

Assessing theoretical uncertainties for cosmological constraints from weak lensing surveys

Journal Article

Author(s):

Tan, Ting; Zürcher, Dominik ; Fluri, Janis ; Refregier, Alexandre; Tarsitano, Federica; Kacprzak, Tomasz

Publication date:

2023-07

Permanent link:

<https://doi.org/10.3929/ethz-b-000614383>

Rights / license:

[Creative Commons Attribution 4.0 International](#)

Originally published in:

Monthly Notices of the Royal Astronomical Society 522(3), <https://doi.org/10.1093/mnras/stad1142>

Funding acknowledgement:

192243 - Cosmological Weak Lensing and Neutral Hydrogen (SNF)

Assessing theoretical uncertainties for cosmological constraints from weak lensing surveys

Ting Tan,^{1,2★} Dominik Zürcher^{1,2}, Janis Fluri,^{2,3} Alexandre Refregier,² Federica Tarsitano^{1,2} and Tomasz Kacprzak^{1,2}

¹*Sorbonne Université, CNRS/IN2P3, Laboratoire de Physique Nucléaire et de Hautes Energies, LPNHE, 4 Place Jussieu, F-75252 Paris, France*

²*Institute for Particle Physics and Astrophysics, Department of Physics, ETH Zürich, Wolfgang Pauli Strasse 27, CH-8093 Zürich, Switzerland*

³*Data Analytics Lab, Department of Computer Science, ETH Zurich Universitätstrasse 6, CH-8006 Zürich, Switzerland*

Accepted 2023 April 6. Received 2023 April 6; in original form 2022 July 7

ABSTRACT

Weak gravitational lensing is a powerful probe, which is used to constrain the standard cosmological model and its extensions. With the enhanced statistical precision of current and upcoming surveys, high-accuracy predictions for weak lensing statistics are needed to limit the impact of theoretical uncertainties on cosmological parameter constraints. For this purpose, we present a comparison of the theoretical predictions for the non-linear matter and weak lensing power spectra, based on the widely used fitting functions (`mead` and `rev-halofit`), emulators (`EuclidEmulator`, `EuclidEmulator2`, `BaccoEmulator`, and `CosmicEmulator`), and N -body simulations (`PKDGRAV3`). We consider the forecasted constraints on the Λ CDM and w CDM models from weak lensing for stage III and stage IV surveys. We study the relative bias on the constraints and their dependence on the assumed prescriptions. Assuming a Λ CDM cosmology, we find that the relative agreement on the S_8 parameter is between 0.2 and 0.3σ for a stage III-like survey between the above predictors. For a stage IV-like survey the agreement becomes 1.4 – 3.0σ . In the w CDM scenario, we find broader S_8 constraints, and agreements of 0.18 – 0.26σ and 0.7 – 1.7σ for stage III and stage IV surveys, respectively. The accuracies of the above predictors therefore appear adequate for stage III surveys, whereas the fitting functions would need improvements for future stage IV surveys. Furthermore, we find that, of the fitting functions, `mead` provides the best agreement with the emulators. We discuss the implication of these findings for the preparation of future weak lensing surveys, and the relative impact of theoretical uncertainties to other systematics.

Key words: gravitational lensing: weak – large-scale structure of Universe – Cosmological parameters.

1 INTRODUCTION

The next generation of wide field cosmological surveys, such as the Vera Rubin Observatory Legacy Survey of Space and Time (LSST¹; Abell et al. 2009), *Euclid*,² and the *Nancy Grace Roman Space Telescope* (NGRST³; Akesson et al. 2019) will map the matter distribution of the local Universe with an unprecedented accuracy. These high-precision measurements present a challenge for the theoretical modelling of cosmological observables. Cosmic shear is a cosmological observable that relies on the distortions of galaxy shapes caused by weak gravitational lensing (e.g. Bartelmann & Schneider 2001). This effect is due to the gravitational deflection of photons by the matter density field along the line of sight. Cosmic shear measures the inhomogeneities in the cosmic density field with high precision and can be used as an unbiased tracer of the matter distribution. It is sensitive to both, the matter distribution of the Universe and the growth of cosmic structure, which is important

for the understanding of the expansion history of the Universe. A commonly used cosmic shear summary statistic is the cosmic shear angular power spectrum, which can be predicted from the matter power spectrum. The modelling of the matter power spectrum on large scales can be derived using perturbation theory (Bernardeau et al. 2002; Crocce & Scoccimarro 2006; Baumann et al. 2012; Crocce, Scoccimarro & Bernardeau 2012; Blas, Garny & Konstandin 2014; Blas et al. 2016; Foreman & Senatore 2016; Nishimichi, Bernardeau & Taruya 2016; Beutler et al. 2017; Cataneo et al. 2019; d’Amico et al. 2020), where the structure formation of the Universe is linear. Some extended perturbation theories (e.g. Chudaykin et al. 2020; D’Amico, Senatore & Zhang 2021) can provide an accurate model up to $k \sim 0.3 h \text{ Mpc}^{-1}$. However, at non-linear, smaller scales, non-linear processes have a strong impact on the matter power spectrum, and perturbation theory is no longer valid.

In this work, we compare the theoretical predictions of the non-linear matter power spectrum, and the associated theoretical uncertainties on cosmological parameters from measurements of the cosmic shear angular power spectrum. The comparison includes some widely used models fitted from N -body simulations using analytical halo models: `halofit` (Smith et al. 2003) is fitted to low resolution, gravity-only N -body simulations, which is known to exhibit a non-

* E-mail: ting.tan@lpnhe.in2p3.fr

¹<https://www.lsst.org>.

²<https://www.cosmos.esa.int/web/euclid/home>.

³<https://roman.gsfc.nasa.gov/>.

negligible mismatch with current state-of-the-art hydrodynamic N -body simulations; `rev-halofit` (Takahashi et al. 2012), developed as the revisited version of `halofit` is used in the analysis of the Dark Energy Survey (DES; Amon et al. 2022); and `mead` (Mead et al. 2015), which is used in the analysis of the Kilo-Degree Survey combined with the VISTA Kilo-Degree Infrared Galaxy Survey (Giblin et al. 2021). Apart from the halo model fitting method, emulators are generated from the interpolation of a suite of N -body simulations, e.g. `CosmicEmulator` (Heitmann et al. 2009, 2013; Lawrence et al. 2017), `BaccoEmulator` (Angulo et al. 2020; Aricò et al. 2021), `EuclidEmulator` (Knabenhans et al. 2019) and its updated version `EuclidEmulator2` (Collaboration et al. 2020), `COSMOPower` (Mancini et al. 2022), and `GP emulator` (Giblin et al. 2019). In this study, `CosmicEmulator`, `BaccoEmulator`, `EuclidEmulator`, and `EuclidEmulator2` are representatively selected in the comparison at the level of the matter power spectrum, and a comparison between `rev-halofit`, `mead` and `EuclidEmulator` is also shown in Knabenhans et al. (2021). In order to estimate the theoretical uncertainties, we look at the weak lensing cosmological parameter constraints, by generating a forecast for a stage III, DES-like survey and a stage IV, *Euclid*-like survey. We take into account the parameters described by the standard Λ CDM cosmological model and the extended w CDM model. As a further investigation, we also discuss the relative impact of theoretical uncertainties compared to other systematics, such as baryonic effects, photometric redshift uncertainty (e.g. Huterer et al. 2006), shear bias (e.g. Bernstein & Jarvis 2002; Hirata et al. 2004), and galaxy intrinsic alignment (e.g. Heavens, Refregier & Heymans 2000).

This paper is organized as follows. In Section 2, we describe the theoretical framework, including three halo-model based fitting functions, `mead`, `halofit`, and `rev-halofit`; four power spectrum emulators extracted from N -body simulations: `CosmicEmulator`, `BaccoEmulator`, `EuclidEmulator` and `EuclidEmulator2`, and one N -body simulation code `PKDGRAV3` (Potter, Stadel & Teyssier 2017). In Section 3, we present the method and the relevant codes used in this study. We summarize our results in Section 4 and our conclusions in Section 5.

2 THEORY

In this section, we describe the theoretical background of the matter power spectrum, weak lensing, and its angular power spectrum, as well as the different predictors of the matter power spectrum that we include in the comparisons.

2.1 Weak lensing

Considering the cosmic density field $\rho(\vec{r})$ at the position \vec{r} , the density contrast $\delta(\vec{r})$ is defined as the relative difference of $\rho(\vec{r})$ to the average density $\bar{\rho}$

$$\delta(\vec{r}) = \frac{\rho(\vec{r}) - \bar{\rho}}{\bar{\rho}}. \quad (1)$$

In Fourier space, the density contrast takes the following form:

$$\delta(\vec{k}) = \int \delta(\vec{r}) \exp(i\vec{k} \cdot \vec{r}) d^3r. \quad (2)$$

Furthermore, the matter power spectrum $P(\vec{k})$ is defined as the correlation of the density contrast in Fourier space (Peebles 2020):

$$\langle \delta(\vec{k}) \delta(\vec{k}') \rangle = (2\pi)^3 \delta_{\text{D}}^{(3)}(\vec{k} + \vec{k}') P(\vec{k}), \quad (3)$$

where δ_{D} is the three-dimensional Dirac delta function. For full-sky surveys, the cosmic shear angular power spectrum is approximately identical to the convergence power spectrum (Bartelmann & Maturi 2016), which can be defined as a weighted integration along the line-of-sight over the matter power spectrum (Bartelmann & Schneider 2001), and simplified using the Kaiser–Limber approximation (Limber 1953; Kaiser 1992, 1998; LoVerde & Afshordi 2008). We follow the formalism of LoVerde & Afshordi (2008), Giannantonio et al. (2012), Kilbinger et al. (2017), Kitching et al. (2017), and Tarsitano et al. (2021) to compute the cross-correlated shear power spectrum with tomographic redshift bins i and j :

$$C_{\gamma}^{ij}(\ell) = \frac{9}{16} \left(\frac{H_0}{c} \right)^4 \Omega_{\text{m}}^2 \int_0^{\chi_{\text{h}}} d\chi P_{\text{NL}} \left(\frac{\ell}{r}, \chi \right) \frac{g_i(\chi) g_j(\chi)}{(\text{ar}(\chi))^2}. \quad (4)$$

Here P_{NL} is the non-linear matter power spectrum, χ is the comoving distance, χ_{h} is the comoving horizon distance, r is the comoving angular diameter distance, Ω_{m} is the total matter density, $a = (1+z)^{-1}$ is the scale factor, and $g(\chi)$ is the lensing efficiency function defined as:

$$g_i(\chi) = 2 \int_{\chi}^{\chi_{\text{h}}} d\chi' n_i(\chi') \frac{r(\chi) r(\chi' - \chi)}{r(\chi')}, \quad (5)$$

with $n_i(\chi)$ being the normalized number density of the observed galaxies at a comoving distance χ .

2.2 Matter power spectrum

The matter power spectrum is a fundamental statistics to study the large-scale structure of the Universe. As seen above, it is, in particular, useful to predict the cosmic shear angular power spectrum. Therefore, it is necessary to have an accurate theoretical model for the matter power spectrum on all scales. On large scales and mildly non-linear scales, the matter power spectrum can be modelled using perturbation theory and some extended theories. On small scales, which are in the non-linear regime, these approaches are not suited to predict the power spectrum with the necessary precision, while other methods are developed with the use of a halo model or simulations.

2.2.1 Analytical predictions

A common way to model the matter power spectrum on these small scales is to empirically fit physically motivated formulas to measurements from N -body simulations, e.g. as done in Hamilton et al. (1991). Furthermore, modelling the density field as a collection of virialized haloes, the matter power spectrum can be approximated analytically using the statistics of haloes, and fitted to simulations or emulators (Ma & Fry 2000; Seljak 2000; Cooray & Sheth 2002).

In this study, we compare three halo-model based fitting functions: `mead`, `halofit`, and `rev-halofit`. `halofit` was built using a series of N -body simulations with a total of $N = 256^3$ particles and the box size from 84 to 240 Mpc h^{-1} . Using the halo model, the matter power spectrum is constructed with two terms, the one-halo term proposed by Ma & Fry (2000), Peacock & Smith (2000), Seljak (2000), Scoccimarro et al. (2001) and a two-halo term (Ma & Fry 2000; Seljak 2000; Scoccimarro et al. 2001) to describe the exclusion effects between dark matter haloes. The one-halo term indicates the correlation of the matter field of one single halo, which dominates on small scales, whereas the two-halo term describes the cross-correlation between different haloes, which has a strong impact on larger scales. Assuming that the haloes are distributed according to the halo mass function (Press & Schechter 1974; Sheth & Tormen 1999), the matter power spectrum modelled with this

approach can achieve a high precision on large scales. However, due to the lack of baryons and the relatively low resolution of the N -body simulations used in their study, `halofit` does not match high-resolution N -body simulations, giving an accuracy at the 5 per cent level at $k = 1 h \text{Mpc}^{-1}$ (Heitmann et al. 2010), and larger differences for $k > 1 h \text{Mpc}^{-1}$, which is insufficient for the non-linear regime. `rev-halofit` is a revised prescription of `halofit`, which provides a more accurate prediction of the matter power spectrum for $k < 30 h \text{Mpc}^{-1}$ and $z < 10$, with a 5 per cent level accuracy at $k = 1 h \text{Mpc}^{-1}$ and 10 per cent level accuracy at $k = 10 h \text{Mpc}^{-1}$. `rev-halofit` uses high-resolution N -body simulations for 16 cosmological models around the Wilkinson Microwave Anisotropy Probe (WMAP) best-fitting cosmological parameters. The N -body simulations were run with the `Gadget-2` N -body code (Springel, Yoshida & White 2001; Springel 2005a), 1024^3 particles in total, and the box size from 320 to $2000 h^{-1} \text{Mpc}$. The power spectrum is fitted using an improved fitting formula with five more model parameters as compared to `halofit`. Several extended methods have been proposed to improve the halo model (Bird, Viel & Haehnelt 2012; Mohammed & Seljak 2014; Seljak & Vlah 2015). Here we only consider `mead` (Mead et al. 2015), which reaches an accuracy at the 5 per cent level for $k = 10 h \text{Mpc}^{-1}$ and $z < 2$. `mead` introduces more physical parameters in addition to the halo model, and is fitted to the ‘Coyote Universe’ (Heitmann et al. 2013) suite of high-resolution simulations, the same simulations used for the generation of `CosmicEmulator`. It also includes massive neutrinos (Mead et al. 2016) and baryonic effects e.g. active galactic nuclei (AGNs) feedback, supernovae explosions, and gas cooling. However, we only consider the dark-matter-only case in this study.

