Euclid: Impact of non-linear and baryonic feedback prescriptions on cosmological parameter estimation from weak lensing cosmic shear


(Affiliations can be found after the references)

Received 3 November 2020 / Accepted 12 April 2021

ABSTRACT

Upcoming surveys will map the growth of large-scale structure with unprecedented precision, improving our understanding of the dark sector of the Universe. Unfortunately, much of the cosmological information is encoded on small scales, where the clustering of dark matter and the effects of astrophysical feedback processes are not fully understood. This can bias the estimates of cosmological parameters, which we study here for a joint analysis of mock Euclid cosmic shear and Planck cosmic microwave background data. We use different implementations for the modelling of the signal on small scales and find that they result in significantly different predictions. Moreover, the different non-linear corrections lead to biased parameter estimates, especially when the analysis is extended into the highly non-linear regime, with the Hubble constant, H0, and the clustering amplitude, σ8, affected the most. Improvements in the modelling of non-linear scales will therefore be needed if we are to resolve the current tension with more and better data. For a given prescription for the non-linear power spectrum, using different corrections for baryon physics does not significantly impact the precision of Euclid, but neglecting these correction does lead to large biases in the cosmological parameters. In order to extract precise and unbiased constraints on cosmological parameters from Euclid cosmic shear data, it is therefore essential to improve the accuracy of the recipes that account for non-linear structure formation, as well as the modelling of the impact of astrophysical processes that redistribute the baryons.

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

1. Introduction

Next-generation surveys (stage IV) of the cosmic large-scale structure will greatly improve both the amount and quality of data for cosmological investigations. For instance, in the coming decade the surveys carried out by Euclid1, the Vera C. Rubin Observatory2, and the Nancy Grace Roman Space Telescope3 will probe scales and redshifts that were previously inaccessable. The correct interpretation of such a large amount of high-quality data, however, poses a challenge for our theoretical modelling.

In this paper we explore the current modelling limitations for one of the most promising probes: cosmic shear, the measurement of the apparent distortions of galaxy shapes caused by the weak lensing (WL) effect of intervening matter between us and distant sources (see Kilbinger 2015, for a recent review). It provides a direct way to trace the distribution of matter, and as a result it can be used to infer the total matter power spectrum, P_m(k, z). In contrast, the galaxy power spectrum, P_g(k, z), estimated from the clustering of galaxies, depends on the galaxy bias and how galaxies occupy non-linear structures.

A complication is that much of the constraining power of the cosmic shear signal relies on the ability to interpret scales far into the non-linear regime, corresponding to wavenumbers k ≈ 7 hMpc−1 (e.g., Huterer & Takada 2005; Sembolini et al. 2011; Taylor et al. 2018b). On those scales, perturbations of the matter density field are no longer small, and linear theory cannot be used to predict the evolution of large-scale structures.

There are theoretical approaches for predicting clustering beyond the limit of linear theory, such as: standard perturbation theory (Blas et al. 2014; see also Bernardeau et al. 2002 for a...
detailed review) that includes higher-order terms; renormalised perturbation theory (Crocce & Scoccimarro 2006; Crocce et al. 2012; Blas et al. 2016); response functions (Nishimichi et al. 2016); effective field theory (Baumann et al. 2012); and the reaction method (Cataneo et al. 2019). These methods are able to achieve accuracies on power spectrum predictions of ≈1% with respect to numerical simulations, up to scales k ≤ 0.3 h Mpc⁻¹ (see e.g., Foreman & Senatore 2016; Beutler et al. 2017; D’Amico et al. 2020). This is sufficient for modelling the mildly non-linear regime, where the baryonic acoustic oscillation peak is located, but these techniques cannot be used to predict the signal in the highly non-linear regime that WL analyses will probe. The reaction method, however, can in principle achieve an accuracy of ≈1% up to scales k ≤ 10 h Mpc⁻¹ when using the results of an emulator as a baseline (Giblin et al. 2019; Cataneo et al. 2020).

The common approach of modelling the non-linear part of the power spectrum instead relies on fitting formulae determined from comparisons to N-body simulations of cold dark matter particles (e.g., Halofit; Smith et al. 2003). While very economical in terms of computational time, these fitting formulae have a limited range of applicability because the simulations they are based on assume a specific model – usually the cosmological constant cold dark matter (ΛCDM) model or minimal extensions with a constant dark energy equation of state parameter allowed to be different from –1. Thus, applying these corrections to models outside the range constrained by the simulations, for example a more general dark energy fluid, may lead to biased results (see e.g., Casarini et al. 2011a; Seo et al. 2012).

Recently, so-called emulators – based on a large suite of simulations, such as the Coyote Universe (Heitmann et al. 2010, 2009, 2014; Lawrence et al. 2010), the Mira Universe (Heitmann et al. 2016; Lawrence et al. 2017), the Euclid Emulator Project (Knabenhans et al. 2019), and the BACCO simulation project (Angulo et al. 2020) – have been proposed as an alternative to fitting formulae. Emulators interpolate between high-resolution simulation runs at key points (nodes) in the cosmological parameter space. The main advantage of an emulator with respect to a fitting formula is that it does not degrade the accuracy of the corrections over the parameter space sampled by the simulations, such as the range in redshift and scales.

Another way to predict the matter power spectrum on small scales is provided by HMCode⁴ (Mead et al. 2015), which is based on the analytical halo model (Peacock & Smith 2000; Seljak 2000; Cooray & Sheth 2002) and tuned to match the Coyote Extended Emulator (Heitmann et al. 2014) results. It has subsequently been improved to include effects of neutrinos, chameleon and Vainshtein screened models, and dynamical dark energy (Mead et al. 2016) ⁵.

All these methods depend on the quality of the simulations on which they are based, in addition to the declared quality of the method itself (for example the accuracy of the fitting formula or the accuracy of the interpolation of the emulators). Restricting ourselves to just the non-linear clustering, the agreement between purely dark matter simulations is limited by the size of the simulated volume, the number of particles employed in the simulation, and the choice of initial conditions (see e.g., Casarini et al. 2015; Schneider & Teyssier 2015). Hence, part of the differences between the methods described above may be attributed to the simulation parameters on which they are based (e.g., when dimension and resolution are insufficient) rather than the methods themselves.

Moreover, our ability to extract cosmological information from WL measurements on small scales is limited further by baryonic feedback processes (Semboloni et al. 2011) because gas cooling, star formation, galactic winds, supernova explosions, and feedback from active galactic nuclei (AGNs) modify the expected distribution of matter on small scales (Jing et al. 2006; Rudd et al. 2008; Casarini et al. 2011b, 2012; van Daalen et al. 2011; Castro et al. 2018; Debackere et al. 2020). Accurate predictions of the matter power spectrum on those scales require hydrodynamical simulations that not only need to reproduce the non-linear clustering of cold dark matter particles, but should also reliably describe the baryonic component. Such hydrodynamical simulations are much more demanding in terms of computational resources, and the impact of baryonic feedback extracted from hydrodynamical runs has to be modelled and incorporated in the reconstruction of the cosmic shear signal (see e.g., Schneider & Teyssier 2015; Schneider et al. 2016, for a method to include the impact of feedback in the data analysis pipeline). In order to reduce the computational power requested to include such effects, methods relying on ‘baryonisation’ algorithms that mimic the effects of astrophysical processes induced by baryons have been proposed (see e.g., Aricò et al. 2021, 2020).

Given the cost of simulating large volumes with high resolution for every necessary point in the parameter space for each specific cosmological model, we are particularly interested in techniques that can drastically reduce the number of simulations (see Linder & White 2005; Francis et al. 2007; Casarini et al. 2009). As an example, in this work we use the PRequα⁶ method, which allows us to determine the non-linear power spectrum of a dynamical dark energy model at a particular redshift with an ensemble of non-linear spectra of simpler constant w models (Casarini et al. 2009, 2016).

In this paper, we investigate the impact of different implementations of the non-linear corrections and baryonic feedback prescriptions on cosmological parameter estimation, adopting cosmic shear measurements from Euclid (Laureijs et al. 2011) as our baseline (for an analysis of the impact of non-linear corrections on other LSS observables see e.g., Safi & Farhang 2020). We note that we do not investigate here other effects that might affect the parameter estimation pipeline, such as the common assumption of a Gaussian likelihood, which is only an approximation. The paper is organised as follows. After describing the various non-linear recipes in Sect. 2, in Sect. 3 we summarise the Euclid specifications relevant for our analysis. As a first example of the impact of non-linear prescriptions, we assess in Sect. 4 the constraining power of the Euclid survey on the dark energy parameters, here assumed to be described by the so-called CPL parameterisation (Chevalier & Polarski 2001; Linder 2003). In Sect. 5 we quantify the shifts in cosmological parameters when a wrong correction pipeline is used, focusing on the combination of Planck and mock Euclid data. In Sect. 6 we examine the impact of baryonic feedback.

2. Available non-linear prescriptions

In this section we describe the techniques that we use in this study to compute the matter power spectrum in the deeply non-linear regime.

