



**HAL**  
open science

## A new model to predict weak-lensing peak counts

Chieh-An Lin, Martin Kilbinger

► **To cite this version:**

Chieh-An Lin, Martin Kilbinger. A new model to predict weak-lensing peak counts: II. Parameter constraint strategies. *Astronomy and Astrophysics - A&A*, 2015, 583, pp.A70. 10.1051/0004-6361/201526659 . cea-01383759

**HAL Id: cea-01383759**

**<https://cea.hal.science/cea-01383759>**

Submitted on 19 Oct 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A new model to predict weak-lensing peak counts

## II. Parameter constraint strategies

Chieh-An Lin (林皆安) and Martin Kilbinger

Service d'Astrophysique, CEA Saclay, Orme des Merisiers, Bât 709, 91191 Gif-sur-Yvette, France  
e-mail: chieh-an.lin@cea.fr

Received 2 June 2015 / Accepted 17 August 2015

### ABSTRACT

*Context.* Peak counts have been shown to be an excellent tool for extracting the non-Gaussian part of the weak lensing signal. Recently, we developed a fast stochastic forward model to predict weak-lensing peak counts. Our model is able to reconstruct the underlying distribution of observables for analysis.

*Aims.* In this work, we explore and compare various strategies for constraining a parameter using our model, focusing on the matter density  $\Omega_m$  and the density fluctuation amplitude  $\sigma_8$ .

*Methods.* First, we examine the impact from the cosmological dependency of covariances (CDC). Second, we perform the analysis with the copula likelihood, a technique that makes a weaker assumption than does the Gaussian likelihood. Third, direct, non-analytic parameter estimations are applied using the full information of the distribution. Fourth, we obtain constraints with approximate Bayesian computation (ABC), an efficient, robust, and likelihood-free algorithm based on accept-reject sampling.

*Results.* We find that neglecting the CDC effect enlarges parameter contours by 22% and that the covariance-varying copula likelihood is a very good approximation to the true likelihood. The direct techniques work well in spite of noisier contours. Concerning ABC, the iterative process converges quickly to a posterior distribution that is in excellent agreement with results from our other analyses. The time cost for ABC is reduced by two orders of magnitude.

*Conclusions.* The stochastic nature of our weak-lensing peak count model allows us to use various techniques that approach the true underlying probability distribution of observables, without making simplifying assumptions. Our work can be generalized to other observables where forward simulations provide samples of the underlying distribution.

**Key words.** gravitational lensing: weak – large-scale structure of Universe – methods: statistical

## 1. Introduction

Weak lensing (WL) is a gravitational deflection effect of light by matter inhomogeneities in the Universe that causes distortion of source galaxy images. This distortion corresponds to the integrated deflection along the line of sight, and its measurement probes the high-mass regions of the Universe. These regions contain structures that formed during the late-time evolution of the Universe, which depends on cosmological parameters, such as the matter density parameter  $\Omega_m$ , the matter density fluctuation  $\sigma_8$ , and the equation of state of dark energy  $w$ . Ongoing and future surveys such as KiDS<sup>1</sup>, DES<sup>2</sup>, HSC<sup>3</sup>, WFIRST<sup>4</sup>, Euclid<sup>5</sup>, and LSST<sup>6</sup> are expected to provide tight constraints on those and other cosmological parameters and to distinguish between different cosmological models, using weak lensing as a major probe.

Lensing signals can be extracted in several ways. A common observable is the cosmic shear two-point-correlation function (2PCF), which has been used to constrain cosmological parameters in many studies, including recent ones (Kilbinger et al. 2013; Jee et al. 2013). However, the 2PCF only retains

Gaussianity, and it misses the rich nonlinear information of the structure evolution encoded on small scales. To compensate for this drawback, several non-Gaussian statistics have been proposed, for example higher order moments (Kilbinger & Schneider 2005; Semboloni et al. 2011; Fu et al. 2014; Simon et al. 2015), the three-point correlation function (Schneider & Lombardi 2003; Takada & Jain 2003; Scoccimarro et al. 2004), Minkowski functionals (Petri et al. 2015), or peak statistics, which is the aim of this series of papers. Some more general work comparing different strategies to extract non-Gaussian information can be found in the literature (Pires et al. 2009, 2012; Bergé et al. 2010).

Peaks, defined as local maxima of the lensing signal, are direct tracers of high-mass regions in the large-scale structure of the Universe. In the medium and high signal-to-noise (S/N) regime, the peak function (the number of peaks as function of S/N) is not dominated by shape noise, and this permits one to study the cosmological dependency of the peak number counts (Jain & Van Waerbeke 2000). Various aspects of peak statistics have been investigated in the past: the physical origin of peaks (Hamana et al. 2004; Yang et al. 2011), projection effects (Marian et al. 2010), the optimal combination of angular scales (Kratichvil et al. 2010; Marian et al. 2012), redshift tomography (Hennawi & Spergel 2005), cosmological parameter constraints (Dietrich & Hartlap 2010; Liu, X. et al. 2014), detecting primordial non-Gaussianity (Maturi et al. 2011; Marian et al. 2011), peak statistics beyond the abundance (Marian et al. 2013), the

<sup>1</sup> <http://kids.strw.leidenuniv.nl/>

<sup>2</sup> <http://www.darkenergysurvey.org/>

<sup>3</sup> <http://www.naoj.org/Projects/HSC/HSCProject.html>

<sup>4</sup> <http://wfirst.gsfc.nasa.gov/>

<sup>5</sup> <http://www.euclid-ec.org/>

<sup>6</sup> <http://www.lsst.org/lsst/>

impact from baryons (Yang et al. 2013; Osato et al. 2015), magnification bias (Liu, J. et al. 2014), and shape measurement errors (Bard et al. 2013). Recent studies by Liu et al. (2015, hereafter LPH15), Liu et al. (2015, hereafter LPL15), and Hamana et al. (2015) have applied likelihood estimation for WL peaks on real data and have shown that the results agree with the current  $\Lambda$ CDM scenario.

Modeling number counts is a challenge for peak studies. To date, there have been three main approaches. The first one is to count peaks from a large number of  $N$ -body simulations (Dietrich & Hartlap 2010; LPH15), which directly emulate structure formation by numerical implementation of the corresponding physical laws. The second family consists of analytic predictions (Maturi et al. 2010; Fan et al. 2010) based on Gaussian random field theory. A third approach has been introduced by Lin & Kilbinger (2015, hereafter Paper I): similar to Kruse & Schneider (1999) and Kainulainen & Marra (2009, 2011a,b), we propose a stochastic process to predict peak counts by simulating lensing maps from a halo distribution drawn from the mass function.

Our model possesses several advantages. The first one is flexibility. Observational conditions can easily be modeled and taken into account. The same is true for additional features, such as intrinsic alignment of galaxies and other observational and astrophysical systematics. Second, since our method does not need  $N$ -body simulations, the computation time required to calculate the model are orders of magnitudes faster, and we can explore a large parameter space. Third, our model explores the underlying probability density function (PDF) of the observables. All statistical properties of the peak function can be derived directly from the model, making various parameter estimation methods possible.

In this paper, we apply several parameter constraint and likelihood methods for our peak-count-prediction model from Paper I. Our goal is to study and compare different strategies and to make use of the full potential of the fast stochastic forward modeling approach. We start with a likelihood function that is assumed to be Gaussian in the observables with constant covariance and then compare this to methods that make fewer and fewer assumptions, as follows.

The first extension of the Gaussian likelihood is to take the cosmology-dependent covariances (CDC, see Eifler et al. 2009) into account. Thanks to the fast performance of our model, it is feasible to estimate the covariance matrix for each parameter set.

The second improvement we adopt is the copula analysis (Benabed et al. 2009; Jiang et al. 2009; Takeuchi 2010; Scherrer et al. 2010; Sato et al. 2011) for the Gaussian approximation. Widely used in finance, the copula transform uses the fact that any multivariate distribution can be transformed into a new one where the marginal PDF is uniform. Combining successive transforms can then give rise to a new distribution where all marginals are Gaussian. This makes weaker assumptions about the underlying likelihood than the Gaussian hypothesis.

Third, we directly estimate the full underlying distribution information in a non-analytical way. This allows us to strictly follow the original definition of the likelihood estimator: the conditional probability of observables for a given parameter set. In addition, we compute the  $p$ -value from the full PDF. These  $p$ -values derived for all parameter sets allow for significance tests and provide a direct way to construct confidence contours.

Furthermore, our model makes it possible to dispose of a likelihood function altogether, using approximate Bayesian computation (ABC, see e.g. Marin et al. 2011) for exploring the parameter space. ABC is a powerful constraining technique

based on accept-reject sampling. Proposed first by Rubin (1984), ABC produces the posterior distribution by bypassing the likelihood evaluation, which may be complex and practically unfeasible in some contexts. The posterior is constructed by comparing the sampled result with the observation to decide whether a proposed parameter is accepted. This technique can be improved by combining ABC with population Monte Carlo (PMC<sup>7</sup>, Beaumont et al. 2009; Cameron & Pettitt 2012; Weyant et al. 2013). Until now, ABC seems to already have various applications in biology-related domains (e.g., Beaumont et al. 2009; Berger et al. 2010; Csilléry et al. 2010; Drovandi & Pettitt 2011), while applications for astronomical purposes are few: morphological transformation of galaxies (Cameron & Pettitt 2012), cosmological parameter inference using type Ia supernovae (Weyant et al. 2013), constraints of the disk formation of the Milky Way (Robin et al. 2014), and strong lensing properties of galaxy clusters (Killedar et al. 2015). Very recently, two papers (Ishida et al. 2015; Akeret et al. 2015) dedicated to ABC in a general cosmological context have been submitted.

The paper is organized as follows. In Sect. 2, we briefly review our model introduced in Paper I, the setting for the parameter analysis, and the criteria for defining parameter constraints. In Sect. 3, we study the impact of the CDC effect. The results from the copula likelihood can be found in Sects. 4, and 5 we estimate the true underlying PDF in a non-analytic way and show parameters constraints without the Gaussian hypothesis. Section 6 focuses on the likelihood-free ABC technique, and the last section is dedicated to a discussion where we summarize this work.

## 2. Methodology

### 2.1. Our model

Our peak-count model uses a probabilistic approach that generates peak catalogs from a given mass function model. This is done by generating fast simulations of halos, computing the projected mass, and simulating lensing maps from which one can extract WL peaks. A step-by-step summary is given as follows:

1. sample halo masses and assign density profiles and positions (fast simulations);
2. compute the projected mass and subtract the mean over the field (ray-tracing simulations);
3. add noise and smooth the map with a kernel; and
4. select local S/N maxima.

Here, two assumptions have been made: (1) only bound matter contributes to number counts; and (2) the spatial correlation of halos has a small impact on WL peaks. Paper I showed that combining both hypotheses gives a good estimation of the peak abundance.

We adopt the same settings as Paper I: the mass function model from Jenkins et al. (2001), the truncated Navarro-Frenk-White halo profiles (Navarro et al. 1996, 1997), Gaussian shape noise, the Gaussian smoothing kernel, and sources at fixed redshift which are distributed on a regular grid. The field of view is chosen such that the effective area after cutting off the border is 25 deg<sup>2</sup>. An exhausted list of parameter values used in this paper can be found in Table 1. Readers are encouraged to read Paper I for their definitions and for detailed explanations for our model.

<sup>7</sup> This algorithm is called PMC ABC by some and SMC (sequential Monte Carlo) ABC by others.

**Table 1.** Parameter values adopted in this study.

| Parameter                        | Symbol                  | Value                    |
|----------------------------------|-------------------------|--------------------------|
| Lower sampling limit             | –                       | $10^{12} M_{\odot}/h$    |
| Upper sampling limit             | –                       | $10^{17} M_{\odot}/h$    |
| NFW inner slope                  | $\alpha$                | 1                        |
| $M$ – $c$ relation parameter     | $c_0$                   | 11                       |
| $M$ – $c$ relation parameter     | $\beta$                 | 0.13                     |
| Source redshift                  | $z_s$                   | 1                        |
| Intrinsic ellipticity dispersion | $\sigma_{\epsilon}$     | 0.4                      |
| Galaxy number density            | $n_g$                   | $25 \text{ arcmin}^{-2}$ |
| Pixel size                       | $\theta_{\text{pix}}$   | 0.2 arcmin               |
| Kernel size                      | $\theta_G$              | 1 arcmin                 |
| Shape noise                      | $\sigma_{\text{pix}}$   | 0.283                    |
| Smoothed noise                   | $\sigma_{\text{noise}}$ | 0.0226                   |
| Effective field area             | –                       | $25 \text{ deg}^2$       |

All computations with our model in this study are performed by our CAMELUS algorithm<sup>8</sup>. A realization (from a mass function to a peak catalog) of a  $25 \text{ deg}^2$  field costs few seconds to generate on a single-CPU computer. The real time cost depends of course on input cosmological parameters, but this still gives an idea about the speed of our algorithm.

## 2.2. Analysis design

Throughout this paper,  $\pi$  denotes a parameter set. To simplify the study, the dimension of the parameter space is reduced to two:  $\pi \equiv (\Omega_m, \sigma_8)$ . The other cosmological parameters are fixed, including  $\Omega_b = 0.047$ ,  $h = 0.78$ ,  $n_s = 0.95$ , and  $w = -1$ . The dark energy density  $\Omega_{\text{de}}$  is set to  $1 - \Omega_m$  to match a flat universe. On the  $\Omega_m$ – $\sigma_8$  plane, we explore a region where the posterior density, or probability, is high, see Fig. 1. We compute the values of three different log likelihoods on the grid points of these zones. The grid size of the center zone is  $\Delta\Omega_m = \Delta\sigma_8 = 0.005$ , whereas it is 0.01 for the rest. This results in a total of 7821 points in the parameter space to evaluate.