2.2.2 Emulators

The fitting functions based on halo models described in Section 2.2.1 can provide accurate non-linear power spectrum predictions for large k -modes and a wide redshift range, which can be used to predict cosmological observables. However, they also have limitations as the precision is not uniform for different cosmological parameters, and it is difficult for fitting functions to give a high precision below the 1 per cent level compared to high-resolution simulations. Power spectrum emulators are constructed following a different approach in which one interpolates the power spectrum from a set of N -body simulations within a certain range of relevant parameters, using interpolation methods, e.g. Gaussian processes regression (Heitmann et al. 2010, 2013; Angulo et al. 2020) or polynomial chaos expansion (Knabenhans et al. 2019; Collaboration et al. 2020). Compared to fitting functions, emulators usually provide consistent precision of the predictions for different k -modes. However, emulators also have limitations: First, the covered parameter space is limited, thus making it difficult to perform a likelihood analysis, for which one needs to explore a wide range of parameter values. Secondly, the ranges of k and redshift are also limited, making it difficult to compute the weak lensing cosmic shear observables for high ℓ s, which require an integration over a large k range.

In this study, we compare four emulators: `CosmicEmulator` (Heitmann et al. 2016), `BaccoEmulator` (Angulo et al. 2020), `EuclidEmulator` (Knabenhans et al. 2019), and `EuclidEmulator2` (Collaboration et al. 2020), which are selected as representatives for different interpolation methods, i.e. `CosmicEmulator` using Gaussian processes regression, `EuclidEmulator` using polynomial chaos expansion, and `BaccoEmulator` using Neural network, and Gaussian processes regression. `CosmicEmulator`

is fitted using a set of the ‘Coyote Universe’ simulations and the ‘Mira-Titan Universe’ (Lawrence et al. 2017) simulations. We use the latest version of the emulator (Heitmann et al. 2016), for which the ‘Mira-Titan Universe’ simulations were run with 3200^3 particles and a simulation volume of $(2100 h^{-1} \text{Mpc})^3$. The `CosmicEmulator` successfully achieves high-precision predictions of the power spectrum within the 4 per cent level for $k_{\text{max}} = 5 h \text{Mpc}^{-1}$ and $z < 2$. It allows for the variation of various parameters, including the matter density Ω_m , the amplitude of density fluctuations σ_8 , the baryon density Ω_b , the scalar spectral index n_s , the dark energy equation of state parameters w_0 , and w_a , the dimensionless Hubble parameter h , the neutrino density Ω_ν , and the redshift z . `EuclidEmulator` uses a different emulation method using N -body simulations generated with the `PKDGRAV3` code (Potter et al. 2017). It uses 100 simulations with 2048^3 particles in a $(1250 h^{-1} \text{Mpc})^3$ simulation volume. The non-linear correction is encoded as a boost factor adding up to the input linear power spectrum, achieving a precision at the 1 per cent level for predictions within the ranges $k < 1 h \text{Mpc}^{-1}$ and $z < 1$. Knabenhans et al. (2019) demonstrated that `EuclidEmulator` agrees with `rev-halofit` at the 8 per cent level. As an updated version of `EuclidEmulator`, `EuclidEmulator2` is extended with dynamical dark energy and massive neutrinos, created with a larger parameter space and a modified version of the `PKDGRAV3` N -body code. `EuclidEmulator2` provides a consistent accuracy with simulations at the 2 per cent level up to $k_{\text{max}} = 10 h \text{Mpc}^{-1}$ for $z < 2$, and slightly lower accuracy for higher redshift $z \sim 3$. However, as `EuclidEmulator2` uses the amplitude of the primordial power spectrum A_s instead of σ_8 as input parameter, we use the following formula to transfer σ_8 into A_s (Hand et al. 2018):

$$A_s = \left(\frac{\sigma_8}{\sigma_{8,0}} \right)^2 \times A_{s,0} \quad (6)$$

in our comparison, where $\sigma_{8,0} = 0.826$ and $A_{s,0} = 2.184 \times 10^{-9}$.

`BaccoEmulator` is another state-of-the-art emulator using an updated version of the `L-Gadget3` code (Springel 2005b; Angulo et al. 2012) with 4320^3 particles in a $(1440 h^{-1} \text{Mpc})^3$ simulation volume. It has a 2 per cent level accuracy over the redshift range $0 < z < 1.5$ and $k < 5 h \text{Mpc}^{-1}$.

2.2.3 N -body simulations

We also include in this study a comparison with a dark-matter-only N -body simulation run with `PKDGRAV3`, which is based on a binary tree algorithm. This code uses fifth order multipole expansions of the gravitational potential between particles and can achieve fast computational speeds with hardware acceleration. A comparison between `PKDGRAV3` and the N -body codes, `Gadget-3`, `Gadget-4`, and `Ramses` is presented in Schneider et al. (2016) and Springel et al. (2021). The `PKDGRAV3` simulations are the same as the ones used for `EuclidEmulator`, with 2048^3 particles in total and the box size of $L = 1250 h^{-1} \text{Mpc}$. The details are presented in Knabenhans et al. (2019).

3 METHOD

In this work, we perform a comparison of predictors of the non-linear matter power spectrum, i.e. halo-model based fitting functions and emulators. We estimate the theoretical uncertainties of these predictors on the parameter constraint level by looking at the weak lensing cosmological parameter constraints from a stage III and a stage IV surveys. For each survey, we perform a comparison using

Table 1. Parameter settings for mock surveys: The stage IV survey is created using a four times larger survey area and galaxy density compared to the stage III survey. A deeper Smail redshift distribution is also used in the stage IV survey.

Survey	Stage III	Stage IV
Survey area [deg ²]	5000	20 000
Galaxy density [arcmin ⁻²]	5	20
Redshift distribution	Smail	Smail
Redshift bins	4	4
Redshift range ⁴	0.025 ~ 3.0	0.025 ~ 3.0

⁴The presented redshift range refers to the considered range used in the generation process of the covariance matrix for the mock surveys. The range differs from the redshift range used for the predictors of the weak lensing power spectrum in Section 4.2, where we use [0.08,2] for the stage III survey and [0.08,3.0] for the stage IV survey.

the standard Λ CDM cosmological model and the extended w CDM model.

3.1 Survey parameters

The estimate of the theoretical uncertainties for cosmological parameters is realized by forecasting the constraints for a stage III and a stage IV surveys. The covariance matrix is estimated from simulations, as described in Section 3.2 below. Table 1 shows the parameter settings used for the generation of the mock galaxy surveys. Martinelli et al. (2021) suggests using $\ell_{\max} = 5000$ for stage IV-like surveys to probe deep into non-linear regime. However, in this study we use a more conservative limit of $\ell_{\max} = 1000$, and do not take into account baryonic effects.

We use Smail et al. (1995) distributions to model the global redshift distribution of the source galaxies for both the stage III and the stage IV surveys. The corresponding formulas and parameter settings for these two distributions are as follows:

$$n(z)_{\text{stageIII}} = z^\alpha \exp \left[- \left(\frac{z}{z_0} \right)^\beta \right], \quad (7)$$

with $\alpha = 1.5$, $\beta = 1.1$, and $z_0 = 0.31$ and

$$n(z)_{\text{stageIV}} = \left(\frac{z}{z_0} \right)^\alpha \exp \left[- \left(\frac{z}{z_0} \right)^\beta \right], \quad (8)$$

with $\alpha = 2.0$, $\beta = 1.5$, and $z_0 = 0.64$ (Martinelli et al. 2021). In both cases the source galaxies are randomly divided into four tomographic bins with equal number of galaxies in each bin, and a Gaussian convolution is performed so that they follow the schema in Amara & Réfrégier (2007). The four tomographic bins are chosen to reduce the computation time and for simplicity. This is a conservative choice for the estimation of theoretical uncertainty, but could be enough for a forecast comparison. As a result of the auto- and cross-combinations of these four redshift bins, we have 10 combinations of auto- and cross-correlations for the cosmic shear measurements (four auto-correlations and six cross-correlations). Fig. 1 shows the global and tomographic redshift distributions used in this study.

3.2 Covariance matrix

An accurate estimate of the survey covariance matrix is crucial for the correct calculation of the likelihood function. We estimate the covariance matrices for the stage III and stage IV survey setups described in Table 1 from numerical simulations, using the NGSF

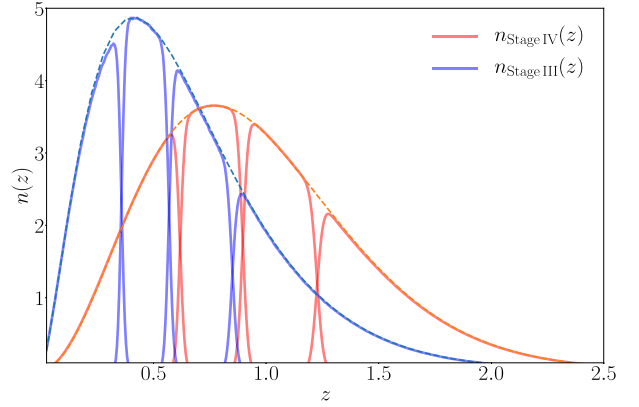


Figure 1. The redshift distributions of the source galaxies. One can see the four tomographic distributions for the stage III and the stage IV surveys. The global distributions, which follows the Smail et al. (1995) model, are shown by the dashed lines.

code described in Zürcher et al. (2021) and Dominiket et al. (2022). We generate a large number ($N = 2000$) of realization of the angular power spectra for each survey setup following the methodology outlined in Zürcher et al. (2021). In the following, we introduce the used N -body simulations, briefly summarize the forward modelling procedure used to generate the angular power spectra and describe the estimation of the covariance matrix. We refer the reader to Zürcher et al. (2021) for a more detailed description of the methodology.

We utilize the 50 independent PKDGRAV3 (Potter et al. 2017) N -body simulations at the fiducial cosmology that were previously used in Dominiket et al. (2022); Zürcher et al. (2021) and generated using the state-of-the-art dark-matter-only N -body code PKDGRAV3. The cosmological parameters in the used simulations are fixed to the (Λ CDM, TT, TE, EE + lowE + lensing) results of Planck 2018 (Aghanim et al. 2020), except for Ω_m and σ_8 which are set to the values found in Troxel et al. (2018). This setup results in $\Omega_{\text{cdm}} = 0.26$, $\sigma_8 = 0.84$, $\Omega_b = 0.0493$, $n_s = 0.9649$, $w = -1$, and $h = 0.6736$. We include three massive neutrino species in all simulations. The neutrinos are modelled as a relativistic fluid (Tram et al. 2019) and a degenerate mass hierarchy with a minimal neutrino mass of $m_\nu = 0.02$ eV per species was chosen. The dark energy density Ω_Λ is adapted for each cosmology to achieve a flat geometry.

Each simulation was run using a unit box with a side-length of $900 \text{ Mpc } h^{-1}$ and 768^3 simulated particles. In order to achieve a simulation volume large enough to cover the redshift range up to $z = 3.0$ the unit box was replicated up to 14 times per dimension depending on the cosmology. While such a replication scheme is known to underpredict the variance of very large, superbox modes (Fluri et al. 2019), it has been demonstrated by Dominiket et al. (2022) that the simulations accurately recover the angular power spectra predicted by the theory code CLASS (Lesgourgues 2011) for $\ell \in [30, 2048]$.

The particle shells from each PKDGRAV3 simulation are combined into tomographic full-sky mass maps using the UFALCON software (Sgier et al. 2019). The particle shells are weighted according to the tomographic redshift distributions shown in Fig. 1. The UFALCON software uses the HEALPIX (Gorski et al. 2005) pixelization scheme to pixelize the sphere. A resolution of $N_{\text{SIDE}} = 1024$ was chosen. UFALCON also makes use of the Born approximation, which is known to deteriorate the accuracy of the produced mass maps. However, Petri, Haiman & May (2017) have demonstrated that the introduced bias is negligible for stage III-like and stage IV-like surveys.

