter in the process of parameter estimation. Similar to what we did to obtain Fig. 10, we bin the realisations of the CDFS according to the number of group-sized haloes in the realisations. For each bin, we obtain the mean shear correlation function and its ratio to the mean shear correlation function of all realisations, $r_\pm(\theta, n_{\mathrm{halo}}) = \xi_\pm(\theta, n_{\mathrm{halo}})/\xi_\pm(\theta)$. The functions r_+ and r_- are shown in Fig. 11. The realisations with fewer (more) haloes than the average generally display a smaller (larger) shear correlation function. We fit the ratios in each bin with a double power law of the form

$$r_\pm(\theta, n_{\mathrm{halo}}) = A_\pm(n_{\mathrm{halo}})\theta^{\alpha_\pm(n_{\mathrm{halo}})} + B_\pm(n_{\mathrm{halo}})\theta^{\beta_\pm(n_{\mathrm{halo}})}. \quad (25)$$

For values of n_{halo} which do not coincide with one of the bin centres, the functions $r_\pm$ are obtained by linear interpolation between the fits for the two adjacent bins. With this, we extend our model for the shear correlation function to $m'_\pm(\theta; \boldsymbol{\pi}, n_{\mathrm{halo}}) = m_\pm(\theta; \boldsymbol{\pi}) \, r_\pm(\theta, n_{\mathrm{halo}})$. In Fig. 12, we show the resulting posterior distributions for $\sigma_8(\Omega_{\mathrm{m}} = 0.25)$ and n_{halo}, keeping all other cosmological parameters fixed and using a uniform prior for n_{halo}. The two-dimensional distribution shows a weak correlation between the two parameters: as expected, a low (high) value of n_{halo} requires a slightly higher (lower) value of σ_8. The marginalised posterior for σ_8 is very similar to the one shown in Fig. 7, where the field selection is not taken into account. However, including n_{halo} increases the MAP estimate of σ_8 by 5% to $\hat{\sigma}_8 = 0.71^{+0.10}_{-0.15}$ for the ICA likelihood and
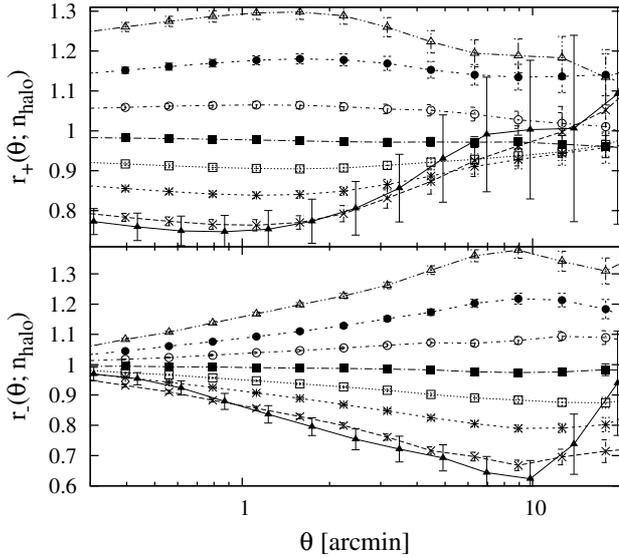
Fig. 11. Ratios r_+ (*upper panel*) and r_- (*lower panel*) of the shear correlation functions in a particular n_{halo}-bin to the average correlation function of all realisations. The lowest (solid) curve represents the bin with $n_{halo} \in [0, 4)$, the second lowest the bin with $n_{halo} \in [4, 8)$, and so on. The highest ratio corresponds to the bin with $n_{halo} \geq 28$. The error bars have been estimated from the field-to-field variation.

Fig. 12. *Upper panel*: posterior density for $\sigma_8(\Omega_m = 0.25)$ and n_{halo} computed using the ICA likelihood, keeping all other cosmological parameters constant. *Lower panels*: marginalised posterior densities of $\sigma_8(\Omega_m = 0.25)$ (*left panel*) and n_{halo} (*right panel*). Solid black curves show the results from using the ICA likelihood, dashed red lines from the Gaussian likelihood and the ray-tracing covariance.

by 10% to $\hat{\sigma}_8 = 0.65^{+0.13}_{-0.20}$ for the Gaussian likelihood and the ray-tracing covariance matrix. The marginalised posterior distribution of n_{halo} shows a weak peak at $\hat{n}_{halo} \approx 13$ (compared to the average of $\bar{n}_{halo} = 18.5$ for all ray-tracing realisations) in the ICA case and even lower values if the Gaussian likelihood is used. Overall, however, the posterior is very shallow.

Having corrected for the field selection, we can now recompute the probabilities given in Table 2 for drawing the CDFS at random. We find for the ICA-MAP estimate Prob($\sigma_8 < 0.71$) = 9.4% for the original shear correlation functions and Prob($\sigma_8 < 0.71$) = 18.5% for the re-centred ones. For the Gaussian likelihood, we find Prob($\sigma_8 < 0.65$) = 6.0% (original) and Prob($\sigma_8 < 0.65$) = 14.9% (re-centred), respectively.

