Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **518,** 1859–1879 (2023)
Advance Access publication 2022 June 28

https://doi.org/10.1093/mnras/stac1671

Parameter inference with non-linear galaxy clustering: accounting for theoretical uncertainties

Mischa Knabenhans [1]★ Thejs Brinckmann [2,3,4]★ Joachim Stadel [1] Aurel Schneider [1] and Romain Teyssier [1,5,6]

[1]*Institute for Computational Science, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland*
[2]*Dipartimento di Fisica e Scienze della Terra, Università degli Studi di Ferrara, via Giuseppe Saragat 1, I-44122 Ferrara, Italy*
[3]*Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Ferrara, Via Giuseppe Saragat 1, I-44122 Ferrara, Italy*
[4]*C.N. Yang Institute for Theoretical Physics and Department of Physics & Astronomy, Stony Brook University, Stony Brook, NY 11794, USA*
[5]*Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ 08544, USA*
[6]*The Program in Applied and Computational Mathematics, Princeton University, Fine Hall, Washington Road, Princeton, NJ 08544-1000 USA*

Accepted 2022 May 27. Received 2022 May 24; in original form 2021 October 23

ABSTRACT

We implement EUCLIDEMULATOR (version 1), an emulator for the non-linear correction of the matter power spectrum, into the Markov chain Monte Carlo forecasting code MONTEPYTHON. We compare the performance of HALOFIT, HMCODE, and EUCLIDEMULATOR1, both at the level of power spectrum prediction and at the level of posterior probability distributions of the cosmological parameters, for different cosmological models and different galaxy power spectrum wavenumber cut-offs. We confirm that the choice of the power spectrum predictor has a non-negligible effect on the computed sensitivities when doing cosmological parameter forecasting, even for a conservative wavenumber cut-off of $0.2\,h\,\mathrm{Mpc}^{-1}$. We find that EUCLIDEMULATOR1 is on average up to 17 per cent more sensitive to the cosmological parameters than the other two codes, with the most significant improvements being for the Hubble parameter of up to 42 per cent and the equation of state of dark energy of up to 26 per cent, depending on the case. In addition, we point out that the choice of the power spectrum predictor contributes to the risk of computing a significantly biased mean cosmology when doing parameter estimations. For the four tested scenarios we find biases, averaged over the cosmological parameters, of between 0.5σ and 2σ (from below 1σ up to 6σ for individual parameters). This paper provides a proof of concept that this risk can be mitigated by taking a well-tailored theoretical uncertainty into account as this allows to reduce the bias by a factor of 2 to 5, depending on the case under consideration, while keeping posterior credibility contours small: the standard deviations are amplified by a factor of ≤ 1.4 in all cases.

Key words: methods: numerical – methods: statistical – cosmological parameters – large-scale structure of Universe.

1 INTRODUCTION

After very successful missions such as Planck[1] (Planck Collaboration 2006; Planck Collaboration VI 2020) and SDSS[2] (Gunn et al. 2006; Abdurro'uf et al. 2022), many ground- and space-based experiments are being conducted or planned in order to learn more about our Universe. DES[3] (The Dark Energy Survey Collaboration 2005; Abbott et al. 2022), KiDS[4] (de Jong et al. 2013; Heymans et al. 2021), the Dark Energy Spectroscopic Instrument[5] (DESI, DESI Collaboration 2016), the European space mission Euclid[6] (Laureijs et al. 2011), the Legacy Survey of Space and Time (LSST) of the Vera C. Rubin observatory[7] (LSST Science Collaboration 2009), the Nancy Grace Roman Space Telescope[8] (Akeson et al. 2019), and the Square-Kilometre Array[9] (SKA, Maartens et al. 2015) are a few well-known examples. They will provide highly accurate measurements of various cosmological observables, giving insight into the nature of dark matter and dark energy, and are likely to measure the total mass of neutrinos.

In order to exploit this wealth of high-quality data that are being produced, they have to be met by equally accurate theoretical predictions. We need to quantify the process of structure formation at high precision and to include all systematics related to the observations. Only if the highly demanding requirements on both sides are met, do we optimize our chances to determine the cosmological parameters with high enough accuracy to potentially reveal new physics.

One approach is extended perturbation theories that are able to accurately model the mildly non-linear scales. See e.g. CLASS-PT (Chudaykin et al. 2020), which was recently used for Euclid forecasts (Chudaykin & Ivanov 2019) and BOSS data analyses (Ivanov, Simonović & Zaldarriaga 2020a, b), and EFTofLSS (publicly

★ E-mail: m_knabenhans@hotmail.com (MK); thejs.brinckmann@gmail.com (TB)

[1]https://www.cosmos.esa.int/web/planck
[2]https://www.sdss.org
[3]www.darkenergysurvey.org
[4]http://kids.strw.leidenuniv.nl/index.php
[5]www.desi.lbl.gov/category/announcements/
[6]sci.esa.int/euclid
[7]https://www.vro.org, www.lsst.org/lsst
[8]http://roman.gsfc.nasa.gov
[9]https://www.skatelescope.org

© 2022 The Author(s)
Published by Oxford University Press on behalf of Royal Astronomical Society

Downloaded from https://academic.oup.com/mnras/article/518/2/1859/6619061 by Dept of Energy/OSTI user on 26 August 2024

available as PyBird, D'Amico, Senatore & Zhang 2021), also recently used for BOSS data analyses (Colas et al. 2020; D'Amico et al. 2020). These approaches allow us to perturbatively compute an accurate galaxy power spectrum to a wavenumber of about $k = 0.3\,h\,\mathrm{Mpc}^{-1}$, exceeding the range of reliability of linear perturbation theory (PT).

However, future surveys are expected to probe much smaller scales, e.g. Euclid will collect accurate data up to $k = 10\,h\,\mathrm{Mpc}^{-1}$. In order to accurately model increasingly non-linear scales, those methods either need drastic improvements, or we need to use a different approach, e.g. other non-linear modelling techniques such as emulation. Cosmic emulators do not suffer from the neuralgic problems at high k values as do perturbation theoretical approaches, nor do they require expensive N-body simulations beyond those used initially to train the emulator. Such emulators have received significant attention in recent years with the development of new publicly available tools and codes (Heitmann et al. 2009, 2010, 2014, 2016; Lawrence et al. 2010, 2017; Mead et al. 2015, 2016; Bird et al. 2019; Cataneo et al. 2019; DeRose et al. 2019; Fluri et al. 2019; McClintock et al. 2019a, b; Nishimichi et al. 2019; Rogers et al. 2019; Valcin et al. 2019; Winther et al. 2019; Zhai et al. 2019; Aricò et al. 2020, 2021; Angulo et al. 2021; Mead, Brieden & Tröster 2021; Spurio Mancini et al. 2022).

We present an improvement on prediction accuracy of galaxy clustering observables due to integration of the publicly available EUCLIDEMULATOR1[10] (Euclid Collaboration 2019) into the Markov chain Monte Carlo (MCMC) sampling package MONTEPYTHON[11] (Audren et al. 2013b; Brinckmann & Lesgourgues 2019). EUCLIDEMULATOR1 is a cosmic emulator that quickly and accurately predicts the non-linear correction (NLC) factor of the matter power spectrum for cold dark matter cosmologies with a varying time-independent Dark energy (DE) equation of state (EoS) parameter w_0 (these are commonly referred to as 'w_0CDM' cosmologies).

Popular methods for including non-linear information of cosmic structure formation in an MCMC run are via HALOFIT (Smith et al. 2003; Takahashi et al. 2012) or HMCODE[12] (Mead et al. 2015, 2016). In this work, we compare these two codes to the cosmic emulator EUCLIDEMULATOR1, which is able to capture non-linear features, e.g. non-linear damping of the Baryon accoustic oscillations (BAO), as accurately as 0.3 per cent or less compared to full N-body simulations over a wide range of both redshifts and k modes – an accuracy level unprecedented by any other cosmic emulator.

The most widely used forecasting method is Fisher matrix forecasts (e.g. Tegmark, Taylor & Heavens 1997). In its traditional form, it is conceptually very simple, but is restricted to certain very idealistic mathematical properties of the corresponding likelihood function (efforts in development of methods trying to overcome this limitation can, e.g. be found in Sellentin, Quartin & Amendola 2014). Further, it involves computations of numerical derivatives that are likely to be unstable. For a discussion of these problems and efforts towards solutions we refer to Euclid Collaboration (2020), Yahia-Cherif et al. (2021), and Bhandari et al. (2021).

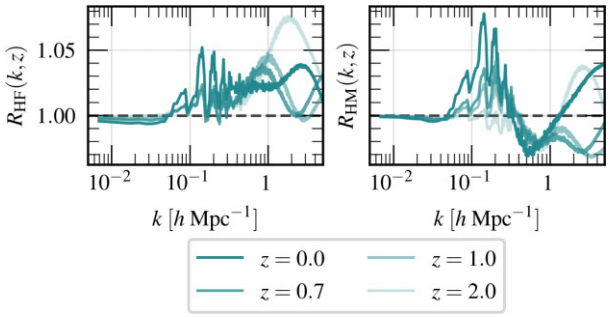

Figure 1. Redshift evolution for power spectrum ratio (as defined in equation 1) between HALOFIT and EUCLIDEMULATOR1 (left panel) and HMCODE and EUCLIDEMULATOR1 (right panel). The evolution, which is only weakly dependent on the cosmology, is shown for the *Euclid* reference cosmology.

We connect measurements and theory in a Bayesian forecasting process through MCMC sampling method (Christensen et al. 2001; Lewis & Bridle 2002; Perotto et al. 2006), which estimates the posterior probability distributions of cosmological parameters. This is a more data-driven, iterative approach that has the advantage of not relying on derivatives or on any assumptions of the shape of the underlying likelihood. This is the most common approach for cosmological parameter inference. Since it provides more reliable results in parameter sensitivity forecasts than classical Fisher analyses, we choose to use the MCMC method in the analysis presented in this paper.

In this work, we investigate the influence of different code implementations (with different strength and weaknesses) on sensitivity forecast and parameter estimation results and we argue that using theoretical uncertainties can help deal with that (also beyond forecasts in which they are used already today). We choose to perform this analysis with the following three codes: HALOFIT and HMCODE because they are very established in the field, and EUCLIDEMULATOR1 as it is a representative of cosmic emulators which, as described above, have become more popular recently. We shall stress, however, that the key take-away message of this work is of qualitative rather than quantitative nature and that these qualitative results are ultimately independent of the set of codes used. Along the way of our investigations we address four main questions:

(i) What are the performance differences of EUCLIDEMULATOR1, HALOFIT, and HMCODE?
(ii) What is the added value of considering (mildly) non-linear scales in parameter forecasts?
(iii) What is the impact of the choice of the (non-linear) predictor model on the parameter estimation result?
(iv) How are the forecasting results affected by different choices of theoretical uncertainty models?

In order to understand (i), we first compare the codes at the level of the matter power spectrum in Section 2.1 (see Figs 1 and 2) before performing a thorough forecast analysis in Section 4.

Traditionally, in galaxy clustering analyses information from non-linear scales is removed as being 'too difficult' to model correctly,

[10]We use version 1 of EUCLIDEMULATOR, referred to as EUCLIDEMULATOR1, which is available at https://github.com/miknab/EuclidEmulator. Note that during the preparation of this manuscript an improved version, EUCLIDEMULATOR2 (Euclid Collaboration 2021), has been made publicly available at https://github.com/miknab/EuclidEmulator2. The latter features additional cosmological parameters, namely, the neutrino mass sum ($\sum m_\nu$) and dynamical DE (w_0, w_a), as well as an expanded redshift and wavenumber range, and an improved python wrapper.

[11]See https://github.com/brinckmann/montepython_public for the latest version 3.5.

[12]Since HMCODE-2020 (Mead et al. 2021) is not yet available within the Boltzmann solver CLASS, we instead use HMCODE-2016. Note that the older version is known to have difficulties in capturing non-linear features correctly, e.g. non-linear damping of the Baryon accoustic oscillations (BAO), which were improved in the newer version. While this may impact some parts of the analysis, it also illustrates the need for accurately accounting for theoretical uncertainties and problems associated with not doing so.

Downloaded from https://academic.oup.com/mnras/article/518/2/1859/6619061 by Dept of Energy/OSTI user on 26 August 2024

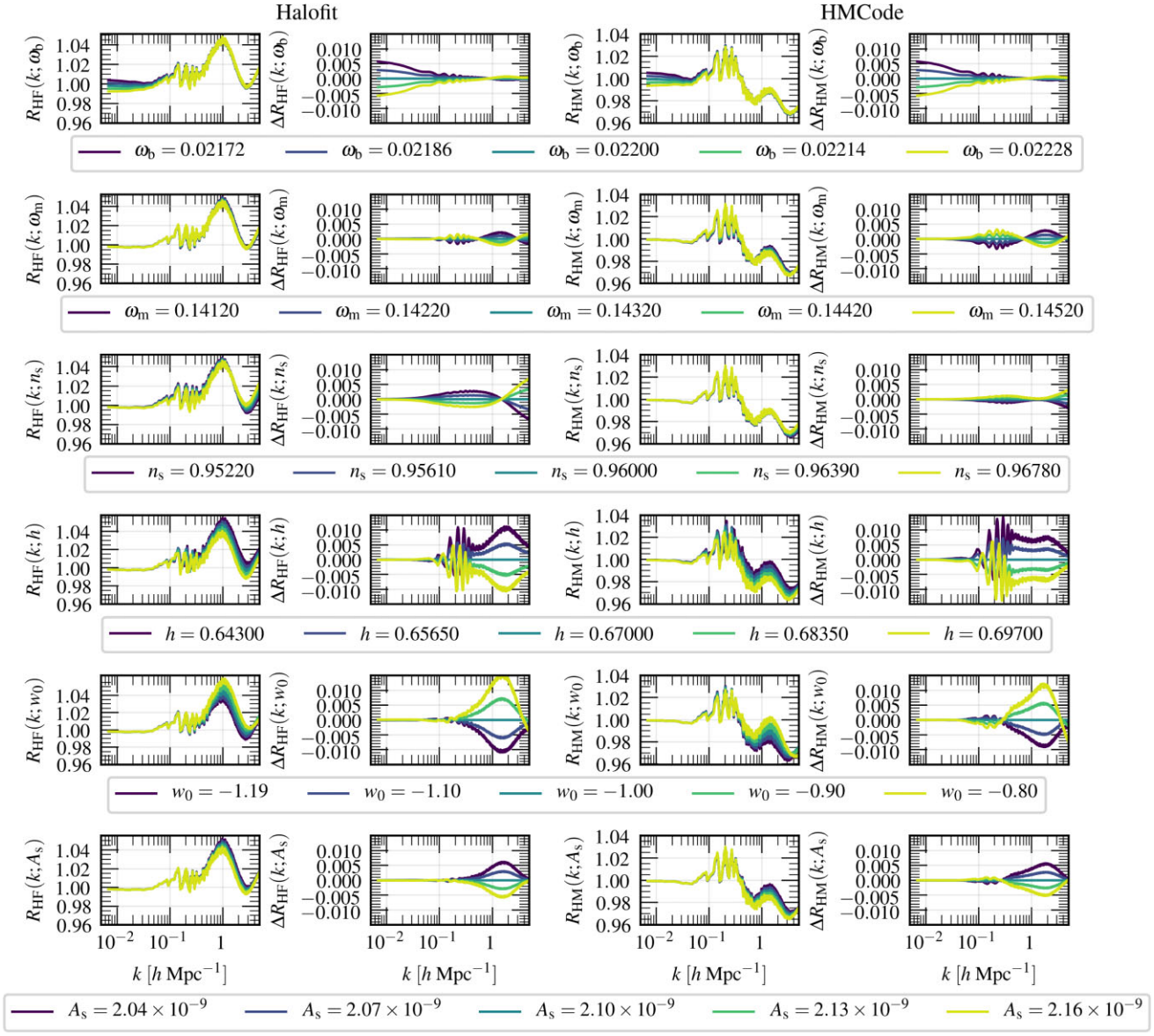

Figure 2. Variation of the non-linear matter power spectrum due to variation of cosmological parameters. Each row shows the variation of one cosmological parameter: $\omega_{\rm b}$ in the first row, $\omega_{\rm m}$ in the second, etc. All parameters that are not varied in any given row are set to their corresponding values in the Euclid Reference Cosmology. Notice that the Euclid Reference cosmology corresponds to the middle parameter values in each of the parameter value legends, i.e. $\omega_{\rm b} = 0.022$, $\omega_{\rm m} = 0.1432$, $n_{\rm s} = 0.96$, $h = 0.67$, $w_0 = -1.0$. $A_{\rm s} = 2.1 \times 10^{-9}$ is always fixed (except for the last row) and $z = 1$ for all panels. The variation is shown for the quantity $R_{\rm model}$ as defined in equation (1) (columns 1 and 3). In order to emphasize (i) which scales are affected the most by the variation of a specific parameter and (ii) how strongly R is impacted by the variation of each parameter, we also plot the ΔR with respect to the R of the Euclid Reference Cosmology (columns 2 and 4). It is evident, for instance, that the variation of h and w_0 has a much larger effect on R than e.g. $\omega_{\rm m}$. Notice that the tested parameter values are taken to be the Euclid Reference Cosmology value (the central value) $\pm i\sigma_{p,\,\rm Planck2018}$, $i \in 1, 2$ where $\sigma_{p,\,\rm Planck2018}$ denotes the standard deviation of each parameter p of the Planck 2018 results (for more details see section 17.14 on p. 301 in Planck Collaboration 2018).

and usually only information from linear and quasi-linear scales is included. However, a wealth of information is available on non-linear scales that may improve constraints. Indeed, a few parameters exhibit very clear signals only in the (mildly) non-linear regime, particularly the DE EoS parameter w_0, which is illustrated in Fig. 2. As such, we address question (ii) in Section 4.2 by increasing the maximum wavenumber included in the analysis. This makes a case for cosmic emulators as the accuracy at non-linear scales is one of their key selling points in contrast to models based on PT.