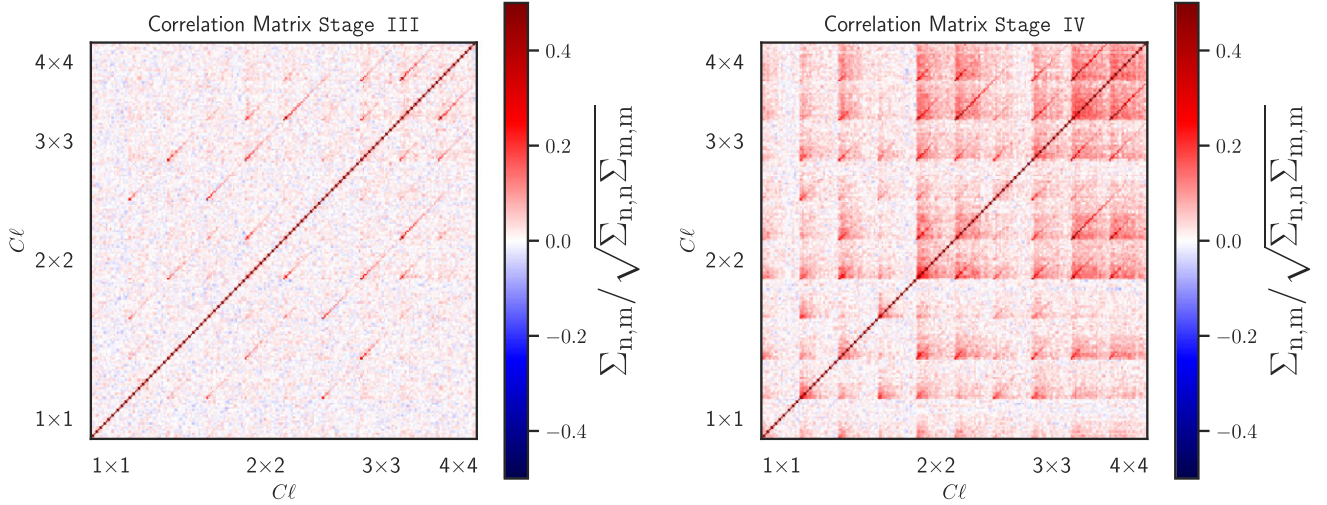


Figure 2. Correlation matrices for the stage III survey (left-hand panel) and the stage IV survey (right-hand panel). The ordering of the redshift tomographic bin combinations for the angular power spectra is 1×1 , 1×2 , 1×3 , 1×4 , 2×2 , 2×3 , 2×4 , 3×3 , 3×4 , and 4×4 , from left to right. For each angular power spectrum, all 20 bins ranging from $\ell = 100$ to $\ell = 1000$ are shown.

The spherical Kaiser–Squires mass mapping technique (Kaiser & Squires 1993; Wallis et al. 2022) is used to obtain the cosmic shear signal from the simulated mass maps. To forward-model a realistic weak lensing survey a shape noise signal must then be added to the cosmic shear signal and an appropriate survey mask must be applied. The survey masks are regularly chosen such that we obtain eight stage III surveys and two stage IV surveys from each full-sky map.

The shape noise signal is obtained in the same way as described in Zürcher et al. (2021). We randomly sample galaxy positions within the survey region until the target source density is reached. The intrinsic ellipticities of the galaxies are then drawn from a probability distribution that was fit to the observed galaxy ellipticities in Troxel et al. (2018) (see Zürcher et al. 2021). The ellipticity of each individual galaxy is rotated by a random phase. Using 5 and 20 shape noise realization per survey patch, we achieve the desired number of $N = 2000$ survey realization for the stage III and stage IV survey setup, respectively.

The tomographic angular power spectra realization $C_{\ell,i}$ are then measured from the forward-modelled surveys using the `anafast` routine of the `HEALPY` software (Zonca et al. 2019) using 20 bins from $\ell_{\min} = 100$ to $\ell_{\max} = 1000$, the same as Sgier et al. (2019), where the index i runs over the number of survey realization N . The covariance matrix Σ is estimated according to

$$\hat{\Sigma} = \frac{1}{N-1} \sum_{i=1}^N (C_{\ell,i} - \bar{C}_{\ell})(C_{\ell,i} - \bar{C}_{\ell})^T, \quad (9)$$

where \bar{C}_{ℓ} indicates the mean of the angular power spectra realization $C_{\ell,i}$. The estimated correlation matrices $C_{n,m} \equiv \Sigma_{n,m} / \sqrt{\Sigma_{n,n} \Sigma_{m,m}}$ are presented in Fig. 2.

3.3 Likelihood analysis

We use a Bayesian likelihood approach to evaluate the cosmological parameter constraints of different predictors. We assume a Gaussian error model and the likelihood is realized by:

$$\log \mathcal{L} = -\frac{1}{2} \sum_{ij} \left(C_{\ell,\text{truth}}^i - C_{\ell,\text{compare}}^i \right)^T \Sigma^{-1} \left(C_{\ell,\text{truth}}^j - C_{\ell,\text{compare}}^j \right) \quad (10)$$

Table 2. The fiducial values for the cosmological parameters and the flat priors for the cosmological parameters that are varied in the analysis.

Parameters	Fiducial values	Priors (stage III survey)	Priors (stage IV survey)
Ω_m	0.291	[0, 0.6]	[0.2, 0.4]
n_s	0.969	[0.3, 2.0]	[0.7, 1.2]
h	0.69	[0.1, 2.5]	[0.4, 0.9]
σ_8	0.826	[0.3, 1.4]	[0.7, 0.95]
w_0	-1.0	[-3.5, 0.5]	[-2.5, 0.5]
Ω_b	0.0473		

Here $C_{\ell,\text{truth}}$ stands for the value of the observable, computed using `PyCosmo` (Refregier et al. 2018; Tarsitano et al. 2021; Moser et al. 2022) with a chosen predictor and the fiducial cosmological parameters, measured by the Wilkinson Microwave Anisotropy Probe satellite (WMAP) 9 (Hinshaw et al. 2013), presented in Table 2. $C_{\ell,\text{compare}}$ is predicted using another predictor for comparison. The cosmology for the observable is different from what is used for the covariance matrix. However, this effect is neglected assuming the covariance matrix parameter independent (Kodwani, Alonso & Ferreira 2018). Σ^{-1} is the unbiased estimate of the inverse covariance matrix (Hartlap Simon & Schneider 2007; Percival et al. 2014) represented as:

$$\Sigma^{-1} = \frac{N - N' - 2}{N - 1} \hat{\Sigma}^{-1}, \quad (11)$$

N is the number of realization generated from the simulations and N' is the total number of data bins, which is given by

$$N' = N_{\text{redshift}} \times N_{\ell}. \quad (12)$$

Here, we have $N = 2000$, $N_{\ell} = 20$, and $N_{\text{redshift}} = 10$.

3.4 Parameter inference

The posterior is sampled efficiently using the Markov Chain Monte Carlo (MCMC) ensemble sampler, `emcee` (Foreman-Mackey et al. 2013). We vary four cosmological parameters $\{\Omega_m, \sigma_8, n_s, \text{ and } h\}$ for the Λ CDM cosmological model and an additional parameter w_0

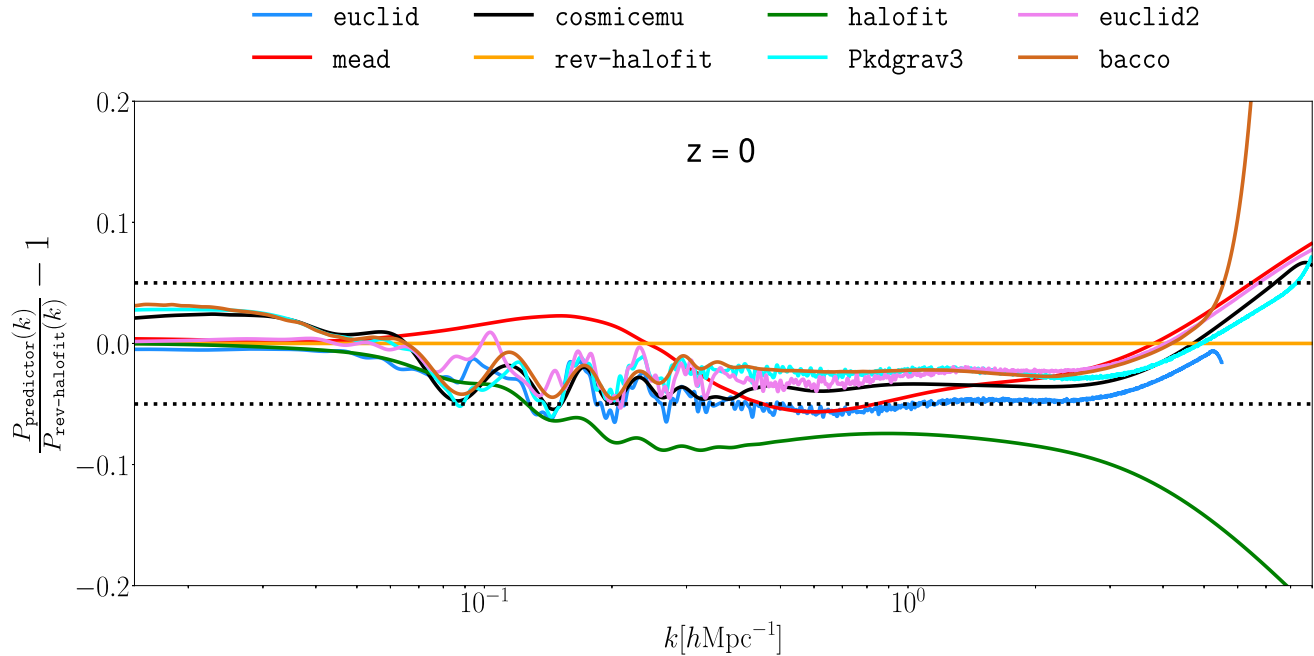


Figure 3. Comparison of dark-matter-only, non-linear $P(k)$ predictions for different predictors at redshift $z = 0$, subtracted and divided by *rev-halofit* as reference.

for the extended Λ CDM model, where we fix $w_a \equiv 0$. Table 2 shows the priors used for these parameters. We run the MCMC chains with 100 walkers per parameter and cut the burn in phase for each run as one-third of the chain length. Each individual chain has more than 100 000 samples. For the visualization of the marginalized posteriors, we use the public *Getdist* (Lewis 2019).

4 RESULTS

We present the results of our comparison of different predictors in this section, including the analysis of the matter power spectrum, the weak lensing power spectrum, and the cosmological parameter constraints based on the stage III and stage IV weak lensing surveys.

4.1 Power spectrum

We use the linear power spectrum predicted by *PyCosmo* and generated the following Eisenstein & Hu (1999) as the input for all predictors. Fig. 3 shows the comparison of dark-matter-only non-linear $P(k)$ predictions from different predictors at redshift $z = 0$, and the comparison for different redshifts ranging from $z = 0$ to $z = 5$ in Appendix A. The results are shown for k ranging from $k = 0.01$ to $9 h\text{Mpc}^{-1}$ using 10 000 bins. *BaccoEmulator* and *CosmicEmulator* are not valid for $z > 3$, so we do not present their comparison for the higher redshift at $z = 5$. Figs 3 and A1 indicate that:

(i) All the predictors except for *halofit* are within the 5 per cent level of accuracy compared to *rev-halofit* for $z < 2$ and $k < 7 h\text{Mpc}^{-1}$ (*BaccoEmulator* is valid for $z < 1.5$ and $k < 5 h\text{Mpc}^{-1}$, see the details in Fig. A1). Note that this is consistent with the comparison of *mead*, *rev-halofit* and *halofit* in Mead et al. (2015).

(ii) *halofit* shows stronger discrepancies compared with the other predictors at small scales for $k > 0.1 h\text{Mpc}^{-1}$ and this discrepancy can reach 20 per cent for $k \sim 10 h\text{Mpc}^{-1}$.

(iii) *mead* and *rev-halofit* show close agreement with the emulators at the 5 per cent level for $k < 9 h\text{Mpc}^{-1}$ and $z < 0.5$. However, at higher redshifts $1 < z < 5$, the discrepancies between *mead* and the emulators can reach 10 per cent for $k > 3 h\text{Mpc}^{-1}$, whereas *rev-halofit* provides a more consistent precision within 5 per cent.

(iv) All the emulators yield an agreement within the 2–3 per cent level compared with the *PKDGRAV3* simulation for $k < 9 h\text{Mpc}^{-1}$ and $z < 1.5$. However, this is not valid at higher redshifts. The disagreement at higher redshifts between emulators and *PKDGRAV3* might be due to the fact that emulators were built by interpolation within a certain parameter range, thus the accuracies could not be ensured beyond this range.

(v) For large scales with $k < 0.5 h\text{Mpc}^{-1}$, the different predictors show a better agreement at higher redshifts.

4.2 Weak lensing power spectrum

We compute the weak lensing shear power spectrum C_ℓ for the stage III and the stage IV survey with different predictors. Limited by the range of k_{max} of the emulators, the C_ℓ s are computed using 20 ℓ -bins spaced linearly between $\ell_{\text{min}} = 100$ and $\ell_{\text{max}} = 1000$ (A further investigation on the impact of varying ℓ_{max} is presented in Appendix B2). The integrated redshift range is $[0.08, 2.0]$ for the stage III survey and $[0.08, 3.0]$ for the stage IV survey. This setting was chosen in order to avoid the instability of emulators for low redshifts, where we found that *EuclidEmulator* and *EuclidEmulator2* predict the C_ℓ s with a discrepancy larger than 10 per cent at $z < 0.08$. This choice differs from the setting used for the generation of the covariance matrix. However, we find that this only changes the discrepancies between different predictors for C_ℓ s by 0.1 per cent, since only 1 per cent of the low-redshift galaxies are missed for the stage III survey and 0.1 per cent of the galaxies for the stage IV survey. Using this redshift range, we have to exclude *CosmicEmulator* from the comparison for the