---

⁴ https://github.com/alexander-mead/HMCode
⁵ During the preparation of this work an update of HMCode was published, as described in Mead et al. (2021).
⁶ https://github.com/luciano-casarini/PRequα
2.1. Halofit

One of the first widely used prescriptions to model the non-linear part of the power spectrum, called Halofit, was developed by Smith et al. (2003). The authors measured the non-linear evolution of the matter power spectra using a large library of cosmological $N$-body simulations with power-law initial spectra (Jenkins et al. 1998).

The Halofit approach is based on the halo model (Peacock & Smith 2000; Seljak 2000; Ma & Fry 2000), in which the density field is described in terms of the distribution of isolated dark matter haloes. The correlations in the field are assumed to arise from the clustering of haloes with respect to each other on large scales, and through the clustering of dark matter particles within the same halo on small scales. The total non-linear power spectrum, $P_{NL}(k)$, can then be decomposed into

$$P_{NL}(k) = P_0(k) + P_1(k),$$

where $P_0(k)$ is the quasi-linear term related to the large-scale contribution to the power spectrum, and $P_1(k)$ describes the contribution from the self-correlation of haloes. These terms are also known as the two-halo and the one-halo term, respectively, and we discuss them in this order below.

Seljak (2000), Ma & Fry (2000), Scoccimarro et al. (2001) proposed to use linear theory filtered by the effective window that corresponds to the distribution of haloes as a function of mass, $n(M)$, convolved with their density profiles, $\tilde{\rho}(k, M)$, and a prescription for their bias with respect to the underlying mass field, $b_1(M)$. The quasi-linear term can then be expressed as

$$P_0(k) = P_L(k) \left[ \frac{1}{k^3} \int dM b_1(M)n(M)\tilde{\rho}(k, M)^2 \right],$$

with $\tilde{\rho}$ the homogeneous background matter density, and $P_L(k)$ the linear power spectrum.

A simpler approach was proposed by Peacock & Smith (2000), who assumed that the quasi-linear term corresponds to pure linear theory, $P_0(k) = P_L(k)$. However, quasi-linear effects must modify the relative correlations of haloes away from linear theory, irrespective of the allowance made for the finite size of the haloes (see Smith et al. 2003, and references therein). Halofit takes then an empirical approach, allowing the quasi-linear term to depend on $n(M)$, and truncating its effects at small scales. If we define the dimensionless power spectrum as

$$\Delta^2(k) \equiv \frac{k^3}{2\pi^2} P(k),$$

the quasi-linear term in Halofit is given by

$$\Delta^2_{NL}(k) = \Delta_L^2(k) \left[ 1 + \frac{\Delta_L^2(k)\beta_1 \gamma_1}{1 + \alpha_1 \Delta_L^2(k)} \right] e^{-f(y)},$$

where $y \equiv k/k_{sr}$, $k_{sr}$ is a non-linear wavenumber related to the spherical collapse model (Press & Schechter 1974; Sheth & Tormen 1999; Jenkins et al. 2001), $\alpha_1$ and $\beta_1$ are coefficients sensitive to the input linear spectrum, and $f(y) = y^4/4 + y^2/8$ governs the decay rate at small scales.

To describe the clustering of matter on small scales, we need a description for $P_1$, which is given by

$$P_1(k) = \frac{1}{\rho^2(2\pi)^3} \int dM n(M)\tilde{\rho}(k, M)^2,$$

In order to model this term, we can use an expression that looks like a shot-noise spectrum on large scales, but progressively vanishes on small scales by the filtering effects of halo profiles and the mass function. A good candidate is

$$\Delta^2_1(k) = \frac{a_n y^3}{1 + b_n y + c_n y^n \gamma^2},$$

where $a_n$, $b_n$, $c_n$, and $\gamma_n$ are dimensionless numbers that depend on the input spectrum.

Cooray & Sheth (2002), however, showed that with this expression the halo model disagrees with low-order perturbation theory in some cases. To solve this, Halofit modifies Eq. (6) to obtain a spectrum steeper than Poisson on the largest scales

$$\Delta^2_1(k) = \frac{\Delta^2_0(k)}{1 + \mu y^{-1} + \nu y^{-2}},$$

where $\mu$ and $\nu$ are, again, coefficients that depend on the input spectrum. Smith et al. (2003) showed that Halofit is able to reproduce the measurements from simulations more accurately, and down to smaller scales, than the halo model.

2.2. Halofit with Bird and Takahashi corrections

All the coefficients used in the Halofit recipe were determined by Smith et al. (2003) from a fit to cold dark matter simulations in boxes of lengths of 256 $h^{-1}$ Mpc containing 256$^3$ particles (Jenkins et al. 1998). As a consequence of the relatively large particle mass, Halofit may not be suitable if we want to test cosmologies with massive neutrinos, or go down to very small scales, where the impact of baryonic interactions is non-negligible. Moreover, the limited simulation volume results in large sample variance (see Casarini et al. 2015; Schneider & Teyssier 2015), and thus may lead to inaccurate results even for a $\Lambda$CDM cosmology (White & Vale 2004; Casarini et al. 2009, 2012; Hilbert et al. 2009; Heitmann et al. 2010).

To address these limitations, Bird et al. (2012) investigated the impact of massive neutrinos, and performed several $N$-body simulations of the matter power spectrum incorporating massive neutrinos with masses between 0.15 and 0.6 eV. They focussed on non-linear scales below $10 h$ Mpc at $z < 3$, and extended the Halofit approximation to account for massive neutrinos. They found that in the strongly non-linear regime Halofit overpredicts the suppression due to the free-streaming of the neutrinos. In particular, the asymptotic behaviour of the non-linear term in Halofit is given by $\Delta^2_{NL} \sim y^{\gamma_n}$, and therefore Bird et al. (2012) adjusted $\gamma_n$ to their $\Lambda$CDM simulations with massive neutrinos. Moreover, they modified the non-linear power spectrum with the ansatz

$$(\Delta_{NL}^2)^2 = \Delta_{NL}^2(1 + Q_\nu),$$

with

$$Q_\nu = \frac{f_\nu}{1 + k^3 m}$$

where $f_\nu = \Omega_\nu/\Omega_m$ is the ratio between the neutrino and total matter energy densities, and $l$ and $m$ are fitted to the simulations. They also modified Eq. (4) to

$$\Delta^2_1(k) = \Delta_1^2(k) \left[ 1 + \frac{\Delta_1^2(k)\beta_1 \gamma_1}{1 + \alpha_1 \Delta_1^2(k)} \right] e^{-f(y)},$$

with

$$Q_\nu = \frac{1 f_\nu}{1 + k^3 m}.$$
with
\[ \hat{\Lambda}_n^2 = \Lambda_n^2 \left( 1 + \frac{pf_0^2k^2}{1 + 1.5k^2} \right), \]  
(11)
\[ \hat{\beta}_n = \beta_n + f_s \langle r + n^2 \rangle, \]  
(12)
where \( p, r, \) and \( s \) are fitted to the simulations.

Another important improvement to the original Halofit was introduced in Takahashi et al. (2012), who updated the fitting parameters using high-resolution \( N \)-body simulations with box sizes of \( L = 300-2000 h^{-1} \) Mpc with \( n_p = 1024^3 \) particles each, for 16 cosmological models around the best fitted cosmological parameters from WMAP data (Hinshaw et al. 2013), including dark energy models with a constant equation of state. This revised version of Halofit provides an accurate prediction of the non-linear matter power spectrum down to \( k \sim 30 h^{-1} \) Mpc\(^{-1}\) and up to \( z \geq 3 \) with an accuracy \( \sim 5-10\% \). In the remainder of this paper we refer to the non-linear prescription that includes the improvements from Takahashi et al. (2012) and Bird et al. (2012) as Halofit, for simplicity.

### 2.3. Halofit with P\( K \)equal

One of the limitations of the standard Halofit approach, even after the corrections by Bird et al. (2012) and Takahashi et al. (2012) are considered, is that it is based on a fit to \( N \)-body simulations with a constant value for the dark energy equation of state. However, given the precision of stage IV surveys, we are particularly interested in determining whether the data prefer an evolving dark energy equation of state. To avoid biases in our non-linear predictions, we could run \( N \)-body simulations that include a time dependence for \( w \), such as the CPL parameterisation (Chevallier & Polarski 2001; Linder 2003) given by

\[ w(a) = w_0 + w_a (1 - a). \]  
(13)

This approach implies the need for significant computational resources. Another option, however, is to map the non-linear power spectra of dark energy models with a constant equation of state to those with a time varying one. In this work we consider the P\( K \)equal1 code (Casarini et al. 2016), which implements the spectral equivalence from Casarini et al. (2009). Francis et al. (2007) showed how predictions for constant \( w \) models at \( z = 0 \) can be related to the power spectra of cosmologies with an evolving equation of state \( w(a) \) given by Eq. (13) with an accuracy \( \sim 0.5\% \) up to \( k = 1 h \) Mpc\(^{-1}\). The P\( K \)equal1 technique achieves this precision also at \( z > 0 \) for a general equation of state \( w = w(a) \) by imposing the equivalence of the distance to the last scattering surface and requiring that the amplitudes of the density fluctuations at the redshift of interest are the same. For a given set of values of \( w_0 \) and \( w_a \), these two conditions yield at each \( z \) a unique value of \( w_{eq} \) and \( \sigma_8_{eq} \) for the constant \( w \) model.