With this (approximate) treatment of the systematic effects caused by the field selection, we can now put the CDFS in context with the results from the WMAP five-year data. For this, we fit the shear correlation function for Ω_m and σ_8, marginalising over h_{100} (with a Gaussian prior centred on $h_{100} = 0.7$ and $\sigma_{h_{100}} = 0.07$, as suggested by the Hubble Key Project; Freedman et al. 2001) and n_{halo} with a uniform prior. We use the WMAP Markov chain for a flat ΛCDM model (lcdm+sz+lens; Dunkley et al. 2009; Komatsu et al. 2009), where again we marginalise over all parameters except Ω_m and σ_8. The resulting posterior distributions for the CDFS only (blue dashed contours), WMAP only (red contours) and the combination of both measurements (thick black contours) are shown in Fig. 13. Clearly, the joint posterior is dominated by the WMAP data; however, the constraints from the CDFS allow us to exclude parameter combinations where both Ω_m and σ_8 are large. We find the MAP estimates $\hat{\Omega}_m = 0.26^{+0.03}_{-0.02}$ and $\hat{\sigma}_8 = 0.79^{+0.04}_{-0.03}$ when marginalising over the other parameter.

Finally, note that the two criteria discussed in this section are not strictly independent. However, it is highly improbable that a single field will contain more than one massive halo above the X-ray flux limit. Therefore, selecting fields without an X-ray-bright cluster prior to performing the steps that lead to Fig. 10 would change the halo numbers that go into the analysis by at
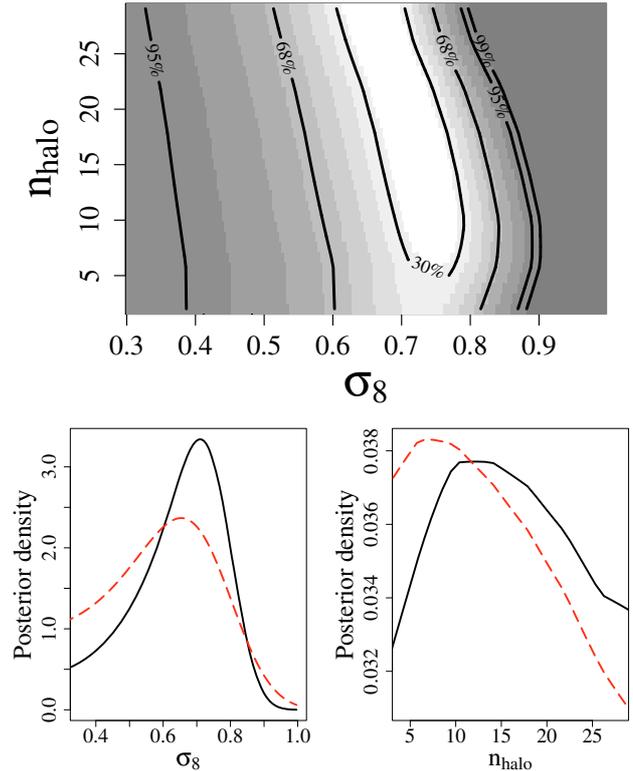
most one and would not significantly influence the foregoing discussion.

5. Summary and discussion

In this paper, we have investigated the validity of the approximation of a Gaussian likelihood for the cosmic shear correlation function, which is routinely made in weak lensing studies. We have described a method to estimate the likelihood from a large set of ray-tracing simulations. The algorithm tries to find a new set of (non-orthogonal) basis vectors with respect to which the components of the shear correlation functions become approximately statistically independent. This then allows us to estimate the high-dimensional likelihood as a product of one-dimensional probability distributions. A drawback of this method is that quite a large sample of realistically simulated correlation functions is required to get good results for the tails of the likelihood. However, this should become less problematic in the near future when increasingly large ray-tracing simulations will become available.

We have investigated how the constraints on matter and vacuum energy density, Hubble parameter and power spectrum normalisation depend on the shape of the likelihood for a survey composed of 0.5 deg × 0.5 deg fields and a redshift distribution similar to the CDFS. We find that if the non-Gaussianity of the likelihood is taken into account, the posterior likelihood becomes more sharply peaked and skewed. When fitting only for Ω_m and σ_8, the maximum of the posterior is shifted towards lower Ω_m and higher σ_8, and the area of the 68% highest
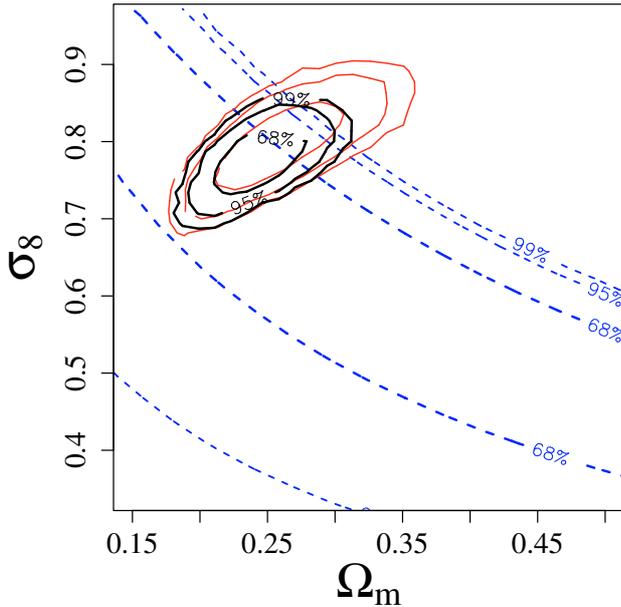
Fig. 13. Posterior density for Ω_m and σ_8, where we have marginalised over the Hubble constant h_{100} and the number of haloes in the field n_{halo}. The dashed blue contours show the 68%, 95% and 99% credible regions resulting from the cosmic shear analysis of the CDFS (using the ICA likelihood), the red contours show the posterior from the WMAP 5-year data (using the flat ΛCDM model). The combined posterior is shown with thick black contours.