For each  $\pi$ , we carry out  $N = 1000$  realizations of a  $25 \text{ deg}^2$  field and determine the associated data vector  $\mathbf{x}^{(k)} = (x_1^{(k)}, \dots, x_d^{(k)})$  for all  $k$  from 1 to  $N$ . These are independent samples drawn from their underlying PDF of observables for a given parameter  $\pi$ . We estimate the model prediction (which is the mean), the covariance matrix, and the inverse matrix (Hartlap et al. 2007), respectively, by following

$$x_i^{\text{mod}} = \frac{1}{N} \sum_{k=1}^N x_i^{(k)}, \quad (1)$$

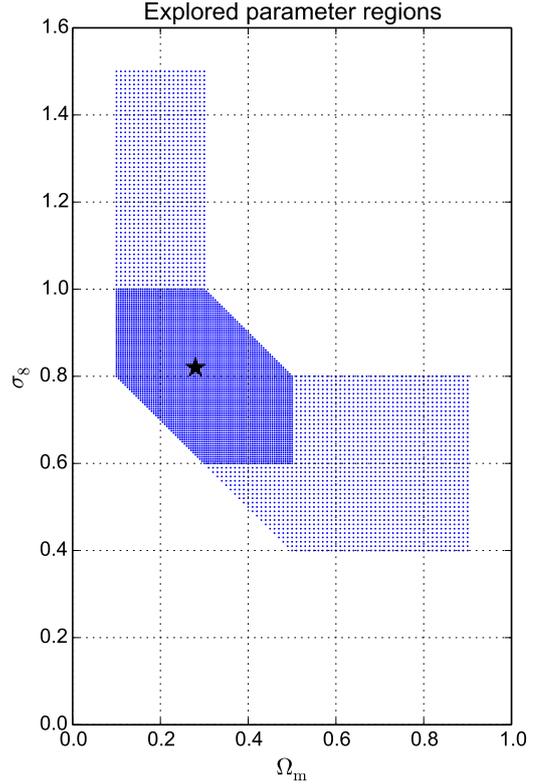
$$\hat{C}_{ij} = \frac{1}{N-1} \sum_{k=1}^N (x_i^{(k)} - x_i^{\text{mod}})(x_j^{(k)} - x_j^{\text{mod}}), \quad \text{and} \quad (2)$$

$$\widehat{\mathbf{C}}^{-1} = \frac{N-d-2}{N-1} \widehat{\mathbf{C}}^{-1}, \quad (3)$$

where  $d$  denotes the dimension of data vector. This results in a total area of  $25 \text{ 000 deg}^2$  for the mean estimation.

In this paper, the observation data  $\mathbf{x}^{\text{obs}}$  are identified with a realization of our model, which means that  $\mathbf{x}^{\text{obs}}$  is derived by a particular realization of  $\mathbf{x}(\pi^{\text{in}})$ . The input parameters chosen are  $\pi^{\text{in}} = (\Omega_m^{\text{in}}, \sigma_8^{\text{in}}) = (0.28, 0.82)$ . The authors would like to highlight that the accuracy of the model is not the aim of this research work, but precision. Therefore, the input choice and the uncertainty of random process should have little impact.

<sup>8</sup> <http://github.com/Linc-tw/camelus>



**Fig. 1.** Location of 7821 points on which we evaluate the likelihoods. In the condensed area, the interval between two grid points is 0.005, while in both wing zones it is 0.01. The black star shows  $(\Omega_m^{\text{in}}, \sigma_8^{\text{in}}) = (0.28, 0.82)$ .

Peak-count information can be combined into a data vector using different ways. Inspired by Dietrich & Hartlap (2010) and LPL15, we studied three types of observables. The first is the abundance of peaks found in each S/N bin (labeled *abd*), in other words, the binned peak function. The second is the S/N values at some given percentiles of the peak cumulative distribution function (CDF, labeled *pct*). The third is similar to the second type, but without taking peaks below a threshold S/N value (labeled *cut*) into account. Mathematically, the two last types of observables can be denoted as  $x_i$ , thereby satisfying

$$p_i = \int_{v_{\text{min}}}^{x_i} n_{\text{peak}}(v) dv, \quad (4)$$

where  $n_{\text{peak}}(v)$  is the peak PDF function,  $v_{\text{min}}$  a cutoff, and  $p_i$  a given percentile. The observable  $\mathbf{x}^{\text{abd}}$  is used by LPL15, while readers find  $\mathbf{x}^{\text{cut}}$  from Dietrich & Hartlap (2010). We would like to clarify that using  $\mathbf{x}^{\text{pct}}$  for analysis could be risky, since this includes peaks with negative S/N. From Paper I, we observe that although high-peak counts from our model agree well with  $N$ -body simulations, predictions for local maxima found in underdensity regions (peaks with  $S/N < 0$ ) are inaccurate. Thus, we include  $\mathbf{x}^{\text{pct}}$  in this paper only to give an idea about how much information we can extract from observables defined by percentiles.

Observable vectors are constructed by the description above with the settings of Table 2. This choice of bins and  $p_i$  is made such that the same component from different types of observables represents about the same information, since the bin center of  $\mathbf{x}^{\text{abd}5}$  roughly correspond to  $\mathbf{x}^{\text{cut}5}$  for the input cosmology  $\pi^{\text{in}}$ . Following LPL15, who discovered in their study that the bin-width choice has a minor impact on parameter constraints if the

**Table 2.** Definition of  $\mathbf{x}^{\text{abd5}}$ ,  $\mathbf{x}^{\text{pct5}}$ , and  $\mathbf{x}^{\text{cut5}}$ .

|                             |            |            |            |            |           |
|-----------------------------|------------|------------|------------|------------|-----------|
| Label                       | abd5       |            |            |            |           |
| Bins on $\nu$               | [3.0, 3.8[ | [3.8, 4.5[ | [4.5, 5.3[ | [5.3, 6.2[ | [6.2, +∞[ |
| $x_i$ for $\pi^{\text{in}}$ | 330        | 91         | 39         | 18         | 15        |
| Label                       | pct5       |            |            |            |           |
| $\nu_{\text{min}}$          | $-\infty$  |            |            |            |           |
| $p_i$                       | 0.969      | 0.986      | 0.994      | 0.997      | 0.999     |
| $x_i$ for $\pi^{\text{in}}$ | 3.5        | 4.1        | 4.9        | 5.7        | 7.0       |
| Label                       | cut5       |            |            |            |           |
| $\nu_{\text{min}}$          | 3          |            |            |            |           |
| $p_i$                       | 0.5        | 0.776      | 0.9        | 0.955      | 0.98      |
| $x_i$ for $\pi^{\text{in}}$ | 3.5        | 4.1        | 4.9        | 5.7        | 6.7       |

**Notes.** Parameters such as  $\nu_{\text{min}}$  and  $p_i$  are used by Eq. (4). As an indication, their values for the input cosmology  $\pi^{\text{in}}$  are also given. They were calculated by averaging over 1000 realizations.

**Table 3.** Correlation matrices of  $\mathbf{x}^{\text{abd5}}$ ,  $\mathbf{x}^{\text{pct5}}$ , and  $\mathbf{x}^{\text{cut5}}$  in the input cosmology. For  $\mathbf{x}^{\text{abd5}}$ , the peak abundance is weakly correlated between bins.

$$\begin{aligned}
 \text{abd5} & \begin{pmatrix} 1 & -0.05 & -0.09 & -0.08 & -0.16 \\ -0.05 & 1 & -0.05 & -0.01 & -0.12 \\ -0.09 & -0.05 & 1 & -0.04 & -0.11 \\ -0.08 & -0.01 & -0.04 & 1 & -0.06 \\ -0.16 & -0.12 & -0.11 & -0.06 & 1 \end{pmatrix} \\
 \text{pct5} & \begin{pmatrix} 1 & 0.62 & 0.29 & 0.15 & 0.11 \\ 0.62 & 1 & 0.58 & 0.36 & 0.25 \\ 0.29 & 0.58 & 1 & 0.66 & 0.43 \\ 0.15 & 0.36 & 0.66 & 1 & 0.59 \\ 0.11 & 0.25 & 0.43 & 0.59 & 1 \end{pmatrix} \\
 \text{cut5} & \begin{pmatrix} 1 & 0.58 & 0.31 & 0.20 & 0.15 \\ 0.58 & 1 & 0.61 & 0.39 & 0.28 \\ 0.31 & 0.61 & 1 & 0.65 & 0.47 \\ 0.20 & 0.39 & 0.65 & 1 & 0.70 \\ 0.15 & 0.28 & 0.47 & 0.70 & 1 \end{pmatrix}
 \end{aligned}$$

estimated number count in each bin is  $\geq 10$ , we chose not to explore different choices of binwidths for  $\mathbf{x}^{\text{abd5}}$ . We also note that  $p_i$  for  $x_i^{\text{cut5}}$  are logarithmically spaced.

By construction, the correlation between terms of percentile-like vectors is much higher than for the case of peak abundance. This tendency is shown in Table 3 for the  $\pi^{\text{in}}$  cosmology. We discovered that  $\mathbf{x}^{\text{pct5}}$  and  $\mathbf{x}^{\text{cut5}}$  are highly correlated, while for  $\mathbf{x}^{\text{abd5}}$ , the highest absolute value of off-diagonal terms does not exceed 17%. A similar result was observed when we binned data differently. This suggests that the covariance should be included in likelihood analyses.

### 2.3. Constraint qualification

In this paper, both Bayesian inferences and likelihood-ratio tests (see, e.g., Theorem 10.3.3 from Casella & Berger 2002) have been performed. To distinguish between these two cases, we call *credible region* the posterior PDF obtained from the Bayesian approach, which differs from the *confidence region*, whose interpretation can be found in Sect. 5.2.

To quantify parameter-constraint contours, we introduce two criteria. Inspired by Jain & Seljak (1997) and Maoli et al. (2001), the first criterion is to determine the error on

$$\Sigma_8 \equiv \sigma_8(\Omega_m/0.27)^\alpha. \quad (5)$$

Since the banana-shaped contour becomes more or less an elongated ellipse in log space,  $\Delta\Sigma_8$  represents the ‘‘thickness’’ of the banana, tilted by the slope  $\alpha$ . Therefore, we first fit  $\alpha$  with the linear relation  $\log \Sigma_8 = \log \sigma_8 + \alpha \log(\Omega_m/0.27)$ , and then calculate the  $1\sigma$  interval of  $\Sigma_8$  on the  $\Omega_m$ - $\Sigma_8$  plane. For a Bayesian approach, this interval is given by the 68% most probable interval from the marginalized likelihood, while for a frequentist approach, significance levels are given by likelihood-ratio tests on the marginalized likelihood. Examples of both approaches are shown by Fig. 2. Since no real data are used in this study, we are not interested in the best fit value, but the  $1\sigma$  interval width  $\Delta\Sigma_8$ .

The second indicator is the figure of merit (FoM) for  $\Omega_m$  and  $\sigma_8$ , proposed by Dietrich & Hartlap (2010). They define a FoM similar to the one from Albrecht et al. (2006) as the inverse of the area of the  $2\sigma$  region.

## 3. Influence of the cosmology-dependent covariance

### 3.1. Formalism

In this section, we examine the cosmology-dependent-covariance (CDC) effect. From our statistic, we estimate the inverse covariance from Eq. (3) for each  $\pi$  from 1000 realizations. By setting the Bayesian evidence to unity,  $P(\mathbf{x} = \mathbf{x}^{\text{obs}}) = 1$ , we write the relation among prior probability  $\mathcal{P}(\pi)$ , the likelihood  $\mathcal{L}(\pi) \equiv P(\mathbf{x}^{\text{obs}}|\pi)$ , and posterior probability  $\mathcal{P}(\pi|\mathbf{x}^{\text{obs}})$  as

$$\mathcal{P}(\pi|\mathbf{x}^{\text{obs}}) = \mathcal{L}(\pi)\mathcal{P}(\pi). \quad (6)$$

Given a model, we write  $\Delta\mathbf{x}(\pi) \equiv \mathbf{x}^{\text{mod}}(\pi) - \mathbf{x}^{\text{obs}}$  as the difference between the model prediction  $\mathbf{x}^{\text{mod}}$  and the observation  $\mathbf{x}^{\text{obs}}$ . Then the Gaussian log-likelihood is given by

$$-2 \ln \mathcal{L}(\pi) = \ln \left[ (2\pi)^d \det(\mathbf{C}) \right] + \Delta\mathbf{x}^T \mathbf{C}^{-1} \Delta\mathbf{x} \quad (7)$$

where  $d$  denotes the dimension of the observable space, and  $\mathbf{C}$  is the covariance matrix for  $\mathbf{x}^{\text{mod}}$ .

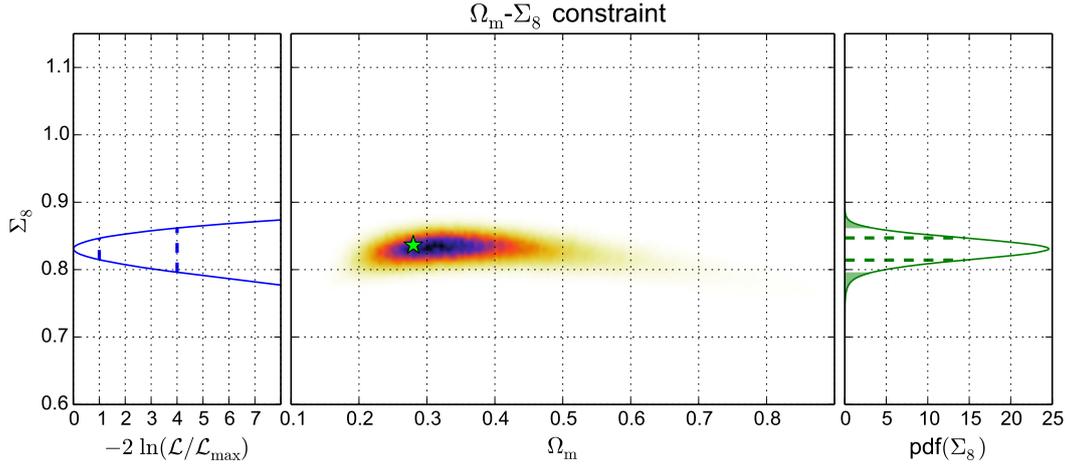
Estimating  $\mathbf{C}$  as an ensemble average is difficult since cosmologists only have one Universe. One can derive  $\mathbf{C}$  from observations with statistical techniques, such as bootstrap or jackknife (LPL15), or from a sufficient number of independent fields of view from  $N$ -body simulations (LPH15) or using analytic calculations. However, the first method only provides the estimation for a specific parameter set  $\mathbf{C}(\pi = \pi^{\text{obs}})$ ; the second method is limited to a small amount of parameters owing to the very high computational time cost; and the third method involves higher order statistics of the observables, which might not be well known. Thus, most studies suppose that the covariance matrix is invariant so ignore the CDC effect. In this case, the determinant term becomes a constant, and likelihood analysis can be summed up as the minimization of  $\chi^2 \equiv \Delta\mathbf{x}^T \mathbf{C}^{-1} \Delta\mathbf{x}$ .