However, once we go beyond linear scales, theoretical modelling uncertainties grow substantially, as different simulation codes increasingly disagree with each other. While the gravity-only calculation is under control up to high k-modes (Schneider et al. 2016; Garrison, Eisenstein & Pinto 2019), baryonic feedback effects are known to affect the matter power spectrum beyond $k \sim 0.1$–$1\,h\,{\rm Mpc}^{-1}$ but remain poorly constrained (see e.g. Van Daalen et al. 2011; Schneider et al. 2016; Chisari et al. 2019; Aricò et al. 2020; Giri & Schneider 2021; Mead et al. 2021).

Sprenger et al. (2019) introduced a theoretical uncertainty increasing with wavenumber in order to account for the growing uncertainties at small cosmological scales, thereby making a step in the direction towards including more non-linear k modes into the analysis rather than just ignoring them as was often done in the past (see Audren et al. 2013a; Baldauf et al. 2016 for earlier

Downloaded from https://academic.oup.com/mnras/article/518/2/1859/6619061 by Dept of Energy/OSTI user on 26 August 2024

work in this direction). An uncertainty like that used in Sprenger et al. (2019) works well for sensitivity forecasts, where we are only interested in the uncertainties of parameter estimates and not the estimates themselves. However, results presented in this paper indicate that in the case of (mock) data analysis parameter estimation we should consider a theoretical uncertainty directly tied to the modelling choices made for the specific code that is used to fit the data, as otherwise the parameter estimates are likely to be biased.

While we discuss theoretical uncertainties and their modelling in Section 2.4, they are also the focus of questions (iii) and (iv), which are addressed in Sections 5.1 and 5.2. Indeed, when we set up a scenario to mimic real data analysis (where the true model is not known), we find large biases in the inferred parameters, in particular when extending the analysis to larger wavenumbers. This is not unexpected, as, indeed, the three codes model non-linear scales differently. It serves as a word of caution that when high-resolution data from future surveys such as Euclid are analysed, we need to carefully account for modelling uncertainties. We make a first attempt at this using an inflated theoretical uncertainty inspired by the agreement between EUCLIDEMULATOR1 and the other two codes, and we show that with such an uncertainty envelope, we get greatly improved agreement between the codes, reducing the bias in the parameter inference without significantly decreasing the sensitivity. We leave it to future work to produce a more robust scheme for accounting for the theoretical modelling uncertainty of current non-linear parameter estimation tools.

This paper is structured as follows: In Section 2, we present a comparison of the three codes analysed in this paper – EUCLIDEMULATOR1, HALOFIT, and HMCODE – at the level of the matter power spectrum and discuss both known and unknown systematics impacting parameter estimations and sensitivity forecasts as well as how they are related to code differences. In this section, we motivate the need for a theoretical uncertainty in order to mitigate unwanted effects due to those systematics. Next, in Section 3 we introduce the analysis methodology employed in this work, including some theoretical considerations as well as the strategy of our computational experiments. The results are presented and discussed in Sections 4 and 5. We finally conclude this paper in Section 6.

2 SYSTEMATIC UNKNOWNS IN NON-LINEAR COSMOLOGICAL STRUCTURE FORMATION

2.1 Code comparison at the matter power spectrum level

Before jumping to the MCMC results, let us first quantify differences between the three non-linear prescriptions for predicting the matter power spectrum $P(k, z; p)$ considered in this paper: HALOFIT (HF), HMCODE (HM) and EUCLIDEMULATOR1 (EE). This will help us understand the differences we may see in the following section.

Due to differences in how the non-linear prescriptions model non-linear structure formation, we expect a difference in how each code responds to changes in cosmological parameters. In the following text, we want to compare the non-linear prescriptions to each other in order to quantify these differences. Taking EUCLIDEMULATOR1 as the reference, we vary one cosmological parameter at a time, keeping all others fixed to a fiducial cosmological model (Euclid Reference Cosmology as defined in Euclid Collaboration 2019 and given in the caption of Fig. 2). We define the model ratio

$$R_{\rm model}(k, z; p) \equiv \frac{P_{\rm model}(k, z; p)}{P_{\rm EE}(k, z; p)}, \tag{1}$$

where model $\in \{\rm HF, HM\}$ and $p \in \{\omega_{\rm b}, \omega_{\rm m}, n_{\rm s}, h, w_0, A_{\rm s}\}$.[13] We consider the difference with respect to the reference cosmological model

$$\Delta R_{\rm model}(k, z; p) = R_{\rm model}(k, z; p) - R_{\rm model}(k, z; p_{\rm ref}) \tag{2}$$

This quantifies how, for a given non-linear prescription, the matter power spectrum changes when we vary one of the cosmological parameters.

In Fig. 1, we show the ratio from equation (1), i.e. how well HALOFIT (left panel) and HMCODE (right panel) agree with EUCLIDEMULATOR1. This is shown at four redshifts ($z = [0, 0.7, 1, 2]$), which includes the ratio today and spans the redshift range of the Euclid galaxy clustering likelihood ($z = 0.7$ to $z = 2$), and for a fixed cosmology (this is also shown for a varying cosmology at $z = 1$ in Fig. 2 columns 1 and 3). We show only the redshift dependence for the reference cosmological model, as the difference from equation (2) changes negligibly across redshift. We see several notable trends:

(i) Both HALOFIT and HMCODE see an overall small shift of the matter power spectrum to smaller scales (larger k), resulting in a disagreement on BAO scales ($k \approx [0.05, 0.5]\,h\,{\rm Mpc}^{-1}$) as seen from the sharp wiggles and the small overall amplitude shift on those scales. For both codes, the disagreement on BAO scales compared to EUCLIDEMULATOR1 increases with smaller redshift. The disagreement on BAO scales compared to EUCLIDEMULATOR1 is obviously highly scale-dependent due to the BAO feature, but it is maximally about 5 per cent for HALOFIT at $z = 0$ (about 2 per cent in the Euclid likelihood redshift range) and about 8 per cent for HMCODE at $z = 0$ (about 4 per cent in the Euclid likelihood redshift range). For HALOFIT, we see a larger amplitude of the matter power spectrum compared to EUCLIDEMULATOR1 across all redshifts considered, while for HMCODE the sign of the amplitude shift depends on redshift: at $z = 2$ the amplitude is shifted to smaller values, while at the other redshifts considered the amplitude of the matter power spectrum is shifted to larger values.

(ii) On small scales ($k \gtrsim 0.5\,h\,{\rm Mpc}^{-1}$), we see a very different behaviour between each of the three codes. Across all redshifts, HALOFIT sees a larger amplitude of the matter power spectrum compared to EUCLIDEMULATOR1, but in a scale-dependent and redshift-dependent way: at $z = 0$ the disagreement compared to EUCLIDEMULATOR1 peaks at 4 per cent at $k \approx 2.5\,h\,{\rm Mpc}^{-1}$ and is about 2–3 per cent across the rest of the interval ($k \approx [0.5, 5]\,h\,{\rm Mpc}^{-1}$); at $z = 0.7$ the disagreement peaks at about 3.5 per cent at $k \approx 0.8\,h\,{\rm Mpc}^{-1}$, dropping to no difference at $k \approx 2\,h\,{\rm Mpc}^{-1}$, before rising to 3 per cent at $k \approx 5\,h\,{\rm Mpc}^{-1}$; at $z = 1$ the disagreement peaks at about 4.5 per cent at $k \approx 1\,h\,{\rm Mpc}^{-1}$, dropping to no difference at $k \approx 3\,h\,{\rm Mpc}^{-1}$, before rising to 1.5 per cent at $k \approx 5\,h\,{\rm Mpc}^{-1}$; at $z = 2$ the disagreement peaks at about 7.5 per cent at $k \approx 2\,h\,{\rm Mpc}^{-1}$, dropping to about 2–3 per cent near either end point of the interval $k \approx [0.5, 5]\,h\,{\rm Mpc}^{-1}$.

(iii) On small scales ($k \gtrsim 0.5\,h\,{\rm Mpc}^{-1}$) for HMCODE, we see a decrease in the amplitude of the matter power spectrum compared to EUCLIDEMULATOR1 for all redshifts of about 1–3 per cent, except for $z = 0$ and $z = 2$, where the disagreement changes from a 3 per cent decrease in amplitude at around $k \approx [0.5, 1]\,h\,{\rm Mpc}^{-1}$ to a 2–4 per cent increase at around $k \approx [2, 5]\,h\,{\rm Mpc}^{-1}$: at $z = 2$ this comes as a peak at 4 per cent at $k \approx 2.5\,h\,{\rm Mpc}^{-1}$ that again decreases to around 1 per cent at $k \approx 5\,h\,{\rm Mpc}^{-1}$, while at $z = 0$ this comes as a steady increase until a peak at 4 per cent at $k \approx 5\,h\,{\rm Mpc}^{-1}$.

[13] EUCLIDEMULATOR1 does not accept $A_{\rm s}$ as input but only σ_8. The conversion for each set of cosmological parameters is done with CLASS.

Downloaded from https://academic.oup.com/mnras/article/518/2/1859/6619061 by Dept of Energy/OSTI user on 26 August 2024

In Fig. 2, we vary the cosmological parameters one at a time at redshift $z = 1$: the baryon abundance ω_b (first row), the matter abundance ω_m (second row), the tilt of the primordial power spectrum n_s (third row), the Hubble parameter h (fourth row), the EoS of DE w_0 (fifth row), and the amplitude of the primordial power spectrum A_s. We show the ratio from equation (1) in columns 1 (HALOFIT) and 3 (HMCODE), and the difference from equation (2) in columns 2 (HALOFIT) and 4 (HMCODE). Notice that the tested parameter values are taken to be the Euclid Reference Cosmology value (the central value) $\pm i\sigma_{p,\,\mathrm{Planck2018}}$, $i \in 1, 2$ where $\sigma_{p,\,\mathrm{Planck2018}}$ denotes the standard deviation of each parameter p of the Planck 2018 results (for more details see section 17.14 on p. 301 in Planck Collaboration 2018).[14]

We see that the difference (equation 2) of HALOFIT and HMCODE compared to EUCLIDEMULATOR1 is fairly small for ω_b, ω_m, and n_s. For ω_b, the difference peaks at about 0.5 per cent on very large, unobservable scales (small k) for a 2σ variation in the parameter, with negligible difference between HALOFIT and HMCODE; for ω_m the difference peaks on observable scales, but with only about 0.4 per cent maximal difference for HMCODE (the difference is smaller for HALOFIT) for a 2σ variation in the parameter; for n_s the difference is largest for HALOFIT (it is much smaller for HMCODE), peaking at around 0.4–0.6 per cent for a 2σ variation in the parameter.

For the two parameters, h and w_0, we see over per cent level differences: for h we see a large difference of HALOFIT and HMCODE compared to EUCLIDEMULATOR1 across BAO and small scales, with the largest difference on BAO scales, where the shift in BAO peak location results in a difference of up to 1.5 per cent for HMCODE and 1.1 per cent for HALOFIT; for w_0 the largest difference is on small scales, ranging from up to a few permille on BAO scales to 1–1.5 per cent (with largest difference compared to EUCLIDEMULATOR1 being for HALOFIT) at around $k \approx 1.5$–$2\,h\,\mathrm{Mpc}^{-1}$.

For the last parameter, A_s, we see again somewhat smaller differences that reach their maximal amplitude of about 0.6 per cent at $k \sim 1.5\,h\,\mathrm{Mpc}^{-1}$ for HALOFIT. For HMCODE the maximal amplitude is marginally smaller (about 0.5 per cent) and it is reached only at $k \sim 2\,h\,\mathrm{Mpc}^{-1}$.

We will see in Section 5 that the differences in Fig. 2 lead to parameter estimation biases when performing a mock data analysis. This will be discussed further there.

2.2 Sensitivity forecasting and parameter estimation at non-linear scales

In the past, galaxy clustering data analyses used to include a conservative non-linear cut-off in k-space, above which all data are discarded. As, however, it is exactly these non-linear scales that modern cosmological surveys aim to exploit, such a cut-off defeats the point of the surveys. Yet, the theory required to analyse the data from non-linear scales is based on imperfect understanding of the non-linear structure evolution in the Universe due to hydrodynamics, galaxy formation, feedback processes etc. As a consequence, the codes implementing the theory (such as HMCODE, HALOFIT, and EUCLIDEMULATOR1 discussed in the previous subsection) tend to increasingly disagree the more non-linear scales are considered.

We will now discuss how such theory and code disagreements can influence cosmological parameter inference on a grander scale. There are two primary tasks/goals in cosmological parameter inference:

(i) Measuring cosmological parameters with high accuracy (low bias);
(ii) Measuring cosmological parameters with high precision (low variance/uncertainties) in a *robust* manner.

For both tasks, we typically compare a given theory to (mock) data by computing (or sampling) a posterior distribution for the cosmological parameters in a Bayesian way: the parameter value is estimated by the maximum posterior probability, while the uncertainties are derived from the credible contours of the posterior. As credible contours can be reduced by simply ignoring sources of uncertainty, robustness plays also a key role in precision measurements: we try to find the credible contours that are simultaneously as small and as correct as possible.

'Parameter estimation' is the task in which the values of the parameters themselves are to be determined, i.e. in which the maximum posterior probability is to be evaluated. In a real-case scenario, the data come from an actual observation/experiment. In past decades, data were mainly collected from the realm of linear structure formation, which stands on firm theoretical grounds such that there was little disagreement about the theories used to analyse the data.

However, modern cosmological experiments have entered the realm of non-linear structure formation and the corresponding data contain a wealth of valuable information which should not be disregarded. Nevertheless, non-linear structure formation is not known at the same level of precision as its linear counterpart and, hence, it is important to remember that the choice of the predictor may lead to slightly different estimates of the parameter values. In other words, the obtained parameter values are not universal but 'remember' the theory/code they were estimated with. A similar result has been found and published by Martinelli et al. (2021), who used weak lensing observables for their analysis. In this paper, we offer a proof-of-concept solution for this problem inspired by work previously published: we suggest that code-dependence is best handled by introducing a 'theoretical uncertainty' into the data analysis (in Audren et al. 2013a; Baldauf et al. 2016; Sprenger et al. 2019 this concept is introduced solely in the context of sensitivity forecasting). Theoretical uncertainties are discussed more deeply in Section 5.2.

'Parameter sensitivity forecasting' (or short sensitivity forecasting) is a common method to project how precisely a given experiment is able to determine the cosmological parameters. As such, it provides guidance in decision-making processes regarding the experimental design of a given survey and verifies to the community that the science goals are achievable.

In contrast to the task of parameter estimation, sensitivity forecasts are usually performed with mock data. These mock data are typically generated using the same code and theory as is used to compute/sample the posterior distribution, thereby we assume that the data are perfectly modelled with appropriate uncertainties. Developments in cosmological structure formation theory, and the ever-increasing accuracy of computational methods able to predict cosmological observables at non-linear scales, allow scientists to better take into account systematics in their analyses. For this reason, it is an interesting exercise in its own right to measure the performance of newly developed methods applied to sensitivity forecasts

[14] Notice that setting $w_0 = w_0^{\mathrm{ref}} - 2\sigma_{\mathrm{p,Planck2018}} = -1.2$ while keeping all other cosmological parameter fixed to their fiducial values results in a σ_8 value outside the parameter space accepted by EUCLIDEMULATOR1 (see Euclid Collaboration 2019 for the definition of the accepted parameter space). For this reason the corresponding w_0 value was manually adjusted to $w_0 = -1.19$. Of course, if one wants to probe parameters outside the allowed parameter range of any given emulator, an emulator with a wider parameter space (but preferably comparable accuracy over the entire space) has to be used.

Downloaded from https://academic.oup.com/mnras/article/518/2/1859/6619061 by Dept of Energy/OSTI user on 26 August 2024

and benchmark them against more established tools. However, even when various sources of uncertainties are taken into account, it is highly unlikely that none have been missed at all.

In the context of sensitivity forecasts, theoretical uncertainties can hence be used in order to (1) fold in uncertainties due to neglected systematics (such as e.g. baryonic physics or scale-dependent galaxy bias) and (2) to deal with unknown uncertainties that have not been modelled. As, however, different codes implement different theories and assumptions and hence take into account different sources of uncertainties, the result of sensitivity forecast does also depend on the choice of code used. A theoretical uncertainty hence makes a sensitivity forecast more robust and less dependent on the modelling choices implemented into the codes used in the forecast.

2.3 Treatment of known systematics: parameter inference with cosmological emulators

There is a growing number of power spectrum predictors available on the market, ranging from full-fledged cosmological simulations over halo model derivatives and PT-based models all the way to purely data-driven approaches. The first kind offers the most accurate predictions but have limited use for sensitivity forecasts and parameter estimations due to their immense computational expense. PT- and halo model-based models are orders of magnitudes cheaper to evaluate than simulations but they suffer from more systematics as they are based on calibrated fitting functions. In the past, they were the choice of power spectrum predictors in parameter inference due to their low computational cost. However, emulators are comparable with (semi-)analytical models in terms of computational expense but often carry less systematics as they are based on simulations. It must be emphasized, though, that the remaining systematics need to be treated carefully. Observations necessarily contain information whose modelling needs to take into account sources of uncertainty such as redshift-space distortions (RSDs), the Alcock–Paczyński effect, galaxy bias, baryonic effects on structure formation etc. (see Section 3.1).

Although this is not a special trait of emulators, many are actually designed to predict correction factors to deal with individual systematics. Examples are the NLC of the matter power spectrum (Euclid Collaboration 2019, 2021), the baryonic emulator (Schneider et al. 2020; Giri & Schneider 2021), the halo bias (Valcin et al. 2019; McClintock et al. 2019a), or various emulators from the BACCO project (Aricò et al. 2020, 2021). This way, multiple individual emulators, each modelling a different effect, can, under certain assumptions, be stacked on top of each other to take a multitude of systematics into account.

It can therefore be anticipated that future parameter inference studies will use emulators as robust predictors that may include uncertainties in a modular way. A possible approach to deal with unknown sources of systematic errors is discussed in the subsequent subsection.