posterior density credible region decreases by about 40% compared to the case of a Gaussian likelihood. For the four-dimensional parameter space, we have conducted a Fisher matrix analysis to obtain lower limits on the errors achievable with a $1500\,\mathrm{deg}^2$ survey. As in the two-dimensional case, we find the most important effect to be that the error bars decrease by 10–40% compared to the Gaussian likelihood. Less severe is the slight tilt of the Fisher ellipses when marginalising over two of the four parameters, particularly when h_{100} is involved.

In the second part of this work, we have presented a re-analysis of the CDFS-HST data. Using the non-Gaussian likelihood, we find $\sigma_8 = 0.68^{+0.09}_{-0.16}$ for $\Omega_m = 0.25$ (keeping all other parameters fixed to their fiducial values), compared to $\sigma_8 = 0.59^{+0.10}_{-0.19}$ obtained from the Gaussian likelihood with a co-variance matrix estimated from the ray-tracing simulations. We have then tried to quantify how (un-)likely it is to randomly select a field with the characteristics of the *Chandra* Deep Field South with a power spectrum normalisation this low. We have used 9600 ray-tracing realisations of the CDFS to estimate the sampling distribution of the ICA-MAP estimator for σ_8. For our fiducial, WMAP5-like cosmology, we find that $\mathrm{Prob}(\sigma_8 \leq 0.68) \approx 5\%$, assuming that the location of the CDFS on the sky was chosen randomly. The fact that the CDFS was selected not to contain an extended X-ray source in the ROSAT All-Sky Survey can lead to a bias of the estimated σ_8 by at most 5%. This is because the clusters excluded by this criterion are rare and mostly at low redshifts, and therefore not very lensing-efficient. The second relevant selection criterion is that the CDFS should not contain any relevant NED source. We model this by selecting only those fields which contain a specific number n_{halo} of group- and cluster-sized haloes. We find that for those realisations for which the number of such haloes is below the average, the estimates of σ_8 can be biased low by about 5–10%. We include this effect in our likelihood analysis by extending our model shear correlation function by a correction factor depending on n_{halo}

and treating n_{halo} as a nuisance parameter. This increases the estimate of σ_8 by 5% to $\hat{\sigma}_8 = 0.71^{+0.10}_{-0.15}$ for the ICA likelihood and by 10% to $\hat{\sigma}_8 = 0.65^{+0.13}_{-0.20}$ for the Gaussian likelihood. This procedure also yields tentative evidence that the number of massive haloes in the CDFS is only $\approx 70\%$ of the average, in qualitative agreement with the findings of Phleps et al. (2007).

Finally, we combine the CDFS cosmic shear results with the constraints on cosmological parameters from the WMAP experiment. We fit for Ω_m and σ_8, where we marginalise over the Hubble constant and take into account the field selection bias by marginalising also over n_{halo}. While the posterior is clearly dominated by the WMAP data, the CDFS still allows us to exclude parts of the parameter space with high values of both Ω_m and σ_8. Assuming a flat Universe, the MAP estimates for these two parameters are $\hat{\Omega}_m = 0.26^{+0.03}_{-0.02}$ and $\hat{\sigma}_8 = 0.79^{+0.04}_{-0.03}$.

Appendix A: Projection pursuit density estimation

In order to have an independent check of the ICA-based likeli-hood estimation algorithm, we employ the method of *projection pursuit density estimation* (PPDE; Friedman et al. 1984). Like our ICA method, PPDE aims to estimate the joint probability density $p(\boldsymbol{x})$ of a random vector $\boldsymbol{x}$, given a set of observations of $\boldsymbol{x}$. As starting point, an initial model $p_0(\boldsymbol{x})$ for the multidi-mensional probability distribution $p(\boldsymbol{x})$ has to be provided, for which a reasonable choice is e.g. a multivariate Gaussian with a covariance matrix estimated from the data. The method then identifies the direction $\boldsymbol{\theta}_1$ along which the marginalised model distribution differs most from the marginalised density of the data points and corrects for the discrepancy along the direction $\boldsymbol{\theta}_1$ by multiplying p_0 with a correction factor. This yields a re-fined density estimate $p_1(\boldsymbol{x})$, which can be further improved by iteratively applying the outlined procedure.