The performance of this method has been tested for several dark energy models (Casarini et al. 2009; Casarini 2010), and also in the presence of gas cooling, star formation and SN feedback (Casarini et al. 2011b). These studies find differences in power spectra between the mapped dynamical dark energy models and the ensemble of equivalent constant \( w \) models that are within 1% up to \( k \sim 2-3 h \) Mpc\(^{-1}\). With this method it is possible to extend both emulators (see Casarini et al. 2016) and fitting formulae (as in this work) to dynamical dark energy models if they are valid for constant \( w \) models.

### 2.4. H\( M \)Code

An alternative approach to predict the non-linear matter power spectrum, called H\( M \)Code, was proposed by Mead et al. (2015). It introduces physically motivated free parameters into the halo model formalism, instead of using empirical fitting functions. Mead et al. (2015) fit these to \( N \)-body simulations with box sizes \( L = 90-1300 h^{-1} \) Mpc and \( n_p = 512^3-1024^3 \) particles for a variety of \( \Lambda \)CDM and \( w \)CDM models (Heitmann et al. 2010). H\( M \)Code also accounts for the effects of baryonic feedback on the power spectrum by fitting the halo model to hydrodynamical simulations that include gas cooling, star formation, as well as supernova and AGN feedback (Schaye et al. 2010; van Daalen et al. 2011).

In Mead et al. (2016) H\( M \)Code was updated to account for massive neutrinos, chameleon and Vainshtein screening mechanisms, as well as evolving dark energy equations of state described by the CPL parameterisation. Throughout the rest of the paper we will consider this latest version when referring to H\( M \)Code. We note, however, that the P\( K \)equal1 approach can also be used with the original H\( M \)Code by applying it to the equivalent \( \{w_{eq}, 0\} \) constant \( w \) model at any redshift \( z \), and imposing the same \( \sigma_8(z) \). This yields the prediction for a given \( \{w_0, w_a\} \) CPL model, with an accuracy similar to that of the fit and the simulations on which H\( M \)Code itself is based.

### 2.5. Comparison

Before we study how the different methods affect the inference of cosmological parameters and related confidence regions, we show by how much the power spectra differ as a function of the wavenumber \( k \) for the prescriptions described above. We note that theoretical approaches based on perturbation theory or the effective field theory of large-scale structure do not yet provide accurate power spectra for the small scales we consider here.

In Fig. 1 we show the relative difference between the various predictions and the one from Halofit. The power spectra were computed using \( \text{CAMB} \) (Lewis et al. 2000) at redshift \( 0 \), where discrepancies are most pronounced. Hence, these represent a worst case scenario as WL probes mostly intermediate redshifts. The upper-left panel shows the relative differences between the linear power spectrum (orange), Halofit with P\( K \)equal1 (blue), and H\( M \)Code (red) for the \( \Lambda \)CDM model using the Planck TTTEEE+lowE+lensing+BAO 2018 mean values for the various cosmological parameters (Planck Collaboration VI 2020). The linear power spectrum diverges from Halofit for \( k > 0.1 h \) Mpc\(^{-1}\), while the disagreement between Halofit and H\( M \)Code is below 7.5% for \( k \leq 10 h \) Mpc\(^{-1}\). While in general non-linear corrections boost the matter power spectrum, Fig. 1 shows how the linear power spectrum takes larger values than the corrected one in the range 0.005 \( \leq k \leq 0.1 \). This feature, which is slightly counter-intuitive though well known, is due to the effect of the quasi-linear term of the halo model, relevant at these scales, that can lead to a suppression of power with respect to the linear case (Smith et al. 2003).

The upper-right panel shows the variation on small scales when we consider parameters around the mean values of \( \Lambda \)CDM. For this comparison we generated 100 power spectra with cosmological parameters drawn from a five-dimensional Gaussian distribution with the diagonal given by five times the 1\( \sigma \) uncertainty quoted by Planck TTTEEE+lowE+lensing+BAO 2018.\footnote{Note that we fix the value of \( \tau \) to its mean value; therefore we are left with a five-dimensional Gaussian instead of the six-dimensional one that would correspond to \( \Lambda \)CDM.}
These results highlight that the discrepancy between Halofit and HMCode can be larger than 10% for $k > 3 \, h \, \text{Mpc}^{-1}$.

The bottom-left panel considers models beyond $\Lambda$CDM, allowing for a constant equation of state parameter, $w$, that can differ from $-1$. The different lines correspond to power spectra when we draw parameters from a six-dimensional Gaussian distribution where we adopt a dispersion of 0.3 for $w$ (but we do require $w < -1/3$). The bands showing the discrepancy between the different non-linear corrections increase slightly (a 10% discrepancy is reached at scales of $k = 1-2 \, h \, \text{Mpc}^{-1}$), but the overall shape remains the same.

Finally, in the bottom-right panel we consider dynamical dark energy models with a dark energy equation of state given by Eq. (13). In this case we add $w_a$ (so that we draw parameters from a seven-dimensional Gaussian) with a dispersion of 1.0 and we require the dark energy equation of state to be always smaller than $-1/3$, that is $w_0 + w_a < -1/3$. The overall shape for the bands is the same for HMCode and the linear spectra, although the discrepancies are significantly larger (already 10% at scales of $k = 0.6 \, h \, \text{Mpc}^{-1}$). We also note that, since we allow $w_a$ to vary, there is a difference between Halofit and Halofit+PKequal.

Lawrence et al. (2017) presented an updated version of COSMIC EMU that includes massive neutrinos. Their high-resolution simulations are interpolated with an accuracy of $\sim 4\%$. Lawrence et al. (2017) compared their COSMIC EMU predictions with Halofit, and HMCode. When massive neutrinos are considered, the different approaches show differences of $\sim 20\%$ and $\sim 15\%$, respectively, in the power spectra for scales above $k = 0.1 \, h \, \text{Mpc}^{-1}$, indicating the need for further improvements.

3. The Euclid Cosmic Shear survey

Our aim is to investigate how the expected constraints on cosmological parameters from Euclid data depend on the recipe that is used to predict the matter power spectrum on non-linear scales, although we note that our finding are also relevant for other stage IV experiments. Euclid an M-class mission of the European Space Agency (Laureijs et al. 2011) that will carry out a spectroscopic and a photometric survey of galaxies over an area of 15,000 deg$^2$. The cosmic shear measurements use high-quality imaging at optical wavelengths, supported by multi-band optical ground-based photometry and near-infrared observations by Euclid. The telescope is designed so that (residual) instrumental sources of bias in the observed cosmic shear signal are subdominant compared to the statistical uncertainties (e.g., Cropper et al. 2013; Euclid Collaboration 2020b). However, to achieve its objectives, it is essential that the signal can be accurately predicted in the non-linear regime. Although this is also relevant to fully exploit the data from the clustering of galaxies and the cross-correlations with the lensing signal, we focus on the cosmic shear case in this paper and defer a more comprehensive study to future work.

We adopt the baseline specifications for the Euclid data, which are described in Euclid Collaboration (2020a, hereafter EC19). The redshift distribution of the sources is given by

$$n(z) \propto \left(\frac{z}{z_0}\right)^2 \exp\left[-\left(\frac{z}{z_0}\right)^{3/2}\right],$$

with $z_0 = 0.9/\sqrt{2}$, resulting in a mean redshift of $\langle z \rangle = 0.96$. The sample is divided into 10 equi-populated redshift bins $n(z)$. 

Fig. 1. Relative difference between the various predicted matter power spectra and the Halofit prediction at $z = 0$ for different cosmologies. Top-left panel: using the mean values of the $\Lambda$CDM parameters given by Planck 2018 (Planck Collaboration VI 2020). Top-right panel: differences when the $\Lambda$CDM parameters are drawn from a five-dimensional Gaussian distribution with the diagonal given by five times the 1$\sigma$ uncertainty quoted by Planck TTTT+EE+lowE+lensing+BAO 2018. Bottom-left panel: differences when also the constant dark energy equation of state is drawn from a Gaussian distribution with a dispersion of 0.3. Bottom-right panel: difference for evolving dark energy models where, alongside $w_0$, also $w_a$ is Gaussian distributed with a dispersion of 1.0. In the bottom panels $w_0$ and $w_a$ (if present) are always chosen in such a way that $w(z) < -1/3$. The feature visible at $k \approx 0.005 \, h \, \text{Mpc}^{-1}$ corresponds the scale at which the non-linear corrections are turned on in CAMB.
(with the index $i$ indicating the tomographic bin). We assume an average number density of galaxies with precise shape measurements of $n_g = 30$ arcmin$^{-2}$. To capture the noise arising from the intrinsic galaxy ellipticities we adopt a dispersion of $\sigma_\epsilon = 0.21$ for each of the two ellipticity components. We also use the same approach as EC19 to compute the WL power spectrum, defined as

$$C_{ij}^\ell(z) = c \int dz \frac{W_i^i(z)W_j^j(z)}{H(z)z^2} P_{\delta \delta} \left[ \frac{f + 1/2}{r(z)} \right]^2,$$