Alternatively, the stochastic characteristic of our model provides a quick and simple way to estimate the covariance matrix  $\mathbf{C}(\pi)$  of each single parameter set  $\pi$ . To examine the impact of the CDC effect in the peak-count framework, we write down the constant-covariance Gaussian (labeled cg), the semi-varying-covariance Gaussian (labeled svg), and the varying-covariance Gaussian (labeled vg) log-likelihoods as

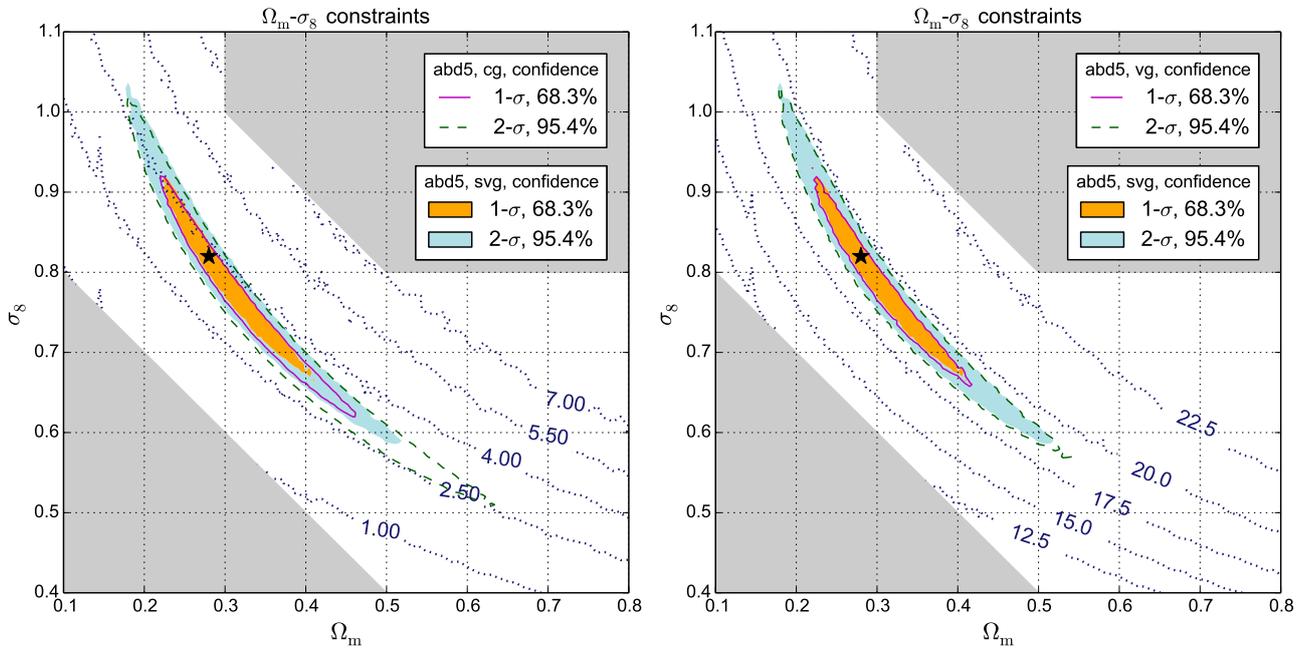
$$L_{\text{cg}} \equiv \Delta\mathbf{x}^T(\pi) \widehat{\mathbf{C}}^{-1}(\pi^{\text{obs}}) \Delta\mathbf{x}(\pi), \quad (8)$$

$$L_{\text{svg}} \equiv \Delta\mathbf{x}^T(\pi) \widehat{\mathbf{C}}^{-1}(\pi) \Delta\mathbf{x}(\pi), \quad \text{and} \quad (9)$$

$$L_{\text{vg}} \equiv \ln \left[ \det \widehat{\mathbf{C}}(\pi) \right] + \Delta\mathbf{x}^T(\pi) \widehat{\mathbf{C}}^{-1}(\pi) \Delta\mathbf{x}(\pi). \quad (10)$$



**Fig. 2.** *Middle panel:* likelihood value using  $\mathbf{x}^{\text{abd5}}$  on the  $\Omega_m$ - $\Sigma_8$  plane. The green star represents the input cosmology  $\pi^{\text{in}}$ . Since  $\log \sigma_8$  and  $\log \Omega_m$  form an approximately linear degeneracy, the quantity  $\Sigma_8 \equiv \sigma_8(\Omega_m/0.27)^\alpha$  allows us to characterize the banana-shape contour thickness. *Right panel:* the marginalized PDF of  $\Sigma_8$ . The dashed lines give the  $1\sigma$  interval (68.3%), while the borders of the shaded areas represent  $2\sigma$  limits (95.4%). *Left panel:* log-value of the marginalized likelihood ratio. Dashed lines in the *left panel* give the corresponding value for  $1$  and  $2\sigma$  significance levels, respectively.



**Fig. 3.** Confidence regions derived from  $L_{\text{cg}}$ ,  $L_{\text{svg}}$ , and  $L_{\text{vg}}$  with  $\mathbf{x}^{\text{abd5}}$ . The solid and dashed lines represent  $L_{\text{cg}}$  in the *left panel* and  $L_{\text{vg}}$  in the *right panel*, while the colored areas are from  $L_{\text{svg}}$ . The black star stands for  $\pi^{\text{in}}$  and gray areas represent the non-explored parameter space. The dotted lines are different isolines, the variance  $\hat{C}_{55}$  of the bin with highest S/N in the *left panel* and  $\ln(\det \hat{C})$  for the *right panel*. The contour area is reduced by 22% when taking the CDC effect into account. The parameter-dependent determinant term does not contribute significantly.

Here, the term  $\hat{C}^{-1}(\pi^{\text{obs}})$  in Eq. (8) refers to  $\hat{C}^{-1}(\pi^{\text{in}})$ , where  $\pi^{\text{in}}$  is described in Sect. 2.2. By comparing the contours derived from different likelihoods, we aim to measure (1) the evolution of the  $\chi^2$  term by substituting the constant matrix with the true varying  $\hat{C}^{-1}$ ; and (2) the impact from adding the determinant term. Therefore,  $L_{\text{svg}}$  is just an illustrative case to assess the influence of the two terms in the likelihood.

### 3.2. The $\chi^2$ term

The lefthand panel of Fig. 3 shows the comparison between confidence regions derived from  $L_{\text{cg}}$  and  $L_{\text{svg}}$  with  $\mathbf{x}^{\text{abd}}$ . It shows a clear difference of the contours between  $L_{\text{cg}}$  and  $L_{\text{svg}}$ . Since

the off-diagonal correlation coefficients are weak (as shown in Table 3), the variation in diagonal terms of  $\mathbf{C}$  plays a major role in the size of credible regions. The isolines for  $\hat{C}_{55}$  are also drawn in Fig. 3. These isolines cross the  $\Omega_m$ - $\sigma_8$  degeneracy lines from  $L_{\text{cg}}$  and thus shrink the credible region. We also find that the isolines for  $\hat{C}_{11}$  and  $\hat{C}_{22}$  are noisy and that those for  $\hat{C}_{33}$  and  $\hat{C}_{44}$  coincide well with the original degeneracy direction.

Table 4 shows the values of both criteria for different likelihoods. We observe that using  $L_{\text{svg}}$  significantly improves the constraints by 24% in terms of FoM. Regarding  $\Delta\Sigma_8$ , the improvement is weak. As a result, using varying covariance matrices breaks down part of the banana-shape degeneracy and shrinks the contour length, but does not reduce the thickness.

**Table 4.**  $\Delta\Sigma_8$ , the error on the parameter (5) and the figure of merit (FoM) for confidence regions are summarized for the different analysis approaches performed in this paper.

|                   | $\mathbf{x}^{\text{abd5}}$ |     | $\mathbf{x}^{\text{pct5}}$ |     | $\mathbf{x}^{\text{cut5}}$ |     |
|-------------------|----------------------------|-----|----------------------------|-----|----------------------------|-----|
|                   | $\Delta\Sigma_8$           | FoM | $\Delta\Sigma_8$           | FoM | $\Delta\Sigma_8$           | FoM |
| $L_{\text{cg}}$   | 0.032                      | 46  | 0.037                      | 31  | 0.065                      | 13  |
| $L_{\text{svg}}$  | 0.031                      | 57  | 0.032                      | 42  | 0.054                      | 21  |
| $L_{\text{vg}}$   | 0.031                      | 56  | 0.032                      | 43  | 0.052                      | 18  |
| $L_{\text{cc}}$   | 0.032                      | 43  | 0.038                      | 33  | 0.056                      | 13  |
| $L_{\text{vc}}$   | 0.033                      | 52  | 0.034                      | 39  | 0.058                      | 16  |
| $L_{\text{true}}$ | 0.033                      | 54  | 0.035                      | 39  | 0.058                      | 17  |
| $p$ -value        | 0.035                      | 39  | 0.037                      | 27  | 0.067                      | 12  |

**Notes.**  $L_{\text{cg}}$ ,  $L_{\text{svg}}$ , and  $L_{\text{vg}}$  are introduced in Sect. 3,  $L_{\text{cc}}$  and  $L_{\text{vc}}$  in Sect. 4, and  $L_{\text{true}}$  and  $p$ -value in Sect. 5. In each case, we take  $\mathbf{x}^{\text{abd5}}$ ,  $\mathbf{x}^{\text{pct5}}$ , or  $\mathbf{x}^{\text{cut5}}$  as data vector as indicated in the table rows.

In the lefthand panels of Fig. 4, we show the same constraints derived from two other observables  $\mathbf{x}^{\text{pct5}}$  and  $\mathbf{x}^{\text{cut5}}$ . We see a similar CDC effect for both. We observe that  $\mathbf{x}^{\text{pct5}}$  has less constraining power than  $\mathbf{x}^{\text{abd5}}$ , and  $\mathbf{x}^{\text{cut5}}$  is outperformed by both other data vectors. This is due to the cutoff value  $\nu_{\text{min}}$ . Introducing a cutoff at  $\nu_{\text{min}} = 3$  decreases the total number of peaks and amplifies the fluctuation of high-peak values in the CDF. When we use percentiles to define observables, the distribution of each component of  $\mathbf{x}^{\text{cut5}}$  becomes wider than the one of the corresponding component of  $\mathbf{x}^{\text{pct5}}$ , and this greater scatter in the CDF enlarges the contours. However, the cutoff also introduces a tilt for the contours. Table 5 shows the best fit  $\alpha$  for the different cases. The difference in the tilt could be a useful tool for improving the constraining power. This has also been observed by Dietrich & Hartlap (2010). Nevertheless, we do not take on any joint analysis since  $\mathbf{x}^{\text{abd5}}$  and  $\mathbf{x}^{\text{cut5}}$  contain essentially the same information.

### 3.3. Impact from the determinant term

The righthand panel of Fig. 3 shows the comparison between  $L_{\text{svg}}$  and  $L_{\text{vg}}$  with  $\mathbf{x}^{\text{abd5}}$ . It shows that adding the determinant term does not result in significant changes of the parameter constraints. The isolines from  $\ln(\det \hat{C})$  explain this, since the gradients are perpendicular to the degeneracy lines. We observe that including the determinant makes the contours slightly larger, but almost negligibly so. The total improvement in the contour area compared to  $L_{\text{cg}}$  is 22%.

However, a different change is seen for  $\mathbf{x}^{\text{pct5}}$  and  $\mathbf{x}^{\text{cut5}}$ . Adding the determinant to the likelihood computed from these observables induces a shift of contours toward the higher  $\Omega_{\text{m}}$  area. In the case of  $\mathbf{x}^{\text{cut5}}$ , this shift compensates for the contour offset from the varying  $\chi^2$  term, but does not improve either  $\Delta\Sigma_8$  or FoM significantly, as shown in Table 4. As a result, using the Gaussian likelihood, the total CDC effect can be summed up as an improvement of at least 14% in terms of thickness and 38% in terms of area.

The results from Bayesian inference is very similar to the likelihood-ratio test. Thus, we only show their  $\Delta\Sigma_8$  and FoM in Table 6 and best fits in Table 7. We recall that a similar analysis was done by Eifler et al. (2009) on shear covariances. Our observations agree with their conclusions: a relatively large impact from the  $\chi^2$  term and negligible change from the determinant term. However, the total CDC effect is more significant in the peak-count framework than for the power spectrum.

## 4. Testing the copula transform

### 4.1. Formalism

Consider a multivariate joint distribution  $P(x_1, \dots, x_d)$ . In general,  $P$  could be far from Gaussian so that imposing a Gaussian likelihood could induce biases. The idea of the copula technique is to evaluate the likelihood in a new observable space where the Gaussian approximation is better. Using a change in variables, individual marginalized distributions of  $P$  can be approximated to Gaussian ones. This is achieved by a series of successive one-dimensional, axis-wise transformations. The multivariate Gaussianity of the transformed distribution is not guaranteed. However, in some cases, this transformation tunes the distribution and makes it more “Gaussian”, so that evaluating the likelihood in the tuned space is more realistic (Benabed et al. 2009; Sato et al. 2011).

From Sklar’s theorem (Sklar 1959), any multivariate distribution  $P(x_1, \dots, x_d)$  can be decomposed into the copula density multiplied by marginalized distributions. A comprehensible and elegant demonstration is given by R ajtschendorf (2009). Readers are also encouraged to follow Scherrer et al. (2010) for detailed physical interpretations and Sato et al. (2011) for a very pedagogical derivation of the Gaussian copula transform.