More formally, the PPDE density estimate is of the form

$$p_M(\boldsymbol{x}) = p_0(\boldsymbol{x}) \prod_{m=1}^{M} f_m(\boldsymbol{\theta}_m \cdot \boldsymbol{x}), \qquad (A.1)$$

where p_M is the estimate after M iterations of the procedure and p_0 is the initial model. The univariate functions f_m are multiplicative corrections to the initial model along the directions $\boldsymbol{\theta}_m$. The density estimate can be obtained iteratively us-ing the relation $p_M(\boldsymbol{x}) = p_{M-1}(\boldsymbol{x}) f_M(\boldsymbol{\theta}_M \cdot \boldsymbol{x})$. At the Mth step of the iteration, a direction $\boldsymbol{\theta}_M$ and a function f_M are chosen to minimise the Kullback-Leibler divergence (Kullback & Leibler 1951) between the actual data density $p(\boldsymbol{x})$ and the density esti-mate $p_M(\boldsymbol{x})$,

$$D_{\mathrm{KL}}[p, p_M] = \int \mathrm{d}\boldsymbol{x}\, p(\boldsymbol{x}) \log \frac{p(\boldsymbol{x})}{p_M(\boldsymbol{x})}, \qquad (A.2)$$

as a goodness-of-fit measure. The Kullback-Leibler diver-gence provides a "distance measure" between two probability

distribution functions, since it is non-negative and zero only if $p \equiv q$, albeit not symmetric. Only the cross term

$$W(\boldsymbol{\theta}_M, f_M) = - \int \mathrm{d}\boldsymbol{x} \, p(\boldsymbol{x}) \log p_M(\boldsymbol{x}) \tag{A.3}$$

of the K-L divergence is relevant for the minimisation, all other terms do not depend on $\boldsymbol{\theta}_M$ and f_M. By using Eq. (A.1), one sees that the minimum of W is attained at the same location as the minimum of

$$w(\boldsymbol{\theta}_M, f_M) = - \int \mathrm{d}\boldsymbol{x} \, p(\boldsymbol{x}) \log f_M(\boldsymbol{\theta}_M \cdot \boldsymbol{x}), \tag{A.4}$$

which is the expectation value of $\log f_M$ with respect to $p(\boldsymbol{x})$. The data density $p(\boldsymbol{x})$ is unknown; however, the data comprise a set of N samples from this distribution. The expectation value of $\log f_M$ can therefore be estimated by

$$\hat{w}(\boldsymbol{\theta}_M, f_M) = -\frac{1}{N} \sum_{i=1}^{N} \log f_M(\boldsymbol{\theta}_M \cdot \boldsymbol{x}_i). \tag{A.5}$$

For fixed $\boldsymbol{\theta}_M$, the minimum of Eq. (A.4) is attained for

$$f_M(\boldsymbol{\theta}_M \cdot \boldsymbol{x}) = \frac{p^{\boldsymbol{\theta}_M}(\boldsymbol{\theta}_M \cdot \boldsymbol{x})}{p_{M-1}^{\boldsymbol{\theta}_M}(\boldsymbol{\theta}_M \cdot \boldsymbol{x})}, \tag{A.6}$$

where $p^{\boldsymbol{\theta}_M}$ and $p_{M-1}^{\boldsymbol{\theta}_M}$ are the marginal densities of the data and of model density from the $(M-1)$st iteration along the direction $\boldsymbol{\theta}_M$, respectively. With this, the iterative process that leads to estimates of $\boldsymbol{\theta}_M$ and f_M schematically consists of:

- choosing a direction $\boldsymbol{\theta}_M$;
- computing the marginal densities $p^{\boldsymbol{\theta}_M}$ and $p_{M-1}^{\boldsymbol{\theta}_M}$;
- computing $f_M(\boldsymbol{\theta}_M \cdot \boldsymbol{x})$ according to Eq. (A.6);
- computing $\hat{w}(\boldsymbol{\theta}_M, f_M)$;
- choosing a new $\boldsymbol{\theta}_M$ that decreases $\hat{w}$;
- continuing from step 2 until a convergence criterion is fulfilled.

To efficiently compute the marginals $p^{\boldsymbol{\theta}_M}$ and $p_{M-1}^{\boldsymbol{\theta}_M}$, Monte Carlo samples of these densities are used. Note that the data already comprise a sample of $p(\boldsymbol{x})$; a sample of $p_{M-1}^{\boldsymbol{\theta}_M}$ can be obtained efficiently by an iterative method: since p_{M-1} is similar to p_{M-2}, a subset of the sample from p_{M-1} can be obtained by rejection sampling from the sample from the $(M-2)$-nd step. The remaining data vectors are then drawn by rejection sampling from p_0. For more technical details of the estimation procedure, we refer the reader to Friedman et al. (1984).

Note that the PPDE technique, although using very similar methodology as our ICA-based procedure, is different in the important point that it does not rely on the assumption that a linear transformation of the data leads to statistical independence of the components of the transformed data vectors. It therefore comprises a good test of the validity of this approximation.