2.4 Treatment of unknown systematics: parameter estimation with theoretical uncertainty

Theoretical uncertainties and their applications in the context of sensitivity forecasts have been studied in previous work (e.g. Audren et al. 2013a; Baldauf et al. 2016; Sprenger et al. 2019). In this work, we propose to extend the use of theoretical uncertainties in order to mitigate the negative impact of theory and code disagreements on parameter estimations due to unknown systematic effects (see Section 2.2). To this end it is natural to start with a theoretical uncertainty readily available such as the one published in Sprenger et al. (2019) and study its effects on parameter estimation. We hence review this particular uncertainty envelope in Section 2.4.1. In addition, we introduce a new theoretical uncertainty in Section 2.4.2 that is tailored to the power spectrum prediction codes used in this paper. The comparison of the two theoretical uncertainties is presented later in the context of a mock parameter estimation in Section 5.2.

2.4.1 *Review of the simulation-based theoretical uncertainty*

We start out reviewing the theoretical uncertainty of Sprenger et al. (2019), which aims to identify the theoretical precision with which we will be able to analyse *Euclid* data using cosmological simulations. This uncertainty envelope, in the following denoted $\sigma_{\mathrm{th,\,S}}$, is roughly given by the agreement between hydrodynamical simulations at the time of publishing of that work, which is expected to improve by the time *Euclid* data are available and being analysed. The envelope $\sigma_{\mathrm{th,\,S}}$ may be sufficient for any traditional forecasting with an aim to predict the sensitivity of future surveys as state-of-the-art tools on average disagree at a level that is mostly captured by $\sigma_{\mathrm{th,\,S}}$ (Sprenger et al. 2019). It was shown in that paper that accounting for uncertainties in the simulations is important with regards to parameter constraints. Indeed, they conclude that taking a carefully constructed uncertainty envelope into consideration one arrives at improved constraints and sensitivities for various cosmological parameters, compared to simply removing all information from those scales, even though a lot of information is removed at small scales due to the theoretical uncertainty.

The approach of Sprenger et al. (2019) takes three uncertainty values at three different wave numbers as inputs,

(i) 0.33 per cent at $k = 0.01\,h\,\mathrm{Mpc}^{-1}$,
(ii) 1.0 per cent at $k = 1\,h\,\mathrm{Mpc}^{-1}$,
(iii) 10.0 per cent at $k = 10\,h\,\mathrm{Mpc}^{-1}$,

and interpolates between these points using the ansatz for the uncertainty envelope

$$\alpha(k,z) = \begin{cases} a_1 \exp c_1 \log_{10}\left[\frac{k}{k_{1\,h\,\mathrm{Mpc}^{-1}}(z)}\right], & \frac{k}{k_{1\,h\,\mathrm{Mpc}^{-1}}(z)} < 0.3 \\ a_2 \exp c_2 \log_{10}\left[\frac{k}{k_{1\,h\,\mathrm{Mpc}^{-1}}(z)}\right], & \frac{k}{k_{1\,h\,\mathrm{Mpc}^{-1}}(z)} > 0.3 \end{cases}, \quad (3)$$

where $\alpha \propto \sigma_{\mathrm{th,\,S}}$ and k depends on the line-of-sight direction μ and the cosmology, $k = k(\mu, \mathrm{cosmo})$ (for more details on this point see Sprenger et al. 2019).

In general, we define

$$k_{k_0}(z) = k_0 \cdot (1+z)^{\frac{2}{2+n_{\mathrm{s}}}}, \quad (4)$$

such that $k_{1\,h\,\mathrm{Mpc}^{-1}}(z)$ is defined accordingly with $k_0 = 1\,h\,\mathrm{Mpc}^{-1}$. In order to interpolate between the points mentioned above the coefficients are set to

$$a_1 = 0.014806,\ a_2 = 0.022047,\ c_1 = 0.75056,\ c_2 = 1.5120\,. \quad (5)$$

See Fig. 3 for an illustration of the uncertainty envelopes, including their redshift dependence. The blue lines represent $\sigma_{\mathrm{th,\,S}}$ at three different redshifts.

In Sprenger et al. (2019) (see their section 5 for details) the theoretical 1σ uncertainty was given as

$$\sigma_{\mathrm{th}}(k,\mu,z) = \left[\frac{V_r(z)}{2(2\pi)^2} k^2 \Delta k \Delta\mu \frac{\Delta z}{\Delta \bar{z}}\right]^{1/2} \alpha(k,\mu,z) P_{\mathrm{gg}}(k,\mu,z) \quad (6)$$

Downloaded from https://academic.oup.com/mnras/article/518/2/1859/6619061 by Dept of Energy/OSTI user on 26 August 2024

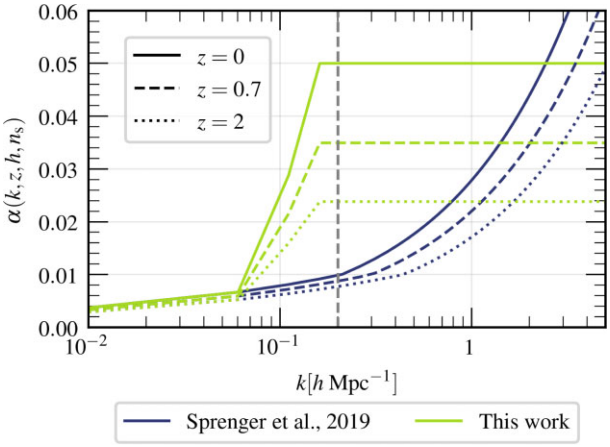

Figure 3. The two different envelope functions α for theoretical uncertainty as defined in equation (3). The solid, dashed, and dotted lines indicate the uncertainty envelope at redshifts $z = 0$, $z = 0.7$, and $z = 2$, respectively, where [0.7,2] is the redshift range of the Euclid likelihood. The blue lines indicate the original theoretical uncertainty $\sigma_{\mathrm{th,\,S}}$ from Sprenger et al. (2019) and the green lines correspond to $\sigma_{\mathrm{th,\,KB}}$, a rough estimate of the agreement between EUCLIDEMULATOR1 and HMCODE, as described in the text in Section 2.4. The vertical dashed grey line indicates $k = 0.2\,h\,\mathrm{Mpc}^{-1}$, which is a popular choice for the upper k cut-off value. In most cases, we use the theoretical uncertainty $\sigma_{\mathrm{th,\,S}}$, but at the end we compare to the other uncertainty for a cross-comparison case.

with redshift correlation length Δz, survey redshift bin width $\Delta\bar{z}$, survey volume per redshift bin $V_r(z)$, and wavenumber correlation length Δk. This uncertainty is then simply added in quadrature to the observational uncertainty in the χ^2 computation.

However, we note that the publicly available likelihood (as of MONTEPYTHON version 3.4) has a bug resulting in an inverted ratio of $(\frac{\Delta z}{\Delta\bar{z}})^{-1}$, which with the chosen redshift correlation length $\Delta z = 1$ and survey redshift bin width $\Delta\bar{z} = 0.1$ gives rise to a factor $(100)^{1/2} = 10$ smaller uncertainty than intended. Since correlation lengths are inherently difficult to determine, and Sprenger et al. (2019) used very conservative correlation lengths, this can largely be interpreted as an adjustment to less conservative correlation lengths that would still be within the range of possible values: i.e. if we use $\Delta z = 0.1$ instead of $\Delta z = 1$, $\Delta\mu = 1$ instead of $\Delta\mu = 2$, and $\Delta k = 0.02\,h\,\mathrm{Mpc}^{-1}$ instead of $\Delta k = 0.05\,h\,\mathrm{Mpc}^{-1}$ then without the bug we would have a factor 2 smaller uncertainty than intended. Although unfortunate, given that the precision of future cosmological simulations that the uncertainty is intended to mimic is not known and Sprenger et al. (2019) instead used the precision near the time of writing, a factor 2 smaller uncertainty (improved precision) is not unrealistic. This has been fixed in the upcoming MONTEPYTHON v3.6, with a flag determining which uncertainty to use, including the option to use the KB theoretical uncertainty from this work.[15]

In contrast to the hydrodynamical simulations mentioned above, HALOFIT and HMCODE exhibit an error on BAO scales that is sufficient to bias the result of data analyses.[16] While we would expect the Euclid Collaboration analysis of future data to employ methods that accurately capture the BAO scales, leaving the Sprenger et al. (2019) theoretical uncertainty robust for forecasts, we find in this paper (cf. Section 5.1) that the bias remains for analyses that employ commonly used current fast methods such as e.g. HALOFIT and HMCODE. Reducing this bias motivates the construction of a new uncertainty envelope that will be the topic of the next subsection.

2.4.2 *Theoretical uncertainty based on non-linear predictors*

EUCLIDEMULATOR1 has been compared to HALOFIT and HMCODE in Fig. 2 (for a comparison between EUCLIDEMULATOR1 and HALOFIT also see Euclid Collaboration 2019). The agreement at linear scales is almost perfect. At intermediate scales the agreement is at the 1–2 per cent level for redshift $z = 2$, while it is at the few per cent level for $z = 0$. For smaller scales the agreement is at the few per cent level for all redshifts $0 \leq z \leq 2$. We therefore construct a new theoretical uncertainty envelope $\sigma_{\mathrm{th,\,KB}}$ based on the disagreement between these codes (see green lines in Fig. 3).

We find that this new uncertainty envelope reduces the bias on (mock) data analyses significantly (discussion in Section 5.2). However, while the disagreement between HALOFIT/HMCODE and EUCLIDEMULATOR1 oscillates considerably in k at intermediate and small scales for all z, for the sake of simplicity we chose to model the uncertainties with a constant close to the maximal disagreement found on scales $0.1\,h\,\mathrm{Mpc}^{-1} \lesssim k \lesssim 1\,h\,\mathrm{Mpc}^{-1}$. Hence, at these scales our uncertainty is somewhat overly conservative. The resulting uncertainty envelope follows the one by Sprenger et al. (2019) up to $k = 0.05\,h\,\mathrm{Mpc}^{-1}$ and is fixed to 5 per cent above $k = 0.15\,h\,\mathrm{Mpc}^{-1}$ at $z = 0$. The intermediate part is modelled by equation (3) using the coefficients $a_2 = 0.2998$ and $c_2 = 2.2517$, with the amplitude fixed to $\sigma_{\mathrm{th,\,S}}$ at $k = 0.05\,h\,\mathrm{Mpc}^{-1}$.

Another question is up to which k_{max} mode these uncertainties can be trusted. Such a k_{max} represents a cut-off scale above which the theoretical uncertainty is de facto infinite. In other words, all information coming from k modes larger than k_{max} is simply discarded. We use two different cut-off points: one ultra-conservative disregarding all scales above $k_{\mathrm{max}} = k_{0,\mathrm{max}} = 0.2\,h\,\mathrm{Mpc}^{-1}$ at $z = 0$, and one informed by our choice of galaxy bias model (see Section 3.1), which is only reliable up to about $k_{\mathrm{max}} = k_{0,\mathrm{max}} \approx 0.4\,h\,\mathrm{Mpc}^{-1}$ at $z = 0$. Both scale with redshift as defined in equation (4).

In this work, emulators are heavily employed. We should hence also consider the impact of the emulation accuracy on the construction of the theoretical uncertainty envelope. The error due to emulation with EUCLIDEMULATOR1 is at sub-per cent level (Euclid Collaboration 2019) and hence negligible compared to the disagreement between the hydrodynamical simulations. As a result, the emulation error can be safely ignored in the construction of the theoretical uncertainty envelope.

3 ANALYSIS METHODOLOGY

In Section 1, we listed the four key questions that guide our efforts in this paper:

(i) What are the performance differences of EUCLIDEMULATOR1, HALOFIT, and HMCODE?

[15] In the `montepython/likelihoods/euclid_pk/euclid_pk.data` likelihood file set `euclid_pk.TheoErrorType='Sprenger_original'` for the theoretical uncertainty as used in Sprenger et al. (2019), `euclid_pk.TheoErrorType='Sprenger'` for the uncertainty as intended by Sprenger et al. (2019), and `euclid_pk.TheoErrorType='KB'` for the KB theoretical uncertainty of this work.

[16] Note this specific issue was mitigated in the latest version of HMCODE (Mead et al. 2021).

Downloaded from https://academic.oup.com/mnras/article/518/2/1859/6619061 by Dept of Energy/OSTI user on 26 August 2024

(ii) What is the added value of considering (mildly) non-linear scales in parameter forecasts?

(iii) What is the impact of the choice of the (non-linear) predictor model on the parameter estimation result?

(iv) How are the forecasting results affected by different choices of theoretical uncertainty models?

In order to answer these questions we run and analyse a set of MCMCs. In this section, we discuss the general aspects of the MCMC analysis methodology as well as the experimental set-up used to generate those MCMCs. The results and interpretations will be reported later in Sections 4 and 5.

3.1 Modelling the galaxy–galaxy power spectra

EUCLIDEMULATOR1 predicts the NLC to the dark matter power spectrum only, which is not an observable quantity. A related quantity that we can measure from observations is the galaxy–galaxy power spectrum $P_{\rm gg}(k,\,\mu,\,z)$, where μ is the cosine between the wave mode $\mathbf{k}$ and the line-of-sight direction $\mathbf{e}_r$. Following the expansion in Sprenger et al. (2019) and using their notation, we have

$$P_{\rm gg}(k,\mu,z) = f_{\rm AP}(z)\times f_{\rm res}(k,\mu,z)\times f_{\rm RSD}(\hat{k},\hat{\mu},z)\times b^2(z) \times P_{\delta\delta}(k,z)\,, \tag{7}$$

where $f_{\rm AP}$ denotes the correction due to the Alcock–Paczyński effect, $f_{\rm res}$ is a correction factor necessary because the resolution is limited, $f_{\rm RSD}$ accounts for the RSDs, and $b(z)$ is the galaxy bias taking into account that galaxies are only biased tracers of the underlying matter distribution. Notice that due to the Alcock–Paczyński effect the parameters k and μ are actually functions of the underlying cosmology. Parameters with a hat denote that they are based on the true/real but unknown cosmology, while those without a hat are based on an assumed underlying cosmology. For a more elaborate explanation of this, see Sprenger et al. (2019).

For the galaxy bias we choose a linear, redshift-dependent bias $b(z) = \beta_0(1+z)^{0.5\beta_1}$ as in Sprenger et al. (2019), where β_0 and β_1 are included as nuisance parameters in the MCMC analysis. Since galaxy bias is known to be better described by a scale-dependent galaxy bias (e.g. Desjacques, Jeong & Schmidt 2018; Giusarma et al. 2018), we choose to cut the analysis at an ultra-conservative $k_{\rm max} = 0.2\,h\,{\rm Mpc}^{-1}$ and a somewhat less conservative $k_{\rm max} = 0.4\,h\,{\rm Mpc}^{-1}$, assuming that any residual inaccuracy of the bias model up to these scales would be picked up by the theoretical uncertainty.

The peculiar motion of galaxies give rise to the so-called Fingers of God (Jackson 1972; Tully & Fisher 1978) and are treated separately as an exponential suppression of power following Taruya, Nishimichi & Saito (2010) and Bull et al. (2015), where the suppression is governed by a nuisance parameter $\sigma_{\rm NL}$ that we marginalize over. The remaining RSD effect is described by the Kaiser formula (Kaiser 1987), which relates the galaxy power spectrum $P_{\rm gg}$ to the dark matter power spectrum $P_{\delta\delta}$. EUCLIDEMULATOR1 directly gives us an NLC factor $B(k,z) = P_{\delta\delta}^{\rm non-linear}(k,z)/P_{\delta\delta}^{\rm linear}(k,z)$, which, setting aside the other correction factors for the moment, allows us to write the RSD and bias correction as (see Appendix A)

$$P_{\rm gg}(k,\mu,z) = b^2(z)\left[B(k,z) + 2\mu^2 f + \mu^4 f^2\right] P_{\delta\delta}^{\rm linear}(k,z)\,. \tag{8}$$

For HALOFIT and HMCODE, which return a non-linear power spectrum, we find identical results expressing this as (see Appendix A)

$$P_{\rm gg}(k,\mu,z) = b^2(z)\left[1+\mu^2 f\right]^2 P_{\delta\delta}^{\rm non-linear}(k,z) \tag{9}$$

The other correction factors are multiplicative and can trivially be included in these expressions. For deeper discussion of all of these correction factors we refer the reader to Sprenger et al. (2019) and references therein.

3.2 Experimental set-up

We run the MCMCs with MONTEPYTHON version 3.1 (Audren et al. 2013b; Brinckmann & Lesgourgues 2019) and CLASS version 2.8 (Blas, Lesgourgues & Tram 2011; Lesgourgues 2011). We use the publicly available Euclid galaxy clustering likelihood[17] from Sprenger et al. (2019), modified according to Sections 3.1 and 2.4 to allow for including NLCs by the EUCLIDEMULATOR1 and a modified theoretical uncertainty scheme.

The likelihood uses 13 redshift bins in the redshift range $0.7 < z < 2$, a sky fraction $f_{\rm sky} = 0.3636$, nine μ bins, and 100 logarithmically spaced bins in k-space from $k_{\rm min} = 0.02\,{\rm Mpc}^{-1}$ to $k_{\rm max} = 0.2$ or $0.4\,h\,{\rm Mpc}^{-1}$ (see the end of Section 2.4 for a discussion of $k_{\rm max}$). For more details we refer the reader to Sprenger et al. (2019).

We combine the Euclid likelihood with the fake Planck likelihood and noise spectra from Brinckmann et al. (2019), which is based on the old 'fake planck bluebook' likelihood with specifications updated to the full mission and omitting the 217 GHz channel, in order to avoid foreground complications. The end result is a likelihood in good agreement with the sensitivity of the Planck 2018 results (Planck Collaboration VI 2020), but with a slightly better sensitivity for the optical depth to reionization, $\sigma(\tau_{\rm reio})$. The fake Planck likelihood considers a maximum multipole of $\ell_{\rm max} = 3000$ for temperature, polarization, and lensing, with the noise exploding long before the maximum multipole (see fig. 1 of Brinckmann et al. 2019).

3.3 MCMC analysis

We use the MCMC technique to compare the agreement between a fiducial and a predicted galaxy–galaxy power spectrum, $P_{\rm gg}$, where the latter takes a cosmology as an input argument. While $P_{\rm gg}$ is the observable galaxy clustering quantity used in this work, the three codes compared to one another all predict the non-observable matter power spectrum $P_{\delta\delta}$. Technical details about how $P_{\rm gg}$ is related to the $P_{\delta\delta}$ and how the NLC predicted by EUCLIDEMULATOR1 is worked into this relation are given in equations (7) and (8).