(15)

where $P_{\delta \delta}$ is the non-linear matter power spectrum, $r(z)$ is the comoving distance to redshift $z$, and the window function $W_i^i$ is defined as

$$W_i^i(z) = W_i^0(z) - \frac{\mathcal{A}_i C_{IA} \Omega_m \Omega_b}{D(z)c} F_{IA}(z) H(z) n_i(z),$$

(16)

with $n_i(z)$ normalised such that $\int dz n_i(z) = 1$. In Eq. (16), the first term corresponds to the usual lensing kernel,

$$W_i^0(z) = \frac{3}{2} \Omega_m \left( \frac{H_0}{c} \right)^2 (1 + z) r(z) \times \int_{z}^{z_{\text{max}}} dz' n_i(z') \left[ 1 - \frac{r(z)}{r(z')} \right],$$

(17)

whilst the second term models the effect of intrinsic alignments (IA). Here $D(z)$ is the growth factor, $C_{IA} = 0.0134$ is a constant so that the normalisation of the model $\mathcal{A}_i$ can be compared to the literature. To describe the dependence of the IA signal as a function of scale, redshift and galaxy luminosity, we adopt the extended non-linear alignment model (Joachimi et al. 2015), and the function $F_{IA}(z)$ is given by

$$F_{IA}(z) = (1 + z)^{\eta_{IA}} [L(z)/L_*]^{\beta_{IA}}.$$  

(18)

where $\langle L(z) \rangle$ and $L_*(z)$ are the redshift-dependent mean and a characteristic luminosity of source galaxies as computed from the luminosity function, respectively. The parameters $\eta_{IA}$, $\beta_{IA}$, and $\mathcal{A}_i$ are free parameters that can be determined observationally. We use $\mathcal{A}_i = \{1.72, -0.41, 2.17\}$ as fiducial values, as was done in EC19, while they are allowed to vary in the parameter estimation.

Our approach for the modelling of IA is phenomenological, but we note that more physically motivated models have been proposed. For instance, Fortuna et al. (2021) used a halo model approach to link direct observations of IA to implications for cosmic shear. Finally, we note that the modelling of the IA signal is linked to non-linear structure formation, but exploring this further is beyond the scope of this paper.

We adopt the same approach presented in EC19 to model the covariance of the WL observable. As usual, it can be split into Gaussian and non-Gaussian contributions (Takada & Hu 2013; Cooray & Hu 2001; Hamilton et al. 2006; Hu & Kravtsov 2003; Kayo et al. 2012). The latter involve the convergence trispectrum, but there are significant uncertainties on how to model this quantity. Relying on the halo-model formalism (see e.g., Cooray & Hu 2001), it was shown in EC19 that the signal-to-noise ratio of the shear power spectrum (with the specifications used in this work) decreases by 30% at $\ell_{\text{max}} = 5000$ when we add the non-Gaussian contribution. This loss of information corresponds to an effective cut at $\ell_{\text{max}} = 1420$ when using only a Gaussian covariance. Following EC19, we decide to avoid uncertainties related to the non-Gaussian modelling and use a Gaussian covariance with a cut at $\ell_{\text{max}} = 1500$. This corresponds to the pessimistic scenario. We also consider the impact of changing this cut in the optimistic settings, where we set it at $\ell_{\text{max}} = 5000$. It is important to note that an $\ell$-cut preserving the signal-to-noise ratio is not directly linked to constraints on the parameters (see e.g., Copeland et al. 2018), but the amplitude of the impact of the non-Gaussian modelling used here is small compared to other assumptions.

### 4. Impact of non-linear corrections on forecast constraints

Given our desire to use measurements on small scales to estimate cosmological parameters, it is essential to assess how the different methods to model the power spectrum on highly non-linear scales affect the results. In the following, we limit ourselves to a simple extension of the $\Lambda$CDM model: We assume that dark energy is dynamical with its equation of state parameterised according to Eq. (13), and that density perturbations for this component are well described by the parameterised post-Friedmann approach, which assumes that the dark energy field remains smooth with respect to matter at the scales of interest (Hu & Sawicki 2007; Hu 2008; Fang et al. 2008).

To evaluate the impact of the different recipes for the non-linear power spectrum on the final parameter estimation from Euclid, we use the Fisher matrix approach (see EC19, for an extensive review of the methodology). For the different methods we compute the figure of merit (FoM) for the parameters $w_0$ and $w_a$, where the FoM is defined as

$$\text{FoM} = \sqrt{\text{det} \mathbf{F}_{w_0 w_a}}.$$  

(19)

and $\mathbf{F}_{w_0 w_a}$ is the Fisher matrix marginalised over all the cosmological parameters except for $w_0$ and $w_a$. This allows us to examine whether degeneracies between cosmological parameters differ when switching from one method to the other. We can also quantify the biases in dark energy parameters and changes in the FoM, which captures the performance of Euclid.

The free parameters for this analysis are: the matter and baryon density parameters $\Omega_m$ and $\Omega_b$; the dark energy parameters, $w_0$ and $w_a$; the spectral index of primordial perturbations, $n_s$; the dimensionless Hubble parameter, $h$; and $\sigma_8$, that is the root mean square of present-day linearly evolved density fluctuations in spheres of $8 h^{-1}$ Mpc radius. The fiducial values of these parameters are listed in Table 1.

<table>
<thead>
<tr>
<th>$\Omega_m$</th>
<th>$\Omega_b$</th>
<th>$w_0$</th>
<th>$w_a$</th>
<th>$n_s$</th>
<th>$h$</th>
<th>$\sigma_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.32</td>
<td>0.05</td>
<td>-1</td>
<td>0</td>
<td>0.960</td>
<td>0.67</td>
<td>0.816</td>
</tr>
</tbody>
</table>

Table 1. Fiducial values for the cosmological parameters considered in the Fisher matrix analysis.

We used the specification of Euclid for WL observations that were detailed in Sect. 3 and computed the FoM from the WL power spectra $C_{ij}^\ell$ for different values of the maximum multipole, $\ell_{\text{max}}$. We considered two different values for $\ell_{\text{max}}$, namely 1500 and 5000, with the latter probing deeper into the non-linear regime. We note that here the same $\ell_{\text{max}}$ applies to all redshift bins, thus leading to a different cut in scales ($k_{\text{max}}$) for each bin. We investigated the difference between this analysis and one considering a $k_{\text{max}}$ rather than a multipole cut in Appendix A. In both cases the minimum multipole was fixed to $\ell_{\text{max}} = 10$, following EC19. We note that the cuts we used here should be considered an approximation; in general, there is no direct mapping between $\ell_{\text{max}}$ and $k_{\text{max}}$, and a more refined approach would...
be needed to convert a cut in multipoles into a cut in wavenumber (Taylor et al. 2018b).

Table 2 lists the resulting FoM values for our baseline $w_0w_a$CDM cosmology for the three different non-linear recipes. The variation in the predicted FoMs is substantial, which is not surprising given the differences we see in Fig. 1: at large $k$ the differences exceed 10% when the parameters are allowed to vary with respect to the fiducial model.

The large variation is caused by two separate effects. Firstly, the fiducial $C_l^{\ell}$ is obtained by integrating up to $k = 30h\,\text{Mpc}^{-1}$ and then inverted when computing the covariance matrix that enters the Fisher matrix forecast. Small differences can become large in the inversion process. Secondly, what is important is not so much the fiducial model itself, but rather the derivatives of the power spectrum with respect to the model parameters. Different non-linear corrections predict different derivatives, thus leading to different Fisher matrix elements. This is supported by the fact that the FoM for HMCode and Halofit+PKequal are very similar. These two prescriptions explicitly take deviations of $w_0$ and $w_a$ from the fiducial $\Lambda$CDM values into account, while this is not the case for the standard Halofit, which is designed to describe $\Lambda$CDM cosmologies and extended to cases where dark energy is described by a constant equation of state. As a consequence, the amplitude of the derivatives of the matter power spectrum with respect to $\{w_0, w_a\}$ are similar between HMCode and Halofit+PKequal, but different for Halofit. This explains why the FoM values in the third and fourth columns are so similar. We also tested the impact of the differences in the covariance matrix and in the derivatives with respect to $w_a$, obtaining the FoM through the described procedure, but keeping fixed to the a reference non-linear prescription either the first or the latter.

It is also worth noting that the FoM is highest for Halofit because in that case $w_a$ impacts the power spectra at non-linear scales through its effect on the linear matter power spectrum, while in the other two methods the non-linear corrections are affected by this parameter as well. Overall, the non-linear corrections dampen the derivatives with respect to $w_a$, leading to weaker constraints on this parameter and to a lower FoM. These results demonstrate that the FoM depends critically on the non-linear model that is used, highlighting the need for (more) accurate prescriptions.

In addition to this, Table 2 shows also that the difference in the FoM between HMCode and Halofit decreases in the optimistic case with respect to the pessimistic one. While this might seem counter intuitive, this effect is due to the fact that the derivatives with respect to $w_a$, which are the main parameters for the variation in the FoM value, tend to vanish at high multipoles. For such a reason, including smaller scales in the analysis reduces the impact of such derivatives, and the main difference between the recipe is now in the covariance matrices. However, they are computed at the parameter fiducial value, where the differences between recipes are less significant. Instead, if we compare Halofit with Halofit+PKequal, the change in moving from the pessimistic to the optimistic case is negligible, due to the different construction of the latter recipe, which is a correction of Halofit itself, rather than being based on a different approach.