Appendix B: Fisher matrix of the ICA likelihood

In this appendix, we give the derivation of Eq. (23). In the general case, the Fisher matrix is given by (e.g. Kendall et al. 1987)

$$\mathsf{F}_{\alpha\beta} = \left\langle \frac{\partial \log L}{\partial \pi_\alpha} \frac{\partial \log L}{\partial \pi_\beta} \right\rangle. \tag{B.1}$$

In our case, the likelihood depends on cosmological parameters only through the difference between data and model vector, i.e. $\boldsymbol{s} = \boldsymbol{\breve{\xi}} - \boldsymbol{\breve{m}}$ (see Eq. (8)). This allows us to write

$$\frac{\partial \log L(\boldsymbol{s}(\boldsymbol{\pi}))}{\partial \pi_\alpha} = \frac{\partial \log L(\boldsymbol{s})}{\partial s_i} \frac{\partial s_i}{\partial \pi_\alpha} \tag{B.2}$$

$$= \frac{\mathrm{d} \log p_{s_i}(s_i)}{\mathrm{d} s_i} \frac{\partial s_i}{\partial \pi_\alpha}, \tag{B.3}$$

where in the last step we have made use of the fact that the likelihood factorises in the ICA basis. The expression for the Fisher matrix then can be written as

$$\mathsf{F}_{\alpha\beta} = \sum_{i,j} \left\langle \frac{\mathrm{d} \log p_{s_i}(s_i)}{\mathrm{d} s_i} \frac{\mathrm{d} \log p_{s_j}(s_j)}{\mathrm{d} s_j} \right\rangle \frac{\partial \breve{m}_i}{\partial \pi_\alpha} \frac{\partial \breve{m}_j}{\partial \pi_\beta}. \tag{B.4}$$

Next, we compute the expectation value on the right hand side and obtain

$$\mathsf{F}_{\alpha\beta} = \sum_{i \neq j} \frac{\partial \breve{m}_i}{\partial \pi_\alpha} \frac{\partial \breve{m}_j}{\partial \pi_\beta} \int \mathrm{d}s_i \frac{\mathrm{d} p_{s_i}(s_i)}{\mathrm{d} s_i} \int \mathrm{d}s_j \frac{\mathrm{d} p_{s_j}(s_j)}{\mathrm{d} s_j} \tag{B.5}$$

$$+ \sum_i \frac{\partial \breve{m}_i}{\partial \pi_\alpha} \frac{\partial \breve{m}_i}{\partial \pi_\beta} \int \mathrm{d}s_i \, p_{s_i}(s_i) \left(\frac{\mathrm{d} \log p_{s_i}(s_i)}{\mathrm{d} s_i} \right)^2. \tag{B.6}$$

The integrals in the first term of the right hand side vanish since the p_{s_i} drop to zero for very large and small values of s_i. This leaves us with

$$\mathsf{F}_{\alpha\beta} = \sum_i \frac{\partial \breve{m}_i}{\partial \pi_\alpha} \frac{\partial \breve{m}_i}{\partial \pi_\beta} \int \mathrm{d}s_i \, p_{s_i}(s_i) \left(\frac{\mathrm{d} \log p_{s_i}(s_i)}{\mathrm{d} s_i} \right)^2. \tag{B.7}$$

The derivatives in Eq. (B.7) can be strongly affected by noise in the estimated $p_{s_i}(s_i)$, in particular in the tails of the distributions. For their numerical computation, we therefore choose the following four-point finite difference operator (Abramowitz & Stegun 1964):

$$\frac{\mathrm{d}p}{\mathrm{d}s} =$$
$$\frac{p(s-2h) - 8p(s-h) + 8p(s+h) - p(s+2h)}{12h} + \mathrm{O}(h^5), \tag{B.8}$$

which we find to be more stable against this problem than its more commonly used two-point counterpart. Because of this potential difficulty, we cross-check our results with the alternative method provided by Eq. (22). This method is significantly slower, but numerically simpler. This is because the derivatives of the log-likelihood in Eq. (22) are on average computed close to the maximum-likelihood point, where the likelihood estimate is well sampled. Reassuringly, we find excellent agreement between the two methods. Finally, we have investigated the influence of the choice of the Kernel function K in Eq. (10), which might affect the computation of the numerical derivatives. Our results prove to be stable against variation of K, provided that we chose a differentiable Kernel function.

Appendix C: Further conclusions for our KSB+ pipeline from the STEP simulations

In this appendix we assume that the reader is familiar with basic KSB notation. For a short introduction and a summary of differences between various implementations see Heymans et al. (2006).

Within the Shear TEsting Programme[2] (STEP) simulated images containing sheared galaxies are analysed in blind tests, in order to test the shear measurement accuracy of weak lensing pipelines. In these analyses the shear recovery accuracy has been quantified in terms of a multiplicative calibration bias m and additive PSF residuals c. From the analysis of the first set of simulations (STEP1, Heymans et al. 2006), which mimic simplified ground-based observations, we find that our KSB+ implementation significantly under-estimates gravitational shear on average if no calibration correction is applied. After the elimination of selection and weighting-related effects this shear calibration bias amounts to a relative under-estimation of $m = -9\%$. According to our testing the largest contribution to this bias originates from the inversion of the P^g-tensor, which describes the response of galaxy ellipticity to gravitational shear. While a full-tensor inversion reduces this bias, it strongly increases the measurement noise (see also Erben et al. 2001) and dependence on galaxy selection criteria. We therefore decided to stick to the commonly applied approximation of $(P^g)^{-1} = 2/\mathrm{Tr}[P^g]$, which we measure from individual galaxies, and correct the shear estimate using a multiplicative calibration factor of $c_{cal} = 1.10$ in the S07 analysis. This average calibration correction was found to be stable to the ~2%-level between different STEP1 simulation subsets. However, note that the bias depends on the details of the KSB implementation, which might explain some of the scatter between the results for different KSB codes in STEP1. In particular, we identified a strong dependence on the choice of the Gaussian filter scale r_g, which is used in the computation of the KSB brightness moments. For example changing from our default $r_g = 1.0\,r_f$, where r_f is the flux radius as measured by SExtractor (Bertin & Arnouts 1996), to $r_g = 0.7\,r_f$, worsens the bias to $m = -17\%$.