Consider a  $d$ -dimensional distribution  $P(\mathbf{x})$ , where  $\mathbf{x} = (x_1, \dots, x_d)$  is a random vector. We let  $P_i$  be the marginalized 1-point PDF of  $x_i$ , and  $F_i$  the corresponding CDF. Sklar’s theorem shows that there is a unique  $d$ -dimensional function  $c$  defined on  $[0, 1]^d$  with uniform marginal PDF, such that

$$P(\mathbf{x}) = c(\mathbf{u})P_1(x_1) \cdots P_d(x_d), \quad (11)$$

where  $u_i \equiv F_i(x_i)$ . The function  $c$  is called the *copula density*. On the other hand, let  $q_i \equiv \Phi_i^{-1}(u_i)$ , where  $\Phi_i$  is the CDF of the normal distribution with the same means  $\mu_i$  and variances  $\sigma_i^2$  as the laws  $P_i$ , such that

$$\Phi_i(q_i) \equiv \int_{-\infty}^{q_i} \phi_i(q') dq', \quad (12)$$

$$\phi_i(q_i) \equiv \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{(q_i - \mu_i)^2}{2\sigma_i^2}\right]. \quad (13)$$

We can then define a new joint PDF  $P'$  in the  $\mathbf{q}$  space that corresponds to  $P$  in  $\mathbf{x}$  space, i.e.  $P'(\mathbf{q}) = P(\mathbf{x})$ . The marginal PDF and CDF of  $P'$  are only  $\phi_i$  and  $\Phi_i$ , respectively. Thus, applying Eq. (11) to  $P'$  and  $\phi_i$  leads to

$$P'(\mathbf{q}) = c(\mathbf{u})\phi_1(q_1) \cdots \phi_d(q_d). \quad (14)$$

By the uniqueness of the copula density,  $c$  in Eqs. (11) and (14) are the same. Thus, we obtain

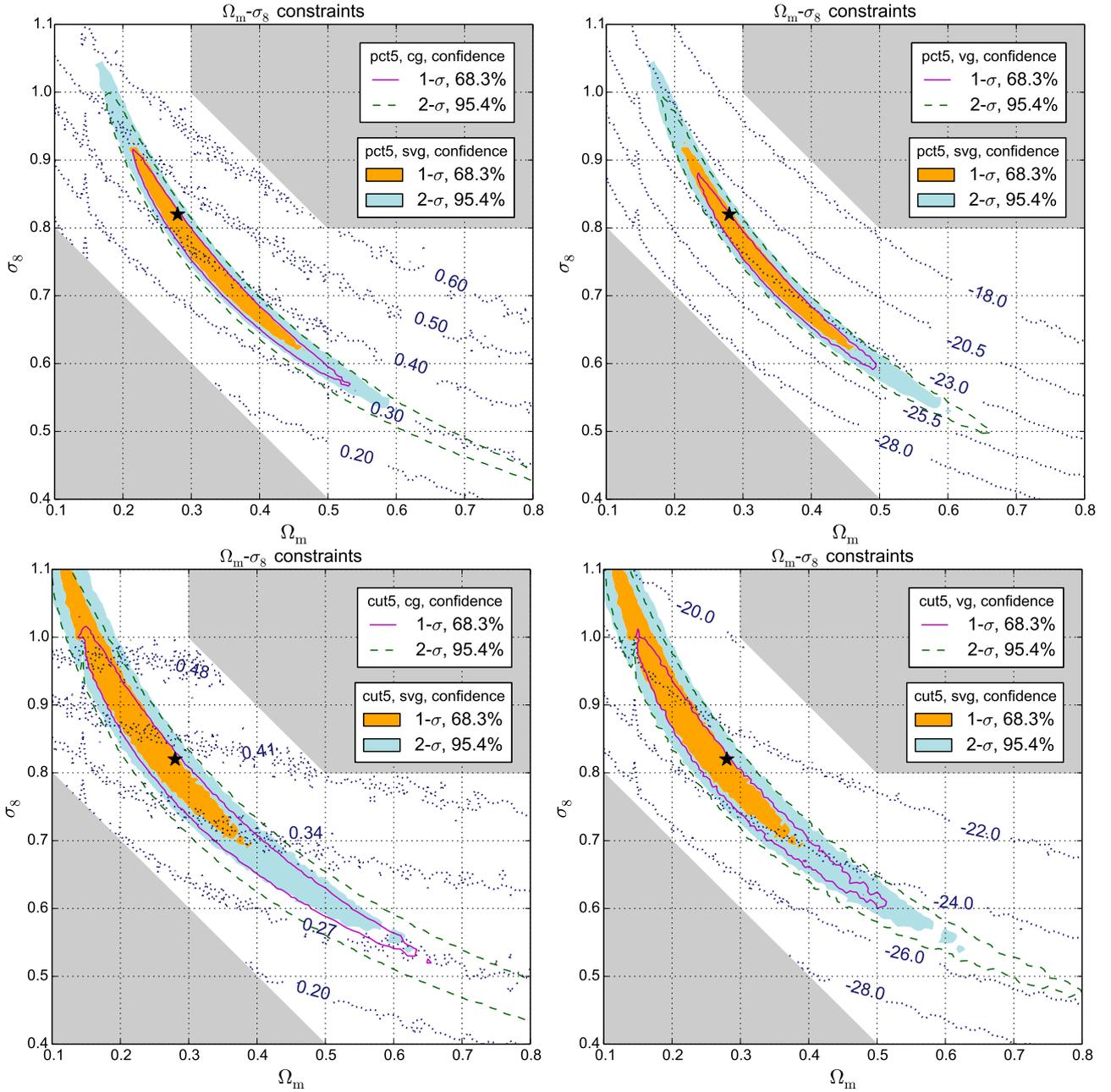
$$P(\mathbf{x}) = P'(\mathbf{q}) \frac{P_1(x_1) \cdots P_d(x_d)}{\phi_1(q_1) \cdots \phi_d(q_d)}. \quad (15)$$

We note that the marginal PDFs of  $P'$  are identical to a multivariate Gaussian distribution  $\phi$  with mean  $\boldsymbol{\mu}$  and covariance  $\mathbf{C}$ , where  $\mathbf{C}$  is the covariance matrix of  $\mathbf{x}$ . The PDF of  $\phi$  is given by

$$\phi(\mathbf{q}) \equiv \frac{1}{\sqrt{(2\pi)^d \det \mathbf{C}}} \exp\left[-\frac{1}{2} \sum_{i,j} (q_i - \mu_i) C_{ij}^{-1} (q_j - \mu_j)\right]. \quad (16)$$

Finally, by approximating  $P'$  to  $\phi$ , one gets the *Gaussian copula transform*:

$$P(\mathbf{x}) = \phi(\mathbf{q}) \frac{P_1(x_1) \cdots P_d(x_d)}{\phi_1(q_1) \cdots \phi_d(q_d)}. \quad (17)$$



**Fig. 4.** Similar to Fig. 3. Confidence regions with  $\mathbf{x}^{\text{pct5}}$  and  $\mathbf{x}^{\text{cut5}}$ . Both *upper panels* are drawn with  $\mathbf{x}^{\text{pct5}}$  and both *lower panels* with  $\mathbf{x}^{\text{cut5}}$ . Both *left panels* are the comparison between  $L_{\text{cg}}$  and  $L_{\text{svg}}$ , and both *right panels* between  $L_{\text{svg}}$  and  $L_{\text{vg}}$ .

**Table 5.** Best fits of  $(\Sigma_8, \alpha)$  from all analyses using the likelihood-ratio test and  $p$ -value analysis (confidence region).

|                   | $\mathbf{x}^{\text{abd5}}$ |          | $\mathbf{x}^{\text{pct5}}$ |          | $\mathbf{x}^{\text{cut5}}$ |          |
|-------------------|----------------------------|----------|----------------------------|----------|----------------------------|----------|
|                   | $\Sigma_8$                 | $\alpha$ | $\Sigma_8$                 | $\alpha$ | $\Sigma_8$                 | $\alpha$ |
| $L_{\text{cg}}$   | $0.831^{+0.016}_{-0.016}$  | 0.54     | $0.822^{+0.018}_{-0.019}$  | 0.54     | $0.800^{+0.030}_{-0.035}$  | 0.45     |
| $L_{\text{svg}}$  | $0.831^{+0.016}_{-0.015}$  | 0.52     | $0.820^{+0.015}_{-0.016}$  | 0.51     | $0.800^{+0.031}_{-0.023}$  | 0.40     |
| $L_{\text{vg}}$   | $0.829^{+0.015}_{-0.015}$  | 0.52     | $0.819^{+0.015}_{-0.016}$  | 0.52     | $0.800^{+0.024}_{-0.028}$  | 0.42     |
| $L_{\text{cc}}$   | $0.830^{+0.016}_{-0.016}$  | 0.54     | $0.825^{+0.018}_{-0.020}$  | 0.54     | $0.807^{+0.025}_{-0.031}$  | 0.46     |
| $L_{\text{vc}}$   | $0.829^{+0.016}_{-0.016}$  | 0.52     | $0.823^{+0.016}_{-0.019}$  | 0.53     | $0.798^{+0.029}_{-0.029}$  | 0.44     |
| $L_{\text{true}}$ | $0.828^{+0.018}_{-0.015}$  | 0.53     | $0.823^{+0.015}_{-0.020}$  | 0.53     | $0.800^{+0.028}_{-0.030}$  | 0.44     |
| $p$ -value        | $0.835^{+0.016}_{-0.019}$  | 0.54     | $0.823^{+0.018}_{-0.018}$  | 0.54     | $0.798^{+0.032}_{-0.034}$  | 0.45     |

**Notes.** The description of  $L_{\text{cg}}$ ,  $L_{\text{svg}}$ , and  $L_{\text{vg}}$  can be found in Sect. 3,  $L_{\text{cc}}$  and  $L_{\text{cg}}$  in Sect. 4, and  $L_{\text{true}}$  and  $p$ -value in Sect. 5. We note that the best fits for  $\Sigma_8$  are indicative since we do not use the real observational data in this study.

**Table 6.** Similar to Table 4, but for credible regions.

|                   | $\mathbf{x}^{\text{abd5}}$ |     | $\mathbf{x}^{\text{pct5}}$ |     | $\mathbf{x}^{\text{cut5}}$ |     |
|-------------------|----------------------------|-----|----------------------------|-----|----------------------------|-----|
|                   | $\Delta\Sigma_8$           | FoM | $\Delta\Sigma_8$           | FoM | $\Delta\Sigma_8$           | FoM |
| $L_{\text{cg}}$   | 0.033                      | 43  | 0.038                      | 31  | 0.066                      | 15  |
| $L_{\text{svg}}$  | 0.031                      | 53  | 0.033                      | 41  | 0.056                      | 20  |
| $L_{\text{vg}}$   | 0.031                      | 53  | 0.032                      | 40  | 0.055                      | 18  |
| $L_{\text{cc}}$   | 0.033                      | 40  | 0.040                      | 30  | 0.071                      | 14  |
| $L_{\text{vc}}$   | 0.033                      | 47  | 0.035                      | 36  | 0.060                      | 16  |
| $L_{\text{true}}$ | 0.034                      | 49  | 0.036                      | 36  | 0.061                      | 17  |
| ABC               | 0.056                      | 31  | 0.044                      | 33  | 0.068                      | 16  |

**Notes.** The description of ABC can be found in Sect. 6.

Why is it more accurate to calculate the likelihood in this way? In the classical case, since the shape of  $P(\mathbf{x})$  is unknown, we approximate it to a normal distribution:  $P(\mathbf{x}) \approx \phi(\mathbf{x})$ . Applying the Gaussian copula transform means that we carry out this approximation in the new space of  $\mathbf{q}$ :  $P'(\mathbf{q}) \approx \phi(\mathbf{q})$ . Since  $q_i = \Phi_i^{-1}(F_i(x_i))$ , at least the marginals of  $P'(\mathbf{q})$  are strictly Gaussian. And Eq. (17) gives the corresponding value in  $\mathbf{x}$  space, while taking  $P'(\mathbf{q}) \approx \phi(\mathbf{q})$  in  $\mathbf{q}$  space. However, in some cases, the copula has no effect at all. We consider  $f(x, y) = 2\phi_2(x, y)\Theta(xy)$  where  $\phi_2$  is the two-dimensional standard normal distribution, and  $\Theta$  is the Heaviside step function. The value of  $f$  is two times  $\phi_2$  if  $x$  and  $y$  have the same sign and 0 otherwise. The marginal PDF of  $f$  and  $\phi_2$  turn out to be the same. As a result, the Gaussian copula transform does nothing and  $f$  remains extremely non-Gaussian. However, if we do not have any prior knowledge, then the result with the copula transformation should be at least as good as the classical likelihood.

By applying Eq. (17) to  $P(\mathbf{x}^{\text{obs}}|\boldsymbol{\pi})$ , one gets the *copula likelihood*:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\pi}) &= \frac{1}{\sqrt{(2\pi)^d \det \mathbf{C}}} \\ &\times \exp \left[ -\frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d (q_i^{\text{obs}} - \mu_i) C_{ij}^{-1} (q_j^{\text{obs}} - \mu_j) \right] \\ &\times \prod_{i=1}^d \left[ \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp \left( -\frac{(q_i^{\text{obs}} - \mu_i)^2}{2\sigma_i^2} \right) \right]^{-1} \prod_{i=1}^d P_i(x_i^{\text{obs}}). \end{aligned} \quad (18)$$

In this paper,  $\mu_i = x_i^{\text{mod}}$ . Including the dependency on  $\boldsymbol{\pi}$  for all relevant quantities, the varying-covariance copula log-likelihood  $L_{\text{vc}}$  is given by

$$\begin{aligned} L_{\text{vc}} &\equiv \ln [\det \widehat{\mathbf{C}}(\boldsymbol{\pi})] \\ &+ \sum_{i=1}^d \sum_{j=1}^d (q_i^{\text{obs}}(\boldsymbol{\pi}) - x_i^{\text{mod}}(\boldsymbol{\pi})) \widehat{C}_{ij}^{-1}(\boldsymbol{\pi}) (q_j^{\text{obs}}(\boldsymbol{\pi}) - x_j^{\text{mod}}(\boldsymbol{\pi})) \\ &- 2 \sum_{i=1}^d \ln \hat{\sigma}_i(\boldsymbol{\pi}) - \sum_{i=1}^d \left( \frac{q_i^{\text{obs}}(\boldsymbol{\pi}) - x_i^{\text{mod}}(\boldsymbol{\pi})}{\hat{\sigma}_i(\boldsymbol{\pi})} \right)^2 \\ &- 2 \sum_{i=1}^d \ln \hat{P}_i(x_i^{\text{obs}}|\boldsymbol{\pi}). \end{aligned} \quad (19)$$

Here,  $\hat{P}_i(\cdot|\boldsymbol{\pi})$  is the  $i$ th marginal  $\boldsymbol{\pi}$ -dependent PDF that we estimate directly from the  $N$  samples  $x_i^{(k)}$ ,  $k = 1 \dots N$

already mentioned in Sect. 2.2, using the kernel density estimation (KDE):

$$\hat{P}_i(x_i) = \frac{1}{N} \sum_{k=1}^N \frac{1}{h_i} K \left( \frac{x_i - x_i^{(k)}}{h_i} \right), \quad (20)$$

where the kernel  $K$  is Gaussian, and the bandwidth  $h_i = (4/3N)^{1/5} \hat{\sigma}_i$  is given by Silverman's rule (Silverman 1986). These are one-dimensional PDF estimations, and the time cost is almost negligible. The term  $\hat{P}_i(x_i^{\text{obs}}|\boldsymbol{\pi})$  should be understood as a one-point evaluation of this function at  $x_i^{\text{obs}}$ . The quantities  $x_i^{\text{mod}}(\boldsymbol{\pi})$ ,  $\hat{\sigma}_i(\boldsymbol{\pi})$ , and  $\widehat{C}_{ij}^{-1}(\boldsymbol{\pi})$  are estimated with the same set following Eqs. (1)–(3). Finally,  $q_i^{\text{obs}}(\boldsymbol{\pi}) = \Phi_i^{-1}(\hat{F}_i(x_i^{\text{obs}}|\boldsymbol{\pi}))$ . We highlight that  $\hat{F}_i(\cdot|\boldsymbol{\pi})$  is the CDF that corresponds to  $\hat{P}_i(\cdot|\boldsymbol{\pi})$ , and  $\Phi_i$  also depends on  $\boldsymbol{\pi}$  via  $\mu_i$  and  $\hat{\sigma}_i$ .

