



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

On the road to percent accuracy II

Citation for published version:

Giblin, B, Cataneo, M, Moews, B & Heymans, C 2019, 'On the road to percent accuracy II: calibration of the non-linear matter power spectrum for arbitrary cosmologies', *Monthly Notices of the Royal Astronomical Society* . <https://doi.org/10.1093/mnras/stz2659>

Digital Object Identifier (DOI):

[10.1093/mnras/stz2659](https://doi.org/10.1093/mnras/stz2659)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Monthly Notices of the Royal Astronomical Society

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



On the road to percent accuracy II: calibration of the non-linear matter power spectrum for arbitrary cosmologies

Benjamin Giblin^{1*}, Matteo Cataneo¹, Ben Moews¹ and Catherine Heymans¹

¹ *Scottish Universities Physics Alliance, Institute for Astronomy, University of Edinburgh, Blackford Hill, Edinburgh, EH9 3HJ, UK*

September 16, 2019

ABSTRACT

We introduce an emulator approach to predict the non-linear matter power spectrum for broad classes of beyond- Λ CDM cosmologies, using only a suite of Λ CDM N -body simulations. By including a range of suitably modified initial conditions in the simulations, and rescaling the resulting emulator predictions with analytical ‘halo model reactions’, accurate non-linear matter power spectra for general extensions to the standard Λ CDM model can be calculated. We optimise the emulator design by substituting the simulation suite with non-linear predictions from the standard HALOFIT tool. We review the performance of the emulator for artificially generated departures from the standard cosmology as well as for theoretically motivated models, such as $f(R)$ gravity and massive neutrinos. For the majority of cosmologies we have tested, the emulator can reproduce the matter power spectrum with errors $\lesssim 1\%$ deep into the highly non-linear regime. This work demonstrates that with a well-designed suite of Λ CDM simulations, extensions to the standard cosmological model can be tested in the non-linear regime without any reliance on expensive beyond- Λ CDM simulations.

Key words: Cosmology: theory – Cosmology: cosmological parameters – Surveys – Methods: statistical

1 INTRODUCTION

Large cosmological data sets have ushered in the era of “precision cosmology”, in which parameters describing the Λ CDM model are known with uncertainties of only a few percent. This increased precision has recently brought to light some level of discordance between early- and late-time cosmological probes. In particular, all three major weak lensing surveys (Hildebrandt et al. 2017; Hikage et al. 2018; Troxel et al. 2018) infer lower values of $\sigma_8\sqrt{\Omega_m}$ compared to those obtained from observations of the temperature and polarisation of the cosmic microwave background (CMB; Planck Collaboration et al. 2018). Significant tensions have also arisen between the value of the Hubble constant derived from the CMB and its direct local measurement via the distance-redshift relation calibrated with Cepheid variables and Type Ia supernovae (Riess et al. 2016, 2018). If not the product of an unlikely statistical fluctuation, then these discrepancies may be ascribed either to unaccounted systematic uncertainties in the analyses or to an incompleteness of the concordance cosmology (see, e.g., Verde et al. 2013; Joudaki et al. 2017; Mörtzell & Dhawan 2018; DES Collaboration et al. 2018).

In the event that the data prefer a revision of the standard cosmological model, accurate predictions of large-scale structure statistics for alternative cosmologies to Λ CDM become very valuable. However, we currently lack a method to predict these observ-

ables for general extensions to Λ CDM and over the range of scales relevant for ongoing and future galaxy surveys (Laureijs et al. 2011; LSST Dark Energy Science Collaboration 2012; Hildebrandt et al. 2017; Abbott et al. 2018; Taylor et al. 2018). This means we cannot take advantage of the wealth of information contained in the non-linear scales without resorting to computationally expensive numerical simulations.

It is however infeasible to simulate all possible combinations of cosmological parameters, even within the Λ CDM paradigm. For current applications, fitting functions calibrated on a small number of N -body simulations are good enough to obtain unbiased constraints (Smith et al. 2003; Takahashi et al. 2012; Mead et al. 2015). On the other hand, future all-sky surveys will place more stringent requirements on our theoretical predictions, which makes emulator techniques a preferable strategy as far as accuracy is concerned. The basic idea is to interpolate between the output of ~ 10 – 100 N -body simulations of cosmologies associated with carefully-chosen input parameters (Habib et al. 2007; Schneider et al. 2008; Heitmann et al. 2009, 2016; Lawrence et al. 2017; Euclid Collaboration et al. 2018; Winther et al. 2019; Harnois-Deraps et al. 2019). Previous implementations of this methodology have been applied either to flat, massless neutrinos (i.e. vanilla) Λ CDM, or to a few specific extensions, that is models with massive neutrinos, evolving dark energy with a linearly parameterised equation of state or $f(R)$ gravity. The computational expense of simulating modified gravity theories, however, can be orders of magnitude larger than that of the

* bengib@roe.ac.uk

standard model (Llinares 2018). Consequently, there currently exists no emulator capable of making non-linear predictions accurate at the level of $\sim 1\%$ across a broad range of beyond- Λ CDM cosmologies. This severely limits our capacity for constraining these alternative models with data probing the growth of structure on cosmic scales.