### 5. Bias on cosmological parameter estimates

A major concern is that inaccuracies in the theoretical predictions on non-linear scales translate into shifts in the inferred cosmological parameters. To quantify the impact of such biases, we create a mock dataset of Euclid WL observations, using the specifications listed in Sect. 3. The fiducial cosmology that we use to generate these mock data is summarised in Table 3, where $A_1$ is the amplitude of the primordial power spectrum, $\omega_{b,0} = \Omega_{b,0}h^2$, $\omega_{c,0} = \Omega_{c,0}h^2$ where $\Omega_{c,0}$ is the present time density parameter for cold dark matter, while for the intrinsic alignment nuisance parameters, also free in this analysis, we assume the fiducial values $\{\eta\Lambda, \beta\Lambda\} = \{1.72, -0.41, 2.17\}$. Again, we follow EC19 in neglecting the uncertainties on the galaxy distributions and shear bias, for simplicity. However, we refer the reader to Tutusaus et al. (2020) and Euclid Collaboration (2020b) for analyses focussed on these systematic uncertainties. In contrast to what was done in Sect. 4, we assume here that the expansion history is provided by a dynamical dark energy, assuming $w_0 = -0.9$ and $w_a = 0.1$. This allows us to assess the impact of the different non-linear recipes when a non-standard model is the true underlying cosmology. This is motivated by the fact that the recipes in Sect. 2 differ in how they account for such extensions of the standard cosmological model.

We adopt the non-linear correction provided by Halofit+PKequal as the reference to which we compare the parameter estimates for the other recipes. We stress that we do not advocate the use of a particular prescription, but rather wish to quantify the shifts in the estimated cosmological parameters that may arise from using a different prediction.

We analyse this mock dataset with a Markov chain Monte Carlo method (MCMC), using the MontePython\footnote{https://github.com/brinckmann/montepython_public} suite (Audren et al. 2013; Brinckmann & Lesgourgues 2019), with a Euclid lensing mock likelihood as presented in
Table 4. Mean values, marginalised 68% errors, and biases in cosmological parameters.

<table>
<thead>
<tr>
<th>θ</th>
<th>ℓ_{max}</th>
<th>θ^*</th>
<th>σ</th>
<th>B</th>
<th>θ^*</th>
<th>σ</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>ω_b,0</td>
<td>1500</td>
<td>0.02244</td>
<td>0.00011</td>
<td>0.08</td>
<td>0.02240</td>
<td>0.00012</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>0.02243</td>
<td>0.00012</td>
<td>0.15</td>
<td>0.02246</td>
<td>0.00012</td>
<td>0.05</td>
</tr>
<tr>
<td>ω_c,0</td>
<td>1500</td>
<td>0.12056</td>
<td>0.00036</td>
<td>0</td>
<td>0.12101</td>
<td>0.00039</td>
<td>1.16</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>0.12054</td>
<td>0.00038</td>
<td>0.053</td>
<td>0.12112</td>
<td>0.00036</td>
<td>1.57</td>
</tr>
<tr>
<td>h</td>
<td>1500</td>
<td>0.6689</td>
<td>0.0069</td>
<td>0.16</td>
<td>0.6702</td>
<td>0.0096</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>0.6683</td>
<td>0.0048</td>
<td>0.36</td>
<td>0.6899</td>
<td>0.0066</td>
<td>3.02</td>
</tr>
<tr>
<td>ln(10^{10}A_s)</td>
<td>1500</td>
<td>3.0591</td>
<td>0.0086</td>
<td>0.09</td>
<td>3.0657</td>
<td>0.0088</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>3.0593</td>
<td>0.0090</td>
<td>0.10</td>
<td>3.0656</td>
<td>0.0086</td>
<td>0.85</td>
</tr>
<tr>
<td>n_s</td>
<td>1500</td>
<td>0.9602</td>
<td>0.0025</td>
<td>0.06</td>
<td>0.9615</td>
<td>0.0027</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>0.9604</td>
<td>0.0023</td>
<td>0.18</td>
<td>0.9556</td>
<td>0.0023</td>
<td>1.90</td>
</tr>
<tr>
<td>w_0</td>
<td>1500</td>
<td>−0.888</td>
<td>0.085</td>
<td>0.14</td>
<td>−0.869</td>
<td>0.099</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>−0.888</td>
<td>0.060</td>
<td>0.21</td>
<td>−1.021</td>
<td>0.064</td>
<td>1.88</td>
</tr>
<tr>
<td>w_a</td>
<td>1500</td>
<td>0.07</td>
<td>0.21</td>
<td>0.14</td>
<td>−0.02</td>
<td>0.25</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>0.07</td>
<td>0.16</td>
<td>0.16</td>
<td>0.29</td>
<td>0.16</td>
<td>1.22</td>
</tr>
<tr>
<td>Ω_m,0</td>
<td>1500</td>
<td>0.3212</td>
<td>0.0065</td>
<td>0.18</td>
<td>0.3209</td>
<td>0.0090</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>0.32164</td>
<td>0.0046</td>
<td>0.36</td>
<td>0.3031</td>
<td>0.0057</td>
<td>2.96</td>
</tr>
<tr>
<td>σ_8</td>
<td>1500</td>
<td>0.7852</td>
<td>0.0058</td>
<td>0.14</td>
<td>0.7938</td>
<td>0.0071</td>
<td>1.09</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>0.7847</td>
<td>0.0041</td>
<td>0.30</td>
<td>0.8080</td>
<td>0.0048</td>
<td>4.62</td>
</tr>
<tr>
<td>Δχ^2</td>
<td>1500</td>
<td>0.60</td>
<td></td>
<td>32.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>1.06</td>
<td></td>
<td>62.34</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes. The values are obtained by fitting mock Planck and Euclid WL data to either HMCode without baryonic feedback or Halofit without PRequal non-linear corrections. The last row shows Δχ^2. By construction χ^2 = 0, unless the configuration of the MCMC sampling does not match the one used to create the fiducial synthetic dataset. The number of degrees of freedom in this case is 11 (the number of free parameters in the model; Ω_m,0 and σ_8 are derived parameters), which enables one to compare Δχ^2 to the corresponding confidence interval.

Sprenger et al. (2019)⁹. With this setting we sample the cosmological parameters listed in Table 3 using either Halofit (switching off PRequal) or HMCode. When using HMCode, the baryonic feedback parameters are fixed to the values fitting the COSMIC EMU dark-matter-only simulations (Heitmann et al. 2014).

As done in Sect. 4, we performed the analysis for ℓ_{max} = 1500 and 5000 in order to explore the impact of the non-linear corrections on the results when including high multipoles. Also in this case, we explore the difference with a scale cut in Appendix A. We complemented the WL mock dataset with TT, TE, EE, and lensing data from a mock Planck likelihood, thus reproducing the sensitivity of the full mission. We note that we did not use real Planck data to avoid a mismatch between the fiducial Euclid cosmology and the actual best fitted Planck values.

This approach enables us to determine the ‘bias’ (B) on cosmological parameters with respect to our fiducial cosmology, that is the offset of the mean of the estimated posterior distribution from the true value, which in turn allows us to quantify the impact of using a different prescription for the non-linear evolution of density perturbations. To assess the significance of a particular bias, it is useful to compare it to the expected statistical uncertainty. We therefore consider the (relative) bias on all the cosmological parameters given by

\[ B(\theta) = \frac{|\theta^* - \theta^{\text{fid}}|}{\sigma}, \]

where θ^* is the mean value of the parameter found with the MCMC analysis, σ is the 68% uncertainty estimated from the chains, and θ^{\text{fid}} is its fiducial value. We note that this estimate implicitly assumes that the posterior distribution can be approximated by a Gaussian, which is indeed valid for constraints based on the combination of Planck and Euclid data.

The statistical uncertainties of Euclid are determined by the survey design (see Sect. 3). To draw reliable conclusions from these data, it is essential that systematic uncertainties are sufficiently small. Ideally biases should vanish, but generally it is too costly to achieve this. A reasonable compromise, however, is to adopt B_{\text{thr}} ≤ 0.1 (see e.g., Sect. 4.1 in Massey et al. 2013), which

⁹ With respect to Sprenger et al. (2019), here we apply a z-independent cut-off at k_{max} = 30 h Mpc^{-1}, and we do not include the theoretical error.
is what we do here\textsuperscript{10}. We note that the intrinsic variance of the mean estimated by the MCMC also contributes to our estimate of $B(\theta)$. We quantified this contribution by computing the scatter of the mean value by bootstrapping the chains. We find that it is always within 1\% of the final error on cosmological parameters.