We are also interested in studying the copula transform under the constant-covariance situation. In this case, we define the constant-covariance copula log likelihood  $L_{\text{cc}}$  as

$$\begin{aligned} L_{\text{cc}} &\equiv \sum_{i=1}^d \sum_{j=1}^d (q_i^{\text{obs}}(\boldsymbol{\pi}) - x_i^{\text{mod}}(\boldsymbol{\pi})) \widehat{C}_{ij}^{-1}(\boldsymbol{\pi}^{\text{in}}) (q_j^{\text{obs}}(\boldsymbol{\pi}) - x_j^{\text{mod}}(\boldsymbol{\pi})) \\ &- \sum_{i=1}^d \left( \frac{q_i^{\text{obs}}(\boldsymbol{\pi}) - x_i^{\text{mod}}(\boldsymbol{\pi})}{\hat{\sigma}_i(\boldsymbol{\pi}^{\text{in}})} \right)^2 - 2 \sum_{i=1}^d \ln \hat{P}_i(x_i^{\text{obs}} - x_i^{\text{mod}}(\boldsymbol{\pi})). \end{aligned} \quad (21)$$

Besides the constant covariance, we also suppose that the distribution of each  $x_i$  around its mean value does not vary with  $\boldsymbol{\pi}$ . In this case,  $\hat{P}_i(\cdot)$  denotes the zero-mean marginal PDF, and it is only estimated once from the 1000 realizations of  $\boldsymbol{\pi}^{\text{in}}$ , as are  $\hat{\sigma}_i$  and  $\widehat{C}_{ij}^{-1}$ . We recall that  $q_i^{\text{obs}}(\boldsymbol{\pi}) = \Phi_i^{-1}(\hat{F}_i(x_i^{\text{obs}} - x_i^{\text{mod}}(\boldsymbol{\pi})))$  where  $\Phi_i$  depends on  $\boldsymbol{\pi}$  implicitly via  $\mu_i$  and  $\hat{\sigma}_i$ .

## 4.2. Constraints using the copula

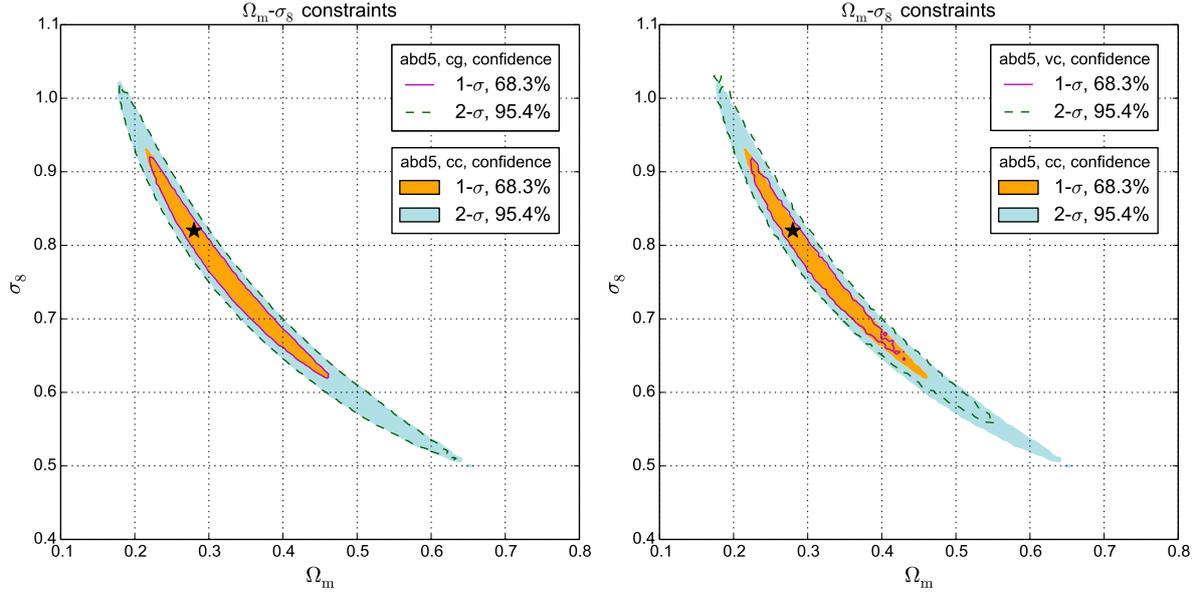
We again use the setting described in Sect. 2.2. We outline two interesting comparisons, which are shown in Fig. 5: between  $L_{\text{cg}}$  and  $L_{\text{cc}}$  in the lefthand panel and between  $L_{\text{cc}}$  and  $L_{\text{vc}}$  in the righthand one, both with  $\mathbf{x}^{\text{abd5}}$ . The lefthand panel shows that, for weak-lensing peak counts, the Gaussian likelihood is a very good approximation. Quantitative results, shown in Tables 4–7, reveal that the Gaussian likelihood provides slightly optimistic  $\Omega_m$ - $\sigma_8$  constraints. We would like to emphasize that the effect of the copula transform is ambiguous, and both tighter or wider constraints are possible. This has already been shown by Sato et al. (2011), who found that the Gaussian likelihood underestimates the constraint power for low  $\ell$  of the lensing power spectrum and overestimates it for high  $\ell$ .

In the righthand panel of Fig. 5, when the CDC effect is taken into account for the copula transform, the parameter constraints are submitted to a similar change to the Gaussian likelihood. Tighter constraints are obtained from  $L_{\text{vc}}$  than from  $L_{\text{cc}}$ . Similar results can be found for  $\mathbf{x}^{\text{pct5}}$  and  $\mathbf{x}^{\text{cut5}}$ . In summary, the copula with varying covariance,  $L_{\text{vc}}$  results in an FoM improvement of at least 10% compared to the Gaussian case with constant covariance,  $L_{\text{cg}}$ .

## 5. Non-analytic likelihood analyses

### 5.1. The true likelihood

In this section, we obtain the parameter constraints in a more direct way. Since our model predictions sample the full PDF, the



**Fig. 5.** Confidence regions derived from copula analyses. *Left panel:* comparison between contours from  $L_{cg}$  (solid and dashed lines) and  $L_{cc}$  (colored areas). *Right panel:* comparison between contours from  $L_{cc}$  (colored areas) and  $L_{vc}$  (solid and dashed lines). The evolution tendency from  $L_{cc}$  to  $L_{vc}$  is similar to the evolution from  $L_{cg}$  to  $L_{vg}$ .

**Table 7.** Similar to Table 5, but for Bayesian inference (credible region).

|            | $\mathbf{x}^{abd5}$       |          | $\mathbf{x}^{pct5}$       |          | $\mathbf{x}^{cut5}$       |          |
|------------|---------------------------|----------|---------------------------|----------|---------------------------|----------|
|            | $\Sigma_8$                | $\alpha$ | $\Sigma_8$                | $\alpha$ | $\Sigma_8$                | $\alpha$ |
| $L_{cg}$   | $0.831^{+0.017}_{-0.016}$ | 0.54     | $0.822^{+0.018}_{-0.020}$ | 0.54     | $0.800^{+0.030}_{-0.036}$ | 0.45     |
| $L_{svg}$  | $0.831^{+0.016}_{-0.015}$ | 0.52     | $0.820^{+0.016}_{-0.017}$ | 0.51     | $0.800^{+0.032}_{-0.024}$ | 0.40     |
| $L_{vg}$   | $0.829^{+0.015}_{-0.015}$ | 0.52     | $0.819^{+0.015}_{-0.017}$ | 0.52     | $0.800^{+0.025}_{-0.029}$ | 0.42     |
| $L_{cc}$   | $0.830^{+0.017}_{-0.017}$ | 0.54     | $0.825^{+0.018}_{-0.022}$ | 0.54     | $0.807^{+0.030}_{-0.041}$ | 0.46     |
| $L_{vc}$   | $0.829^{+0.016}_{-0.016}$ | 0.52     | $0.823^{+0.016}_{-0.019}$ | 0.53     | $0.798^{+0.030}_{-0.030}$ | 0.44     |
| $L_{true}$ | $0.828^{+0.019}_{-0.015}$ | 0.53     | $0.823^{+0.015}_{-0.021}$ | 0.53     | $0.800^{+0.030}_{-0.032}$ | 0.44     |
| ABC        | $0.819^{+0.030}_{-0.025}$ | 0.50     | $0.817^{+0.022}_{-0.022}$ | 0.51     | $0.799^{+0.034}_{-0.034}$ | 0.42     |

**Notes.** The description of ABC can be found in Sect. 6. We note that the best fits for  $\Sigma_8$  are indicative since we do not use the real observational data in this study.

PDF-Gaussianity assumption is no longer necessary. This allows us to go back to the true definition of the log-likelihood:

$$L_{true} \equiv -2 \ln \hat{P}(\mathbf{x}^{obs} | \boldsymbol{\pi}), \quad (22)$$

where  $\hat{P}$  is estimated from our  $N$  realizations  $\mathbf{x}^{(k)}$  ( $\boldsymbol{\pi}$ -dependent) using the kernel density estimation technique. The multivariate estimation is performed by

$$\hat{P}(\mathbf{x}) = \frac{1}{N} \sum_{k=1}^N K(\mathbf{x} - \mathbf{x}^{(k)}), \quad (23)$$

$$K(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d |\det \mathbf{H}|}} \exp \left[ -\frac{1}{2} \mathbf{x}^T \mathbf{H}^{-1} \mathbf{x} \right], \quad (24)$$

where

$$\sqrt{\mathbf{H}_{ij}} = \begin{cases} \left[ \frac{4}{(d+2)N} \right]^{\frac{1}{d+4}} \hat{\sigma}_i & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases} \quad (25)$$

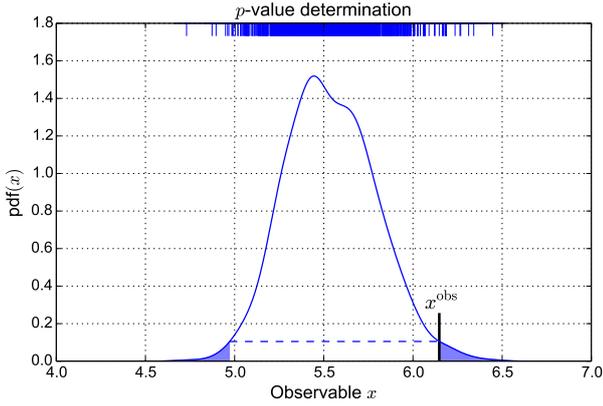
The evaluation of this non-analytic likelihood gets very noisy when the observable dimension  $d$  increases. In this case, a larger  $N$  will be required to stabilize the constraints. As in previous sections, we perform both the likelihood-ratio test and Bayesian inference with this likelihood.

## 5.2. $p$ -value analysis

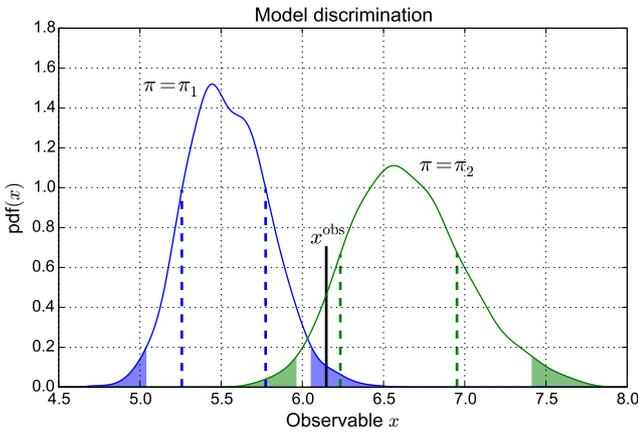
Another non-analytic technique is the  $p$ -value analysis. This frequentist approach provides the significance level by directly determining the  $p$ -value associated with a observation  $\mathbf{x}^{obs}$ . The  $p$ -value is defined as

$$p \equiv 1 - \int d^d \mathbf{x} \hat{P}(\mathbf{x} | \boldsymbol{\pi}) \times \Theta \left( \hat{P}(\mathbf{x} | \boldsymbol{\pi}) - \hat{P}(\mathbf{x}^{obs} | \boldsymbol{\pi}) \right), \quad (26)$$

where  $\Theta$  denotes the Heaviside step function. The integral extends over the region where  $\mathbf{x}$  is more probable than  $\mathbf{x}^{obs}$  for a given  $\boldsymbol{\pi}$ , as shown by Fig. 6. Thus, the interpretation of Eq. (26) is that if we generated  $N$  universes, then at least  $(1-p)N$  of them should have an observational result “better” than  $\mathbf{x}^{obs}$ . In this context, “better” refers to a more probable realization. The



**Fig. 6.** Example for the  $p$ -value determination. The  $x$ -axis indicates a one-dimensional observable, and the  $y$ -axis is the PDF. The PDF is obtained from a kernel density estimation using the  $N = 1000$  realizations. Their values are shown as bars in the rug plot at the top of the panel. The shaded area is the corresponding  $p$ -value for given observational data  $x^{\text{obs}}$ .