The average calibration correction likewise proved to be robust for the second set of image simulations (STEP2, Massey et al. 2007a), which also mimics ground-based data but takes into account more complex PSFs and galaxy morphologies by applying the shapelets formalism (Massey et al. 2004). Yet, the STEP2 analysis revealed a significant magnitude dependence of the shear recovery accuracy for our implementation, with a strong deterioration at faint magnitudes. In this analysis we applied the same signal-to-noise cut $S/N > 4.0$ as in STEP1 (KSB S/N as defined in Erben et al. 2001), where we however ignored the strong noise correlations present in the STEP2 data, which was added to mimic the influence of drizzle.

In the case of uncorrelated noise the dispersion of the sum over the pixel values of N pixels scales as

$$\sigma_N = \sqrt{N}\sigma_1, \tag{C.1}$$

where σ_1 is the dispersion computed from single pixel values. Drizzling, or convolution in the case of the STEP2 simulations, reduces σ_1 but introduces correlations between neighbouring pixels. The signal-to-noise of an object is usually defined as the ratio of the summed object flux convolved with some window or weight function, divided by an rms estimate for the noise in an equal area convolved with the same weight function. If the noise estimate is computed from σ_1 and scaled according to Eq. (C.1), the correlations are neglected and the noise estimate is too small.

In order to estimate the effective influence of the noise correlations in STEP2, we use a pure noise image which was provided together with the simulated images. We compute the rms of the
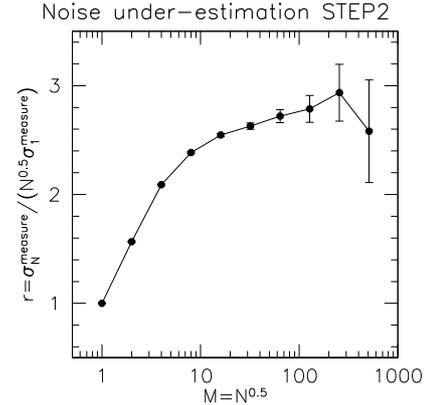
Fig. C.1. Estimate of the effective influence of the noise correlations in the STEP2 simulations: plotted is the ratio of the pixel value dispersion $\sigma_N^{measure}$ measured from large areas of $N = M^2$ pixels to the estimate from the normal single pixel dispersion $\sqrt{N}\sigma_1^{measure}$ as a function of M, determined from an object-free STEP2 image. In the absence of noise correlations $r = 1$ for all M. The value $r \simeq 2.8$ for $M \to \infty$ gives the factor by which the signal-to-noise is over-estimated when measured from the single pixel dispersion $\sigma_1^{measure}$ ignoring the correlations.

pixel sum $\sigma_N^{measure}$ in independent quadratic sub-regions of the image with side length $M = \sqrt{N}$ and determine the ratio

$$r = \frac{\sigma_N^{measure}}{\sqrt{N}\sigma_1^{measure}}, \tag{C.2}$$

which in the absence of correlated noise would be equal to 1 for all N. In the presence of noise correlations it will for large N converge to the factor by which $\sigma_1^{measure}$ under-estimates the uncorrelated σ_1. This can be understood as drizzling or convolution typically re-distributes pixel flux within a relatively small area. As soon as this kernel is much smaller than the area spanned by M^2 pixels, the correlations become unimportant for the area pixel sum. The measured $r(M)$ is plotted in Fig. C.1. Extrapolating to $M \to \infty$ we estimate that ordinary noise measures based on the single pixel dispersion, which ignore the noise correlation, will over-estimate the signal-to-noise of objects by a factor $r \simeq 2.8$. Hence, our original selection criterion $S/N > 4.0$ corresponds to a very low true cut $S/N^{true} \gtrsim 1.4$ including much noisier objects than in STEP1.