In this paper, we perform sensitivity forecasts and mock parameter estimations instead of real-world data analysis. Hence, we first have to construct a fiducial model. As will be explained in more detail below, our MCMC runs can be split into two classes, which we refer to as 'autocomparison' (ac) and 'cross-comparison' (cc) tests (see Section 5 for an explanation why we do this). In the case of autocomparisons, the fiducial model is set up with the same cosmological model, and cut-off scale as the model used to perform the theoretical predictions during the MCMC. For cross-comparisons, on the other hand, we compute a fiducial spectrum with EUCLIDEMULATOR1 in all cases and use either HALOFIT or HMCODE for the theoretical predictions. Notice that the choice of the theoretical uncertainty model does not affect the computation of the fiducial result.

We let the MCMCs run on the zBox4 cluster based at the University of Zurich.[18] We run 12 chains simultaneously for each MCMC. For

[17] Note that in cosmological inference the term 'likelihood' is typically a package containing all modelling of observables plus the likelihood computation.

[18] https://www.ics.uzh.ch/~stadel/doku.php?id=zbox:zbox4

Downloaded from https://academic.oup.com/mnras/article/518/2/1859/6619061 by Dept of Energy/OSTI user on 26 August 2024

any information about how this works in detail, we refer the reader to Audren et al. (2013b) and Brinckmann & Lesgourgues (2019).

To determine convergence we use the standard criterion by Gelman & Rubin (1992), $R - 1 < 0.01$, where R is the Gelman–Rubin coefficient.[19] We let the MCMCs run until this convergence criterion is satisfied in all parameters.

4 SENSITIVITY FORECASTS

Traditionally, different power spectrum predictors are compared at the power spectrum level as is done e.g. in Lawrence et al. (2010), Bird, Viel & Haehnelt (2012), Mead et al. (2016), Lawrence et al. (2017), Euclid Collaboration (2019), Angulo et al. (2021), Euclid Collaboration (2021), and Mead et al. (2021). However, it is not often studied directly how the differences between the predictors impact parameter forecasts in the end (see, however, Martinelli et al. 2021 for a complementary analysis for cosmic shear). We close part of this gap in this section by analysing the code dependence of forecasting results on the three codes studied in this paper, leaving investigations of this aspect with other codes to future work. Notice, though, that similar analyses have been performed in the context of spectroscopic galaxy clustering with PT models, e.g. in Osato et al. (2019), Markovič, Bose & Portsidou (2019), and Pezzotta et al. (2021).

Code dependencies in forecasting results could manifest themselves in different ways: uncertainty contours could be smaller or show different degeneracy directions. Ultimately, we would want to understand the reasons behind any differences, which could be due to mistakes in the modelling, e.g. underestimating uncertainties, not accurately modelling uncertainties (e.g. due to baryonic effects), or incorrectly modelling non-linear effects, or it could be due to new physical signals, e.g. more accurate modelling of non-linear effects introducing new signals that were neglected with other methods.

In this section, we perform an autocomparison MCMC for each code in which we try to fit a fiducial power spectrum with the same predictor code as the one used to compute the fiducial.[20] In such a comparison, no significant difference in the best-fitting cosmologies between the different codes is expected and differences in the posterior probabilities are limited to different sensitivity contours.

4.1 Sensitivity to the cosmological parameters

We first compare the sensitivity to the cosmological parameters for Euclid galaxy clustering plus (fake) Planck cosmic microwave background for each of the three different predictors looked at in this paper. To this end, we analyse four different cases: Λ cold dark matter (CDM) and w_0CDM as well as two different cut-off scales, $k_{\rm max} = 0.2\,h\,{\rm Mpc}^{-1}$ and $k_{\rm max} = 0.4\,h\,{\rm Mpc}^{-1}$.

The 1D marginalized posterior standard deviations for each parameter in all four cases and for all three predictors are summarized in Table B1. The relative differences of these standard deviations are graphically represented in Fig. 4 for better understanding.

Clearly, this relative difference varies dramatically depending on the considered parameter: for instance, the sensitivity of EUCLIDEMULATOR1 to $\omega_{\rm m}$ in the case of a ΛCDM cosmology cut-off at $k_{\rm max} = 0.2\,h\,{\rm Mpc}^{-1}$ is almost the same compared to the corresponding sensitivity of HMCODE (the relative difference is around 1 per cent). However, $\sigma_{\rm EE}$ for h in the case of a w_0CDM cosmology cut-off at $k_{\rm max} = 0.4\,h\,{\rm Mpc}^{-1}$ is reduced by ~ 70 per cent compared to the corresponding $\sigma_{\rm HM}$ value. In other words: EUCLIDEMULATOR1 is roughly 70 per cent more sensitive to h than HMCODE. The more complex the cosmological model and also the higher the cut-off scale $k_{\rm max}$, the bigger the effect, i.e. the smaller the uncertainty obtained from a EUCLIDEMULATOR1-based fit compared to a HALOFIT- or HMCODE-based fit.

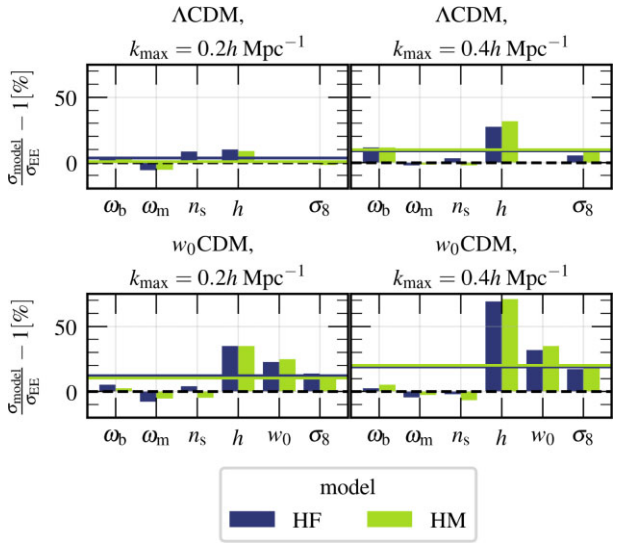

Figure 4. Relative differences (in per cent) of 1D marginalized posterior standard deviations (σ) for each cosmological parameter between the two models HALOFIT (HF) and HMCODE (HM), respectively, with respect to EUCLIDEMULATOR1 (EE) for all four tested cases (given in the panel titles). The solid horizontal lines represent the average over all parameters. Notice that a positive ratio implies that $\sigma_{\rm EE} < \sigma_{\rm model}$. In all four cases, on average, the 1D marginalized posterior standard deviations computed with EUCLIDEMULATOR1 are smaller than those computed with either of the two alternative models. The more complex the cosmological model and also the higher the cut-off scale $k_{\rm max}$, the bigger the effect, i.e. the smaller the uncertainty obtained from a EUCLIDEMULATOR1-based fit compared to a HALOFIT- or HMCODE-based fit. The averaged relative differences vary between 0.9 per cent (HMCODE versus EUCLIDEMULATOR1, ΛCDM, $k_{\rm max} = 0.2\,h\,{\rm Mpc}^{-1}$) and 20.4 per cent (HMCODE versus EUCLIDEMULATOR1, w_0CDM, $k_{\rm max} = 0.4\,h\,{\rm Mpc}^{-1}$) depending on the case.

In the same Fig. 4, we also show the average values of those relative differences over all parameters (for each studied case). These averaged values range from only 0.9 per cent (HMCODE versus EUCLIDEMULATOR1, ΛCDM, $k_{\rm max} = 0.2\,h\,{\rm Mpc}^{-1}$) to 20.4 per cent (HMCODE versus EUCLIDEMULATOR1, w_0CDM, $k_{\rm max} = 0.4\,h\,{\rm Mpc}^{-1}$) depending on the case. This quantity serves as a proxy for the volume (to the $1/d$ power, where d is the dimension of the parameter space under investigation) of the 1σ credible region of the posterior distribution and thus for the sensitivity to the cosmological parameters. As is stated above and clearly visible in Fig. 4, this average value is always positive, meaning that on average the 1D marginalized posterior standard deviations computed from a EUCLIDEMULATOR1-based posterior is smaller than the corresponding value of a posterior based on either HALOFIT or HMCODE. It also becomes clear that the effect is bigger the more complex the cosmological model under consideration is and the higher the cut-off scale $k_{\rm max}$.

[19] Note that the accuracy of the Gelman–Rubin convergence criterium depends on the number of chains as well as the length of the chains, but with 12 chains per run $R - 1 < 0.01$ leads to exceedingly good convergence.

[20] This is the standard approach to forecasting and has been shown to provide reliable uncertainties compared to a randomly generated fiducial, as long as the difference cosmology is not extreme (Perotto et al. 2006).

Downloaded from https://academic.oup.com/mnras/article/518/2/1859/6619061 by Dept of Energy/OSTI user on 26 August 2024

Table 1. Cosmological (top seven rows) and nuisance parameters of the fiducial cosmology used for all MCMC forecasting runs. Rows 8 to 10 list the varying RSD and galaxy bias nuisance parameters of the *Euclid* likelihood (see Section 3.1), while the bottom two rows contain baryonic feedback nuisance parameters of HMCODE, which remain fixed. The fiducial cosmology coincides with the Euclid Reference Cosmology (as defined in the Euclid Collaboration 2019). Notice that the minimal and maximal values are broadly equal to the parameter box of EUCLIDEMULATOR1 (keep in mind that $\Omega_{\rm cdm}h^2$ is not a direct input parameter to EUCLIDEMULATOR1).

Parameter	Fiducial value	Min	Max
$\Omega_{\rm b}h^2$	0.022	0.0215	0.0235
$\Omega_{\rm cdm}h^2$	0.121	0.1091	0.1311
$n_{\rm s}$	0.96	0.9283	1.0027
h	0.67	0.6155	0.7307
w_0	0.9619	−1.25	−0.75
σ_8	0.83	0.7591	0.8707
$z_{\rm reio}$	8.24	0.0	none
$\sigma_{\rm NL}$	7.0	4.0	10.0
$\beta_0^{\rm Euclid}$	1.0	none	none
$\beta_1^{\rm Euclid}$	1.0	none	none
η_0	0.603	fixed	
$c_{\rm min}$	3.13	fixed	

EUCLIDEMULATOR1 models the matter power spectrum much more accurately than either of the halo model-based models, particularly for $0.1\,h\,{\rm Mpc}^{-1} \lesssim k \lesssim 5\,h\,{\rm Mpc}^{-1}$, as is shown in Euclid Collaboration (2019). However, it is not guaranteed that a more accurate model produces smaller posterior contours (only less biased ones). The fact that we still observe smaller contours here may be explained by the fact that EUCLIDEMULATOR1 models the BAOs more accurately than the other two predictors, i.e. scales that carry lots of valuable information about the cosmology. It should, however, not be forgotten that EUCLIDEMULATOR1 does not take baryonic effects into account, while HMCODE can. But we note that we did not take them into account even for the HMCODE runs, instead fixing the nuisance parameters related to baryonic physics to the default values (see Table 1).

Answer to question #1: EUCLIDEMULATOR1 models the $P(k)$ more accurately than HALOFIT or HMCODE and it also leads, on average, to better sensitivity for the cosmological parameters than the latter two codes, with the largest improvements seen for h and w_0 as EUCLIDEMULATOR1 is better at modelling the effect of these parameters on the shape of the matter power spectrum.

4.2 The effect of including mildly non-linear scales

In order to establish the importance of mildly non-linear scales in cosmological parameter estimations, we perform autocomparison tests for two different cut-off scales: $k_{\rm max} = 0.2\,h\,{\rm Mpc}^{-1}$ and $k_{\rm max} = 0.4\,h\,{\rm Mpc}^{-1}$. Although all predictors used in this study could actually reach more strongly non-linear scales, we do not go beyond $0.4\,h\,{\rm Mpc}^{-1}$ because we use only a very simple, linear galaxy bias model which is not valid at small scales.

We quantify the effect of including mildly non-linear scales on the posterior sampling by comparing the sizes of the resulting contours. Of course, a more in-depth analysis employing a more sophisticated galaxy bias model needs to be conducted in order to draw conclusion on the effect of including strongly non-linear scales. As this paper is meant to convey a proof of concept, we leave these analyses to future work.

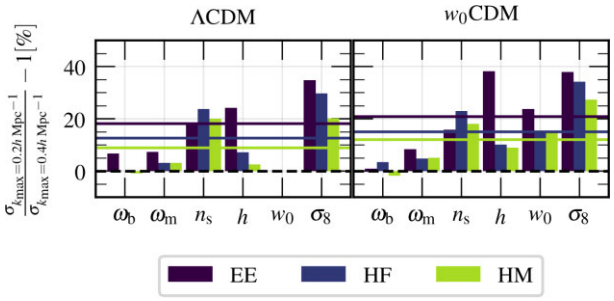

Figure 5. Relative differences (in per cent) of 1D marginalized posterior standard deviations (σ) per cosmological parameter between the two different cut-off scales $k_{\rm max} = 0.2\,h\,{\rm Mpc}^{-1}$ and $k_{\rm max} = 0.4\,h\,{\rm Mpc}^{-1}$ for each of the three predictors. The averaged relative differences range between 9.2 per cent and 18.2 per cent for ΛCDM and from 12.1 per cent up to 20.9 per cent in the w_0CDM case. From this plot it becomes most evident that including smaller scales leads to smaller uncertainties.

To this end, we interpret again the results presented in Table B1. This time, however, we shall not compare the different prediction models but perform all comparisons for each of the predictors separately. In Fig. 5, we plot the relative differences (again in per cent) of the 1D marginalized posterior standard deviations between the two different cut-off scales for each of the power spectrum prediction models. Positive values imply that $\sigma_{k_{\rm max}=0.4\,h\,{\rm Mpc}^{-1}} < \sigma_{k_{\rm max}=0.2\,h\,{\rm Mpc}^{-1}}$. This is the case for almost all parameters for all predictors: the effect of increasing the cut-off scale is largest for EUCLIDEMULATOR1 where the posterior standard deviation averaged over all parameters is reduced by 18.2 per cent in the ΛCDM scenario and by 20.9 per cent in the w_0CDM case. The minimal effect found is 9.2 per cent (ΛCDM) and 12.1 per cent (w_0CDM), respectively. As above, we have represented the average effect (over all parameters) as solid horizontal lines.

Answer to question #2: The uncertainties can be significantly reduced by increasing $k_{\rm max}$ from 0.2 to $0.4\,h\,{\rm Mpc}^{-1}$. This is not only true for EUCLIDEMULATOR1 (though for it the effect is largest) but for all tested prediction models. The magnitude of the effect depends strongly on the model and on the predictor under consideration. This observation does not come as a surprise: it is expected that there is valuable information stored in the galaxy–galaxy power spectrum on (mildly) non-linear scales that can be leveraged in order to estimate the cosmological parameters. The effect is expected to grow as the cut-off scale increases. We reiterate, however, that increasing $k_{\rm max}$ beyond $\sim 0.4\,h\,{\rm Mpc}^{-1}$ requires a more sophisticated galaxy bias model than is considered in this paper. So we leave this analysis to future work.

5 MOCK PARAMETER ESTIMATION

Usually, for forecasting we use the same (non-linear) prediction model for cosmological parameter estimations as is used to compute the fiducial cosmology, i.e. autocomparison tests are the common choice (like in the previous section). This is fine as long as we are interested in the sensitivity of the model response to the choice of parameters (Perotto et al. 2006). We can, however, mimic the situation of data analysis by e.g. taking EUCLIDEMULATOR1 to be the model that produces the 'observed' data which we try to fit with different predictors, which we refer to as the 'cross-comparison' case. I.e. in this section, we always produce the 'observed' mock data with EUCLIDEMULATOR1. Running such cross-comparison tests using differ-

Downloaded from https://academic.oup.com/mnras/article/518/2/1859/6619061 by Dept of Energy/OSTI user on 26 August 2024

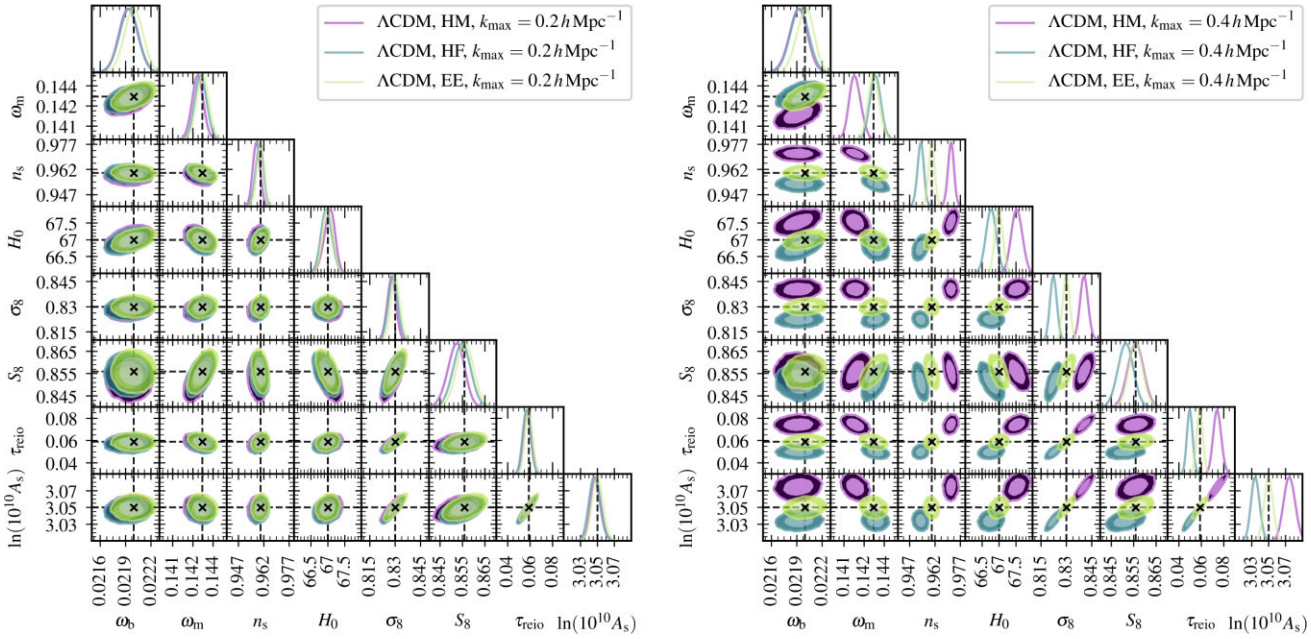

Figure 6. ΛCDM cross comparison of HMCODE and HALOFIT versus EUCLIDEMULATOR1 considering $k \leq 0.2\,h\,{\rm Mpc}^{-1}$ (left triangle plot) and $k \leq 0.4\,h\,{\rm Mpc}^{-1}$ (right triangle plot), respectively. We find that all the three best-fitting cosmologies agree relatively well with each other in the left panel as mostly linear scales are considered. In the right panel, however, clear differences in the posterior distributions emerge due to disagreements in the power spectrum predictions by the three different codes.

ent predictors for the fitting are likely to lead to different best-fitting cosmologies (as we will also confirm later), as different codes encode different physics and differ in the way they are implemented. Such tests can hence be used to emphasize that any cosmological parameter estimation result significantly depends on the code used, unless proper care is taken to account for uncertainties on non-linear scales.