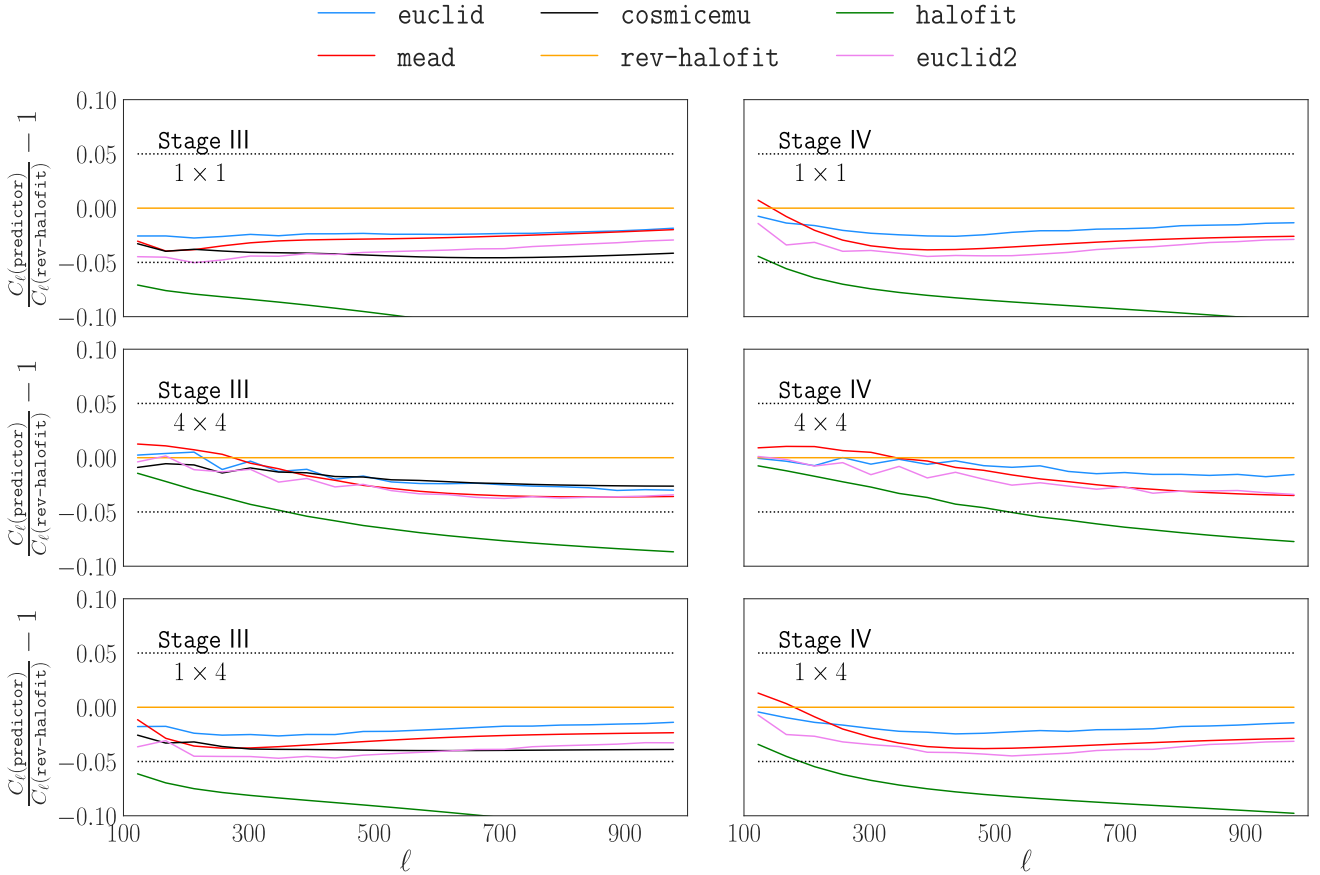


Figure 4. The comparison of weak lensing shear C_ℓ s for different predictors. Each C_ℓ is multiplied by $\ell(\ell + 1)/2\pi$. The upper two panels in each column show the auto-correlated C_ℓ s for the first, and the fourth redshift bin and the bottom ones show the cross correlated C_ℓ s between these two bins. The left-hand panels show the plots for the stage III survey and the right-hand side shows the stage IV survey results.

stage IV survey as it allows only up to $z = 2.0$ (BaccoEmulator is also excluded due to the redshift range up to $z = 1.5$). The comparison is shown in Fig. 4, with the left-hand panels showing the results for the stage III survey and the right-hand side showing the stage IV survey results. In the individual panels, we present $C_\ell \ell(\ell + 1)/2\pi$ for each predictor and illustrate the comparison by subtracting and dividing *rev-halofit* as the reference. In Fig. 4 the first row shows the comparison of the auto-correlated C_ℓ s for the redshift bins 1×1 , the second row for 4×4 , and the bottom row shows the cross correlated C_ℓ s for 1×4 . From Fig. 4, one can infer that:

- (i) All the predictors, except for *halofit*, yield an agreement at the 5 per cent level, both for the auto and cross C_ℓ . This is consistent with our results for $P(k)$.
- (ii) *mead* shows a good agreement with *CosmicEmulator*, *EuclidEmulator2*, and *EuclidEmulator*, whereas *rev-halofit* exhibits a larger discrepancy.
- (iii) The comparison of C_ℓ for different predictors does not show a significant difference between the stage III and the stage IV survey.

4.3 Cosmological parameters constraints

The comparison of the weak lensing cosmological parameter constraints for different predictors is present in this section. As indicated in Section 3, we consider a stage III and a stage IV surveys. For each survey, we perform a comparison using the standard Λ CDM

cosmological model and the extended w CDM model. A summary of the constraints on $\{S_8, \Omega_m, w_0\}$ is presented in Table 3, and the constraints on $\{S_8, \Omega_m, n_s, h, w_0\}$ in Table B1.

4.3.1 Λ CDM cosmology constraints

We present the two-dimensional 68 per cent and 95 per cent confidence level contours of the posterior distributions for the Λ CDM model in Figs 5 and 6 for the stage III and stage IV survey setup, respectively. The parameters $\{\Omega_m, \sigma_8, n_s, h\}$ are varied in the MCMC analysis. We additionally compute the constraints on S_8 , and summarize the shifts in S_8 in Fig. 9, presenting the median values of the posteriors and the error bars indicating the 68 per cent confidence limits of the constraints. For two different predictors, the significance of disagreement is computed by dividing the difference of their means by their combined uncertainties. One can infer from the posterior distributions in Fig. 9 and Table B1 that the agreement on S_8 between different predictors is less than 0.6σ for the stage III survey ($0.2 - 0.3\sigma$ if *halofit* excluded), while being much larger for the stage IV survey. This is caused by the higher constraining power of the stage IV survey. More specifically, the agreements are generally on the $1.4 - 6.1\sigma$ level ($1.4 - 3.0\sigma$ if *halofit* excluded). *mead* shows good agreement with *CosmicEmulator*, *EuclidEmulator*, and *EuclidEmulator2* for the stage III survey while it only agrees well with *EuclidEmulator2* for the stage IV survey. The constraints on h do not show significant

Table 3. Numerical constraints on the cosmological parameters corresponding to the contours in Figs 5, 6, 7, and 8. For each predictor, the σ s show the theoretical discrepancies for each parameter, compared to the reference one.

Survey cosmology	Predictor ref: rev-halofit	S_8	(σ)	Ω_m	(σ)	w_0	(σ)
Stage III Λ CDM	rev-halofit	$0.8147^{+0.0241}_{-0.0203}$		$0.288^{+0.0817}_{-0.0662}$			
	mead	$0.8035^{+0.0269}_{-0.0202}$	0.33	$0.2996^{+0.0848}_{-0.0698}$	0.11		
	halofit	$0.7946^{+0.0292}_{-0.0201}$	0.57	$0.2884^{+0.0783}_{-0.074}$	0.0		
	euclid	$0.8083^{+0.0256}_{-0.0201}$	0.2	$0.2987^{+0.0831}_{-0.0709}$	0.1		
	cosmicemu	$0.8047^{+0.0285}_{-0.018}$	0.29	$0.2916^{+0.0789}_{-0.0741}$	0.03		
	euclid2	$0.8031^{+0.0269}_{-0.0177}$	0.34	$0.2887^{+0.0835}_{-0.0679}$	0.01		
Stage III w CDM	rev-halofit	$0.8165^{+0.0433}_{-0.0661}$		$0.2846^{+0.092}_{-0.09}$		$-0.9242^{+0.4704}_{-2.294}$	
	mead	$0.7947^{+0.0497}_{-0.0588}$	0.26	$0.31^{+0.0824}_{-0.1022}$	0.18	$-1.139^{+0.647}_{-2.2626}$	0.09
	halofit	$0.7879^{+0.0517}_{-0.0612}$	0.34	$0.2968^{+0.0787}_{-0.1011}$	0.09	$-1.1333^{+0.6581}_{-2.3122}$	0.09
	euclid	$0.7977^{+0.0545}_{-0.0542}$	0.22	$0.3049^{+0.085}_{-0.1017}$	0.15	$-1.1886^{+0.7187}_{-2.1508}$	0.11
	cosmicemu	$0.7982^{+0.0504}_{-0.0572}$	0.22	$0.2931^{+0.0921}_{-0.0969}$	0.06	$-1.1408^{+0.6926}_{-2.3046}$	0.09
	euclid2	$0.8018^{+0.0461}_{-0.0627}$	0.18	$0.2896^{+0.0928}_{-0.0877}$	0.04	$-1.0254^{+0.5498}_{-2.2745}$	0.04
Stage IV Λ CDM	rev-halofit	$0.8135^{+0.0023}_{-0.0024}$		$0.2915^{+0.0077}_{-0.0084}$			
	mead	$0.8028^{+0.0027}_{-0.0026}$	2.96	$0.3008^{+0.0094}_{-0.0074}$	0.87		
	halofit	$0.7944^{+0.002}_{-0.0029}$	6.11	$0.2856^{+0.0097}_{-0.0064}$	0.46		
	euclid	$0.8094^{+0.0018}_{-0.003}$	1.37	$0.2917^{+0.0079}_{-0.0084}$	0.02		
	euclid2	$0.8058^{+0.0017}_{-0.0032}$	2.62	$0.2926^{+0.0079}_{-0.0084}$	0.1		
Stage IV w CDM	rev-halofit	$0.8127^{+0.0079}_{-0.0063}$		$0.2909^{+0.0095}_{-0.0086}$		$-1.0127^{+0.1171}_{-0.1046}$	
	mead	$0.7968^{+0.0067}_{-0.0069}$	1.73	$0.2979^{+0.0106}_{-0.0092}$	0.53	$-1.106^{+0.1107}_{-0.1163}$	0.61
	halofit	$0.7902^{+0.007}_{-0.0073}$	2.39	$0.2856^{+0.0093}_{-0.0096}$	0.42	$-1.0646^{+0.1069}_{-0.1197}$	0.35
	euclid	$0.8061^{+0.0073}_{-0.0072}$	0.68	$0.2908^{+0.0088}_{-0.0094}$	0.01	$-1.046^{+0.1142}_{-0.1288}$	0.22
	euclid2	$0.7996^{+0.0078}_{-0.0069}$	1.31	$0.2901^{+0.0099}_{-0.0094}$	0.06	$-1.0965^{+0.1288}_{-0.1255}$	0.51

discrepancies for both surveys, while n_s reveals discrepancies of several σ s for different predictors for the stage IV survey.

4.3.2 w CDM cosmology constraints

We consider the constraining power of weak lensing surveys on dark energy parameters by adopting a time-dependent dynamical dark energy equation of state, the CPT-parametrization (Chevallier & Polarski 2001; Linder 2003), as an extension to the Λ CDM model. The equation-of-state parameter is given by

$$w(a) = w_0 + w_a(1 - a), \quad (13)$$

where we use a fixed $w_a = 0$ and a free w_0 . We present the two-dimensional marginal posterior distributions for the w CDM cosmology parameters in Figs 7 and 8, for the stage III and the stage IV surveys, respectively. Taking into account the dark energy model changes the shape and the contour size of the posterior distributions, decreasing the constraining power on the cosmological parameters. The discrepancies in S_8 between predictors are generally smaller compared with the Λ CDM model due to the decrease in constraining power: $0.18 - 0.34\sigma$ for the stage III survey and $0.7 - 2.4\sigma$ for the stage IV survey ($0.18 - 0.26\sigma$ and $0.7 - 1.7\sigma$ if halofit is excluded, respectively). mead shows relatively good agreement with EuclidEmulator and EuclidEmulator2 for both the stage III and the stage IV surveys. rev-halofit agrees with all the predictors within 0.3σ for the stage III survey, and shows discrepancies at the $0.7 - 2.4\sigma$ level for the stage IV survey. Furthermore, we also consider the case with both free w_0 and w_a

(With a flat prior $[-2, 1]$). Compared with the case with a fixed w_a , this setting gives a tiny impact on the discrepancies between different predictors for $\{S_8, n_s, h\}$. However, it obtains weaker constraints on $\{\Omega_m, w_0\}$, resulting in the good agreements between different predictors. The discrepancies on w_a are within 0.5σ .

4.4 Systematic effects

In this study, we include dark-matter-only predictions, without any consideration of baryonic effects, which can have a strong impact on small scales (Jing et al. 2006; Rudd, Zentner & Kravtsov 2008), and the computation of the matter power spectrum (van Daalen et al. 2011; Casarini et al. 2012; Castro et al. 2018; Debackere, Schaye & Hoekstra 2020). Current studies of halo-model based fitting functions already include other systematics, i.e. massive neutrino and baryonic effects like AGN feedback and gas cooling. The inclusion of these systematics will significantly reduce the constraining power, and might alleviate the discrepancies between the predictors. The impact of taking into account the baryonic effects on cosmic shear can be found in Semboloni et al. (2011) and Martinelli et al. (2021), which indicates that including different baryonic models leads to discrepancies with $<0.5\sigma$ on cosmological parameter constraints for $\ell_{\max} = 1500$, and more significant biases (A few σ s) for higher $\ell_{\max} \sim 5000$. However, it does not broaden significantly the constraints in both cases. In our scenario where ℓ_{\max} is fixed to 1000, it can be foreseen that including baryons will involve a non-negligible impact on the agreements between different predictors.