We plot the dependence of our STEP2 shear estimate on the (uncorrected) S/N in Fig. C.2. For $S/N \lesssim 7$, corresponding to $S/N^{true} \lesssim 2.5$, a significant deterioration of the shear signal occurs, with a mean calibration bias $\langle m \rangle \sim -10\%$ and a large scatter between the different PSF models. We conclude that this approximately marks the limit down to which our KSB+ implementation can reliably measure shear. If we apply a modified cut $S/N > 7.0$ to the STEP2 galaxies, the resulting magnitude and size dependence of the shear calibration bias is $\lesssim \pm 5\%$ (top panels in Fig. C.3). The remaining galaxies are best corrected with a slightly reduced calibration factor $c_{cal} = 1.08$, which we apply for the plots shown in this appendix and the updated shear analysis presented in this paper. The difference between the calibration corrections derived from STEP1 and STEP2 agrees with the estimated ~2% accuracy. Note that the error increases for the highly elliptical PSFs D and E ($e^* \simeq 12\%$) in STEP2, for which in addition significant PSF anisotropy residuals occur (bottom panels in Fig. C.3). This should however not affect our analysis given that typical ACS PSF ellipticities rarely exceed $e^* \simeq 7\%$, see e.g. S07.
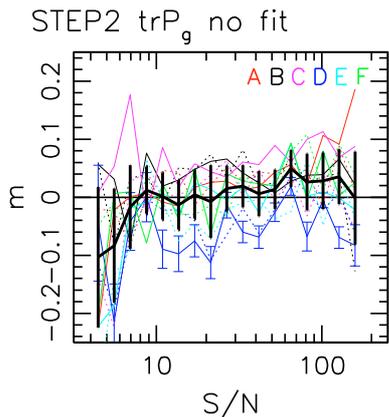
Fig. C.2. Calibration bias m as a function of the uncorrected KSB signal-to-noise S/N for the TS analysis of the STEP2 simulations. Thin solid (dashed) lines show γ_1 (γ_2) estimates for individual PSFs, where we show individual error-bars only for one PSF for clarity. The bold solid line and error-bars show the mean and standard deviation of the individual PSF estimates and shear components. Note the deterioration of the shear estimate for the STEP2 galaxies with $S/N \lesssim 7$ ($S/N^{\text{true}} \lesssim 2.5$). For this plot an adapted calibration correction of 1.08 was applied.
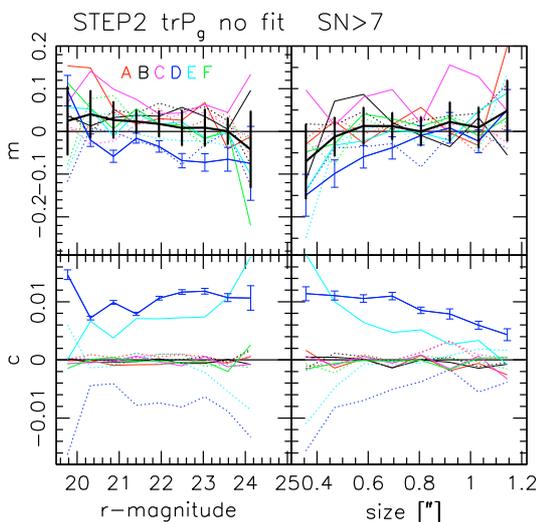

Fig. C.3. Calibration bias m and PSF residuals c as a function of input galaxy magnitude and size for our refined analysis of the STEP2 data. Thin solid (dashed) lines show γ_1 (γ_2) estimates for individual PSFs, where we include individual error-bars only for one PSF for clarity. Bold solid lines and error-bars show the mean and standard deviation of the individual PSF estimates and shear components. In this plot only galaxies with $S/N > 7$ ($S/N^{\text{true}} > 2.5$) are taken into account, which strongly reduces the deterioration found in Massey et al. (2007a) for faint magnitudes. For this plot an adapted calibration correction of 1.08 was applied.

References

Abramowitz, M., & Stegun, I. A. 1964, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, ninth edn. (New York: Dover)
Benjamin, J., Heymans, C., Semboloni, E., et al. 2007, MNRAS, 381, 702
Bernstein, G. M., & Jarvis, M. 2002, AJ, 123, 583
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Chiu, K.-C., Liu, Z.-Y., & Xu, L. 2003, in Proc. 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), Nara, Japan, 751
Comon, P., Jutten, C., & Hérault, J. 1991, Signal Processing, 24, 11
Cooray, A., & Hu, W. 2001, ApJ, 554, 56