**Fig. 7.** PDF from two different models (denoted by  $\pi_1$  and  $\pi_2$ ) and the observation  $x^{\text{obs}}$ . The dashed lines show  $1\sigma$  intervals for both models, while the shaded areas are intervals beyond the  $2\sigma$  level. In this figure, model  $\pi_1$  is excluded at more than  $2\sigma$ , whereas the significance of the model  $\pi_2$  is between 1 and  $2\sigma$ .

significance level is determined by the chi-squared distribution with  $d = 2$  deg of freedom, for two free parameters,  $\Omega_m$  and  $\sigma_8$ . As Fig. 7 shows, this provides a straightforward way to distinguish different cosmological models.

As in Sect. 4.2, we used KDE to estimate the multivariate PDF and numerically integrated Eq. (26) to obtain the  $p$ -value. Monte Carlo integration is used for evaluating the five-dimensional integrals.

### 5.3. Parameter constraints

Figure 8 shows the confidence contours from  $L_{\text{true}}$  and  $p$ -value analysis with observables  $x^{\text{abd5}}$ . We notice that these constraints are very noisy. This is due to a relatively low number of realizations to estimate the probability and prevents us from making definite conclusions. Nevertheless, the result from the left-hand panel reveals good agreement between constraints from two likelihoods. This suggests that we may substitute  $L_{\text{true}}$  with the CDC-copula likelihood to bypass the drawback of noisy estimation from  $L_{\text{true}}$ . In the righthand panel, the result from  $p$ -value analysis seems to be larger. We reduced the noise for the  $p$ -value

analysis by combining  $x^{\text{abd5}}$  into a two-component vector. In this case, the  $p$ -value is evaluated using the grid integration. This data compression technique does not significantly enlarge but visibly smooths the contours.

In the  $L_{\text{true}}$  case, the probability information that we need is local since the likelihood  $P$  is only evaluated at  $x^{\text{obs}}$ . For  $p$ -value analysis, one needs to determine the region where  $P(x) < P(x^{\text{obs}})$  and integrate over it, so a more global knowledge of  $P(x)$  is needed in this case. We recall that KDE smooths, thus the estimation is always biased (Zamboni & Dias 2012). Other estimations, for example using the Voronoi-Delaunay tessellation (see, e.g., Schaap 2007; Guio & Achilleos 2009), could be an alternative to the KDE technique. As a result, observable choice, data compression, and density estimation need to be considered jointly for all non-analytic approaches.

Recent results from CFHTLenS (LPH15) and Stripe-82 (LPL15) resulted in  $\Delta\Sigma_8 \sim 0.1$ , about 2–3 times larger than this study. However, we would like to highlight that redshift errors are not taken into account here and that the simulated galaxy density used in this work is much higher. Also, we choose  $z_s = 1$ , which is higher than the median redshift of both surveys ( $\sim 0.75$ ). All these factors contribute to our smaller error bars.

## 6. Approximate Bayesian computation

### 6.1. PMC ABC algorithm

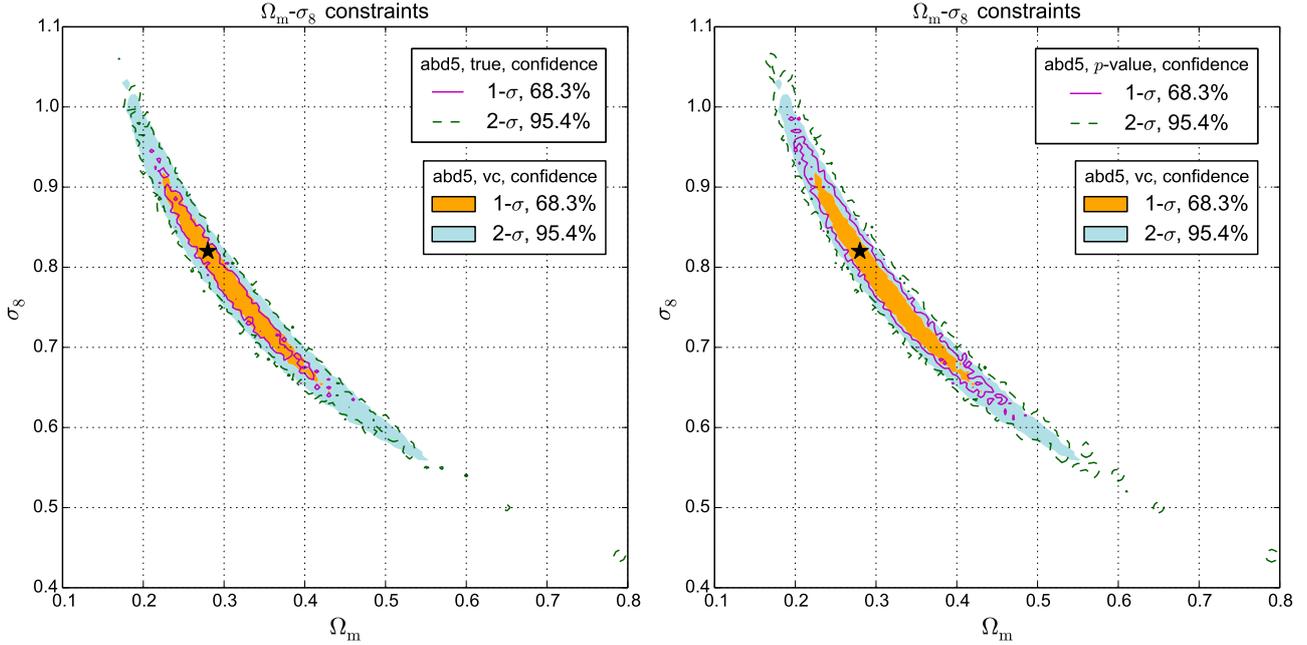
In the previous section, we presented parameter constraints derived from directly evaluating the underlying PDF. Now, we want to move a step further and bypass the likelihood estimation altogether.

Based on an accept-reject rule, approximate Bayesian computation (ABC) is an algorithm that provides an approximate posterior distribution of a complex stochastic process when evaluating the likelihood is expensive or unreachable. There are only two requirements: (1) a stochastic model for the observed data that samples the likelihood function of the observable; and (2) a measure, called summary statistic, to perform model comparison. We present below a brief description of ABC. Readers can find detailed reviews of ABC in Marin et al. (2011) and Sect. 1 of Cameron & Pettitt (2012).

The idea behind ABC can be most easily illustrated in the case of discrete data as follows. Instead of explicitly calculating the likelihood, one first generates a set of parameters  $\{\pi_i\}$  as samples under the prior  $\mathcal{P}(\pi)$ , and then for each  $\pi_i$  simulates a model prediction  $X$  sampled under the likelihood function  $P(\cdot|\pi_i)$ . (Here we put  $X$  in the upper case to emphasize that  $X$  is a random variable.) Keeping only those  $\pi_i$  for which  $X = x^{\text{obs}}$ , the distribution of the accepted samples  $\mathcal{P}_{\text{ABC}}(\pi)$  equals the posterior distribution of the parameter  $\mathcal{P}(\pi|x^{\text{obs}})$  given the observed data, since

$$\begin{aligned} \mathcal{P}_{\text{ABC}}(\pi) &= \sum_X P(X|\pi)\mathcal{P}(\pi)\delta_{X,x^{\text{obs}}} \\ &= P(x^{\text{obs}}|\pi)\mathcal{P}(\pi) \\ &= \mathcal{P}(\pi|x^{\text{obs}}), \end{aligned} \quad (27)$$

where  $\delta_{X,x^{\text{obs}}}$  is Kronecker's delta. Therefore,  $\{\pi_i\}$  is an independent and identically distributed sample from the posterior. It is sufficient to perform a *one-sample test*: using a single realization  $X$  for each parameter  $\pi$  to obtain a sample under the posterior.



**Fig. 8.** *Left panel:* confidence regions derived from  $L_{vc}$  (colored areas) and  $L_{true}$  (solid and dashed lines) with  $\mathbf{x}^{abd5}$ . *Right panel:* confidence regions derived from  $L_{vc}$  (colored areas) and  $p$ -value analysis (solid and dashed lines). The contours from  $L_{true}$  and  $p$ -value analysis are noisy due to a relatively low  $N$  for probability estimation. We notice that  $L_{vc}$  and  $L_{true}$  yield very similar results.

ABC can also be adapted to continuous data and parameters, where obtaining a strict equality  $\mathbf{X} = \mathbf{x}^{obs}$  is practically impossible. As a result, sampled points are accepted with a *tolerance level*  $\epsilon$ , say  $|\mathbf{X} - \mathbf{x}^{obs}| \leq \epsilon$ . What is retained after repeating this process is an ensemble of parameters  $\boldsymbol{\pi}$  that are compatible with the data and that follow a probability distribution, which is a modified version of Eq. (27),

$$\mathcal{P}_\epsilon(\boldsymbol{\pi}|\mathbf{x}^{obs}) = A_\epsilon(\boldsymbol{\pi})\mathcal{P}(\boldsymbol{\pi}), \quad (28)$$

where  $A_\epsilon(\boldsymbol{\pi})$  is the probability that a proposed parameter  $\boldsymbol{\pi}$  passes the one-sample test within the error  $\epsilon$ :

$$A_\epsilon(\boldsymbol{\pi}) \equiv \int d\mathbf{X} P(\mathbf{X}|\boldsymbol{\pi}) \mathbb{1}_{|\mathbf{X} - \mathbf{x}^{obs}| \leq \epsilon}(\mathbf{X}). \quad (29)$$

The Kronecker delta from Eq. (27) has now been replaced with the indicator function  $\mathbb{1}$  of the set of points  $\mathbf{X}$  that satisfy the tolerance criterion. The basic assumption of ABC is that the probability distribution (28) is a good approximation of the underlying posterior, such that

$$\mathcal{P}_\epsilon(\boldsymbol{\pi}|\mathbf{x}^{obs}) \approx \mathcal{P}(\boldsymbol{\pi}|\mathbf{x}^{obs}). \quad (30)$$

Therefore, the error can be separated into two parts: one from the approximation above and the other from the estimation of the desired integral,  $\mathcal{P}_\epsilon$ . For the latter, gathering one-sample tests of  $A_\epsilon$  makes a Monte Carlo estimation of  $\mathcal{P}_\epsilon$ , which is unbiased. This ensures the use of the one-sample test.

A further addition to the ABC algorithm is a reduction in the complexity of the full model and data. This is necessary in cases of very large dimensions, for example, when the model produces entire maps or large catalogs. The reduction of data complexity is done with the so-called *summary statistic*  $s$ . For instance, in our peak-count framework, a complete data set  $\mathbf{x}$  is a peak catalog with positions and S/N values, and the summary statistic  $s$  is chosen here to be  $s(\mathbf{x}) = \mathbf{x}^{abd5}$ ,  $\mathbf{x}^{pct5}$ , or  $\mathbf{x}^{cut5}$ , respectively, for the three cases of observables introduced in Sect. 2.2. As a remark, if this summary statistic is indeed *sufficient*, then Eq. (30)

will no longer be an approximation. The true posterior can be recovered when  $\epsilon \rightarrow 0$ .

For a general comparison of model and data, one chooses a metric  $D$  adapted to the summary statistic  $s$ , and the schematic expression  $|\mathbf{X} - \mathbf{x}^{obs}| \leq \epsilon$  used above is generalized to  $D(s(\mathbf{X}), s(\mathbf{x}^{obs})) \leq \epsilon$ . We highlight that the summary statistic can have a low dimension and a very simple form. In practice, it is motivated by computational efficiency, and it seems that a simple summary can still produce reliable constraints (Weyant et al. 2013).

The integral of Eq. (28) over  $\boldsymbol{\pi}$  is smaller than unity, and the deficit only represents the probability that a parameter is rejected by ABC. This is not problematic since density estimation will automatically normalize the total integral over the posterior. However, a more subtle issue is the choice of the tolerance level  $\epsilon$ . If  $\epsilon$  is too high,  $A(\boldsymbol{\pi})$  is close to 1, and Eq. (30) becomes a bad estimate. If  $\epsilon$  is too low,  $A(\boldsymbol{\pi})$  is close to 0, and sampling becomes extremely difficult and inefficient. How, then, should one choose  $\epsilon$ ? This can be done by applying the iterative importance sampling approach of population Monte Carlo (PMC; for applications to cosmology, see Wraith et al. 2009) and combine it with ABC (Del Moral et al. 2006; Sisson et al. 2007). This joint approach is sometimes called SMC ABC, where SMC stands for sequential Monte Carlo; we refer to it as PMC ABC. The idea of PMC ABC is to iteratively reduce the tolerance  $\epsilon$  until a stopping criterion is reached.

Algorithm 1 details the steps for PMC ABC. We let  $\text{pdf}(\mathbf{x}|\boldsymbol{\pi})$  be a probabilistic model for  $\mathbf{x}$  given  $\boldsymbol{\pi}$ . PMC ABC requires a prior  $\rho(\boldsymbol{\pi})$ , a summary statistic  $s(\mathbf{x})$  that retains only partial information about  $\mathbf{x}$ , a distance function  $D$  based on the dimension of  $s(\mathbf{x})$ , and a shutoff parameter  $r_{stop}$ . We denote  $\mathcal{N}(\boldsymbol{\pi}, \mathbf{C})$  as a multivariate normal with mean  $\boldsymbol{\pi}$  and covariance matrix  $\mathbf{C}$ ,  $K(\boldsymbol{\pi}, \mathbf{C}) \propto \exp(-\boldsymbol{\pi}^T \mathbf{C}^{-1} \boldsymbol{\pi} / 2)$  a Gaussian kernel,  $\text{cov}(\boldsymbol{\pi}_i, w_i)$  the weighted covariance for the set  $\{\boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_p\}$  with weights  $\{w_1, \dots, w_p\}$ , and  $p$  the number of *particles*, i.e., the number of sample points in the parameter space.