5.1 Code dependence of best-fitting cosmology

The third of our key questions (see Section 3) is concerned with the choice of the power spectrum prediction code and its effect on the best-fitting cosmology. In this subsection, we analyse and interpret the results from cross-comparison tests, where we compare power spectrum fits computed with HALOFIT and HMCODE with a fiducial power spectrum computed with EUCLIDEMULATOR1. The resulting posteriors for the ΛCDM case are shown in Fig. 6 (left panel: $k_{\rm max} = 0.2\,h\,{\rm Mpc}^{-1}$; right panel: $k_{\rm max} = 0.4\,h\,{\rm Mpc}^{-1}$), while the ones for the w_0CDM case are shown in Fig. 7 (left panel: $k_{\rm max} = 0.2\,h\,{\rm Mpc}^{-1}$; right panel: $k_{\rm max} = 0.4\,h\,{\rm Mpc}^{-1}$). The numerical results for mean posterior cosmologies and the associated 68 per cent credible contours can be found in Table B2.

Looking at the figures showing the cases for $k_{\rm max} = 0.2\,h\,{\rm Mpc}^{-1}$, we can easily see that for both cosmological models, all codes largely agree on the best-fitting cosmology if only $k \leq k_{\rm max} = 0.2\,h\,{\rm Mpc}^{-1}$ are considered. More concretely, for these cases the fiducial cosmology always lies within the 1σ credible region. The uncertainties are broadly comparable, as only for h (or equivalently H_0) is the sensitivity of EUCLIDEMULATOR1 considerably better than that of the other two codes. This agreement is not surprising at all, remembering that the three considered power spectrum prediction codes agree almost perfectly on linear scales (Euclid Collaboration 2019, 2021).

However, increasing the cut-off scale $k_{\rm max}$ to $0.4\,h\,{\rm Mpc}^{-1}$ paints a different picture (see the right panels of Figs 6 and 7): As on these scales the power spectra predicted by the codes start to deviate more significantly from one another, in this case clear biases between the best-fitting cosmologies computed with the three different predictors are seen.[21] The values vary between 1σ and 6σ depending on the parameter, but irrespective of the cosmological model. The averaged biases are about 0.5σ if a cut-off scale of $k_{\rm max} = 0.2\,h\,{\rm Mpc}^{-1}$ is employed and it is 1–2σ if $k_{\rm max} = 0.4\,h\,{\rm Mpc}^{-1}$. This aspect is visualized in a more condensed way in Fig. 8. In this figure, the distances between the mean cosmology $\mu_{\rm mod}$ of each model to the fiducial cosmology are given in units of the 1D marginalized standard deviations $\sigma_{\rm mod}$ for each cosmological model, cut-off scale, and cosmological parameter. The RMS distance (averaged over all cosmological parameters) is again indicated by the solid horizontal lines. It is most evident that increasing the cut-off scale from 0.2 to $0.4\,h\,{\rm Mpc}^{-1}$ has a major effect on this distance, both on average as well as for some of the individual cosmological parameters. This result is independent of the power spectrum predictor or cosmological model.

We trace back the significant biases observed in the right panel plots in Figs 6 and 7 to the differences in the power spectrum predictions between the different predictor codes. To do so we first look at the quantities

$$Q_{\rm mod}(k;p) \equiv \frac{P_{\rm mod}(k;p)}{P_{\rm EE}(k;p_{\rm fid})}\,, \tag{10}$$

and

$$\Delta Q_{\rm mod}(k;p) \equiv Q_{\rm mod}(k;p) - Q_{\rm mod}(k;p_{\rm fid})\,, \tag{11}$$

shown in Fig. 9. Here mod $\in \{{\rm HF}, {\rm HM}\}$, the set of investigated cosmological parameters is $p \in \{\omega_{\rm b}, \omega_{\rm m}, n_{\rm s}, h, w_0, A_{\rm s}\}$, $p_{\rm fid}$

[21] Notice that similar findings have been reported in Martinelli et al. (2021) in the context of cosmic shear analyses.

Downloaded from https://academic.oup.com/mnras/article/518/2/1859/6619061 by Dept of Energy/OSTI user on 26 August 2024

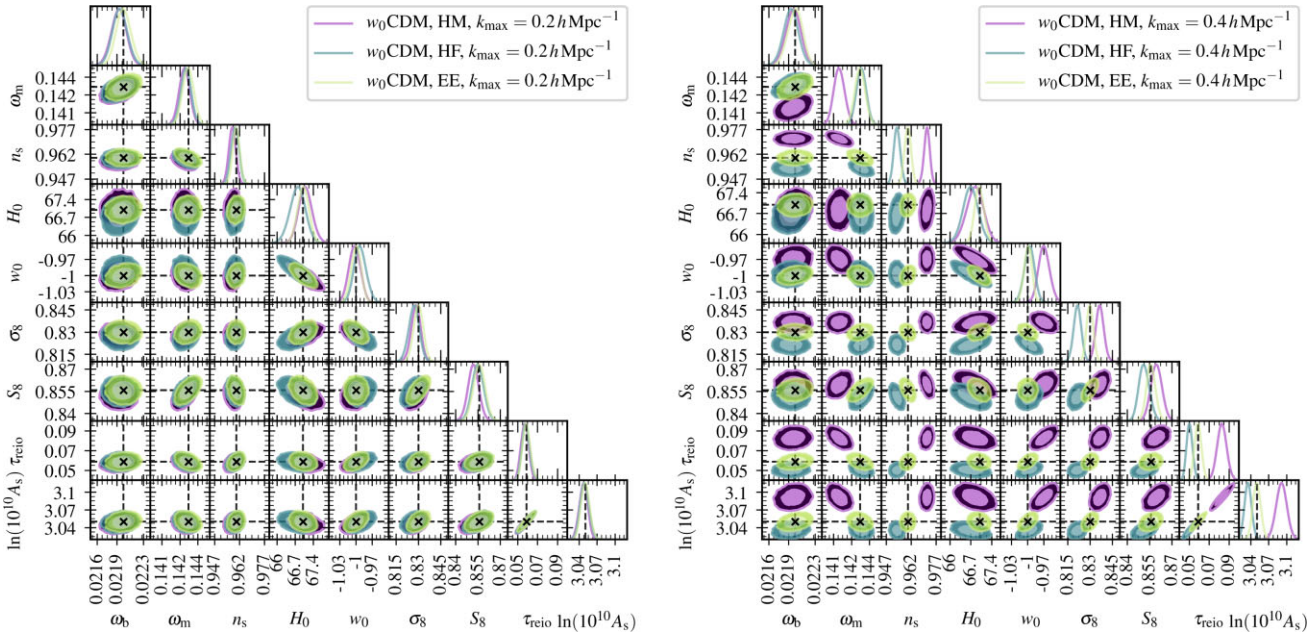

Figure 7. w_0CDM cross comparison of HMCODE and HALOFIT versus EUCLIDEMULATOR1 considering $k \leq 0.2\,h\,\mathrm{Mpc}^{-1}$ (left triangle plot) and $k \leq 0.4\,h\,\mathrm{Mpc}^{-1}$ (right triangle plot), respectively. Just as in Fig. 6 we see good agreement between the posteriors in the case with the lower cut-off scale, while increasing that scale leads to emergent differences.

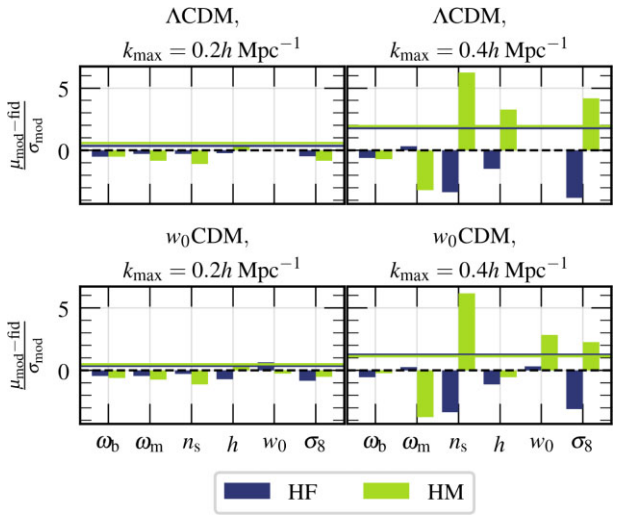

Figure 8. Differences between mean (μ) and fiducial cosmology per cosmological parameter for HALOFIT and HMCODE in the cross-comparison set-up and for both cut-off scales. The differences are given in units of the respective 1D marginalized posterior standard deviations (σ). The solid horizontal lines indicate the RMS distances averaged over all parameters in each panel. This plot shows nicely how different implementations of (even mildly) non-linear physics manifest themselves as larger biases in parameter estimation analyses. Moreover, although the RMS distance for HMCODE and HALOFIT happen to be almost identical, the biases in each cosmological parameter can vary drastically (as e.g. in the case of n_s).

corresponds to the respective parameter set to its fiducial value as given in Table 1. The redshift is kept fixed at $z = 1$. The quantity Q differs from quantity R defined in equation (1) by keeping the denominator fixed to the fiducial cosmology, while it is also varied in the quantity R.

Since the biases grow significantly as we increase the cut-off scale from $k = 0.2\,h\,\mathrm{Mpc}^{-1}$ to $k = 0.4\,h\,\mathrm{Mpc}^{-1}$, we shall focus primarily on this region in the following discussion (marked by the vertical indigo bands in Fig. 9). The argument is most eye-catchingly illustrated considering the example of n_s (third row on Fig. 9): while in general (i.e. over the entire k range taken into account for the MCMCs) the quantities Q are very similar for HALOFIT and for HMCODE, they show a difference in behaviour between $k = 0.2\,h\,\mathrm{Mpc}^{-1}$ and $k = 0.4\,h\,\mathrm{Mpc}^{-1}$. The curves turn upward for HALOFIT but downward for HMCODE. In other words, at the small-scale end of the considered k range, HALOFIT tends to overestimate the power spectrum, while HMCODE underestimates it. The overestimation for HALOFIT is however alleviated if the n_s value is decreased compared to the fiducial value. Similarly, the underestimation in the case of HMCODE is reduced if n_s is somewhat increased. Hence, the MCMC favours lower n_s values for HALOFIT and higher ones for HMCODE when fitting the fiducial cosmology. This aligns perfectly with what we observe in Figs 6 and 7. The same argument can be applied to all other cosmological parameters, albeit with complex parameter degeneracies (made more complex by the inclusion of mock Planck data) resulting in shifts for most parameters.

E.g. changing ω_m (second row on Fig. 9) helps adjust for the differences on BAO scales for HMCODE (note the posterior for ω_m remains largely unchanged for HALOFIT compared to EUCLIDEMULATOR1), while also modulating the small-scale amplitude shift from changing n_s with a similar but somewhat different scale dependence: where increasing n_s increases power on small scales and decreases power on large scales, instead decreasing ω_m decreases power across the observable range, except for the very largest scales where the impact is negligible, with a larger effect on smaller scales.

Therefore, the combination of increased n_s and decreased ω_m that we see for HMCODE results in a different impact on the matter power spectrum than changing n_s alone, giving a flatter increase in power from large to smaller scales.

Also, we see only a very small shift in the mean of each posterior for H_0, but instead see inflated posteriors. We can understand this

Downloaded from https://academic.oup.com/mnras/article/518/2/1859/6619061 by Dept of Energy/OSTI user on 26 August 2024

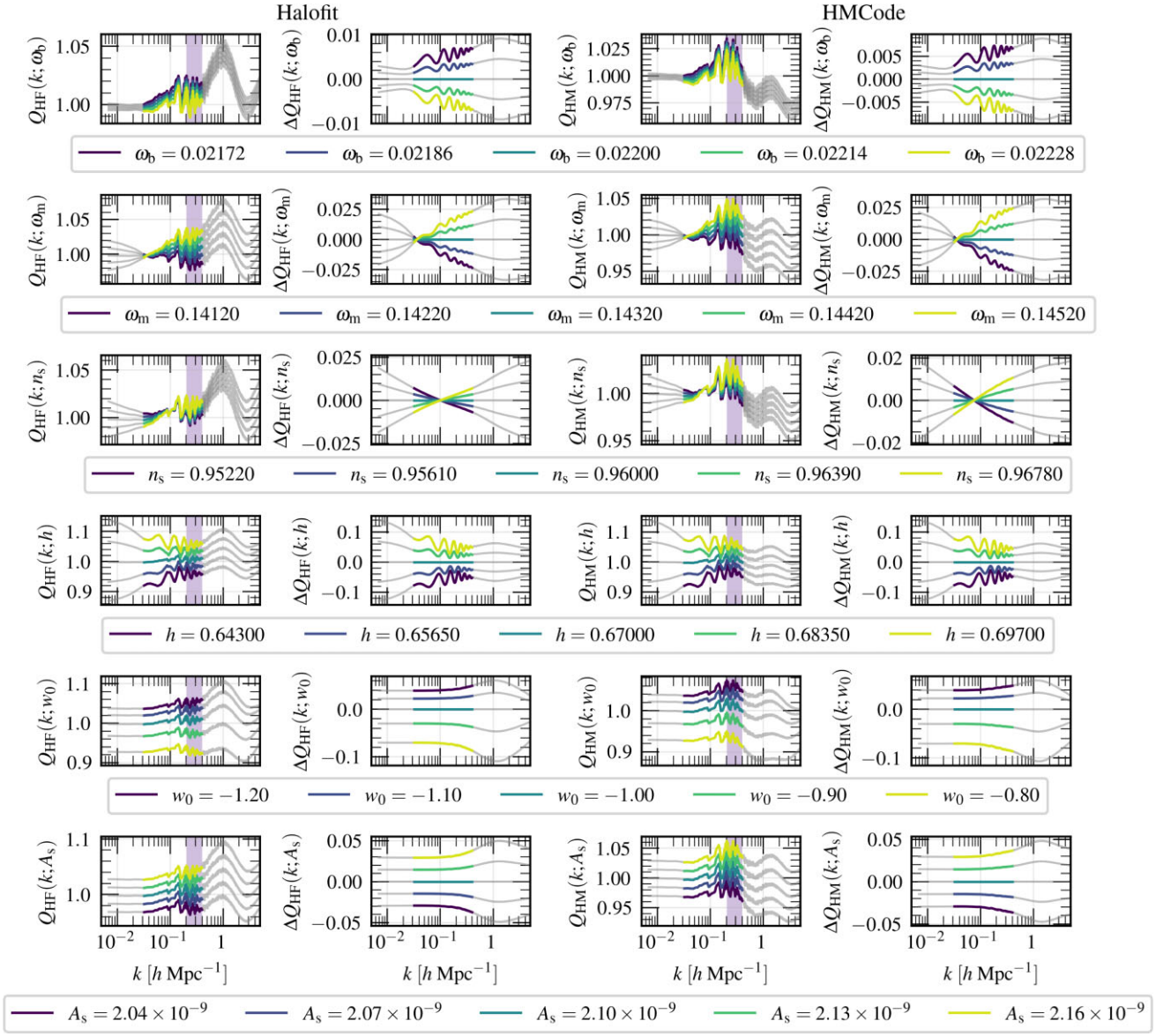

Figure 9. This figure shows the change of the power spectrum predicted by either HALOFIT or HMCODE as individual cosmological parameters are varied while all other parameters are fixed to the fiducial cosmology. The comparison is shown in terms of the quantity Q_{model}, being defined in the main text as the ratio of a power spectrum predicted by model $\in \{\mathrm{HALOFIT}, \mathrm{HMCODE}\}$ with respect to the power spectrum produced by EUCLIDEMULATOR1 at the fiducial cosmology. The curves are shown plotted in the full range $k \in [0.005, 5]$; however, only the colored parts are actually taken into account in the MCMCs and hence only those k modes have an effect on the biases observed in the MCMC posteriors in this section. The shaded indigo bands mark the region between $k = 0.2\,h\,\mathrm{Mpc}^{-1}$ and $k = 0.4\,h\,\mathrm{Mpc}^{-1}$, which is responsible for the larger biases observed in the analysis with the larger cut-off scale. Notice the differences and similarities between this figure and Fig. 2.

as the mock Planck data not allowing for a considerably different mean value from the fiducial value, but that the sampler attempts to adjust for differences on BAO scales by varying H_0, resulting in a decrease in sensitivity to H_0 compared to the EUCLIDEMULATOR1 case, where the non-linear prescription matches the fiducial model, due to washing out information from the placement of the BAO peaks.

Finally, we note that ω_b is largely unaffected by biases, as this parameter is very well determined by the mock Planck likelihood and cannot easily be shifted.

Answer to question #3: While the choice of the power spectrum predictor only has a relatively minor impact on the best-fitting cosmology and its uncertainties when mostly linear scales are considered, the impact grows as increasingly larger k modes are taken into account for the fit. We found this result independently of the fact whether a simple ΛCDM or a more complex w_0CDM model was considered.