The results are reported in Table 4 and the posteriors are presented in Fig. 2. Although the optical depth $\tau$ is a free parameter in the model, it is effectively constrained by the Planck measurements alone. We therefore do not report its value here. Oppositely, as a comparison, we can see how constraints on other parameters tighten thanks to the inclusion of Euclid data. For instance, Planck Collaboration VI (2020) report for $\sigma_8$ a 68\% confidence level (C.L.) interval equal to 1.5\% or 1\% of the parameter value in a $w_0w_a$CDM cosmology, when Planck is respectively combined with RSD and WL data or BAO and SN data. Instead, with Planck-TT and Euclid-WL only we obtain here 0.7\%.

As expected, the biases are larger for HMCode compared to Halofit without $\text{P\text{K}equal}$. We also find that for Halofit, increasing the range from $\ell_{\text{max}} = 1500$ to $\ell_{\text{max}} = 5000$ does not increase the bias significantly, whereas the bias strongly depends on the $\ell$-range for HMCode, both in amplitude and in sign. This can be explained by looking at Fig. 1: at scales larger than a few $h\text{Mpc}^{-1}$, HMCode systematically over-predicts the power with respect to Halofit.

We find that in the Halofit case, the biases in the cosmological parameters approximately satisfy $B(\theta) \lesssim B_{\text{thr}}$ when $\ell_{\text{max}} = 1500$, while for HMCode the biases for almost all the parameters exceed this threshold. When setting $\ell_{\text{max}} = 5000$, the parameters estimated in the HMCode case are all biased significantly more than the acceptable threshold (except for $\omega_{0,0}$).

and now also the Halofit case exhibits biases larger than $B_{\text{thr}}$ for $h$, $\Omega_{m,0}$ and $\sigma_8$.

In order to correct for the significant mismatch in the non-linear prescriptions, HMCode increases $\omega_{c,0}$, $h$, $\ln(10^{10}A_s)$, and $w_a$, while at the same time the values for $n_s$ and $w_0$ are decreased. This tweaking of parameters increases the amplitude of the linear matter power spectrum at scales $0.2\ h\text{Mpc}^{-1} \lesssim k \lesssim 2\ h\text{Mpc}^{-1}$, where HMCode has a lack of power with respect to Halofit+$\text{PKequal}$ (see Fig. 1). As the scales around $0.2\ h\text{Mpc}^{-1}$ are those that mainly contribute to the estimate of $\sigma_8$, this explains the large bias observed for this (derived) parameter, $B(\sigma_8) \gtrsim 5$.

Overall, the $\Delta \chi^2 \lesssim 1$ indicates that replacing Halofit + $\text{PKequal}$ with Halofit-only does not have a strong impact on the results as it is well within the range of the statistical uncertainties\textsuperscript{11}. On the other hand, using HMCode leads to a significantly higher $\Delta \chi^2$, highlighting how the difference between the two non-linear prescriptions cannot be fully compensated by modifying the background quantities and the linear growth.

It is worth to noting that for both Halofit and HMCode the parameters that are most significantly biased are $H_0$ and $\sigma_8$. These are the parameters that currently show tension between high- and low-redshift measurements (e.g., Riess et al. 2019; Hildebrandt 2020; Spurio Mancini et al. 2019). Our results imply that the Euclid cosmic shear measurements have the statistical power to resolve this, but only if we can accurately model the non-linear scales.

\section{Impact of baryons}

Up to this point, we have limited our study to the impact of changing the recipe that is used to compute the non-linear...
evolution of cold dark matter perturbations. On the small scales of interest, however, baryons collapse into the dark matter haloes to form stars, or are heated up, or even expelled into the intergalactic medium. These processes modify the matter distribution, and it is therefore important to account for baryonic physics when computing the matter power spectrum $P_{\delta\delta}(k, z)$ (e.g., van Daalen et al. 2011; Casarini et al. 2012; Castro et al. 2018; Debakere et al. 2020). This can be done by multiplying the non-linear power spectrum, $P_{\delta\delta}(k, z)$ – computed using one of the non-linear prescriptions discussed above – with $B(k, z)$, a ‘baryon correction model’ (BCM) that captures the baryonic effects (e.g., Semboloni et al. 2011), so that

$$P_{e+b}(k, z) = P_{\delta\delta}(k, z)B(k, z),$$

(21)

where $P_{e+b}(k, z)$ is the corrected power spectrum. The function $B(k, z)$ can be estimated by fitting Eq. (21) to power spectra obtained from hydrodynamical simulations that include baryons.

The challenge is that baryonic effects cannot (yet) be incorporated into cosmological simulations from first principles. The different implementations that have been used, not surprisingly, lead to a variety of possible BCM prescriptions. Here we consider three recent proposals.

The first one presented in Harnois-Déraps et al. (2015, HD15 hereafter), is based on three scenarios of the OverWhelmingly Large hydrodynamical simulations (Schaye et al. 2010). These were used to calibrate the power spectra for $z < 1.5$. It is able to reproduce the simulated results with an accuracy better than 2% for scales $k < 1 \, h \text{Mpc}^{-1}$. The functional form of $B(k, z)$ is given by

$$B(k, z) = 1 - A_{\text{HD15}}(z) \exp \left[ \left| B_{\text{HD15}}(z) \left( x(k) - C_{\text{HD15}}(z) \right) \right| \right] + D_{\text{HD15}}(z) x(k) \exp \left[ E_{\text{HD15}}(z) x(k) \right],$$

(22)

with $x(k) \equiv \log_{10}(k/[h \, \text{Mpc}^{-1}])$ and $X_{\text{HD15}}(z)$ are polynomial functions of redshift given in Harnois-Déraps et al. (2015).

As a second model, we consider the results obtained by Schneider & Teyssier (2015, ST15 hereafter), who accounted for the effects of baryons following a different approach. They start from a suite of DM only N-body simulations and modify the density field in such a way that it mimics the effects of a particular feedback recipe. They achieve this by explicitly modelling the main constituents of the haloes, which are dark matter, hot gas in hydrostatic equilibrium, ejected gas and stars. The model parameters are set to resemble SZ and X-ray observations. The resulting modifications to the power spectrum are shown to be well reproduced by defining $B(k, z)$ as

$$B(k, z) = \frac{1 + (k/k_s(z))^2}{[1 + (k/k_g(z))^2]} \mathcal{G}(z) + \left[ 1 + (k/k_s(z))^2 \right] \left[ 1 - \mathcal{G}(z) \right],$$

(23)

with $k_s(z)$ and $\mathcal{G}(z)$ auxiliary functions provided in Schneider & Teyssier (2015). We set the model parameters to the following fiducial values

$$\{k_s, \log M_c, z_b, \tau_p\} = \{67 \, h \, \text{Mpc}^{-1}, 13.8, 2.3, 0.17\}.$$  

(24)

We note that these are different from those of Schneider & Teyssier (2015) since we have updated them to the best fitting values obtained using the more recent Horizon-AGN simulations (as done in Chisari et al. 2018).

Chisari et al. (2018, hereafter Ch18) found that the ST15 model performs well at low redshift, but its accuracy degrades for larger $z$. Fitting the Horizon-AGN simulations, they therefore proposed the third model we will consider here, with

$$B(k, z) = \frac{[1 + k/k_s(z)]^2}{[1 + k/k_s(z)]^2 - \epsilon^2},$$

(25)

where $k_s$ is no longer a constant, but a function of $z$ instead. The detailed form is given in Chisari et al. (2018).

We can now quantify the impact of the choice of BCM on the FoM by comparing it to the results for the dark-matter-only forecasts. To this end, we use Halofit as our benchmark model to compute $P_{\delta\delta}(k, z)$, which is consistent with what is done in the quoted papers. Our results are presented in Table 5.

<table>
<thead>
<tr>
<th>$\ell_{\max}$</th>
<th>Halofit</th>
<th>HD15</th>
<th>ST15</th>
<th>Ch18</th>
</tr>
</thead>
<tbody>
<tr>
<td>1500</td>
<td>23</td>
<td>22</td>
<td>21</td>
<td>22</td>
</tr>
<tr>
<td>5000</td>
<td>44</td>
<td>37</td>
<td>41</td>
<td>41</td>
</tr>
</tbody>
</table>

Table 5. FoM for $w_0$ and $w_a$ parameters estimated using Euclid specifications for WL and the three BCM prescriptions compared to the Halofit forecast with no baryons.
in all three analysis are $h$, $w_0$ and $w_a$. Overall, however, all three cases produce very similar results, as can be seen in the left panel of Fig. 3.

For $\ell_{\text{max}} = 5000$, we find that the biases are very large when BCM effects are neglected; $B(\theta) > B_{\text{th}}$ for all parameters, with $B \gtrsim 5$ for the dark energy parameters $w_0$ and $w_a$, $B \gtrsim 3$ for $h$, and $B \gtrsim 4.5$ for $n_s$. As expected, the biases in the power spectrum amplitude, the baryon density, and the cold dark matter density are the less significant because these are all well constrained by the Planck measurements.

Ignoring BCM effects could lead to a false detection of a time-varying dark energy equation of state. Moreover, with the current tension between $H_0$ measurements between CMB and late-time probes, an unbiased measurement of $H_0$ will also be...
crucial. Our results confirm earlier work (e.g., Sembolini et al. 2011) that correctly modelling the impact of baryonic feedback on the power spectrum is essential for the analysis of Euclid data.