---

**Algorithm 1** Population Monte Carlo approximate Bayesian computation

---

**Require:**

number of particles  $p$   
prior  $\rho(\cdot)$   
summary statistic  $s(\cdot)$   
distance  $D(\cdot, \cdot)$   
shutoff parameter  $r_{\text{stop}}$

set  $t = 0$

**for**  $i = 1$  to  $p$  **do**

generate  $\pi_i^{(0)}$  from  $\rho(\cdot)$  and  $\mathbf{x}$  from pdf  $(\cdot|\pi_i^{(0)})$

set  $\delta_i^{(0)} = D(s(\mathbf{x}), s(\mathbf{x}^{\text{obs}}))$  and  $w_i^{(0)} = 1/p$

**end for**

set  $\epsilon^{(1)} = \text{median}(\delta_i^{(0)})$  and  $\mathbf{C}^{(0)} = \text{cov}(\pi_i^{(0)}, w_i^{(0)})$

**while** success rate  $\geq r_{\text{stop}}$  **do**

$t \leftarrow t + 1$

**for**  $i = 1$  to  $p$  **do**

**repeat**

generate  $j$  from  $\{1, \dots, p\}$  with weights  $\{w_1^{(t-1)}, \dots, w_p^{(t-1)}\}$

generate  $\pi_i^{(t)}$  from  $\mathcal{N}(\pi_j^{(t-1)}, \mathbf{C}^{(t-1)})$  and  $\mathbf{x}$  from pdf  $(\cdot|\pi_i^{(t)})$

set  $\delta_i^{(t)} = D(s(\mathbf{x}), s(\mathbf{x}^{\text{obs}}))$

**until**  $\delta_i^{(t)} \leq \epsilon^{(t)}$

set  $w_i^{(t)} \propto \frac{\rho(\pi_i^{(t)})}{\sum_{j=1}^p w_j^{(t-1)} K(\pi_i^{(t)} - \pi_j^{(t-1)}, \mathbf{C}^{(t-1)})}$

**end for**

set  $\epsilon^{(t+1)} = \text{median}(\delta_i^{(t)})$  and  $\mathbf{C}^{(t)} = \text{cov}(\pi_i^{(t)}, w_i^{(t)})$

**end while**

---

In the initial step, PMC ABC accepts all particles drawn from the prior and defines an acceptance tolerance before starting the first iteration. The tolerance is given by the median of the distances of the summary statistic between the observation and the stochastic model generated from each particle. Then, each iteration is carried out by an importance-sampling step based on weights determined by the previous iteration. To find a new particle, a previous point  $\pi_j^{(t-1)}$  is selected according to its weight. A candidate particle is drawn from a proposal distribution, which is a normal law centered on  $\pi_j^{(t-1)}$  with a covariance equal to the covariance of all particles from the previous iteration. With a model generated using the candidate particle, we accept the new particle if the distance between the model and the observation is shorter than the tolerance, and reject it otherwise. After accepting  $p$  particles, the success rate, defined as the ratio of accepted particles to total tries, is updated. The iterations continue until the success rate decreases below the shutoff value. Instead of defining a minimal tolerance (Beaumont et al. 2002; Fearnhead & Prangle 2010; Weyant et al. 2013), we use a simplified stopping criterion that is based on the selection efficiency of the algorithm. Since McKinley et al. (2009) prove that the stopping criterion has very little impact on the estimated posterior, the choice of the tolerance level is instead a question of computational power.

## 6.2. Settings for ABC

McKinley et al. (2009) studied the impact of the various choices necessary for ABC, by comparing Markov chain Monte Carlo (MCMC) ABC and PMC ABC. The authors concluded that (1) increasing the number of simulations beyond one for a given

parameter does not seem to improve the posterior estimation (similar conclusion found by Bornn et al. 2014); (2) the specific choice of the tolerance level does not seem to be important; (3) the choice of the summary statistic and the distance is crucial; and (4) PMC ABC performs better than MCMC ABC. Therefore, exploring a sufficient summary statistic to represent the whole data set becomes an essential concern for the ABC technique.

To solve the optimal data compression problem, Blum et al. (2013) provide a series of methods in their review for selecting the ideal summaries, as well as methods of reducing the data dimension. Leaving a detailed study of optimal choice for the future work, we adopted a straightforward summary statistic in this work, defined as  $s(\mathbf{x}) = \mathbf{x}^{\text{type5}}$ , and the distance as

$$D(\mathbf{x}^{\text{type5}}, \mathbf{y}^{\text{type5}}) = \sqrt{\sum_{i=1}^5 \frac{(x_i^{\text{type5}} - y_i^{\text{type5}})^2}{C_{ii}}}. \quad (31)$$

This is simply a weighted Euclidean distance, where the weight  $C_{ii}^{-1}$  is needed to level out the values of the different S/N.

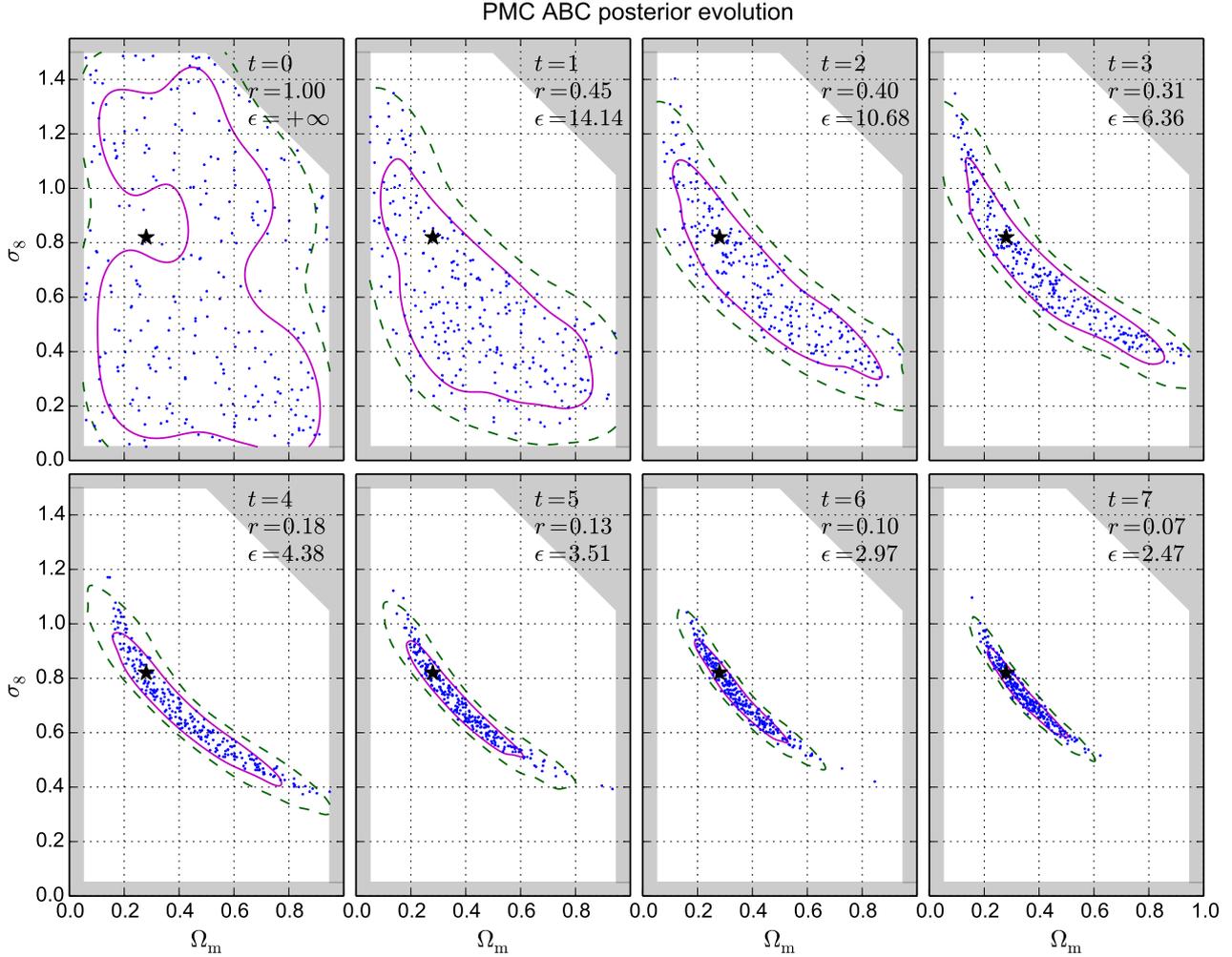
The prior  $\rho$  is chosen to be flat. We set  $p = 250$  and  $r_{\text{stop}} = 0.02$ . In this condition, we can easily compare the computational cost with the analyses presented in previous sections. If  $\tau$  is the time cost for one model realization, the total time consumption is  $7821 \times 1000 \times \tau$  for our likelihood-based analyses in Sects. 3–5, and  $250 \times 1 \times (\sum_t r_t^{-1}) \times \tau$  for ABC where  $r_t$  is the acceptance rate of the  $t$ -th iteration. For  $\mathbf{x}^{\text{abd5}}$ ,  $\sum_t r_t^{-1} \approx 102$  ( $0 \leq t \leq 9$ ), so the computation time for ABC is drastically reduced by a factor of  $\sim 300$  compared to the likelihood analyses. ABC is faster by a similar factor compared to Monte-Carlo sampling, since typically the number of required sample points is  $O(10^4)$ , the same order of magnitude as our number of grid points.

## 6.3. Results from ABC

Figure 9 shows the iterative evolution of the PMC ABC particles. We drew the position of all 250 particles and credible regions for the first eight iterations. The summary statistic is  $\mathbf{x}^{\text{abd5}}$ . The credible regions were drawn from the posterior estimated on a grid using KDE with the ABC particles as sample points  $\mathbf{x}^{(k)}$  in Eq. (23). We ignored the particle weights for this density estimate. We find that the contours stabilize for  $t \geq 8$ , which corresponds to an acceptance rate of  $r = 0.05$ . At these low acceptance rates, corresponding to a small tolerance, the probability of satisfying the tolerance criterion  $D(s(\mathbf{X}), s(\mathbf{x}^{\text{obs}})) \leq \epsilon$  is low even though  $\mathbf{X}$  is sampled from parameters in the high-probability region, and accepting a proposed particle depends mainly on random fluctuations due to the stochasticity of the model.

In Fig. 10, we show the weights of particles sampled at the final iteration  $t = 8$ . The weight is visualized by both color and size of the circle. The figure shows that points farther away from the maximum have larger weights as constructed by Algorithm 1. Since those points are more isolated, their weights compensate for the low density of points, avoiding undersampling the tails of the posterior and overestimating the constraining power.

In Fig. 11 we show the comparison of credible regions between  $L_{\text{cg}}$  and PMC ABC with  $s(\mathbf{x}) = \mathbf{x}^{\text{abd5}}$ . The FoM and the best-fit ABC results are presented in Tables 6 and 7, respectively. The figure shows good agreement between the two cases, and thus validates the performance of PMC ABC. The broader contours from ABC might be caused by a bias of KDE. The same



**Fig. 9.** Evolution of particles and the posterior from the PMC ABC algorithm. We show results from the first 8 iterations ( $0 \leq t \leq 7$ ). Particles are given by blue dots. Solid lines are  $1\sigma$  contours, and dashed lines are  $2\sigma$  contours. White areas represent the prior. The corresponding accept rate  $r$  and tolerance level  $\epsilon$  are also given. We set  $\epsilon^{(0)} = +\infty$ .

reason might be responsible for the slight shift of the contours in the tails of the distribution, which do not follow the particles exactly, which are best visible in the two lefthand panels in the lower row of Fig. 9.

## 7. Summary and discussion

Our model for weak-lensing peak counts, which provides a direct estimation of the underlying PDF of observables, leads to a wide range of possibilities for constraining parameters. To summarize this work, we

- compared different data vector choices;
- studied the dependence of the likelihood on cosmology;
- explored the full PDF information of observables;
- proposed different constraint strategies; and
- examined them with two criteria.

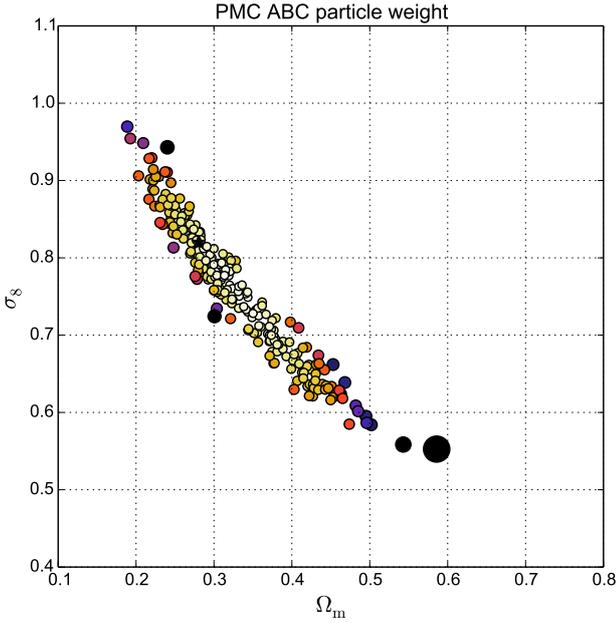
In this paper, we performed three different series of analyses – the Gaussian likelihood, the copula likelihood, and non-analytic analyses – by using three different data vectors: one based on the peak PDF and two on the CDF. We defined two quantitative criteria:  $\Delta\Sigma_8$ , which represents the error bar on the parameter  $\Sigma_8 = \sigma_8(\Omega_m/0.27)^\alpha$  and is a measure of the width of the

$\Omega_m$ - $\sigma_8$  degeneracy direction; and FoM, which is the area of the  $\Omega_m$ - $\sigma_8$  contour. Both Bayesian and frequentist approaches were followed. Although the interpretations are different, the results are very similar.