5.2 Effect of theoretical uncertainty models on the posterior distribution

In the previous subsection, we have shown that the choice of the power spectrum predictor can, in some cases, have a quite significant effect on the resulting best-fitting cosmology. These differences clearly cannot be physical but must be due to modelling and implementation differences. In order to account for this, as was discussed above, one has to include yet another source of uncertainty into the modelling: a theoretical uncertainty (see Section 2.4). However,

Downloaded from https://academic.oup.com/mnras/article/518/2/1859/6619061 by Dept of Energy/OSTI user on 26 August 2024

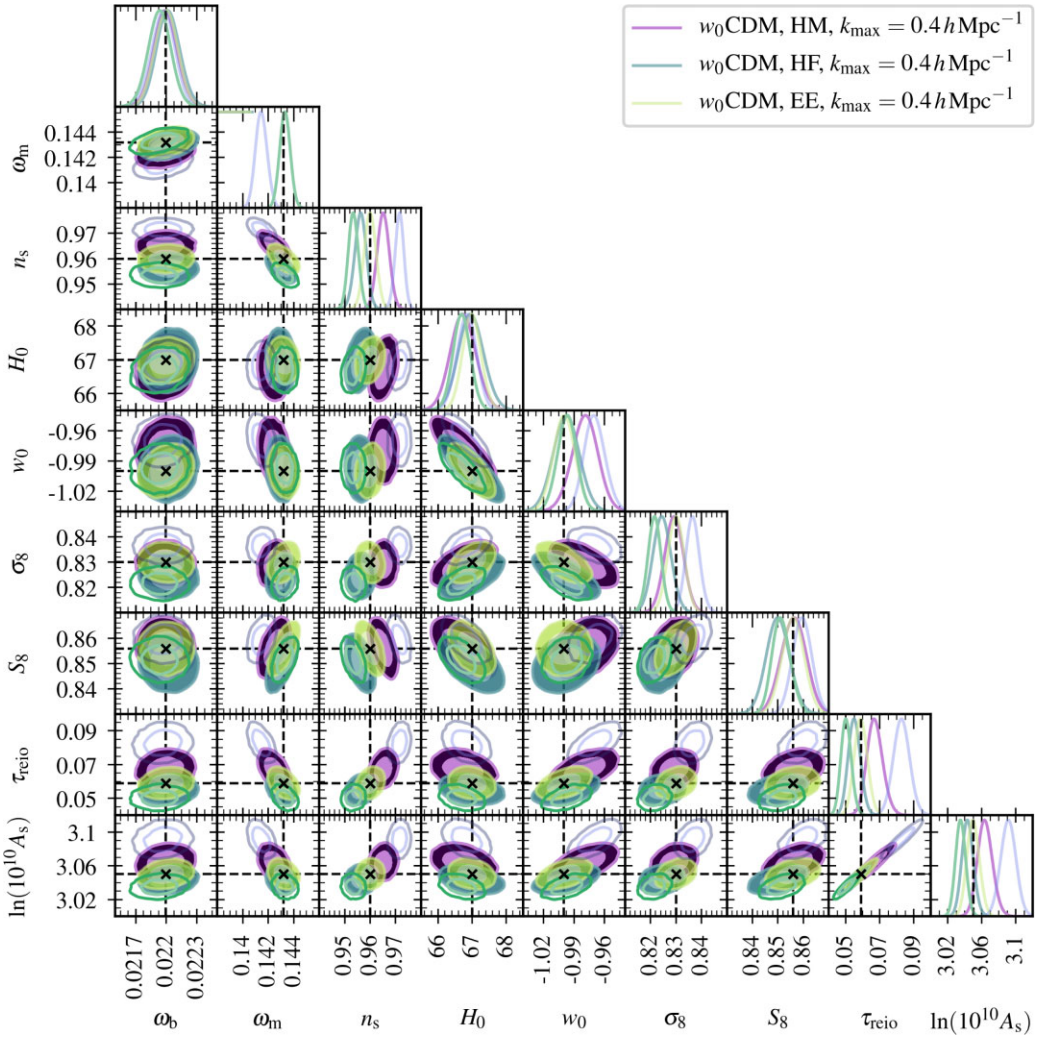

Figure 10. w_0CDM cross-comparison of HMCODE and HALOFIT versus EUCLIDEMULATOR1 ($k_{\max} = 0.4\,h\,\mathrm{Mpc}^{-1}$) using the new uncertainty envelope $\sigma_{\mathrm{th,\,KB}}$ (filled contours) compared to the theoretical uncertainty envelope $\sigma_{\mathrm{th,\,S}}$ as given in Sprenger et al. (2019). Notice that the empty contours correspond to the contours shown in Fig. 7. This plot nicely demonstrates how a carefully modelled theoretical uncertainty is capable of making the best-fitting cosmology less model-dependent while keeping its credible contours reasonably tight. On the other hand, if a insufficient theoretical uncertainty is used, the resulting best-fitting cosmology may end being significantly biased.

a priori such an uncertainty is not known and hence it has to be modelled. For this reason it is a valid question to ask what the impact of the choice for a theoretical uncertainty model is. To answer this question we ran two sets of MCMCs for EUCLIDEMULATOR1, HALOFIT, and HMCODE (restricting ourselves to the w_0CDM model and the case where $k_{\max} = 0.4\,h\,\mathrm{Mpc}^{-1}$): one set with the theoretical uncertainty envelope as given in Sprenger et al. (2019) referred to as '$\sigma_{\mathrm{th,\,S}}$' and one set with the envelope described in Section 2.4 referred to as '$\sigma_{\mathrm{th,\,KB}}$'. More concretely, we first compute a fiducial galaxy power spectrum P_{gg} based on EUCLIDEMULATOR1. In a second step, we run MCMCs for all three predictors, always comparing to the same fiducial. This means that in the case of EUCLIDEMULATOR1 we perform an autocomparison test, while for HALOFIT and HMCODE cross-comparisons are performed.

The different choices of the theoretical uncertainty manifest themselves in different distances between the best-fitting cosmologies and in different contours. Since $\sigma_{\mathrm{th,\,KB}}$ was modelled with specific information about the codes being used in this analysis while $\sigma_{\mathrm{th,\,S}}$ is a rather generic envelope, we expect the posterior distributions featuring $\sigma_{\mathrm{th,\,KB}}$ to agree significantly better with each other than the ones computed with $\sigma_{\mathrm{th,\,S}}$. Specifically, the more aggressive theoretical uncertainty $\sigma_{\mathrm{th,\,KB}}$ down-weights the non-linear portion and BAO scales where the codes disagree relative to the linear portion of the galaxy power spectrum where the codes agree very well, so the bias is reduced. While this test mostly serves as a sanity check, it helps us understand the often neglected impact of modelling choices on the final cosmological parameter estimation.

At this point we shall reiterate: It is always possible to implement so much theoretical uncertainty into an MCMC such that any two codes would agree with each other. Hence, while there is no real challenge in making the best-fitting cosmologies of MCMCs based on EUCLIDEMULATOR1, HALOFIT and HMCODE agree with each other, the difficulty lies in doing so without significantly inflating the credibility contours.

Our test results are shown in Fig. 10 and more compactly in Fig. 11 (for numerical results see Table B3). The filled contours in Fig. 10 represent the MCMCs based on $\sigma_{\mathrm{th,\,KB}}$ and the empty contours those based on $\sigma_{\mathrm{th,\,S}}$. The empty contours correspond to the contours shown

Downloaded from https://academic.oup.com/mnras/article/518/2/1859/6619061 by Dept of Energy/OSTI user on 26 August 2024

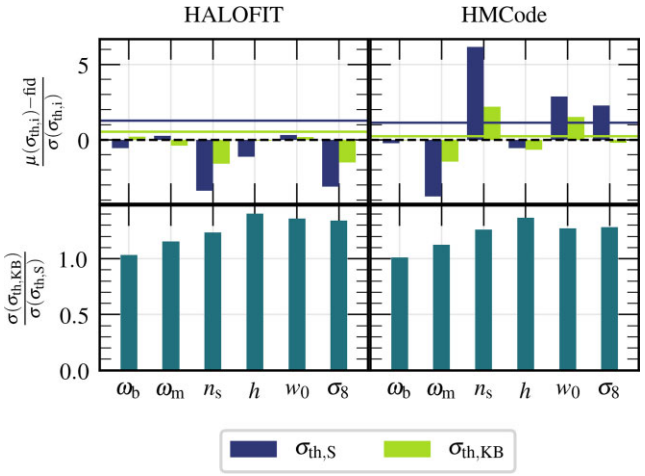

Figure 11. *Upper row:* Differences between mean (μ) and fiducial (fid) cosmology per cosmological parameter for HALOFIT and HMCODE in the cross-comparison set-up for both theoretical uncertainty envelopes $\sigma_{\mathrm{th,\,i}}$ tested in this paper: $i \in \{\mathrm{S, KB}\}$. Here, only results for w_0CDM cosmologies with a cut-off scale of $k_{\mathrm{max}} = 0.4\,h\,\mathrm{Mpc}^{-1}$ are shown. The differences are again given in units of the respective 1D marginalized posterior standard deviations (σ). The blue bars in this plot correspond to the bars (both green and blue) in the lower right panel of Fig. 8. Notice how $\sigma_{\mathrm{th,\,KB}}$ is able to reduce the RMS distance w.r.t. $\sigma_{\mathrm{th,\,S}}$ by a factor of $\sim$2 in the case of HALOFIT and by a factor of $\sim$5 in the case of HMCODE. *Bottom row:* Ratios between the σ's measured with the two theoretical uncertainty envelopes. The fact that these ratios deviate only moderately from unity emphasizes the fact that the $\sigma_{\mathrm{th,\,KB}}$ is modelled carefully such that the agreement between the mean and the fiducial cosmology could be improved without drastically blowing up the uncertainty contours. The most extreme case is that for the parameter h and HALOFIT, where the theoretical uncertainty envelope causes σ to increase by a factor of merely 1.4 compared to the contour computed with $\sigma_{\mathrm{th,\,S}}$. In all other cases σ is increased by an even smaller factor.

in Fig. 7. It is evident that the filled contours agree significantly better with each other than the empty ones while their areas are not much larger. The upper panels of Fig. 11 show again the RMS distances (biases) between the mean cosmologies in units of the 68 per cent 1D marginalized posterior standard deviations σ (similar to Fig. 8). Clearly, the biases are smaller over all parameters when $\sigma_{\mathrm{th,\,KB}}$ is used compared to employing $\sigma_{\mathrm{th,\,S}}$. However, in order to paint the complete picture we need to take into account that not only the biases but also the uncertainties (i.e. the σ's themselves) change when different uncertainty envelopes are used. The ratio between the standard deviations σ depending on the uncertainty envelope $\sigma_{\mathrm{th,\,i}}$, $i \in \{\mathrm{S, KB}\}$ is shown in the lower panel. As expected, this ratio is always greater or equal to unity implying that the uncertainties are somewhat increased when $\sigma_{\mathrm{th,\,KB}}$ is used compared to using $\sigma_{\mathrm{th,\,S}}$. Note, however, how the maximal ratio is only $\sim$1.4. This shows that a carefully modelled uncertainty envelope is capable of significantly reducing biases while not inflating the uncertainty contours by a lot.

Answer to question #4: The choice of the theoretical uncertainty envelope σ_{th} can have a significant impact on the resulting estimate of the best-fitting cosmological parameter values. A theoretical uncertainty which is not modelled with enough care may lead to a considerable bias in the best-fitting cosmology. On the other hand, a carefully modelled theoretical uncertainty is able to account for model specifics and make the resulting posterior less dependent on the chosen power spectrum predictor while keeping the posterior credibility contours reasonably tight. We hence strongly suggest that in future cosmological parameter estimation studies more focus shall be put on the aspect of theoretical uncertainties in order to make the results more model-independent.

6 CONCLUSION

In this paper, we have studied cosmological parameter inference results based on galaxy clustering observables in both ΛCDM and w_0CDM cosmologies focusing on the following four major questions:

(i) What are the performance differences of EUCLIDEMULATOR1, HALOFIT, and HMCODE?
(ii) What is the added value of considering (mildly) non-linear scales in parameter forecasts?
(iii) What is the impact of the choice of the (non-linear) predictor model on the parameter estimation result?
(iv) How are the forecasting results affected by different choices of theoretical uncertainty models?

To answer these questions, we have run multiple MCMCs with the code MONTEPYTHON. The three different power spectrum predictors used for this work are EUCLIDEMULATOR1, HALOFIT, and HMCODE. Notice that EUCLIDEMULATOR1 allows to predict the NLC only. We take advantage of this fact by modifying the Kaiser formula as described in Appendix A.

We start out our analysis (see Section 4.1) by a simple comparison of the MCMC posteriors based on the three different power spectrum predictors. To do so, we let MONTEPYTHON estimate the parameters of a fiducial model, where both the fits and the fiducial are computed with the same predictor (referred to as 'autocomparison' tests). It is therefore expected that the best-fitting cosmology recovers the fiducial cosmology almost perfectly in all cases. However, the posterior contours of EUCLIDEMULATOR1 tend to be tighter than the contours of HALOFIT and HMCODE. The magnitude of this effect varies significantly depending on the cosmological model, the non-linear prescription, and the k-mode cut-off employed. The minimal value is 2.2 per cent in the case of HMCODE versus EUCLIDEMULATOR1 for ΛCDM with $k_{\mathrm{max}} = 0.2\,h\,\mathrm{Mpc}^{-1}$, while the maximum value is 27.9 per cent in the case of HALOFIT versus EUCLIDEMULATOR1 for w_0CDM with $k_{\mathrm{max}} = 0.4\,h\,\mathrm{Mpc}^{-1}$. These results are shown in Fig. 4. This observation together with the knowledge that the physics of non-linear clustering is more accurately modelled in EUCLIDEMULATOR1 than in the halo model-based models points at the fact that parameter forecasts based on EUCLIDEMULATOR1 are subject to smaller uncertainties than those based on the other two predictors.

We analyse the effect of including (mildly) non-linear scales on the posterior probability distribution of the MCMCs in Section 4.2. Concretely, we run two sets of MCMCs, one considering only modes $k \leq k_{\mathrm{max}} = 0.2\,h\,\mathrm{Mpc}^{-1}$ (mostly linear scales) and one including also modes $k \leq k_{\mathrm{max}} = 0.4\,h\,\mathrm{Mpc}^{-1}$ (linear and mildly non-linear scales). We find that doing so reduces the 1D marginalized posterior standard deviations per parameter for all three different predictors (see Fig. 5). This finding confirms that there is valuable information at smaller spatial scales that can be leveraged in order to learn more about the cosmological parameters (clearly this is the motivation for many modern cosmological surveys such as *Euclid* to investigate data from those smaller scales). The magnitude of the effect varies between $\sim$ 9 per cent and $\sim$ 25 per cent (on average over all tested cosmological parameters) depending on the predictor and the cosmological model under consideration. Notice that we did not include scales above $0.4\,h\,\mathrm{Mpc}^{-1}$ as we used only a simple linear

Downloaded from https://academic.oup.com/mnras/article/518/2/1859/6619061 by Dept of Energy/OSTI user on 26 August 2024

galaxy bias model and we neglected baryonic effects in our analysis. The validity of such a simple model is known to break at even more non-linear scales. We thus consider the results presented in this paper as a proof of concept and leave a more sophisticated analysis of the impact of strongly non-linear scales on the posterior uncertainties to future work. Nevertheless, from these observations we can confirm that extending the range of considered k modes beyond linear scales is certainly beneficial to a relevant degree for future cosmological parameter forecasts. In addition we find that, when more non-linear scales are considered in such an analysis, then the results depend on the prescription used to model the physics at these scales, unless a proper theoretical uncertainty is used.

In Section 5, we go one step further and take on the view point of a mock data analysis than merely a parameter forecast. In a data analysis, we need to consider the possibility that the best-fitting cosmology is biased compared to the true cosmology underlying the data. We approach the question about the bias caused by the non-linear model, which is employed in the parameter estimation analysis, by a set of MCMCs referred to as 'cross-comparison tests' (see Section 5.1). For these tests, we always compute the fiducial power spectrum with EUCLIDEMULATOR1. This fiducial power spectrum then serves as mock fiducial data which we try to fit with HALOFIT and HMCODE. Doing so we make the following interesting finding (Fig. 8): as long as only linear scales are considered, the biases between the best-fitting cosmologies from all different codes are rather small (within the 1σ credible region). However, as soon as smaller scales are included in the MCMCs the biases grow significantly, up to 6σ. The take-away message from this finding is that the best-fitting cosmology found in a parameter estimation may severely depend on the choice of the power spectrum predictor and hence the best-fitting cosmology depends on the non-linear model, especially when more and more non-linear scales are taken into account. A similar finding was presented by Martinelli et al. (2021) for the complementary case of cosmic shear, cautioning that care has to be taken when it comes to non-linear predictors and modelling in future analyses of high-resolution large-scale structure data.

The interpretation from the previous paragraph calls for a more model-independent approach in parameter inference initiatives. We propose that this can be achieved by including not just an observational but also a carefully modelled theoretical uncertainty into the parameter estimation (the use of theoretical uncertainties in the field of parameter sensitivity forecasts is already known state of the art). We test two different theoretical uncertainty envelopes in Section 5.2: a rather generic one and a carefully modelled one which is actually informed by the agreement of the three power spectrum predictors compared with each other. We find that the latter envelope leads to much smaller biases in the best-fitting cosmologies while only marginally inflating the credible contours. Using our new theoretical uncertainty $\sigma_{\mathrm{th,\,KB}}$, the RMS distance between the mean and the fiducial cosmology can be reduced by a factor of $\sim$2 for HALOFIT and even by a factor of $\sim$5 for HMCODE. This result is visualized in Figs 10 and 11. We thus advocate for the consideration of carefully modelled theoretical uncertainty envelopes in future data analyses as we demonstrate that they render the results much less model-dependent.

We conclude by stating that we have found the tightest posterior with the smallest bias between the best-fitting and the fiducial cosmology by taking into account not just linear but also mildly non-linear scales and by including a well-informed theoretical uncertainty envelope into our model.