In this paper we complement the work of Cataneo et al. (2019, referred to as ‘C19’ henceforth), who formulated an efficient method to compute the non-linear matter power spectrum in a broad range of extensions to the standard cosmology with the desired percent accuracy. The C19 strategy consists of rescaling the non-linear power spectrum of a *tailored* Λ CDM-like model – the *pseudo* cosmology – by an analytical function encapsulating the non-linear physics beyond the vanilla cosmology – the *halo model reaction*. By construction the pseudo cosmology shares the linear power spectrum of the *real* beyond- Λ CDM cosmology.

For Λ CDM extensions exhibiting a scale-independent linear growth of structure, the corresponding pseudo non-linear power spectrum can be obtained from a Λ CDM emulator by matching the amplitude of mass fluctuations, σ_8 , to that of the real cosmology at the redshift of interest. In general, however, structures in alternative cosmological scenarios grow at different rates on different scales, which makes traditional emulators unfit for this task. Here, we present a new Gaussian process emulator designed to compute the pseudo non-linear matter power spectrum of arbitrary cosmologies, including both scale-independent and scale-dependent linear growth. We show that a suite of ~ 1000 Λ CDM-like simulations is enough to predict with percent accuracy the pseudo non-linear power spectrum of popular cosmologies which have scale-dependent linear growth, such as $f(R)$ gravity and massive neutrino models. We also obtain similar performance for synthetic cosmological models created by rescaling the linear Λ CDM power spectra by smooth arbitrary functions.

The combination of the C19 halo model reactions with the emulator developed in this work will hence provide accurate predictions of the non-linear matter power spectrum in a broad class of cosmological models, including modified gravity, dark energy and massive neutrinos, for use in cosmic shear analyses (see, e.g., Troxel et al. 2018; Hikage et al. 2018; Hildebrandt et al. 2018). The fact that galaxies are biased tracers of the underlying matter distribution means that additional emulator complexity, including parameters to describe the halo occupation distribution, would be required for this methodology to be used in future clustering analyses (see, e.g., Zhai et al. 2019). There is also the potential to use this technique to model the impact of baryonic feedback.

This paper is organised as follows. In Section 2 we examine the expected range of deviations in the linear matter power spectrum of viable extensions to the Λ CDM cosmology, and outline the C19 reaction framework for modelling the non-linear power spectrum. In Section 3 we detail the emulator methodology for arbitrary input linear power spectra, where we define ‘arbitrary’ to mean any power spectra which deviates smoothly from the standard model approximately within the range allowed by observational constraints. We also describe the construction of our suite of surrogate numerical simulations used as a training set for the emulator, which we design to encompass a reasonable range of Λ CDM extensions. In Section 4 we present the performance of our emulator and determine the minimum number of simulations we need to reach the accuracy required by the high-quality data from the next generation of all-sky imaging surveys. Finally, we conclude and discuss the impact of our findings in Section 5.

2 BEYOND VANILLA Λ CDM

The late-time phenomenology of the concordance model – a globally flat Universe with a matter-energy content dominated by cold dark matter, baryonic matter and the cosmological constant – is entirely described by five *base* parameters, $\pi_\Lambda \equiv \{\Omega_b h^2, \Omega_m h^2, h, n_s, \ln(10^{10} A_s)\}$, where Ω_b is the current fraction of energy-density in baryonic matter, Ω_m is the present-day total matter energy-density fraction, h is the dimensionless Hubble constant, and n_s and A_s are, respectively, the slope and amplitude of the primordial scalar power spectrum. Under the assumption of the standard cosmology, measurements of the CMB constrain the radiation content (i.e. photons and three massless neutrino species) to better than one part in a thousand (Fixsen 2009).