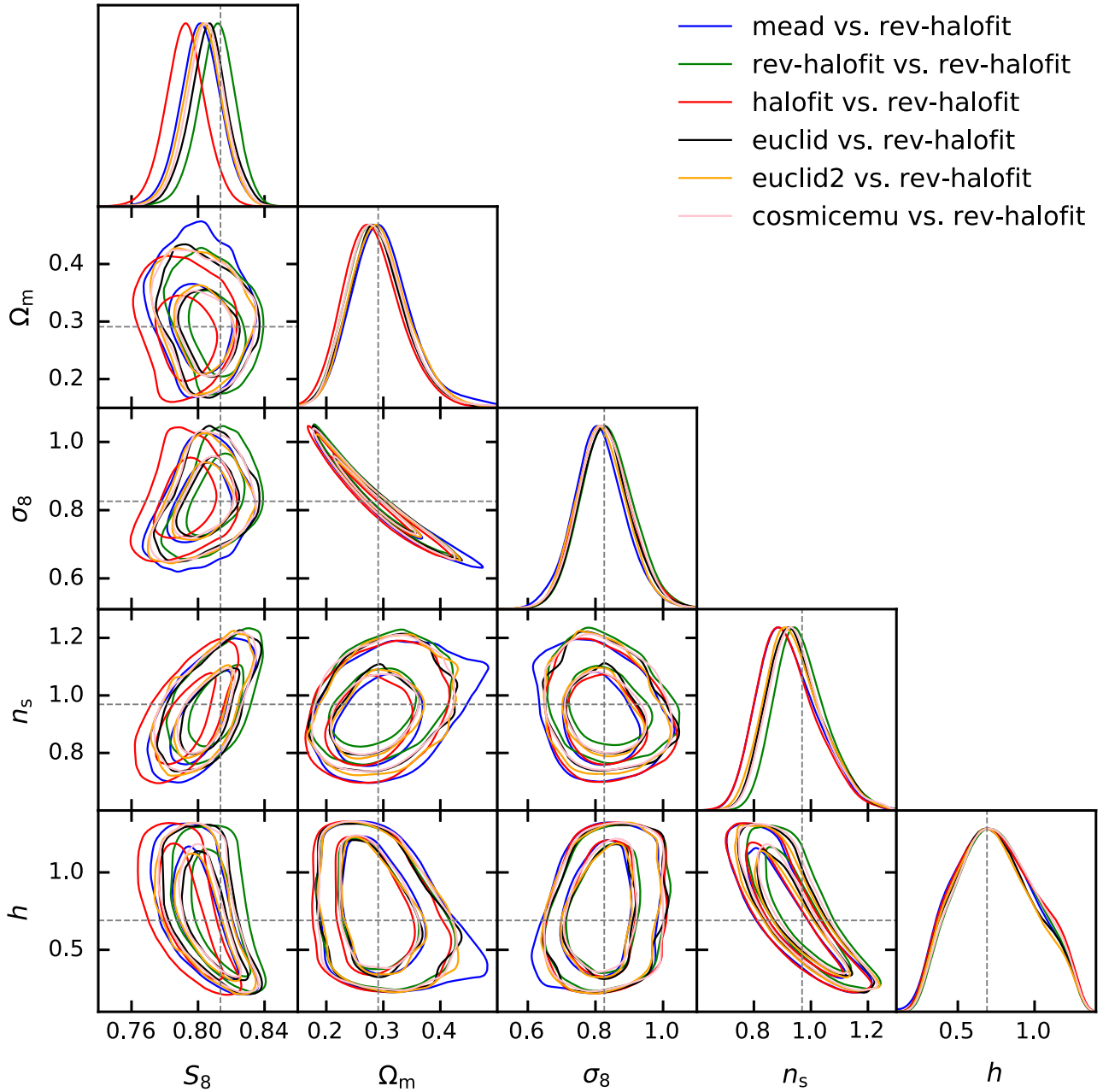


Figure 5. Cosmological parameter constraints for the stage III survey in the Λ CDM model. For each constraint, $C_{\ell, \text{truth}}$ is predicted using the first predictor shown in the legend, and $C_{\ell, \text{compare}}$ computed using the second predictor, as indicated in Section 3.3. For the stage III survey, we set $C_{\ell, \text{truth}}$ with the halo-model based fitting functions (rev-halofit, mead, and halofit) and three emulators (EuclidEmulator, EuclidEmulator2, and CosmicEmulator), and compare with predictions from only the fitting functions (in this figure only rev-halofit).

In practice, there are also other sources of uncertainties in weak lensing experiments, such as photometric redshift uncertainty (Huterer et al. 2006; Choi et al. 2016; Hildebrandt et al. 2020), shear bias (Bernstein & Jarvis 2002; Hirata et al. 2004; Bernstein 2010; Melchior & Viola 2012; Refregier et al. 2012), and galaxy intrinsic alignment (Heavens et al. 2000; Hirata & Seljak 2004; Bridle & King 2007; Joachimi et al. 2011; Fluri et al. 2019). These systematic effects will contribute to the total error budget and broaden the constraints on cosmological parameters. In our analysis, we computed the impact of theoretical uncertainties and compared them to statistical errors. This is useful to allocate a given budget to this source of error, independently of the choices in the treatment

of the other systematics. However, it is also useful to estimate the fraction of the theoretical statistical errors compared with these systematic errors, in order to study their contribution to the total error budget.

For this purpose, we estimate the impact of these systematics by considering other works which have carried out measurements and forecasts for stage III and stage IV surveys. For DES-like stage-III surveys, we can infer from Secco et al. (2022) and Amon et al. (2022), that the constraining power on S_8 will be decreased by ~ 20 per cent when considering the intrinsic alignment models, and less than ~ 5 per cent when considering the photometric redshift uncertainties and shear bias. For LSST-like stage-IV surveys, Krause,

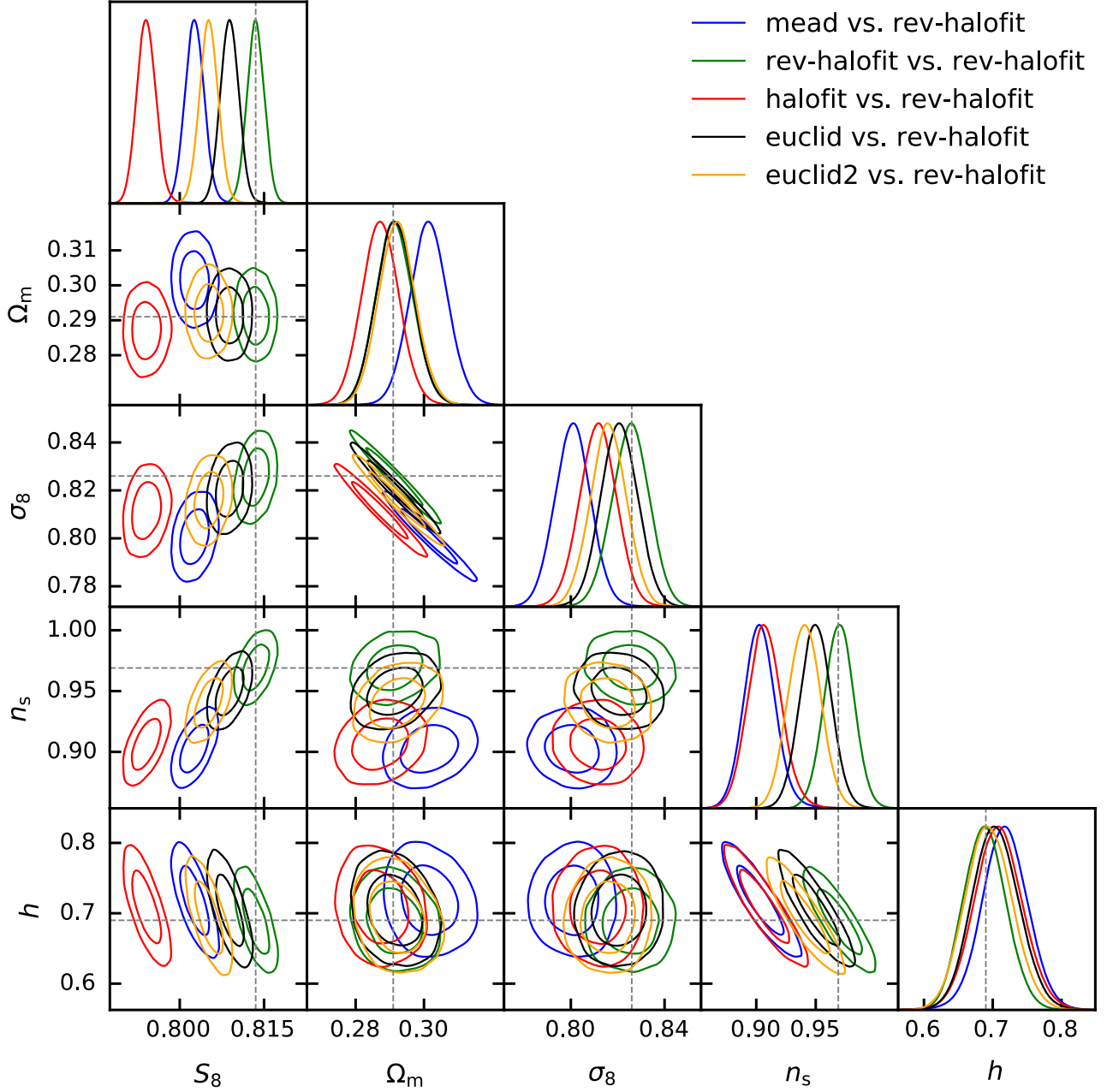


Figure 6. Cosmological parameter constraints of the stage IV survey in the Λ CDM model. Only two emulators, i.e. `EuclidEmulator` and `EuclidEmulator2`, are chosen for $C_{\ell, \text{truth}}$, as `CosmicEmulator` does not provide a sufficient redshift range for the stage IV survey.

Eifler & Blazek (2016) shows that the constraints for Ω_m and σ_8 could be broadened when considering different systematics by: ~ 40 per cent (pessimistic LSST photo- z errors), ~ 50 per cent (optimistic LSST photo- z errors & non-linear intrinsic alignment (IA NLA) model), and ~ 100 per cent (pessimistic LSST photo- z errors & IA NLA model). In this case, the significance of the discrepancies between different predictors will be reduced by 25 per cent – 50 per cent, while still significant with the smallest between mead and `EuclidEmulator` larger than 0.6σ . In practice, the inclusion of all these systematics, as well as theoretical uncertainties will be needed to estimate the total error budget of specific weak lensing measurements.

5 CONCLUSIONS

The different halo-model based fitting functions and emulators have been widely used for the prediction of non-linear power spectrum to study the large-scale structure of the Universe. It is essential to understand their advantages, limitations, and theoretical uncertainties for different surveys and cosmologies. From our results, we conclude that:

- (i) Compared with PKDGRAV3 simulations, the halo-model based fitting functions, except `halofit`, yield a 5 – 10 per cent level accuracy for the matter power spectrum $P(k)$ for $k < 9 h \text{Mpc}^{-1}$ and $z < 2$, while emulators show better precision at the 2 per cent level.

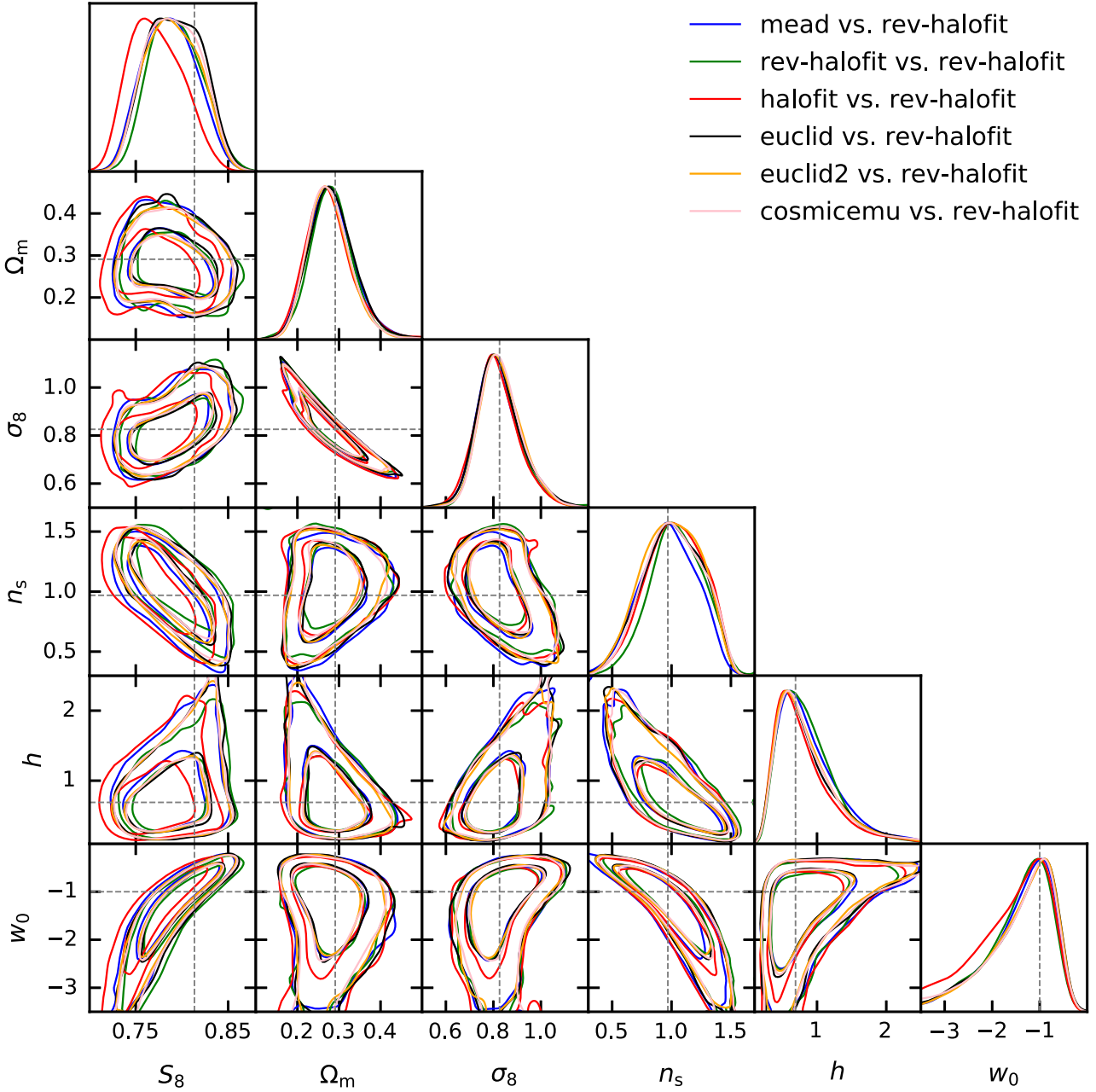


Figure 7. Cosmological parameter constraints of the stage III survey in the w CDM cosmological model. Including w_0 reduces significantly the constraining power, yielding much broader contours than the Λ CDM model.