7. Conclusions

Forthcoming surveys of the large-scale structure will deliver datasets of exquisite quality that will allow us to pursue what is usually referred to as precision cosmology. To interpret correctly these measurements, significant improvements in the underlying theoretical predictions are called for: We need to ensure that errors in the modelling of density fluctuations on non-linear scales do not introduce biases in the inferred cosmological parameters that are larger than the expected statistical uncertainties. This is a particular concern for cosmic shear tomography, given that one has to integrate the matter power spectrum \( P_{\delta \delta}(k,z) \) deep into the non-linear regime, where baryonic physics complicates matters even further. Motivated by these concerns, we have investigated how different popular prescriptions that account for these effects influence both the accuracy and the precision with which Euclid can infer cosmological parameters using cosmic shear alone.

We used Fisher matrix forecasts to quantify the impact of three different non-linear recipes on the dark energy FoM. The recipes that we considered are the revised implementation of Halofit, Halofit+PKequal, and the HMCode prescription. These differ significantly from one another when the cosmological parameters are left free to take values other than the fiducial ones. As a consequence, the derivatives of \( P_{\delta \delta}(k,z) \) that enter the Fisher matrix are changed, leading to quite discrepant FoM values. In particular, we find that the Halofit case provides the higher FoM because of the different role played by \( w_0 \) in the non-linear corrections. Although we have explicitly considered the case of Euclid, this result is generic for cosmic shear tomography analyses, although the size of this effect will depend on the details of the survey of interest. Hence our findings highlight the importance of choosing the most reliable non-linear model, in order to compute realistic estimates of the expected performance of a particular WL survey.

While it is important to quantify the precision, that is how tight the constraints will be, it is perhaps even more important to establish the accuracy of the results: We need to be confident that an incorrect choice of theoretical ingredients does not introduce an unacceptably large bias, that is a deviation from the (unknown) true value. Whether or not the bias is too large also depends on the precision with which that parameter can be measured. We therefore define \( B = |\delta - \delta_{\text{fid}}|/\sigma \), and adopt a threshold of \( B < B_{\text{thr}} \approx 0.1 \).

To study the accuracy with which cosmological parameters can be determined, we created mock data with a given prescription for non-linearities and/or baryon physics, and fitting these with a different model. This allowed us to address the issues of the choice of the non-linear recipe and the baryon correction model separately. To examine the impact of the recipe used to compute the power spectra on non-linear scales, we created a mock dataset using Halofit+PKequal that comprises Euclid cosmic shear and Planck CMB data. We emphasise that the choice of Halofit+PKequal is arbitrary as we do not know which of the non-linear corrections better describes the true small-scale evolution. We fitted these with Halofit or HMCode.

We find that \( B \leq B_{\text{thr}} \) if Halofit is used for \( \ell_{\text{max}} = 1500 \), while for \( \ell_{\text{max}} = 5000 \), some cosmological parameters, namely \( h, \Omega_{\text{m}0} \), and \( \sigma_8 \), exceed the threshold. This is not surprising given the similarities of the two models when one only looks at \( P_{\delta \delta} \) rather than at its derivatives. In contrast, comparing the theoretical spectra obtained using HMCode based corrections with the assumed Halofit+PKequal fiducial leads to strong biases, with almost all parameters already biased by more than \( B_{\text{thr}} \) if we restrict the analysis to \( \ell_{\text{max}} = 1500 \). Including very non-linear scales (\( \ell_{\text{max}} = 5000 \)), we find that \( B > 1 \) for all parameters, except for \( \omega_{\text{b}0} \), which is actually tightly constrained by Planck. In particular, the estimate of \( w_0 \) shifts towards its ACDM value even if the mock data were created using \([w_0, w_a] = (-0.9, 0.1)\). What is even more interesting is that the most biased parameters are \( h (B = 3.02) \) and \( \sigma_8 (B = 4.62) \); the values of both of these are currently debated. The sensitivity of these parameters to the adopted prescription for the non-linear power spectrum highlights the need for further improvements, which may already be needed to correctly interpret current data.

It is also essential that the changes to the power spectrum caused by baryon physics have to be taken into account. For a fixed non-linear recipe and baryon correction model prescription, severe biases are found when fitting the mock data with the right non-linear correction, but not accounting for the presence of baryons, in line with earlier work (e.g., Sembolini et al. 2011). In the most constraining setting (\( \ell_{\text{max}} = 5000 \)), for all the three cases we considered, we find significant biases for all the cosmological parameters, except for \( \{\omega_{\text{b}0}, \ln (10^9\Omega_{\text{m}})\} \), which are actually constrained by the Planck data rather than by cosmic shear. While this work was near to completion, Schneider et al. (2020) presented a similar analysis, but their method to account for baryons is very different from the one we have adopted here. They use instead a model for the baryonification of dark matter only simulations to determine the matter power spectrum (Schneider et al. 2019). Notwithstanding these differences, which make a straightforward comparison impossible, their conclusions are in agreement with what we found here.

As a final remark, we remind the reader that our results refer to the case where cosmic shear is used as the only probe. This is, however, only part of the information that future surveys will provide. Indeed, the same data used to do cosmic shear tomography (WL) can and will be used to compute the photometric galaxy clustering (GCph) and to cross correlate the shear and density fields (XC). As was shown in EC19 and further investigated in Tutusaus et al. (2020), it is the joint use of WL+GCph+XC that is needed to achieve \( \sim 1\% \) errors on \([w_0, w_a]\), rather than any single probe by itself. It is therefore worth considering whether, and to which extent, the results we have obtained here change if all the three probes are considered.

For instance, EC19 limited the GCph and XC to smaller multipole compared to what was used for WL, reducing their sensitivity to the small-scale corrections. This diminishes the impact of errors in the prediction of \( P_{\delta \delta}(k,z) \) in the large \( k \) regime, but at the expense of larger statistical uncertainties. This motivates extending our work to quantify the impact of modelling the small-scale power spectra for the joint probes. Such a study, which is beyond the scope of our initial exploration, would provide guidance on how to proceed in order to exploit the high-quality data that stage IV surveys will provide.

Acknowledgements. D. F. M. thanks the Research Council of Norway for their support, and the resources provided by UNINETT Sigma2 – the National Infrastructure for High Performance Computing and Data Storage in Norway. M. M. has received the support of a fellowship from “la Caixa” Foundation (ID 100010434), with fellowship code LCF/BQ/PI19/11690015, and the support of the Spanish Agencia Estatal de Investigación through the grant “IFT Centro de Excelencia Severo Ochoa SEV-2016-0597”. This paper is based upon work from the COST action CA15117 (CANTATA), supported by COST (European Cooperation in Science and Technology). M. A. acknowledges the computing resources from the COST action CA15117 (CANTATA), supported by COST (European Cooperation in Science and Technology).
support of INFN-CNAF. Stefano Camera acknowledges support from the Italian Ministry of Education, University and Research (MIUR) through Rita Levi Montalcini project ‘PROMETHEUS – Probing and Relating Observables with Multimessenger Experiments To Help Enlightening the Universe’s Structure’, and the ‘Departments of Excellence 2018–2022’ Grant awarded by MIUR (L. 222/2016). The work of SC is supported by the Belgian Fund for Research FNRS-F.R.S. I.T. acknowledges support from the Spanish Ministry of Science, Innovation and Universities through grant ESP2017-89386-C3-1-R. I.T. and T.K. acknowledge funding from the H2020 programme of the European Commission through grant 776247. AP is a UK Research and Innovation Future Leaders Fellow, grant MR/S016066/1. The Euclid Consortium acknowledges the European Space Agency and a number of agencies and institutes that have supported the development of Euclid, in particular the Academy of Finland, the Agenzia Spaziale Italiana, the Belgian Science Policy, the Canadian Euclid Consortium, the Danish Space Research Institute, the Netherlandse Onderzoekschool Voor Astronomie, the Norwegian Space Agency, the Romanian Space Agency, the State Secretariat for Education, Research and Innovation (SERI) at the Swiss Space Office (SSO), and the United Kingdom Space Agency. A complete and detailed list is available on the Euclid web site (http://www.euclid-ec.org).
Appendix A: Comparison between scales and multipole cuts

In the analyses performed in this paper, we considered two cases, with different cuts in multipoles used; the optimistic $\ell_{\text{max}} = 5000$ cut represents the situation in which data for all scales are included in the analysis, while the $\ell_{\text{max}} = 1500$ mimics the removal of non-linear scales that could be performed in the data analysis in order to reduce the impact of small-scale modelling.

However, our approach can be seen as a first approximation as a constant multipole cut corresponds to different redshifts. In this subsection we investigate a less simplistic analysis in order to remove completely the dependence on the non-linear scales.

We assume $k_{\text{max}} = 0.25 h$ Mpc$^{-1}$ and, using the Limber approximation considered to express the $C_{\ell}(\ell)$ in Eq. (15), we convert this cut in a maximum multipole for each redshift bin:

$$\ell_{\text{max}}(z) = k_{\text{max}} \Delta z(z) = \frac{1}{\Delta z(z)},$$

where $\Delta z(z)$ is the mean redshift of each redshift bin.