ACKNOWLEDGEMENTS

MK acknowledges support from the Swiss National Science Foundation (SNF) grant 200020_149848. Simulations were performed on the zBox4+ cluster at the University of Zurich. TB was supported by the grant DOE DE-SC0017848, through the INFN project 'GRANT73/Tec-Nu', and from the COSMOS network (www.cosmosnet.it) through the ASI (Italian Space Agency) Grants 2016-24-H.0 and 2016-24-H.1-2018. Further, the authors are very grateful for Tim Sprenger's helpful discussions and inputs.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding authors.

REFERENCES

Abbott T. M. C. et al., 2022, Phys. Rev. D, 105, 023520
Abdurro'uf K. et al., 2022, ApJS, 259, 39pp
Akeson R. et al., 2019, preprint (arXiv:1902.05569)
Angulo R. E., Zennaro M., Contreras S., Aricò G., Pellejero-Ibañez M., Stücker J., 2021, MNRAS, 507, 5869
Aricò G., Angulo R. E., Hernández-Monteagudo C., Contreras S., Zennaro M., Pellejero-Ibañez M., Rosas-Guevara Y., 2020, MNRAS, 495, 4800
Aricò G., Angulo R. E., Contreras S., Ondaro-Mallea Lurdes P.-I. M., Zennaro M., 2021, MNRAS, 506, 4070
Audren B., Lesgourgues J., Bird S., Haehnelt M., Viel M., 2013a, J. Cosmol. Astropart. Phys., 026, 23
Audren B., Lesgourgues J., Benabed K., Prunet S., 2013b, J. Cosmol. Astropart. Phys., 2013, 001
Baldauf T., Mirbabayi M., Simonović M., Zaldarriaga M., 2016, preprint (arXiv:1602.00674)
Bhandari N., Leonard C. D., Rau M. M., Mandelbaum R., 2021, preprint (arXiv:2101.00298)
Bird S., Viel M., Haehnelt M. G., 2012, MNRAS, 420, 2551
Bird S., Rogers K. K., Peiris H. V., Verde L., Font-Ribera A., Pontzen A., 2019, J. Cosmol. Astropart. Phys., 2019, 050
Blas D., Lesgourgues J., Tram T., 2011, J. Cosmol. Astropart. Phys., 07, 34
Brinckmann T., Lesgourgues J., 2019, Phys. Dark Univ., 24, 100260
Brinckmann T., Hooper D. C., Archidiacono M., Lesgourgues J., Sprenger T., 2019, J. Cosmol. Astropart. Phys., 01, 059
Bull P., Ferreira P. G., Patel P., Santos M. G., 2015, ApJ, 803, 1
Cataneo M., Lombriser L., Heymans C., Mead A. J., Barreira A., Bose S., Li B., 2019, MNRAS, 488, 2121
Chisari N. E. et al., 2019, Open J. Astrophys., 2, 4
Christensen N., Meyer R., Knox L., Luey B., 2001, Class. Quantum Gravity, 18, 2677
Chudaykin A., Ivanov M. M., 2019, J. Cosmol. Astropart. Phys., 11, 34
Chudaykin A., Ivanov M. M., Philcox O. H. E., Simonović M., 2020, Phys. Rev. D, 102, 63533
Colas T., d'Amico G., Senatore L., Zhang P., Beutler F., 2020, J. Cosmol. Astropart. Phys., 06, 1
D'Amico G., Gleyzes J., Kokron N., Markovic K., Senatore L., Zhang P., Beutler F., Gil-Marín H., 2020, J. Cosmol. Astropart. Phys., 05, 5
D'Amico G., Senatore L., Zhang P., 2021, J. Cosmol. Astropart. Phys., 2021, 6
de Jong J. T. A., Kleijn G. A. V., Kuijken K. H., Valentijn E. A., 2013, Exp. Astron., 35, 25
DeRose J. et al., 2019, ApJ, 875, 69
DESI Collaboration, 2016, preprint (arXiv:1611.00036)
Desjacques V., Jeong D., Schmidt F., 2018, Phys. Rep., 733, 1
Euclid Collaboration, 2019, MNRAS, 484, 5509
Euclid Collaboration, 2020, A&A, 642, A191
Euclid Collaboration, 2021, MNRAS, 505, 2840
Fluri J., Kacprzak T., Lucchi A., Refregier A., Amara A., Hofman T., Schneider A., 2019, Phys. Rev. D, 100, 22

Downloaded from https://academic.oup.com/mnras/article/518/2/1859/6619061 by Dept of Energy/OSTI user on 26 August 2024

Garrison L. H., Eisenstein D. J., Pinto P. A., 2019, MNRAS, 485, 3370
Gelman A., Rubin D. B., 1992, Stat. Sci., 7, 457
Giri S. K., Schneider A., 2021, J. Cosmol. Astropart. Phys., 2021, 046
Giusarma E., Vagnozzi S., Ho S., Ferraro S., Freese K., Kamen-Rubio R., Luk K.-B., 2018, Phys. Rev. D, 98, 10
Gunn J. et al., 2006, AJ, 131, 2332
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., 2014, ApJ, 780, 111
Heitmann K. et al., 2016, ApJ, 820, 108
Heymans C. et al., 2021, A&A, 646, A140
Ivanov M. M., Simonović M., Zaldarriaga M., 2020a, J. Cosmol. Astropart. Phys., 05, 42
Ivanov M. M., Simonović M., Zaldarriaga M., 2020b, Phys. Rev. D, 101, 83504
Jackson J. C., 1972, MNRAS, 156, 1P
Kaiser N., 1987, MNRAS, 227, 1
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Lawrence E., Heitmann K., White M., Higdon D., Wagner C., Habib S., Williams B., 2010, ApJ, 713, 1322
Lawrence E. et al., 2017, ApJ, 847, 12
Lesgourgues J., Tram T., 2011, J. Cosmol. Astropart. Phys., 9, 32
Lewis A., Bridle S., 2002, Phys. Rev. D, 66, 1
LSST Science Collaboration, 2009, preprint (arXiv:0912.0201)
Maartens R., Abdalla F., Jarvis M., Santos M. G., 2015, Proc. Sci., Overview of Cosmology with the SKA. SISSA, Trieste, PoS#16
Markovič K., Bose B., Portsidou A., 2019, Open J. Astrophys., 2, 13
Martinelli M. et al., 2021, A&A, 649, A100
McClintock T. et al., 2019a, preprint (arXiv:1907.13167)
McClintock T. et al., 2019b, ApJ, 872, 53
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
Mead A. J., Brieden S., Tröster T., 2021, MNRAS, 502, 1401
Nishimichi T. et al., 2019, ApJ, 884, 29
Osato K., Nishimichi T., Bernardeau F., Taruya A., 2019, Phys. Rev. D, 99, 063530
Perotto L., Lesgourgues J., Hannestad S., Tu H., Wong Y., 2006, J. Cosmol. Astropart. Phys., 2006, 013
Pezzotta A., Crocce M., Eggemeier A., Sánchez A. G., Scoccimarro R., 2021, Phys. Rev. D, 104, 043531
Planck Collaboration, 2006, preprint (arXiv:astro-ph/0604069)
Planck Collaboration, 2018, Planck 2018 Results: Cosmological Parameter Tables. https://wiki.cosmos.esa.int/planck-legacy-archive/images/b/be/Baseline_params_table_2018_68pc.pdf
Planck Collaboration VI, 2020, A&A, 641, 67
Rogers K. K., Peiris H. V., Pontzen A., Bird S., Verde L., Font-Ribera A., 2019, J. Cosmol. Astropart. Phys., 02, 1
Schneider A. et al., 2016, J. Cosmol. Astropart. Phys., 04, 047
Schneider A., Stoira N., Refregier A., Weiss A. J., Knabenhans M., Stadel J., Teyssier R., 2020, J. Cosmol. Astropart. Phys., 2020, 19
Sellentin E., Quartin M., Amendola L., 2014, MNRAS, 441, 1831
Smith R. E. et al., 2003, MNRAS, 341, 1311
Sprenger T., Archidiacono M., Brinckmann T., Clesse S., Lesgourgues J., 2019, J. Cosmol. Astropart. Phys., 1902, 47
Spurio Mancini A., Piras D., Alsing J., Joachimi B., Hobson M. P., 2022, MNRAS, 511, 1771
Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, ApJ, 761, 152
Taruya A., Nishimichi T., Saito S., 2010, Phys. Rev. D, 82, 063522
Tegmark M., Taylor A., Heavens A., 1997, ApJ, 480, 22
The Dark Energy Survey Collaboration, 2005, preprint (arXiv:astro-ph/0510346)
Tully R. B., Fisher J. R., 1978, in Longair M. S., Einasto J., eds, Proc. IAU Symp. 79, The Large Scale Structure of the Universe. Kluwer, Dordrecht, p. 31
Valcin D., Villaescusa-Navarro F., Verde L., Raccanelli A., 2019, J. Cosmol. Astropart. Phys., 2019, 057
Van Daalen M. P., Schaye J., Booth C. M., Dalla Vecchia C., 2011, MNRAS, 415, 3649
Winther H. A., Casas S., Baldi M., Koyama K., Li B., Lombriser L., Zhao G.-B., 2019, Phys. Rev. D, 100, 1
Yahia-Cherif S. et al., 2021, A&A, 649, A52
Zhai Z. et al., 2019, ApJ, 874, 12

APPENDIX A: QUASI-NON-LINEAR GALAXY–GALAXY POWER SPECTRUM

In Section 3 of the main text we introduced how the NLC computed by EUCLIDEMULATOR1 can be used to derive the observable galaxy–galaxy power spectrum. However, we glossed over the fact that the relation we used is not fully self-consistent. This aspect is analysed in more depth in this appendix.

A1 Review of linear RSDs modelling

Equation (7) explains how the dark matter power spectrum $P_{\delta\delta}(k, z)$ is related to the observable galaxy–galaxy power spectrum $P_{\mathrm{gg}(k,\,\mu,\,z)}$. It is repeated here for convenience:

$$P_{\mathrm{gg}}(k, \mu, z) = f_{\mathrm{AP}}(z) \times f_{\mathrm{res}}(k, \mu, z) \times f_{\mathrm{RSD}}(\hat{k}, \hat{\mu}, z) \times b^2(z) \times P_{\delta\delta}(k, z)\,. \tag{A1}$$

Further, in the same section we showed how the NLC B computed by EUCLIDEMULATOR1 can be incorporated into that equation. In this section, we discuss the relation between P_{gg} and $B(k, z)$ in more detail, thereby focusing on the RSDs (f_{RSD}) and ignore the galaxy bias (b), Alcock–Paczynski (f_{AP}), and resolution (f_{res}) correction factors.

The observable galaxy–galaxy power spectrum depends on the direction of observation (see equation A1) because in addition to the redshift signal due to the Hubble expansion there is a contribution to the redshift distance caused by the peculiar velocities of the galaxies, i.e.

$$s = cz = H_0 d + v_r, \tag{A2}$$

where s denotes the redshift distance to a galaxy, c is the speed of light, z is the measured redshift of a galaxy, H_0 the Hubble parameter at present day, d is the proper distance to the galaxy, and v_r is the peculiar radial velocity. On large, linear scales the RSD effect is described by the Kaiser formula (Kaiser 1987), which relates the galaxy power spectrum P_{gg} to the dark matter power spectrum $P_{\delta\delta}$, the dark matter velocity power spectrum $P_{\theta\theta}$ (θ denoting the velocity divergence), and the dark matter mass-velocity cross-power spectrum $P_{\delta\theta}$:

$$P_{\mathrm{gg}}(k, \mu, z) = P_{\delta\delta}(k, z) + 2\mu^2 \beta P_{\delta\theta}(k, z) + \mu^4 \beta^2 P_{\theta\theta}(k, z)\,, \tag{A3}$$

where $\beta := f/b(z)$ is the ratio of the growth rate f and the galaxy bias b. Although equation (A3) still holds in the non-linear regime, in practice the computation of $P_{\mathrm{gg}}^{\mathrm{non-linear}}$ with this equation requires some means to compute $P_{\delta\delta}^{\mathrm{non-linear}}$, $P_{\delta\theta}^{\mathrm{non-linear}}$, and $P_{\theta\theta}^{\mathrm{non-linear}}$. Unfortunately, EUCLIDEMULATOR1, HALOFIT, and HMCODE do only provide the capability to compute $P_{\delta\delta}^{\mathrm{non-linear}}$. However, to linear

Downloaded from https://academic.oup.com/mnras/article/518/2/1859/6619061 by Dept of Energy/OSTI user on 26 August 2024

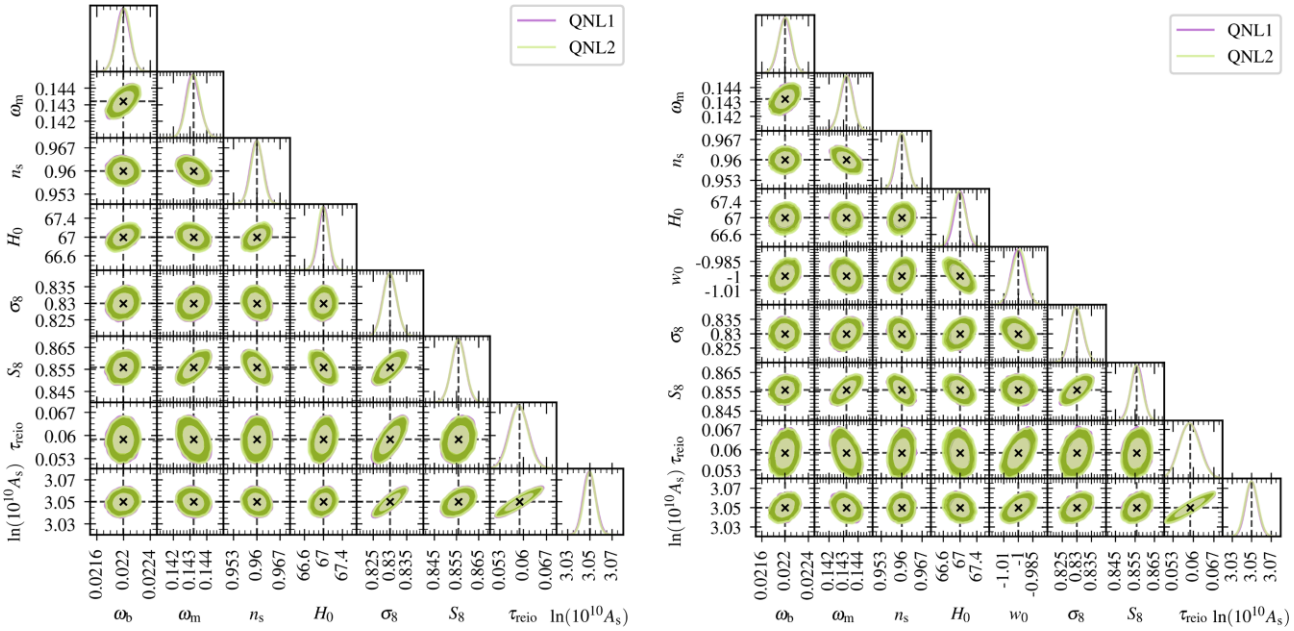

Figure A1. QNL1 versus QNL2 for ΛCDM (left panel) and w_0CDM (right panel). From these plots it becomes obvious that the two different QNL models (see equations 8 and 9) do only very marginally affect the posterior contours. The differences are so small that we neglect them throughout this entire paper.

order the divergence of the velocity field is equal to the mass density field:

$$\theta(k) \approx \theta_{\rm lin}(k) = \frac{ik\delta(x)}{aHf} = \delta(k) \tag{A4}$$

such that $P_{\theta\theta}^{\rm linear} = P_{\delta\theta}^{\rm linear} = P_{\delta\delta}^{\rm linear}$. It is thus possible to simplify equation (A3) to

$$\begin{aligned} P_{\rm gg}^{\rm linear}(k,\mu,z) = {} & P_{\delta\delta}^{\rm linear}(k,z) \\ & + 2\mu^2 f P_{\delta\delta}^{\rm linear}(k,z) \\ & + \mu^4 f^2 P_{\delta\delta}^{\rm linear}(k,z) \\ = {} & \left[1+\mu^2 f\right]^2 P_{\delta\delta}^{\rm linear}(k,z) \end{aligned} \tag{A5}$$

On smaller scales, the Fingers of God effect (Jackson 1972; Tully & Fisher 1978) adds to the RSD and is accounted for by an additional pre-factor (Bull et al. 2015) such that in total

$$P_{\rm gg}^{\rm linear}(k,\mu,z) = f_{\rm RSD} P_{\delta\delta}^{\rm linear}(k,z) \tag{A6}$$

with

$$f_{\rm RSD} = \left[1+\mu^2 f\right]^2 \exp\left(-\hat{k}^2\hat{\mu}^2\sigma_{\rm NL}^2\right). \tag{A7}$$

A2 Quasi-non-linear extension of the RSDs

In the previous subsection we summarized the well-known Kaiser formula that describes the RSDs of the linear power spectrum. Because we are interested in analyses including (mildly) non-linear scales in this paper, a generalization of that Kaiser formula to the non-linear galaxy–galaxy power spectrum is required. The obvious generalization is (compare to equation A6)

$$P_{\rm gg}^{\rm QNL1}(k,\mu,z) = f_{\rm RSD} P_{\delta\delta}^{\rm non-linear}(k,z)\,, \tag{A8}$$

where the superscript 'QNL' stands for 'quasi-non-linear'. Notice however that this generalization breaks self-consistency as the relation equation (A4) cannot be generalized to the non-linear power spectrum. Let us expand the non-linear matter power spectrum into its linear contribution and the NLC:

$$P_{\rm gg}^{\rm QNL1}(k,\mu,z) = f_{\rm RSD} P_{\delta\delta}^{\rm linear}(k,z) B(k,z)\,. \tag{A9}$$

Here, we started with equation (A3), linearized this equation using equation (A4), factored out $P_{\delta\delta}$ in equation (A5), and replaced it by its non-linear counterpart. We shall refer to this as the 'QNL1' model.