For the weak lensing shear power spectrum C_ℓ , all the predictors, except for *halofit*, show a 5 per cent level mutual agreement.

(ii) For the stage III survey with a Λ CDM cosmology, the agreement on S_8 between different predictors are within 0.6σ , and within 0.2σ for other cosmological parameters (0.3σ and 0.2σ if we exclude *halofit*, respectively). This indicates the applicability of the studied predictors for the stage III surveys.

(iii) For the stage IV survey using a Λ CDM cosmology, the disagreements on S_8 are increased to several σ s, with the largest discrepancy of 6.1σ between *rev-halofit* and *halofit*, and the best agreement between *mead* and *EuclidEmulator2*.

(iv) If w_0 is taken into account for the w CDM cosmology, we get weaker constraints on S_8 , and the discrepancies between different

predictors are reduced to $0.2 - 0.3\sigma$ and $0.7 - 2.4\sigma$ for the stage III and the stage IV surveys, respectively ($0.18 - 0.26\sigma$ and $0.7 - 1.7\sigma$ if we exclude *halofit*, respectively). If w_a is taken into account, we get very similar constraints on S_8 compared to the w_0 -only case.

(v) The accuracy of the current fitting function models and emulators therefore appear sufficient for stage III surveys. However, for the future IV surveys, our results suggest that the fitting function models are currently not sufficiently accurate, and would need further improvements in the future. For emulators, it is required to explore wider ranges of cosmological parameters, k -modes, and redshifts, while pursuing consistent precision with reliable hydrodynamic N -body simulations.

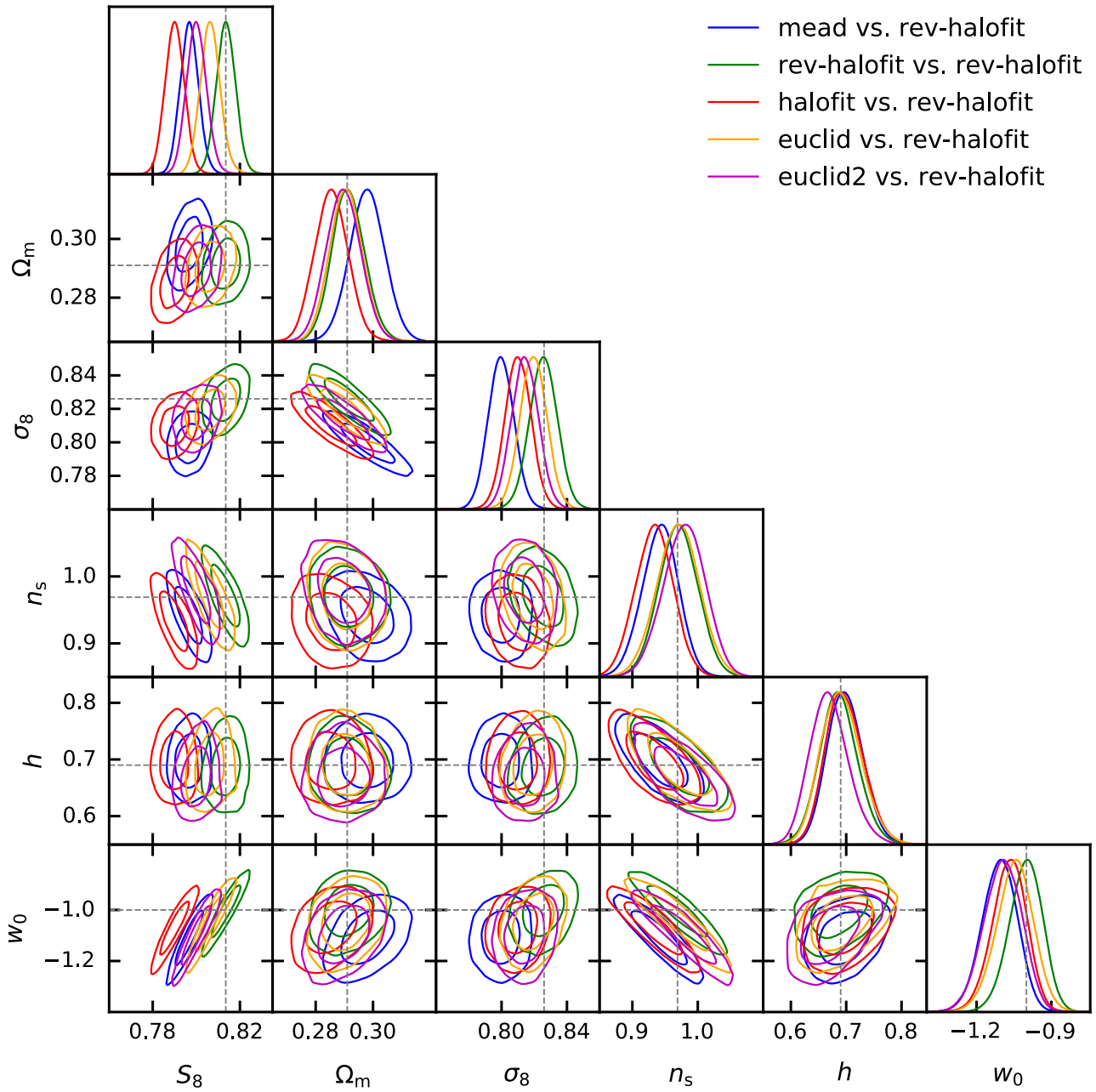


Figure 8. Cosmological parameter constraints of the stage IV survey in the w CDM cosmological model. The discrepancies between the predictors are alleviated, taking into account a simple w CDM cosmological model with a varying w_0 .

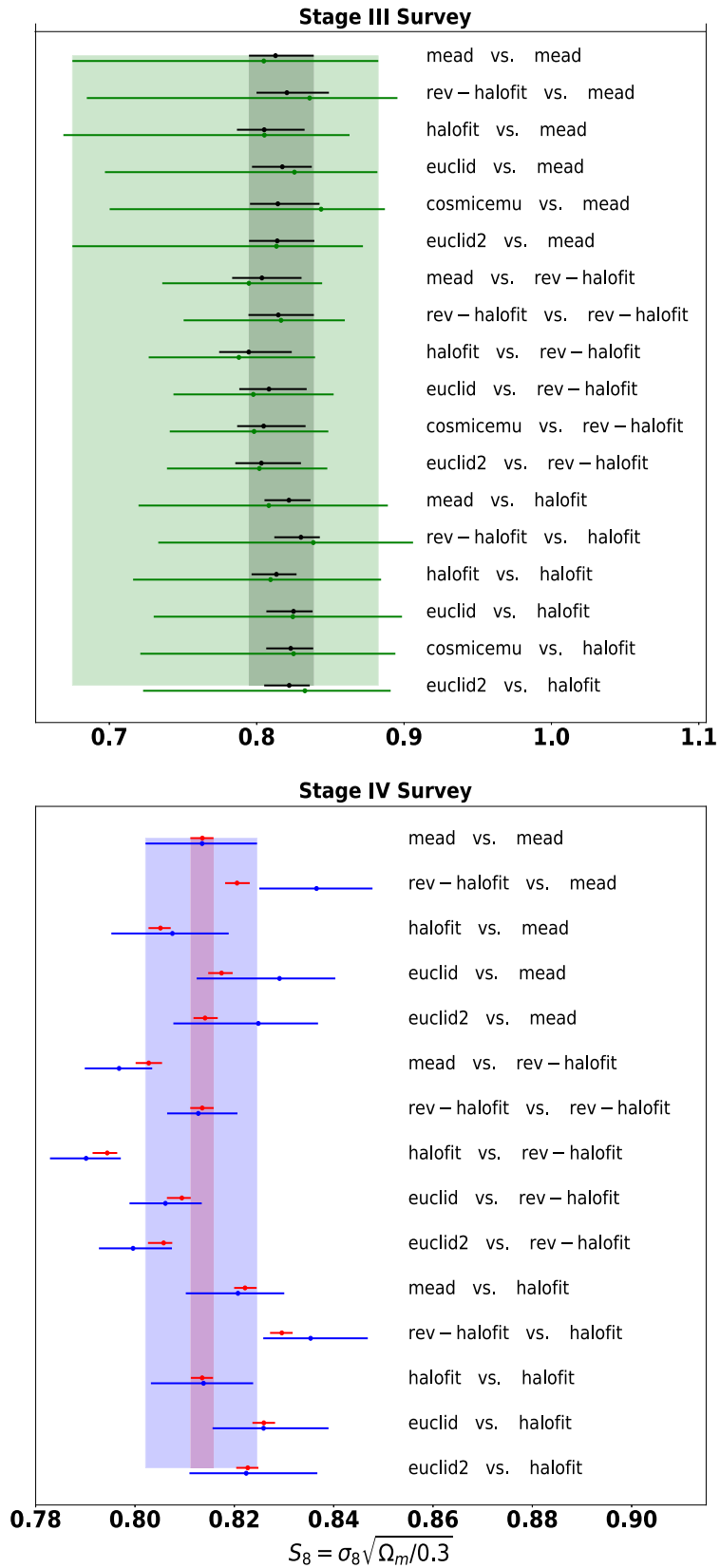


Figure 9. Deviations of the parameter constraints on S_8 . The upper plot shows the result for the stage III survey, for the Λ CDM model (black) and the w CDM model (green), respectively. The lower plot shows the stage IV survey, for the Λ CDM (red) and w CDM (blue), respectively.

(vi) Taking into account other systematic effects such as baryonic effects, photometric redshift uncertainty, shear bias, and galaxy intrinsic alignment will broaden the parameters constraints by 40 per cent – 100 per cent from stage IV weak lensing surveys. This will tend to reduce the significance of the discrepancies between the different predictor. The theoretical uncertainties however remain non-negligible and need to be included in the total error budget for future surveys.

ACKNOWLEDGEMENTS

This work was supported in part by grant 200021_192243 from the Swiss National Science Foundation.

We thank Mischa Knabenhans from University of Zürich for the distribution of PKDGRAV3. We further thank Aurel Schneider from the University of Zürich for the useful discussions regarding this project and the covariance matrix for a stage IV survey. We would also like to thank Uwe Schmitt from ETH Zürich for his support with the GitLab server and development of PyCosmo.

The Collaborating Institutions are the Eidgenössische Technische Hochschule (ETH) Zürich, Ecole Polytechnique, the Laboratoire de Physique Nucléaire et des Hautes Energies of Sorbonne University.

DATA AVAILABILITY

Most of the analysis in this work is down on the Euler cluster⁵ operated by ETH Zurich. Here follows the computational codes used in this study: PyCosmo (Refregier et al. 2018; Tarsitano et al. 2021; Moser et al. 2022) is used as the main tool where all the non-linear codes are implemented for the computation of auto (cross) power spectra, galaxy redshift distribution counts, and observable of cosmic shear. It is also extended to include interfaces with the emulators. AnaFast is used for computation of power spectra from simulations, and all the the maps (masks, weight, shear, and mass) in pipeline are in HealPix format. We use Emcee-3.0.2 (Foreman-Mackey et al. 2013) for the sampling of parameter space and Getdist (Lewis 2019) for the plotting of likelihood contours and Uhammer for the simplification of Emcee running. Some of the results in this paper have been derived using the healpy and HEALPix packages (Gorski et al. 1999). In this study, we made use of the functionalities provided by numpy (van der Walt, Colbert & Varoquaux 2011), scipy (Virtanen et al. 2020), and matplotlib (Hunter 2007).

REFERENCES

Abell P. A. et al., 2009, preprint (arXiv:0912.0201)
 Aghanim N. et al., 2020, *A&A*, 641, A6
 Akeson R. et al., 2019, preprint (arXiv:1902.05569)
 Amara A., Réfrégier A., 2007, *MNRAS*, 381, 1018
 Amon A. et al., 2022, *Phys. Rev. D*, 105, 023514
 Angulo R. E., Springel V., White S. D. M., Jenkins A., Baugh C. M., Frenk C. S., 2012, *MNRAS*, 426, 2046
 Angulo R. E., Zennaro M., Contreras S., Aricó G., Pellejero-Ibañez M., Stöcker J., 2020, *MNRAS*, 507, 5869
 Aricó G., Angulo R. E., Contreras S., Ondaro-Mallea L., Pellejero-Ibañez M., Zennaro M., 2021, *MNRAS*, 506, 4070
 Bartelmann M., Maturi M., 2016, *Scholarpedia*, 12, 32440
 Bartelmann M., Schneider P., 2001, *Phys. Rep.*, 340, 291