We apply this approach to our MCMC analysis, and compare the results we obtain with those of the $\ell_{\text{max}} = 1500$ case. In Fig. A.1 we show the comparison in the results for the analysis performed using Halofit (left panel) and HMCode (right panel). In both cases we find that the $k_{\text{max}}$ analysis leads to broader constraints with respect to the $\ell_{\text{max}}$ case; this is due to the fact that the chosen $k_{\text{max}}$ translates into a much more agressive cut in multipoles, especially at low redshift. Therefore, a significative amount of information is lost with respect to the $\ell_{\text{max}} = 1500$ case.

Concerning the bias found on cosmological parameters, $B(\theta)$ does not change significantly in the Halofit analysis, while for HMCode the biases are slightly enhanced with respect to the $\ell_{\text{max}} = 1500$ case, with $w_0$, $w_a$, $h$ and $\sigma_8$ now reaching $B \approx 1$. This apparent increase however is mostly due to the loss of constraining power when the scale cut is implemented; as it can be seen in Fig. A.1, the marginalised posterior distributions for the cosmological parameters now exhibit non-Gaussian features while Eq. (20) implicitly assumes Gaussian distributions.

In Fig. A.1 it is shown how the scale cut case produces contours closer to the expected fiducial values, a result supported also by the change in $\Delta \chi^2$ moving from the $\ell_{\text{max}}$ to the scale cut, which changes from 32 to 14 in the HMCode case.

We performed the same analysis at the Fisher matrix level finding a still significant dependence of the results on the adopted non-linear recipe. We indeed get $\text{FoM} = (2.93, 3.59, 1.59)$ for Halofit, HMCode, and Halofit + PKequal, respectively. While the severe decrease in the FoM is expected given that we are removing a large part of the data, it is somewhat surprising to still find such a variety of values. This is, however, a consequence of the integrated nature of the WL $\sigma_{\ell_i}^2(\ell)$. To understand what is going on, we focus on the case $i = j = 5$ giving $\ell_{\text{max}} \approx 500$. Because of the photo-z broadening of the lensing kernel, the integral giving $C_{\ell\ell}(\ell_{\text{max}})$ gets contributions from the redshift range (0.1, 2.6). Over this range, the argument $k_{5}(z)$ of the matter power spectrum $P_m(k,z)$ feeding the integral is larger than $k_{\text{max}}$ for $z < 0.84$ so that which non-linear recipe is adopted still matters. Such an argument can be repeated for all the bins combinations and the multipoles thus explaining why the FoM is still dependent on the non-linear recipe even with these very conservative scale cuts, a result in agreement with what was discussed in Taylor et al. (2018a). Therefore, in order to remove completely the dependence on the non-linear description from the analysis, different approaches are needed, for example using band powers rather than a $C_{\ell\ell}(\ell)$ analysis (Joachimi et al. 2021).

![](image.png)

**Fig. A.1.** One-dimensional posterior distributions, and 68% and 95% confidence level marginalised contours for the dark energy parameters ($w_0$ and $w_a$) and the parameters $h$ and $\sigma_8$. Here we compare results obtained with a multipole cut at $\ell_{\text{max}} = 1500$ (green contours) with those related to the analysis with the scales cut $k_{\text{max}} = 0.25 h$ Mpc$^{-1}$ (red contours). The mock data of Euclid cosmic shear assume Halofit+PKequal non-linear corrections as reference, while the parameter estimation is performed with either HMCode (right panel), or Halofit (left panel). Black dashed lines mark the fiducial model.
Appendix B: MCMC results validation

The results of this paper have been obtained using both Fisher matrix and MCMC codes. The Fisher analysis relies on one of the codes used in EC19, which has passed through a careful validation procedure that included intensive comparisons between different Fisher matrix codes. Our MCMC analysis relies on a public Montepython likelihood for Euclid WL (Brinckmann & Lesgourgues 2019), adapted to the Euclid specifications of EC19. It has been further modified to include different models of baryonic feedback effects. This likelihood code was first used in Sprenger et al. (2019). In contrast to the Fisher code, it has not been validated against other codes. In this Appendix, we therefore present a comparison between validated Fisher forecasts and the MCMC ones.

In the Euclid-only case, our analysis reveals some deviations that are attributed to the intrinsic limitation of any Fisher analysis, due to the non-Gaussianity of the posterior distribution, as well as some important parameter degeneracies. Nevertheless, the impact on the forecasts becomes negligible when Planck constraints are included in the analysis. In this case, we find that the forecasts on cosmological parameters obtained with the Fisher and MCMC methods agree very well. This therefore validates our MCMC approach against the Fisher codes used in EC19.

The Euclid forecasts for WL in EC19 are accompanied by a series of public Fisher matrices, corresponding to different setups and cosmologies. In order to validate our Montepython likelihood and MCMC analysis, we compared forecasts obtained for both the pessimistic and optimistic setups, and using the same cosmological parameters (in particular, $\Omega_0$ and $\Omega_m$ instead of $\omega_b$ and $\omega_c$). We performed these comparisons for Euclid only, and in combination with Planck. In the latter case, we used the mock Planck likelihood available in Montepython that accurately reproduces the Planck limits on cosmological parameters. We construct a covariance matrix from the MCMC chains in the Planck-only case. Its inverse provides a Fisher matrix that can be added to the validated Euclid Fisher matrices. We checked that the Planck-only case constrains the standard cosmological parameters well, with close-to-Gaussian two-dimensional posterior distributions. This is a good indication that one can safely use the Planck covariance matrix for the Fisher analysis. In contrast to the standard cosmological parameters, most of the constraining power for the dark energy parameters $w_0$ and $w_a$ comes from Euclid, not Planck. As a consequence, the corresponding entries in the Planck Fisher matrix are not relevant and do not significantly impact the Euclid + Planck forecasts for these parameters.

Fig. B.1. Fisher matrix vs MCMC forecasts, for Euclid WL only with $\ell_{\text{max}} = 1500$. In red, one-dimensional marginalised posterior distributions, with two-dimensional 68% and 95% confidence level marginalised contours, for varying cosmological parameters and fixed DE and IA parameters. Black curves represent the corresponding Fisher ellipses, from the validated Fisher matrix of EC19.
For Euclid only, the marginalised two-dimensional posterior distributions and the Fisher contours obtained in the case $\ell_{\text{max}} = 1500$ for five varying cosmological parameters ($\Omega_{m0}$, $\Omega_{b0}$, $h$, $n_s$, and $\sigma_8$), with DE and IA parameters fixed to their fiducial values, are shown in Fig. B.1. We find that the directions and widths of all the Fisher ellipses are well recovered by the MCMC approach. However, we also find two significant differences. First, the MCMC contours in the plane ($\Omega_{b0}$, $h$) display a banana shape, leading to a more stringent constraint on $\Omega_{b0}$ compared to the Fisher method. In turn, this affects two-dimensional contours between $\Omega_{b0}$ and other parameters. Second, the marginalised posterior distribution for $h$ has a significant non-Gaussian shape, falling more rapidly at lower values than at higher values. Consequently, some of the contours are less extended on one side compared to the Fisher ellipses. This also arises from the degeneracy between $h$ and $\Omega_{b0}$, with these parameters being poorly constrained with Euclid WL only. These features cannot be recovered by the Fisher analysis that relies on the assumption that the posterior distribution is Gaussian, which is not the case for strong variations of $h$ and $\Omega_{b0}$. Nonetheless, these results show that both the MCMC and Fisher methods provide consistent results, even if the agreement between them is limited by the intrinsic limitation of the Fisher analysis. Adding dark energy parameters in the analysis further degrades the agreement between the Fisher and MCMC forecasts, for similar reasons. Here, we have only shown the case in which $w_0$ and $w_a$ remain fixed, for a better illustration of the effect of parameter degeneracies on the two method comparison.

For Planck + Euclid WL, we find a very good agreement between Fisher and MCMC forecasts, even when the DE and IA parameters are varying. This is shown in Fig. B.2 representing the MCMC marginalised two-dimensional posterior distributions, which fit well to all the Fisher ellipses. We find that the differences in the forecasts are less than 10% for all the cosmological parameters. This therefore validates the approach we have used throughout the paper. It also illustrates the importance of adding CMB data to break parameter degeneracies. Compared to the Euclid WL only case, the differences between the Fisher and MCMC methods are suppressed. The few remaining differences can be either due to a limited convergence of the MCMC chains, to uncertainties induced by the binning method, or to a slightly non-Gaussian likelihood function for Planck-only that induces small differences between MCMC posteriors and the corresponding Planck Fisher ellipses extracted from the covariance matrix of the MCMC chains. Similar results have been obtained for the optimistic setup with $\ell_{\text{max}} = 5000$. 

![Fig. B.2. Fisher matrix vs MCMC forecasts, for Planck+Euclid WL, with $\ell_{\text{max}} = 1500$. In red, one-dimensional marginalised posterior distributions, with two-dimensional 68% and 95% confidence level marginalised contours, for varying cosmological parameters including $w_0$ and $w_a$ and IA parameters. Black curves represent the corresponding Fisher ellipses, from the validated Fisher matrix of EC19 combined with a Planck-only Fisher matrix.](image-url)