Davison, A. C. 2003, Statistical Models, Cambridge Series in Statistical and Probabilistic Mathematics (Cambridge University Press)
Dunkley, J., Komatsu, E., Nolta, M. R., et al. 2009, ApJS, 180, 306
Eifler, T., Schneider, P., & Hartlap, J. 2009, A&A, 502, 721
Erben, T., van Waerbeke, L., Bertin, E., Mellier, Y., & Schneider, P. 2001, A&A, 366, 717
Freedman, W. L., Madore, B. F., Gibson, B. K., et al. 2001, ApJ, 553, 47
Friedman, J., Stuetzle, W., & Schroeder, A. 1984, J. Am. Stat. Assoc., 79, 599
Fu, L., Semboloni, E., Hoekstra, H., et al. 2008, A&A, 479, 9
Gelman, A., Carlin, J. B., Stern, H., & Rubin, D. B. 2004, Bayesian Data Analysis (Chapman & Hall/CRC)
Giacconi, R., Rosati, P., Tozzi, P., et al. 2001, ApJ, 551, 624
Giavalisco, M., Ferguson, H. C., Koekemoer, A. M., et al. 2004, ApJ, 600, L93
Grazian, A., Fontana, A., de Santis, C., et al. 2006, A&A, 449, 951
Hartlap, J., Simon, P., & Schneider, P. 2007, A&A, 464, 399
Hastie, T., Tibshirani, R., & Friedman, J. 2001, The Elements of Statistical Learning (Springer)
Heymans, C., Brown, M. L., Barden, M., et al. 2005, MNRAS, 361, 160
Heymans, C., van Waerbeke, L., Bacon, D., et al. 2006, MNRAS, 368, 1323
Hilbert, S., Hartlap, J., White, S. D. M., & Schneider, P. 2009, A&A, 499, 31
Hoekstra, H., Franx, M., Kuijken, K., & Squires, G. 1998, ApJ, 504, 636
Hoekstra, H., Mellier, Y., van Waerbeke, L., et al. 2006, ApJ, 647, 116
Hyvärinen, A. & Oja, E. 1997, Neural Computation, 9(7), 1438
Hyvärinen, A. & Oja, E. 2000, Neural Networks, 13(4–5), 411
Hyvärinen, A., Karhunen, J., & Oja, E. 2001, Independent Component Analysis (Wiley Interscience)
Jain, B., Seljak, U., & White, S. 2000, AJ, 530, 547
Joachimi, B., Schneider, P., & Eifler, T. 2008, A&A, 477, 43
Jutten, C., & Hérault, J. 1991, Signal Processing, 24, 1
Kaiser, N., & Pan-STARRS Collaboration 2005, in BAAS, 37, 465
Kaiser, N., Squires, G., & Broadhurst, T. 1995, ApJ, 449, 460
Kendall, M. G., Stuart, A., & Ord, J. K. (eds.) 1987, Kendall's advanced theory of statistics (New York, NY, USA: Oxford University Press, Inc.)
Komatsu, E., Dunkley, J., Nolta, M. R., et al. 2009, ApJS, 180, 330
Kuijken, K. 2006, A&A, 456, 827
Kullback, S., & Leibler, R. A. 1951, Annals of Mathematical Statistics, 22, 79
Liang, J.-J., Fang, K.-T., Hickernell, F. J., & Li, R. 2001, Math. Comput., 70, 337
Luppino, G. A., & Kaiser, N. 1997, ApJ, 475, 20
Massey, R., Refregier, A., Conselice, C. J., David, J., & Bacon, J. 2004, MNRAS, 348, 214
Massey, R., Heymans, C., Bergé, J., et al. 2007a, MNRAS, 376, 13
Massey, R., Rhodes, J., Leauthaud, A., et al. 2007b, ApJS, 172, 239
Miller, L., Kitching, T. D., Heymans, C., Heavens, A. F., & van Waerbeke, L. 2007, MNRAS, 382, 315
Peacock, J. A., & Dodds, S. J. 1996, MNRAS, 280, 19
Phleps, S., Wolf, C., Peacock, J. A., Meisenheimer, K., & van Kampen, E. 2007, in Cosmic Frontiers, ed. N. Metcalfe, & T. Shanks, ASP Conf. Ser., 379, 327
Press, W., et al. 1992, Numerical Recipes in C (Cambridge University Press)
R Development Core Team. 2007, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria
Refregier, A., & Bacon, D. 2003, MNRAS, 338, 48
Reiprich, T. H., & Böhringer, H. 2002, ApJ, 567, 716
Rix, H.-W., Barden, M., Beckwith, S. V. W., et al. 2004, ApJS, 152, 163
Schneider, P. 2006, in Saas-Fee Advanced Course 33: Gravitational Lensing: Strong, Weak and Micro, ed. G. Meylan, P. Jetzer, P. North, P. Schneider, C. S. Kochanek, & J. Wambsganss, 269
Schneider, P., van Waerbeke, L., Kilbinger, M., & Mellier, Y. 2002, A&A, 396, 1
Schrabback, T., Erben, T., Simon, P., et al. 2007, A&A, 468, 823
Scoccimarro, R., Zaldarriaga, M., & Hui, L. 1999, ApJ, 527, 1
Scott, D. W. 1992, Multivariate Density Estimation: Theory, Practice, and Visualization (New York: John Wiley & Sons)
Semboloni, E., Mellier, Y., van Waerbeke, L., et al. 2006, A&A, 452, 51
Semboloni, E., van Waerbeke, L., Heymans, C., et al. 2007, MNRAS, 375, 6
Semboloni, E., Tereno, I., van Waerbeke, L., & Heymans, C. 2009, MNRAS, 397, 608
Silverman, B. W. 1986, Density Estimation (London: Chapman and Hall)
Smail, I., Hogg, D. W., Yan, L., & Cohen, J. G. 1995, ApJ, 449, 105
Smith, R. E., Peacock, J. A., Jenkins, A., et al. 2003, MNRAS, 341, 1311
Springel, V. 2005, MNRAS, 364, 1105
Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nature, 435, 629
Takada, M., & Jain, B. 2009, MNRAS, 395, 2065
Venables, W., & Ripley, B. 2002, Modern Applied Statistics with S (Springer)
Wolf, C., Meisenheimer, K., Kleinheinrich, M., et al. 2004, A&A, 421, 913