Baumann D., Nicolis A., Senatore L., Zaldarriaga M., 2012, *J. Cosmol. Astropart. Phys.*, 2012, 051
 Bernardeau F., Colombi S., Gaztañaga E., Scoccimarro R., 2002, *Phys. Rep.*, 367, 1
 Bernstein G. M., 2010, *MNRAS*, 406, 2793
 Bernstein G., Jarvis M., 2002, *AJ*, 123, 583
 Beutler F. et al., 2017, *MNRAS*, 464, 3409
 Bird S., Viel M., Haehnelt M. G., 2012, *MNRAS*, 420, 2551
 Blas D., Garny M., Konstandin T., 2014, *J. Cosmol. Astropart. Phys.*, 2014, 010
 Blas D., Garny M., Ivanov M. M., Sibiryakov S., 2016, *J. Cosmol. Astropart. Phys.*, 2016, 052
 Bridle S., King L., 2007, *New J. Phys.*, 9, 444
 Casarini L., Bonometto S. A., Borgani S., Dolag K., Murante G., Mezzetti M., Tornatore L., La Vacca G., 2012, *A&A*, 542, A126
 Castro T., Quartin M., Giocoli C., Borgani S., Dolag K., 2018, *MNRAS*, 478, 1305
 Cataneo M., Lombriser L., Heymans C., Mead A., Barreira A., Bose S., Li B., 2019, *MNRAS*, 488, 2121
 Chevallier M., Polarski D., 2001, *Int. J. Mod. Phys.*, 10, 213
 Choi A. et al., 2016, *MNRAS*, 463, 3737
 Chudaykin A., Ivanov M. M., Philcox O. H., Simonović M., 2020, *Phys. Rev. D*, 102, 063533
 Collaboration E. et al., 2020, *MNRAS*, 505, 2840
 Cooray A., Sheth R., 2002, *Phys. Rep.*, 372, 1
 Crocce M., Scoccimarro R., 2006, *Phys. Rev. D*, 73, 063519
 Crocce M., Scoccimarro R., Bernardeau F., 2012, *MNRAS*, 427, 2537
 d'Amico G., Gleyzes J., Kokron N., Markovic K., Senatore L., Zhang P., Beutler F., Gil-Mañán H., 2020, *J. Cosmol. Astropart. Phys.*, 2020, 005
 D'Amico G., Senatore L., Zhang P., 2021, *J. Cosmol. Astropart. Phys.*, 2021, 006
 Dominik Z. et al., 2022, *MNRAS*, 511, 2075
 Debackere S. N., Schaye J., Hoekstra H., 2020, *MNRAS*, 492, 2285
 Eisenstein D. J., Hu W., 1999, *ApJ*, 511, 5
 Fluri J., Kacprzak T., Lucchi A., Refregier A., Amara A., Hofmann T., Schneider A., 2019, *Phys. Rev. D*, 100, 063514
 Foreman S., Senatore L., 2016, *J. Cosmol. Astropart. Phys.*, 2016, 033
 Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306
 Giannantonio T., Porciani C., Carron J., Amara A., Pillepich A., 2012, *MNRAS*, 422, 2854
 Giblin B., Cataneo M., Moews B., Heymans C., 2019, *MNRAS*, 490, 4826
 Giblin B. et al., 2021, *A&A*, 645, A105
 Gorski K. M., Wandelt B. D., Hansen F. K., Hivon E., Banday A. J., 1999, *The HEALPix Primer*, preprint (arXiv:astro-ph/9905275)
 Gorski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, *ApJ*, 622, 759
 Hamilton A. J. S., Kumar P., Lu E., Matthews A., 1991, *ApJ*, 374, L1
 Hand N., Feng Y., Beutler F., Li Y., Modi C., Seljak U., Slepian Z., 2018, *AJ*, 156, 160
 Hartlap J., Simon P., Schneider P., 2007, *A&A*, 464, 399
 Heavens A., Refregier A., Heymans C., 2000, *MNRAS*, 319, 649
 Heitmann K., Higdon D., White M., Habib S., Williams B. J., Lawrence E., Wagner C., 2009, *ApJ*, 705, 156
 Heitmann K., White M., Wagner C., Habib S., Higdon D., 2010, *ApJ*, 715, 104
 Heitmann K., Lawrence E., Kwan J., Habib S., Higdon D., 2013, *ApJ*, 780, 111
 Heitmann K. et al., 2016, *ApJ*, 820, 108
 Hildebrandt H. et al., 2020, *A&A*, 633, A69
 Hinshaw G. et al., 2013, *ApJS*, 208, 19
 Hirata C. M., Seljak U., 2004, *Phys. Rev. D*, 70, 063526
 Hirata C. M. et al., 2004, *MNRAS*, 353, 529
 Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
 Huterer D., Takada M., Bernstein G., Jain B., 2006, *MNRAS*, 366, 101
 Jing Y., Zhang P., Lin W., Gao L., Springel V., 2006, *ApJ*, 640, L119
 Joachimi B., Mandelbaum R., Abdalla F., Bridle S., 2011, *A&A*, 527, A26

⁵<https://scicomp.ethz.ch/wiki/Euler>.

- Kaiser N., 1992, *ApJ*, 388, 272
- Kaiser N., 1998, *ApJ*, 498, 26
- Kaiser N., Squires G., 1993, *ApJ*, 404, 441
- Kilbinger M. et al., 2017, *MNRAS*, 472, 2126
- Kitching T. D., Alsing J., Heavens A. F., Jimenez R., McEwen J. D., Verde L., 2017, *MNRAS*, 469, 2737
- Knabenhans M. et al., 2019, *MNRAS*, 484, 5509–5529
- Knabenhans M., Brinckmann T., Stadel J., Schneider A., Teyssier R., 2021, *MNRAS*, 518, 1859
- Kodwani D., Alonso D., Ferreira P., 2018, preprint (arXiv:1811.11584)
- Krause E., Eifler T., Blazek J., 2016, *MNRAS*, 456, 207
- Lawrence E. et al., 2017, *ApJ*, 847, 50
- Lesgourgues J., 2011, The Cosmic Linear Anisotropy Solving System (CLASS) III: Comparison with CAMB for Lambda-CDM (arXiv:1104.2934)
- Lewis A., 2019, GetDist: a Python package for analysing Monte Carlo samples, preprint (arXiv:1910.13970)
- Limber D. N., 1953, *ApJ*, 117, 134
- Linder E. V., 2003, *Phys. Rev. Lett.*, 90, 091301
- LoVerde M., Afshordi N., 2008, *Phys. Rev. D*, 78, 123506
- Ma C.-P., Fry J. N., 2000, *ApJ*, 543, 503
- Mancini A. S., Piras D., Alsing J., Joachimi B., Hobson M. P., 2022, *MNRAS*, 511, 1771
- Martinelli M. et al., 2021, *A&A*, 649, A100
- Mead A. J., Peacock J. A., Heymans C., Joudaki S., Heavens A. F., 2015, *MNRAS*, 454, 1958
- Mead A. J., Heymans C., Lombriser L., Peacock J. A., Steele O. I., Winther H. A., 2016, *MNRAS*, 459, 1468
- Melchior P., Viola M., 2012, *MNRAS*, 424, 2757
- Mohammed I., Seljak U., 2014, *MNRAS*, 445, 3382
- Moser B., Lorenz C., Schmitt U., Réfrégier A., Fluri J., Sgier R., Tarsitano F., Heisenberg L., 2022, *Astron. Comput.*, 40, 100603
- Nishimichi T., Bernardeau F., Taruya A., 2016, *Phys. Lett. B*, 762, 247
- Peacock J. A., Smith R. E., 2000, *MNRAS*, 318, 1144
- Peebles P., 2020, Large-Scale Structure of the Universe by Phillip James Edwin Peebles. Princeton University Press
- Percival W. J. et al., 2014, *MNRAS*, 439, 2531
- Petri A., Haiman Z., May M., 2017, *Phys. Rev. D*, 95, 123503
- Potter D., Stadel J., Teyssier R., 2017, *Comput. Astrophys. Cosmol.*, 4, 2
- Press W. H., Schechter P., 1974, *ApJ*, 187, 425
- Refregier A., Kacprzak T., Amara A., Bridle S., Rowe B., 2012, *MNRAS*, 425, 1951
- Refregier A., Gamper L., Amara A., Heisenberg L., 2018, *Astronomy and computing*, 25, 38
- Rudd D. H., Zentner A. R., Kravtsov A. V., 2008, *ApJ*, 672, 19
- Schneider A. et al., 2016, *J. Cosmol. Astropart. Phys.*, 2016, 047
- Scoccimarro R., Sheth R. K., Hui L., Jain B., 2001, *ApJ*, 546, 20
- Secco L. F. et al., 2022, *Phys. Rev. D*, 105, 023515
- Seljak U., 2000, *MNRAS*, 318, 203
- Seljak U. c. v., Vlah Z., 2015, *Phys. Rev. D*, 91, 123516
- Semboloni E., Hoekstra H., Schaye J., van Daalen M. P., McCarthy I. G., 2011, *MNRAS*, 417, 2020
- Sgier R. J., Réfrégier A., Amara A., Nicola A., 2019, *J. Cosmol. Astropart. Phys.*, 2019, 044
- Sheth R. K., Tormen G., 1999, *MNRAS*, 308, 119
- Smail I., Hogg D. W., Yan L., Cohen J. G., 1995, *ApJ*, 449, L105
- Smith R. E. et al., 2003, *MNRAS*, 341, 1311
- Springel V., 2005a, *MNRAS*, 364, 1105
- Springel V., 2005b, *MNRAS*, 364, 1105
- Springel V., Yoshida N., White S. D., 2001, *New Astron.*, 6, 79
- Springel V., Pakmor R., Zier O., Reinecke M., 2021, *MNRAS*, 506, 2871
- Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, *ApJ*, 761, 152
- Tarsitano F. et al., 2021, *Astronomy and Computing*, 36, 100484
- Tram T., Brandbyge J., Dakin J., Hannestad S., 2019, *J. Cosmol. Astropart. Phys.*, 2019, 022
- Troxel M. A. et al., 2018, *Phys. Rev. D*, 98, 043528
- van Daalen M. P., Schaye J., Booth C., Dalla Vecchia C., 2011, *MNRAS*, 415, 3649
- Virtanen P. et al., 2020, *Nat. Methods*, 17, 261
- Wallis C. G., McEwen J. D., Kitching T. D., Leistedt B., Plouviez A., 2022, *MNRAS*, 509, 4480
- van der Walt S., Colbert S. C., Varoquaux G., 2011, *Comput. Sci. Eng.*, 13, 22
- Zonca A., Singer L. P., Lenz D., Reinecke M., Rosset C., Hivon E., Gorski K. M., 2019, *J. Open Source Softw.*, 4, 1298
- Zürcher D., Fluri J., Sgier R., Kacprzak T., Refregier A., 2021, *J. Cosmol. Astropart. Phys.*, 2021, 028

APPENDIX A: POWER SPECTRUM COMPARISON

In this section, we present the comparison of the non-linear power spectrum for all redshifts, as shown in Fig. A1.

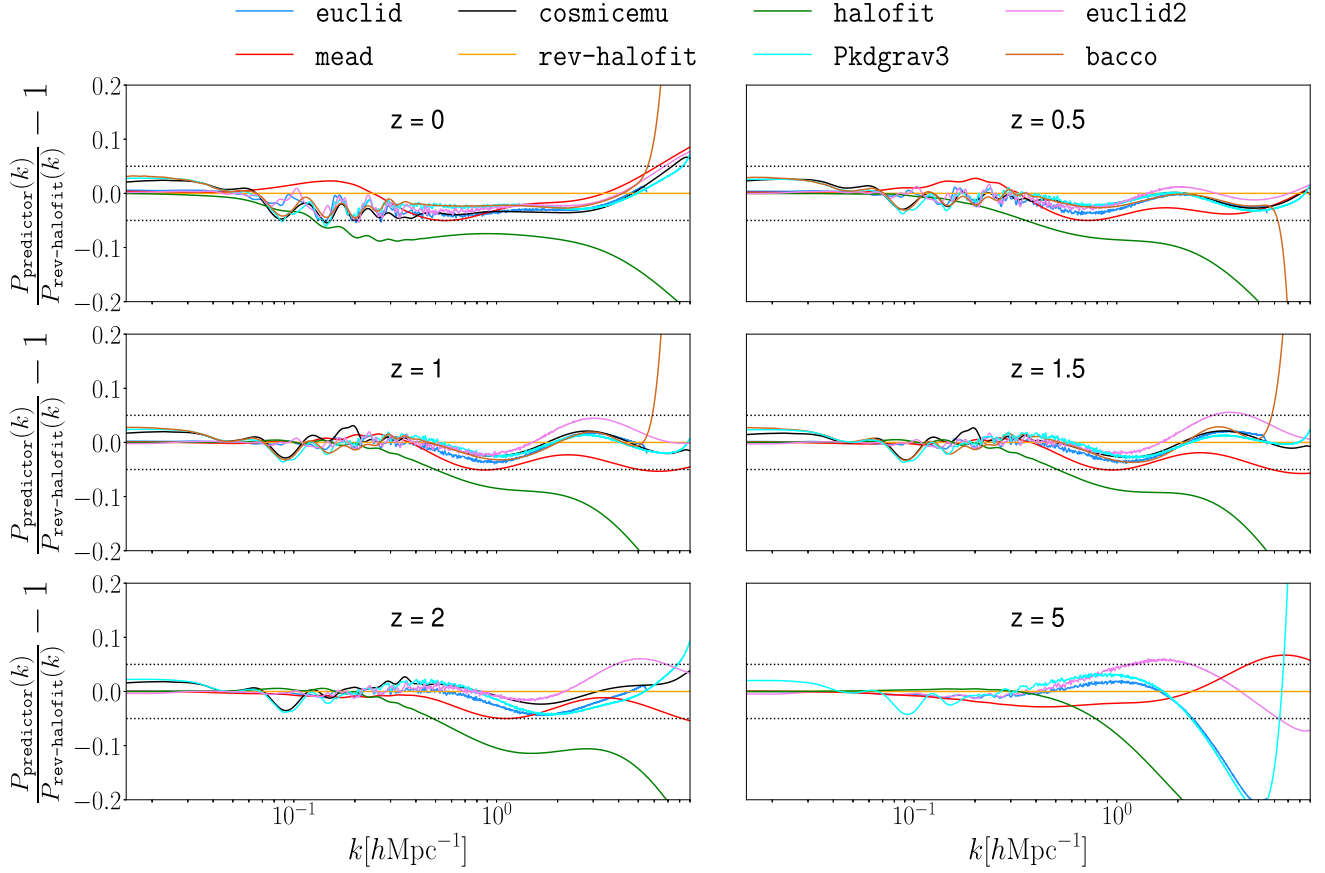


Figure A1. The comparison of the dark-matter-only non-linear $P(k)$ of different predictors at different redshifts ($z = 0, 0.5, 1, 1.5, 2,$ and 5), subtracted and divided by *rev-halofit* as reference. *BaccoEmulator* and *CosmicEmulator* are not valid for $z > 3$, so we do not take them into comparison for $z = 5$.