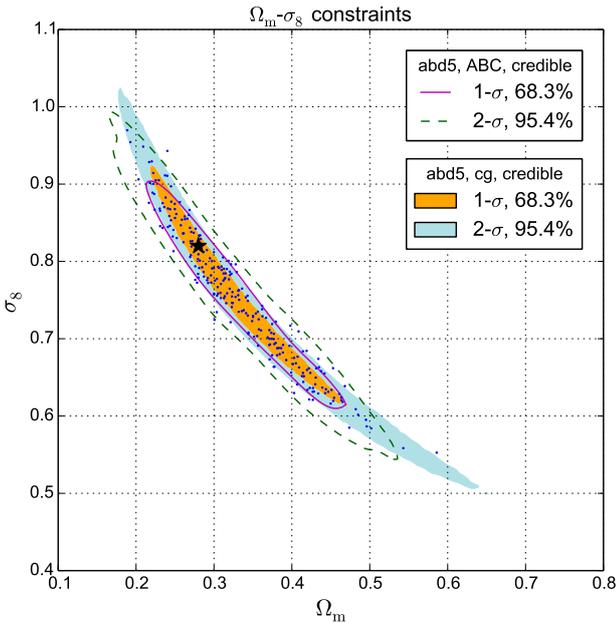
We studied the cosmology-dependent-covariance (CDC) effect by estimating the true covariance for each parameter set. We found that the CDC effect can increase the constraining power up to 22%. The main contribution comes from the additional variation of the  $\chi^2$  term, and the contribution from the determinant term is negligible. These observations confirm a previous study by Eifler et al. (2009).

We also performed a copula analysis, which makes weaker assumptions than Gaussianity. In this case, the marginalized PDF is Gaussianized by the copula transform. The result shows that the difference with the Gaussian likelihood is small. This is dominated by the CDC effect if a varying covariance is taken into account.

Discarding the Gaussian hypothesis on the PDF of observables, we provided two straightforward ways of using the full PDF information. The first one is the true likelihood. The direct evaluation of the likelihood is noisy owing to the high statistical fluctuations from the finite number of sample points. However, we find that the varying-covariance copula likelihood, noted as  $L_{vc}$  above, seems to be a good approximation to the truth. The



**Fig. 10.** Weights of particles from  $t = 8$  with  $s(x) = x^{\text{abd5}}$ . The weight is represented by the size and the color at the same time.



**Fig. 11.** Comparison between credible regions derived from  $L_{cg}$  (colored areas) and ABC (solid and dashed lines).

second method is to determine the  $p$ -value for a given parameter set directly, and this approach gives us more conservative constraints. We outline that both methods are covariance-free, avoiding non-linear effects caused by the covariance inversion.

At the end we showed how approximate Bayesian computation (ABC) derives cosmological constraints using the accept-reject sampling. Combined with importance sampling, this method requires less computational resources than all the others. We proved that by reducing the computational time by a factor of 300, ABC is able to yield consistent constraints from weak-lensing peak counts. Furthermore, Weyant et al. (2013) show in their study that ABC is able to perform unbiased constraints using contaminated data, demonstrating the robustness of this algorithm.

A comparison between different data vectors is done in this study. Although we find for all analyses that  $x^{\text{abd5}}$  outperforms  $x^{\text{pct5}}$  by 20–40% in terms of FoM, this is not necessarily true in general when we use a different percentile choice. Actually, the performance of  $x^{\text{pct}}$  depends on the correlation between its different components. However, the  $x^{\text{pct}}$  family is not recommended in practice because of model biases induced for very low peaks ( $S/N < 0$ ). In addition, our study shows that the  $x^{\text{cut}}$  family is largely outperformed by  $x^{\text{abd}}$ . Thus, we conclude that  $x^{\text{abd}}$  seems to be good candidates for peak-count analysis, while the change in the contour tilt from  $x^{\text{cut}}$  could be interesting when combining with other information.

The methodology that we show for parameter constraints can be applied to all fast stochastic forward models. Flexible and efficient, this approach possesses a great potential whenever the modeling of complex effects is desired. Our study displays two different parameter-constraint philosophies. On the one hand, parametric estimation (Sects. 3 and 4), under some specific hypotheses such as Gaussianity, only requires some statistical quantities such as the covariances. However, the appropriateness of the likelihood should be examined and validated to avoid biases. On the other hand, non-analytic estimation (Sects. 5 and 6) is directly derived from the PDF. The problem of inappropriateness vanishes, but instead the uncertainty and bias of density estimation become drawbacks. Depending on modeling pertinence, an aspect may be more advantageous than another. Although not studied in this work, a hybrid approach using semi-analytic estimator could be interesting. This solicits more detailed studies of trade-off between the inappropriateness of analytic estimators and the uncertainty of density estimation.

*Acknowledgements.* This work is supported by Région d’Île-de-France in the framework of a DIM-ACAV thesis fellowship. We also acknowledge the support from the French national program for cosmology and galaxies (PNCG). The authors wish to acknowledge the anonymous referee for reviewing the paper. Chieh-An Lin would like to thank Karim Benabed, Ewan Cameron, Yu-Yen Chang, Yen-Chi Chen, Cécile Chenot, François Lanusse, Sandrine Pires, Fred Ngolè, Florent Sureau, and Michael Vespe for useful discussions and suggestions on diverse subjects.

## References

- Akeret, J., Refregier, A., Amara, A., Seehars, S., & Hasner, C. 2015, *JCAP*, **08**, 043
- Albrecht, A., Bernstein, G., Cahn, R., et al. 2006, ArXiv e-prints [[arXiv:astro-ph/0609591](https://arxiv.org/abs/astro-ph/0609591)]
- Bard, D., Kratochvil, J. M., Chang, C., et al. 2013, *ApJ*, **774**, 49
- Beaumont, M. A., Zhang, W., & Balding, D. J. 2002, *Genetics*, **162**, 2025
- Beaumont, M. A., Cornuet, J.-M., Marin, J.-M., & Robert, C. P. 2009, *Biometrika*, **96**, 983
- Benabed, K., Cardoso, J.-F., Prunet, S., & Hivon, E. 2009, *MNRAS*, **400**, 219
- Bergé, J., Amara, A., & Réfrégier, A. 2010, *ApJ*, **712**, 992
- Berger, J. O., Fienberg, S. E., Raftery, A. E., & Robert, C. P. 2010, *Proc. National Academy Sci.*, **107**, 157
- Blum, M. G. B., Nunes, M. A., Prangle, D., & Sisson, S. A. 2013, *Statistical Science*, **28**, 189
- Bornn, L., Pillai, N., Smith, A., & Woodard, D. 2014, ArXiv e-prints [[arXiv:1404.6298](https://arxiv.org/abs/1404.6298)]
- Cameron, E., & Pettitt, A. N. 2012, *MNRAS*, **425**, 44
- Casella, G., & Berger, R. L. 2002, *Statistical Inference* (Duxbury)
- Csilléry, K., Blum, M. G., Gaggiotti, O. E., & François, O. 2010, *Trends in Ecology & Evolution*, **25**, 410
- Del Moral, P., Doucet, A., & Jasra, A. 2006, *J. Roy. Statist. Soc.: Series B (Statistical Methodology)*, **68**, 411
- Dietrich, J. P., & Hartlap, J. 2010, *MNRAS*, **402**, 1049
- Drovandi, C. C., & Pettitt, A. N. 2011, *Biometrics*, **67**, 225
- Eifer, T., Schneider, P., & Hartlap, J. 2009, *A&A*, **502**, 721
- Fan, Z., Shan, H., & Liu, J. 2010, *ApJ*, **719**, 1408 (FSL10)
- Fearnhead, P., & Prangle, D. 2010, ArXiv e-prints [[arXiv:1004.1112](https://arxiv.org/abs/1004.1112)]
- Fu, L., Kilbinger, M., Erben, T., et al. 2014, *MNRAS*, **441**, 2725

- Guio, P., & Achilleos, N. 2009, *MNRAS*, **398**, 1254
- Hamana, T., Takada, M., & Yoshida, N. 2004, *MNRAS*, **350**, 893
- Hamana, T., Sakurai, J., Koike, M., & Miller, L. 2015, *PASJ*, **67**, 34
- Hartlap, J., Simon, P., & Schneider, P. 2007, *A&A*, **464**, 399
- Hennawi, J. F., & Spergel, D. N. 2005, *ApJ*, **624**, 59
- Ishida, E. E. O., Vitenti, S. D. P., Penna-Lima, M., et al. 2015, *Astron. Comput.*, **13**, 1
- Jain, B., & Seljak, U. 1997, *ApJ*, **484**, 560
- Jain, B., & Van Waerbeke, L. 2000, *ApJ*, **530**, L1
- Jee, M. J., Tyson, J. A., Schneider, M. D., et al. 2013, *ApJ*, **765**, 74
- Jenkins, A., Frenk, C. S., White, S. D. M., et al. 2001, *MNRAS*, **321**, 372
- Jiang, I.-G., Yeh, L.-C., Chang, Y.-C., & Hung, W.-L. 2009, *AJ*, **137**, 329
- Kainulainen, K., & Marra, V. 2009, *Phys. Rev. D*, **80**, 123020
- Kainulainen, K., & Marra, V. 2011a, *Phys. Rev. D*, **83**, 023009
- Kainulainen, K., & Marra, V. 2011b, *Phys. Rev. D*, **84**, 063004
- Kilbinger, M., & Schneider, P. 2005, *A&A*, **442**, 69
- Kilbinger, M., Fu, L., Heymans, C., et al. 2013, *MNRAS*, **430**, 2200
- Killedar, M., Borgani, S., Fabjan, D., et al. 2015, *MNRAS*, submitted [[arXiv:1507.05617](https://arxiv.org/abs/1507.05617)]
- Kratochvil, J. M., Haiman, Z., & May, M. 2010, *Phys. Rev. D*, **81**, 043519
- Kruse, G., & Schneider, P. 1999, *MNRAS*, **302**, 821
- Lin, C.-A., & Kilbinger, M. 2015, *A&A*, **576**, A24 (Paper I)
- Liu, J., Haiman, Z., Hui, L., Kratochvil, J. M., & May, M. 2014, *Phys. Rev. D*, **89**, 023515
- Liu, J., Petri, A., Haiman, Z., et al. 2015, *Phys. Rev. D*, **91**, 063507 (LPH15)
- Liu, X., Wang, Q., Pan, C., & Fan, Z. 2014, *ApJ*, **784**, 31
- Liu, X., Pan, C., Li, R., et al. 2015, *MNRAS*, **450**, 2888 (LPL15)
- Maoli, R., Van Waerbeke, L., Mellier, Y., et al. 2001, *A&A*, **368**, 766
- Marian, L., Smith, R. E., & Bernstein, G. M. 2010, *ApJ*, **709**, 286
- Marian, L., Hilbert, S., Smith, R. E., Schneider, P., & Desjacques, V. 2011, *ApJ*, **728**, L13
- Marian, L., Smith, R. E., Hilbert, S., & Schneider, P. 2012, *MNRAS*, **423**, 1711
- Marian, L., Smith, R. E., Hilbert, S., & Schneider, P. 2013, *MNRAS*, **432**, 1338
- Marin, J.-M., Pudlo, P., Robert, C. P., & Ryder, R. J. 2011, ArXiv e-prints [[arXiv:1101.0955](https://arxiv.org/abs/1101.0955)]
- Maturi, M., Angrick, C., Pace, F., & Bartelmann, M. 2010, *A&A*, **519**, A23
- Maturi, M., Fedeli, C., & Moscardini, L. 2011, *MNRAS*, **416**, 2527
- McKinley, T., Cook, A. R., & Deardon, R. 2009, *Int. J. Biostatistics*, **5**, A24
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, *ApJ*, **462**, 563
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, *ApJ*, **490**, 493
- Osato, K., Shirasaki, M., & Yoshida, N. 2015, *ApJ*, **806**, 186
- Petri, A., Liu, J., Haiman, Z., et al. 2015, *Phys. Rev. D*, **91**, 103511
- Pires, S., Leonard, A., & Starck, J.-L. 2012, *MNRAS*, **423**, 983
- Pires, S., Starck, J.-L., Amara, A., Réfrégier, A., & Teyssier, R. 2009, *A&A*, **505**, 969
- Robin, A. C., Reylé, C., Fliri, J., et al. 2014, *A&A*, **569**, A13
- Rubin, D. B. 1984, *Ann. Statist.*, **12**, 1151
- Rüschendorf, L. 2009, *J. Stat. Planning and Inference*, **139**, 3921
- Sato, M., Ichiki, K., & Takeuchi, T. T. 2011, *Phys. Rev. D*, **83**, 023501
- Schaap, W. E. 2007, Ph.D. Thesis, Kapteyn Astronomical Institute
- Scherrer, R. J., Berlind, A. A., Mao, Q., & McBride, C. K. 2010, *ApJ*, **708**, L9
- Schneider, P., & Lombardi, M. 2003, *A&A*, **397**, 809
- Soccimarro, R., Sefusatti, E., & Zaldarriaga, M. 2004, *Phys. Rev. D*, **69**, 103513
- Semboloni, E., Schrabback, T., Van Waerbeke, L., et al. 2011, *MNRAS*, **410**, 143
- Silverman, B. W. 1986, Density estimation for statistics and data analysis (Chapman & Hall)
- Simon, P., Semboloni, E., Van Waerbeke, L., et al. 2015, *MNRAS*, **449**, 1505
- Sisson, S. A., Fan, Y., & Tanaka, M. M. 2007, *Proc. National Academy of Sciences*, **104**, 1760
- Sklar, A. 1959, *Publ. Inst. Statist. Univ. Paris*, **8**, 229
- Takada, M., & Jain, B. 2003, *ApJ*, **583**, L49
- Takeuchi, T. T. 2010, *MNRAS*, **406**, 1830
- Weyant, A., Schafer, C., & Wood-Vasey, W. M. 2013, *ApJ*, **764**, 116
- Wraith, D., Kilbinger, M., Benabed, K., et al. 2009, *Phys. Rev. D*, **80**, 023507
- Yang, X., Kratochvil, J. M., Wang, S., et al. 2011, *Phys. Rev. D*, **84**, 043529
- Yang, X., Kratochvil, J. M., Huffenberger, K., Haiman, Z., & May, M. 2013, *Phys. Rev. D*, **87**, 023511
- Zambom, A. Z., & Dias, R. 2012, ArXiv e-prints [[arXiv:1212.2812](https://arxiv.org/abs/1212.2812)]