A second, self-consistent option to generalize equation (A6) is to keep the linear versions of the second and third term in equation (A3), i.e. keep $P_{\delta\theta}^{\rm linear} = P_{\theta\theta}^{\rm linear} = P_{\delta\delta}^{\rm linear}$ and only use the non-linear version of $P_{\delta\delta}$ in the first term of equation (A3). This model, referred to as 'QNL2', is thus given by

$$\begin{aligned} & P_{\rm gg}^{\rm QNL2}(k,\mu,z) \\ & = P_{\delta\delta}^{\rm non-linear}(k,z) + 2\mu^2 f P_{\delta\theta}^{\rm linear}(k,z) + \mu^4 f^2 P_{\theta\theta}^{\rm linear}(k,z) \\ & = P_{\delta\delta}^{\rm non-linear}(k,z) + 2\mu^2 f P_{\delta\delta}^{\rm linear}(k,z) + \mu^4 f^2 P_{\delta\delta}^{\rm linear}(k,z) \end{aligned} \tag{A10}$$

Notice that self-consistency is not broken at the price of ignoring the non-linear contributions in $P_{\delta\theta}$ and $P_{\theta\theta}$. If we substitute in the NLC the linear matter power spectrum can be factored out again:

$$P_{\rm gg}^{\rm QNL2}(k,\mu,z) = \left[B(k,z) + 2\mu^2 f + \mu^4 f^2\right] P_{\delta\delta}^{\rm linear}(k,z) \tag{A11}$$

Adding the Fingers of God for RSDs at small scales we arrive at

$$P_{\rm gg}^{\rm QNL2}(k,\mu,z) = f_{\rm RSD}^{\rm QNL2} P_{\delta\delta}^{\rm linear}(k,z) \tag{A12}$$

with the following expression for the RSD correction factor:

$$f_{\rm RSD}^{\rm QNL2} = \left[B(k,z) + 2\mu^2 f + \mu^4 f^2\right] \exp\left(-\hat{k}^2\hat{\mu}^2\sigma_{\rm NL}^2\right) \tag{A13}$$

Having two options to generalize the Kaiser formula, the natural question arises by how much $P_{\rm gg}^{\rm QNL1}$ and $P_{\rm gg}^{\rm QNL2}$ differ from one another. Results of a test investigating this questions are shown in Fig. A1. As can be seen from these results, for the purpose of parameter forecasting there is virtually no difference between the QNL1 and QNL2 model. For this reason we treat the two models as one for the purposes described in the remaining analyses.

Downloaded from https://academic.oup.com/mnras/article/518/2/1859/6619061 by Dept of Energy/OSTI user on 26 August 2024

More concretely, we have used model QNL2 in combination with EUCLIDEMULATOR1 because it is slightly more self-consistent than model QNL1. The latter, however, is simpler to implement when using HALOFIT or HMCODE. Whenever we worked with the latter two codes, we hence applied QNL1, including for the fiducial model in the cross-comparison results of Section 5.

APPENDIX B: MCMC RESULTS

In this appendix, we provide the numerical results of the MCMCs (see tables) as well as the contour plots of the posterior probability distributions of the cosmological parameters as obtained in auto-comparison experiments (i.e. the sensitivity forecasting-like set-up discussed in Section 4).

B1 Autocomparison results

In Section 4, we described the impact of the choice of the power spectrum predictor on forecasting results. As explained, in this setting only the various sensitivities are of interest as the mean posterior cosmologies coincide with the fiducial cosmology almost perfectly by construction. In the main text we provided the sensitivity information in the form of bar plots (see Figs 4 and 5). As this representation may be somewhat unusual, we provide the more common contour plots here in this appendix (Figs B1 and B2).

B2 Cross-comparison results

We have discussed the impact of the choice of the power spectrum predictor on the final parameter estimation results in Section 5. There we have provided the graphical summaries of the 'cross-comparison' tests. Here, however, we provide the underlying numerical results in tabular form in order to assert full transparency. Table B2 lists the mean cosmology and the 68 per cent credibility limits per cosmological parameter and cosmological model for both power spectrum predictors, HALOFIT and HMCODE, with the fiducial cosmology given by EUCLIDEMULATOR1. For all results in that table the uncertainty envelope by Sprenger et al. (2019) was used. In Table B3, on the other hand, we show the same quantities for the test with our new uncertainty envelope employed that we have shown as lime-green lines in Fig. 3. Notice that for this experiment the results shown for EUCLIDEMULATOR1 correspond to an autocomparison test, while those for HALOFIT and for HMCODE are cross-comparison tests.

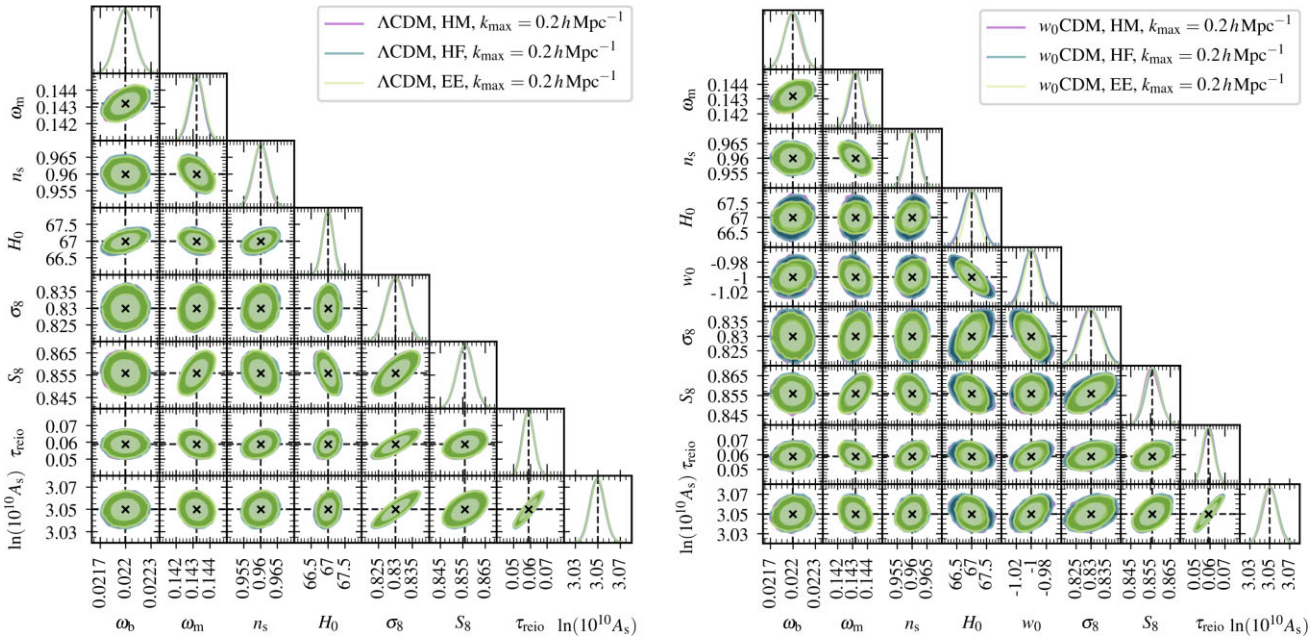

Figure B1. Autocomparisons for ΛCDM (left) and w_0CDM (right) with cut-off scale $k_{max} = 0.2\,h\,Mpc^{-1}$.

Downloaded from https://academic.oup.com/mnras/article/518/2/1859/6619061 by Dept of Energy/OSTI user on 26 August 2024

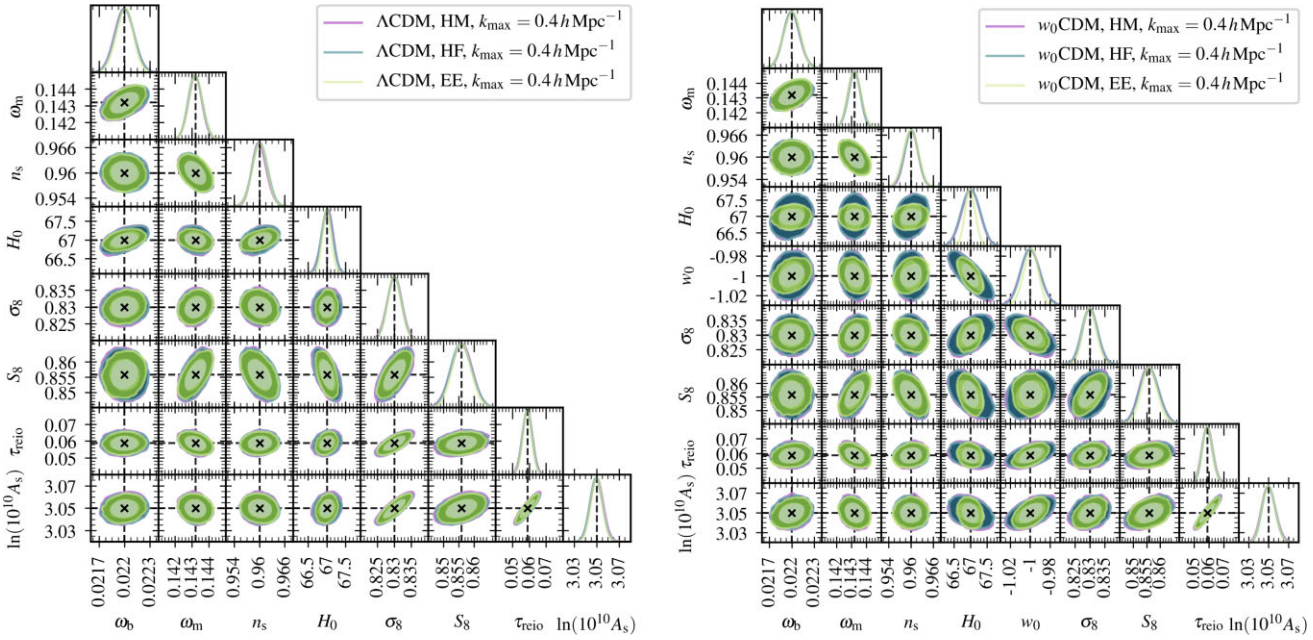

Figure B2. Autocomparisons for ΛCDM (left) and w_0CDM (right) with cut-off scale $k_{\max} = 0.4\,h\,\mathrm{Mpc}^{-1}$.

Table B1. Table of 1D marginalized posterior standard deviations (σ) for autocomparisons. Each column reports about a power spectrum predictor that was used to both create the fiducial model and to sample the posterior probability surface. In each row of the table the standard deviation of such a posterior along a specific cosmological parameter axis and for a specific cosmological model and cut-off scale is denoted. Comparing the different predictors with each other it turns out that the posterior standard deviations of EUCLIDEMULATOR1 are usually smaller than those of HALOFIT or HMCODE. The exception is ω_m (and in some cases n_s), which is itself an input parameter to EUCLIDEMULATOR1, while in HALOFIT and HMCODE it is derived from ω_b and ω_cdm. While in general the size of the posterior contours seem to be comparable for simple ΛCDM cosmologies and low cut-off scales, it can be seen that the difference in σ between EUCLIDEMULATOR1 and HALOFIT and HMCODE is larger for the case of w_0CDM cosmologies and the larger cut-off scale. The relative difference is sometimes as large as 70 per cent along a single parameter as in the case of the Hubble parameter h for w_0CDM and $k_{\max} = 0.4\,h\,\mathrm{Mpc}^{-1}$ (more clearly visible in the lower right panel of Fig. 4). From this one can conclude that EUCLIDEMULATOR1 has a substantial positive effect on the confidence of a parameter estimation compared to the other two predictors.

			EUCLIDEMULATOR (EE)	HALOFIT (HF)	HMCODE (HM)
ΛCDM	$0.2\ h\,\mathrm{Mpc}^{-1}$	$\sigma(\omega_\mathrm{b})$	1.11e-04	1.16e-04	1.15e-04
		$\sigma(\omega_\mathrm{m})$	4.49e-04	4.22e-04	4.24e-04
		$\sigma(n_\mathrm{s})$	2.16e-03	2.34e-03	2.15e-03
		$\sigma(h)$	1.49e-03	1.64e-03	1.62e-03
		$\sigma(\sigma_8)$	2.79e-03	2.83e-03	2.74e-03
	$0.4\ h\,\mathrm{Mpc}^{-1}$	$\sigma(\omega_\mathrm{b})$	1.04e-04	1.16e-04	1.16e-04
		$\sigma(\omega_\mathrm{m})$	4.18e-04	4.09e-04	4.11e-04
		$\sigma(n_\mathrm{s})$	1.83e-03	1.89e-03	1.79e-03
		$\sigma(h)$	1.20e-03	1.53e-03	1.58e-03
		$\sigma(\sigma_8)$	2.07e-03	2.18e-03	2.28e-03
w_0CDM	$0.2\ h\,\mathrm{Mpc}^{-1}$	$\sigma(\omega_\mathrm{b})$	1.12e-04	1.18e-04	1.15e-04
		$\sigma(\omega_\mathrm{m})$	4.74e-04	4.37e-04	4.48e-04
		$\sigma(n_\mathrm{s})$	2.26e-03	2.35e-03	2.15e-03
		$\sigma(h)$	2.24e-03	3.02e-03	3.02e-03
		$\sigma(w_0)$	9.36e-03	1.15e-02	1.17e-02
		$\sigma(\sigma_8)$	3.23e-03	3.68e-03	3.58e-03
	$0.4\ h\,\mathrm{Mpc}^{-1}$	$\sigma(\omega_\mathrm{b}$	1.11e-04	1.14e-04	1.17e-04
		$\sigma(\omega_\mathrm{m})$	4.37e-04	4.17e-04	4.26e-04
		$\sigma(n_\mathrm{s})$	1.95e-03	1.91e-03	1.82e-03
		$\sigma(h)$	1.62e-03	2.74e-03	2.77e-03
		$\sigma(w_0)$	7.56e-03	9.97e-03	1.02e-02
		$\sigma(\sigma_8)$	2.34e-03	2.74e-03	2.81e-03

Downloaded from https://academic.oup.com/mnras/article/518/2/1859/6619061 by Dept of Energy/OSTI user on 26 August 2024

Table B2. Mean posterior cosmologies and 68 per cent credible intervals for HALOFIT and HMCODE, resulting from cross-comparison MCMCs where EUCLIDEMULATOR1 data serve as mock data. The fiducial values for the cosmological parameters are given in the first data column. Notice that for all results presented in this table the theoretical uncertainty by Sprenger et al. (2019) is used (for results based on the new theoretical uncertainty introduced in this paper see Table B3).

			Fiducial	HALOFIT (HF)	HMCODE (HM)
ΛCDM	0.2 $h\,\mathrm{Mpc}^{-1}$	ω_b	0.0220	0.02194 ± 0.00012	0.02194 ± 0.00011
		ω_m	0.1432	0.14307 ± 0.00044	0.14284 ± 0.00044
		n_s	0.9600	0.9593 ± 0.0024	0.9577 ± 0.0022
		h	0.6700	0.6696 ± 0.0017	0.6705 ± 0.0017
		σ_8	0.8300	0.8287 ± 0.0028	0.8278 ± 0.0027
	0.4 $h\,\mathrm{Mpc}^{-1}$	ω_b	0.0220	0.02193 ± 0.00011	0.02192 ± 0.00012
		ω_m	0.1432	0.14333 ± 0.00040	0.14183 ± 0.00043
		n_s	0.9600	0.9535 ± 0.0019	0.9715 ± 0.0018
		h	0.6700	0.6676 ± 0.0016	0.6753 ± 0.0016
		σ_8	0.8300	0.8223 ± 0.0020	0.8405 ± 0.0025
w_0CDM	0.2 $h\,\mathrm{Mpc}^{-1}$	ω_b	0.0220	0.02194 ± 0.00012	0.02193 ± 0.00012
		ω_m	0.1432	0.14300 ± 0.00044	0.14287 ± 0.00045
		n_s	0.9600	0.9593 ± 0.0024	0.9576 ± 0.0021
		h	0.6700	0.6679 ± 0.0031	0.6710 ± 0.0027
		w_0	−1.0000	-0.993 ± 0.011	-1.003 ± 0.010
		σ_8	0.8300	0.8270 ± 0.0036	0.8283 ± 0.0034
	0.4 $h\,\mathrm{Mpc}^{-1}$	ω_b	0.0220	0.02193 ± 0.00011	0.02197 ± 0.00012
		ω_m	0.1432	0.14331 ± 0.00042	0.14146 ± 0.00046
		n_s	0.9600	0.9535 ± 0.0019	0.9715 ± 0.0019
		h	0.6700	0.6669 ± 0.0027	0.6684 ± 0.0029
		w_0	−1.0000	-0.9969 ± 0.0097	-0.970 ± 0.011
		σ_8	0.8300	0.8216 ± 0.0027	0.8367 ± 0.0029

Table B3. Mean posterior cosmologies and 68 per cent credible intervals for HALOFIT and HMCODE, resulting from cross-comparison MCMCs where EUCLIDEMULATOR1 data serve as mock data. The fiducial values for the cosmological parameters are given in the first data column. All results presented in this table are based on the new theoretical uncertainty given in Section 2.4.2.

			EUCLIDEMULATOR (EE)	HALOFIT (HF)	HMCODE (HM)
w_0CDM	0.4 $h\,\mathrm{Mpc}^{-1}$	ω_b	0.02200 ± 0.00011	0.02202 ± 0.00012	0.02199 ± 0.00012
		ω_m	0.14320 ± 0.00050	0.14301 ± 0.00048	0.14245 ± 0.00052
		n_s	0.9600 ± 0.0023	0.9562 ± 0.0024	0.9651 ± 0.0024
		h	0.6700 ± 0.0028	0.6697 ± 0.0039	0.6674 ± 0.0039
		w_0	-1.000 ± 0.011	-0.998 ± 0.013	-0.979 ± 0.014
		σ_8	0.8300 ± 0.0033	0.8246 ± 0.0037	0.8293 ± 0.0038

This paper has been typeset from a TEX/LATEX file prepared by the author.

Downloaded from https://academic.oup.com/mnras/article/518/2/1859/6619061 by Dept of Energy/OSTI user on 26 August 2024