APPENDIX B: COSMOLOGICAL PARAMETER CONSTRAINTS

B1 Summary of constraints

The summary of constraints on $\{S_8, \Omega_m, n_s, h, w_0\}$ is concluded in this section, shown in Table B1.

B2 Different ℓ_{\max}

We investigate the variation of constraints on S_8 with different ℓ_{\max} (800 or 1000) and the results are summarized in Fig. B1. We only consider the stage IV survey with the Λ CDM model, since it gives the largest discrepancies between different predictors. When ℓ_{\max} is reduced from 1000 to 800 where we have less non-

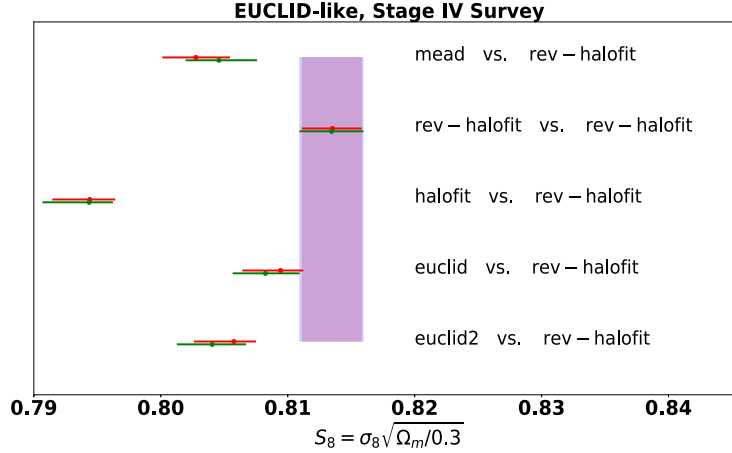
linear effect information, the marginalized 1D constraints on S_8 are broadened by 5 per cent. With this change, *rev-halofit* shows better agreements with *mead*, and larger discrepancies with *EuclidEmulator* and *EuclidEmulator2*.

B3 w CDM with free w_a

We present in Fig. B2 the cosmological parameter constraints of the stage IV survey in the w CDM cosmological model, with both varying w_0 and w_a . Compared with the case with a fixed w_a , this setting gives a tiny impact on the discrepancies between different predictors for $\{S_8, n_s, h\}$. However, it obtains weaker constraints on $\{\Omega_m, w_0\}$, resulting in the good agreements between different predictors. The discrepancies on w_a are within 0.5σ .

Table B1. Complete numerical constraints on the cosmological parameters corresponding to the contours in Figs 5, 6, 7, and 8. For each predictor, the σ s show the theoretical discrepancies for each parameter, compared to the reference one.

Survey cosmology	Predictor ref: rev-halofit	S_8	(σ)	Ω_m	(σ)	n_s	(σ)	h	(σ)	w_0	(σ)
Stage III	rev-halofit	$0.8147^{+0.0241}_{-0.0203}$		$0.288^{+0.0817}_{-0.0662}$		$0.9741^{+0.2489}_{-0.1475}$		$0.6736^{+0.5642}_{-0.4172}$			
	mead	$0.8035^{+0.0269}_{-0.0202}$	0.33	$0.2996^{+0.0848}_{-0.0698}$	0.11	$0.9144^{+0.2425}_{-0.1562}$	0.21	$0.6859^{+0.5779}_{-0.4358}$	0.02		
Λ CDM	halofit	$0.7946^{+0.0292}_{-0.0201}$	0.57	$0.2884^{+0.0783}_{-0.074}$	0.0	$0.9196^{+0.2566}_{-0.1662}$	0.18	$0.6777^{+0.6198}_{-0.4339}$	0.01		
	euclid	$0.8083^{+0.0256}_{-0.0201}$	0.2	$0.2987^{+0.0831}_{-0.0709}$	0.1	$0.9547^{+0.2428}_{-0.1562}$	0.07	$0.6644^{+0.5612}_{-0.4159}$	0.01		
	cosmicemu	$0.8047^{+0.0285}_{-0.018}$	0.29	$0.2916^{+0.0789}_{-0.0741}$	0.03	$0.9332^{+0.2613}_{-0.1399}$	0.14	$0.7452^{+0.5542}_{-0.486}$	0.1		
	euclid2	$0.8031^{+0.0269}_{-0.0177}$	0.34	$0.2887^{+0.0835}_{-0.0679}$	0.01	$0.9184^{+0.2496}_{-0.1316}$	0.19	$0.7467^{+0.5503}_{-0.4853}$	0.1		
Stage III	rev-halofit	$0.8165^{+0.0433}_{-0.0661}$		$0.2846^{+0.092}_{-0.09}$		$0.9164^{+0.5799}_{-0.3511}$		$0.7868^{+0.9823}_{-0.5347}$		$-0.9242^{+0.4704}_{-2.294}$	
	mead	$0.7947^{+0.0497}_{-0.0588}$	0.26	$0.31^{+0.0824}_{-0.1022}$	0.18	$0.9768^{+0.5012}_{-0.4514}$	0.08	$0.6527^{+1.0402}_{-0.4233}$	0.11	$-1.139^{+0.647}_{-2.2626}$	0.09
w CDM	halofit	$0.7876^{+0.0517}_{-0.0612}$	0.34	$0.2968^{+0.0787}_{-0.1011}$	0.09	$0.9919^{+0.4913}_{-0.4914}$	0.1	$0.6197^{+1.1863}_{-0.3779}$	0.13	$-1.1333^{+0.6581}_{-2.3122}$	0.09
	euclid	$0.7977^{+0.0545}_{-0.0542}$	0.22	$0.3049^{+0.085}_{-0.1017}$	0.15	$1.032^{+0.4723}_{-0.4813}$	0.15	$0.6209^{+1.1393}_{-0.3873}$	0.13	$-1.1886^{+0.7187}_{-2.1508}$	0.11
	cosmicemu	$0.7982^{+0.0504}_{-0.0572}$	0.22	$0.2931^{+0.0921}_{-0.0969}$	0.06	$1.0031^{+0.5093}_{-0.4918}$	0.11	$0.6688^{+1.2915}_{-0.4301}$	0.08	$-1.1408^{+0.6926}_{-2.3046}$	0.09
	euclid2	$0.8018^{+0.0461}_{-0.0627}$	0.18	$0.2896^{+0.0928}_{-0.0877}$	0.04	$0.9272^{+0.5729}_{-0.3921}$	0.02	$0.7461^{+1.0126}_{-0.5134}$	0.04	$-1.0254^{+0.5498}_{-2.2745}$	0.04
Stage IV	rev-halofit	$0.8135^{+0.0023}_{-0.0024}$		$0.2915^{+0.0077}_{-0.0084}$		$0.9696^{+0.0178}_{-0.0192}$		$0.6889^{+0.0481}_{-0.0433}$			
	mead	$0.8028^{+0.0027}_{-0.0026}$	2.96	$0.3008^{+0.0094}_{-0.0074}$	0.87	$0.9021^{+0.0193}_{-0.0189}$	2.48	$0.7181^{+0.0495}_{-0.0441}$	0.45		
Λ CDM	halofit	$0.7944^{+0.002}_{-0.0029}$	6.11	$0.2856^{+0.0097}_{-0.0064}$	0.46	$0.9054^{+0.0203}_{-0.0197}$	2.3	$0.7134^{+0.0494}_{-0.05}$	0.35		
	euclid	$0.8094^{+0.0018}_{-0.003}$	1.37	$0.2917^{+0.0079}_{-0.0084}$	0.02	$0.9497^{+0.0198}_{-0.0193}$	0.72	$0.7058^{+0.0505}_{-0.0479}$	0.25		
	euclid2	$0.8058^{+0.0017}_{-0.0032}$	2.62	$0.2926^{+0.0079}_{-0.0084}$	0.1	$0.9402^{+0.0195}_{-0.0206}$	1.07	$0.6958^{+0.0519}_{-0.0475}$	0.1		
	rev-halofit	$0.8127^{+0.0079}_{-0.0063}$		$0.2909^{+0.0095}_{-0.0086}$		$0.9741^{+0.045}_{-0.0555}$		$0.6884^{+0.0599}_{-0.052}$		$-1.0127^{+0.1171}_{-0.1046}$	
Stage IV	mead	$0.7968^{+0.0067}_{-0.0069}$	1.73	$0.2979^{+0.0106}_{-0.0092}$	0.53	$0.9426^{+0.0431}_{-0.0466}$	0.45	$0.698^{+0.0563}_{-0.0449}$	0.13	$-1.106^{+0.1107}_{-0.1163}$	0.61
	halofit	$0.7902^{+0.007}_{-0.0073}$	2.39	$0.2856^{+0.0093}_{-0.0096}$	0.42	$0.9306^{+0.047}_{-0.0479}$	0.6	$0.6986^{+0.0606}_{-0.0507}$	0.13	$-1.0646^{+0.1069}_{-0.1197}$	0.35
w CDM	euclid	$0.8061^{+0.0073}_{-0.0072}$	0.68	$0.2908^{+0.0088}_{-0.0094}$	0.01	$0.9671^{+0.0574}_{-0.053}$	0.09	$0.6968^{+0.0609}_{-0.0604}$	0.1	$-1.046^{+0.1142}_{-0.1288}$	0.22
	euclid2	$0.7996^{+0.0078}_{-0.0069}$	1.31	$0.2901^{+0.0099}_{-0.0094}$	0.06	$0.9791^{+0.0515}_{-0.0588}$	0.07	$0.6711^{+0.0657}_{-0.0548}$	0.21	$-1.0965^{+0.1288}_{-0.1255}$	0.51

**Figure B1.** Deviations of the parameter constraints on S_8 , for the stage IV survey, with the Λ CDM model, and different ℓ_{\max} [$\ell_{\max} = 1000$ (red) and $\ell_{\max} = 800$ (green)].

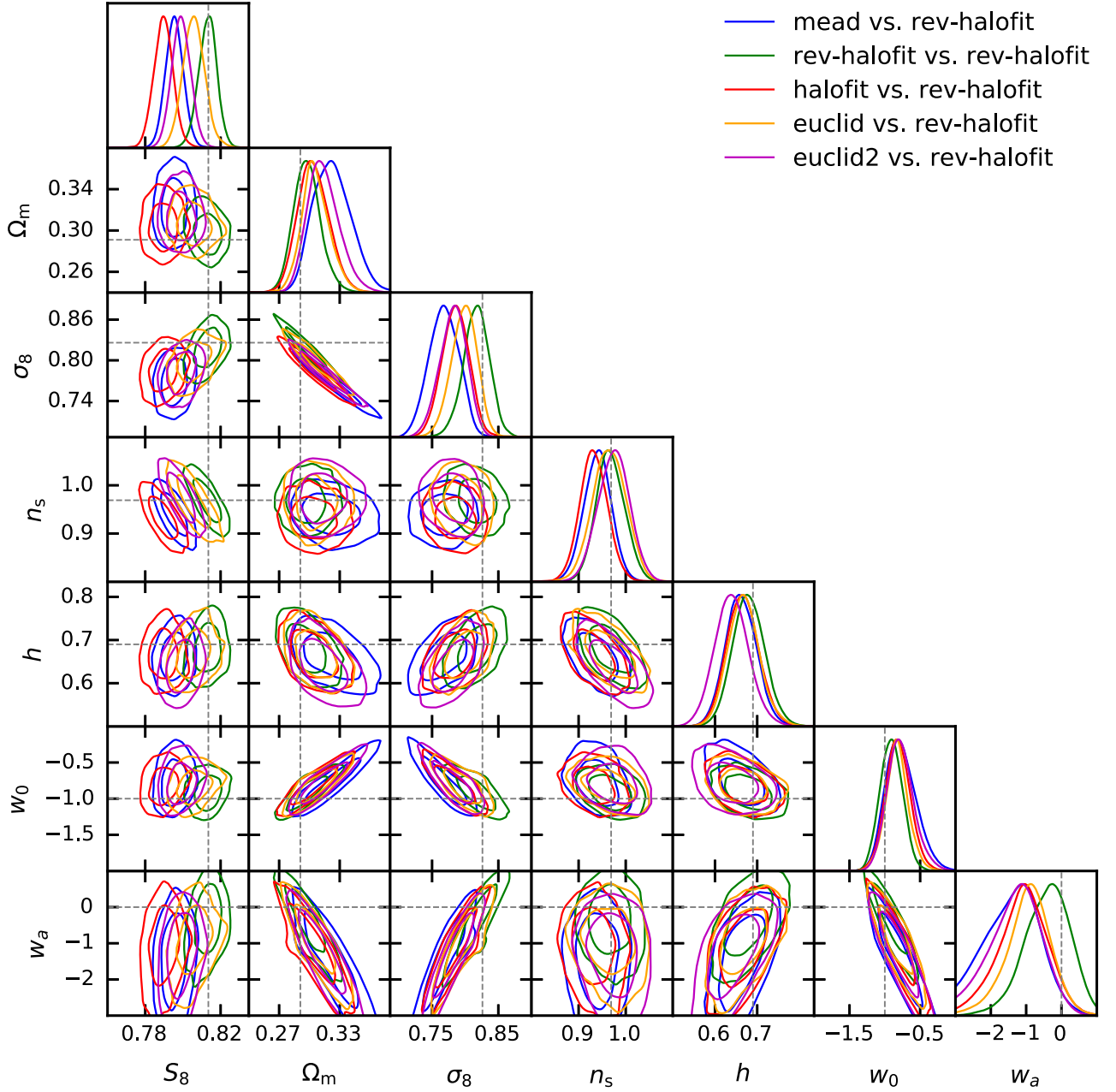


Figure B2. Cosmological parameter constraints of the stage IV survey in the w CDM cosmological model. The discrepancies between the predictors are alleviated, taking into account a simple w CDM cosmological model with both varying w_0 and w_a .

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