Changes to the dark sector and to the law of gravity have profound implications for the formation of structures in the Universe over a wide range of scales. Departures from the standard cosmology, therefore, leave distinctive features on the statistical quantities measuring the clustering of matter, such as the two-point correlation function, or its Fourier transform, the power spectrum $P(k)$. The landscape of Λ CDM extensions is exceptionally vast, and in this work we only consider a family of $f(R)$ gravity theories (Hu & Sawicki 2007) and massive neutrino cosmologies (Lesgourgues & Pastor 2006), both with a Λ CDM background. However, with little or no modification our methodology (see Section 3) can be applied to a much broader class of models, including non-standard dark matter candidates (Marsh & Silk 2014; Schneider 2015; Cyr-Racine et al. 2016; Marsh 2016; Poulin et al. 2016; Hložek et al. 2017; Dakin et al. 2019; Thomas et al. 2019), extra relativistic degrees of freedom (Baumann et al. 2018), and Horndeski theories (Zumalacárregui et al. 2017).

What makes models like $f(R)$ gravity and massive neutrino cosmologies interesting is their scale-dependent linear growth of structure. More specifically, $f(R)$ theories enhance the growth on scales comparable to, or smaller than, the Compton wavelength, which at $z = 0$ reads (Hu & Sawicki 2007)

$$\lambda_{C0} \approx 30 \sqrt{\frac{(n+1)}{4-3\Omega_m} \frac{|f_{R0}|}{10^{-4}}} h^{-1} \text{Mpc}, \quad (1)$$

where f_{R0} defines the extent of the departure from General Relativity, recovered for $f_{R0} = 0$. We set $n = 1$ and use $|f_{R0}| = 10^{-4}, 10^{-5}, 10^{-6}$ (referred to as F4, F5 and F6 in the following), which brackets the range of values giving rise to interesting cosmological behaviours, with F4 already strongly disfavoured by current data (Terukina et al. 2014; Lombriser 2014; Cataneo et al. 2015; Liu et al. 2016; Alam et al. 2016).

On the other hand, the high thermal velocities of massive neutrinos prevent them from clustering on scales smaller than the free-streaming length, λ_{FS} . This results in the suppression of the growth of structure on scales smaller than the maximum free-streaming length, which is defined at the time of the non-relativistic transition (after recombination for $m_\nu \lesssim 0.5$ eV), and is given by (Lesgourgues & Pastor 2006)

$$\lambda_{FS}^{\max} \approx 350 \sqrt{\frac{1}{\Omega_m} \frac{1 \text{ eV}}{m_\nu}} h^{-1} \text{Mpc}. \quad (2)$$

For simplicity, we will consider cosmologies including two massless neutrinos and one massive neutrino with $m_\nu = 0.05, 0.1, 0.2, 0.4$ eV, where the smallest value is close to the minimum sum of neutrino masses consistent with neutrino oscillation experiments (Forero et al. 2014), and the largest value is outside the 95% confidence region found by Planck Collaboration et al. (2018).

Table 1. The massive neutrino, $f(R)$ gravity and hybrid models used in this work to design and test our emulator scheme, each labelled according to the sum of neutrino masses ($\sum m_\nu$) and/or the strength of the deviation from standard gravity ($|f_{R0}|$). We consider two base Λ CDM cosmologies: (A) $\Omega_b h^2 = 0.0225$, $\Omega_m h^2 = 0.1382$, $h = 0.6898$, $n_s = 0.969$, $\ln(10^{10} A_s) = 3.195$; and (B) $\Omega_b h^2 = 0.0173$, $\Omega_m h^2 = 0.0864$, $h = 0.6$, $n_s = 0.969$, $\ln(10^{10} A_s) = 3.807$. Note that (B) sits several standard deviations away from the [Planck Collaboration et al. \(2018\)](#) best fit cosmology, thus working as stress test for our method.

model	$\sum m_\nu$	$ f_{R0} $	base parameters
MNU_0.05	0.05 eV	–	(A)
MNU_0.1	0.1 eV	–	(A)
MNU_0.2	0.2 eV	–	(A)
MNU_0.4	0.4 eV	–	(A)
xMNU_0.05	0.05 eV	–	(B)
xMNU_0.1	0.1 eV	–	(B)
xMNU_0.2	0.2 eV	–	(B)
xMNU_0.4	0.4 eV	–	(B)
F4	–	10^{-4}	(A)
F5	–	10^{-5}	(A)
F6	–	10^{-6}	(A)
F4-MNU_0.4	0.4 eV	10^{-4}	(A)
F5-MNU_0.2	0.2 eV	10^{-5}	(A)
F6-MNU_0.1	0.1 eV	10^{-6}	(A)

In addition, for all these extensions we keep the base cosmological parameters fixed to their Λ CDM values, and compensate for the presence of massive neutrinos by reducing the cold dark matter density as $\Omega_c \rightarrow \Omega_m - \Omega_b - \Omega_\nu$, where

$$\Omega_\nu h^2 = \frac{\sum m_\nu}{93.14 \text{ eV}}. \quad (3)$$

Table 1 lists all models, collectively referred to as *physical models*, used to test our emulation scheme described in Section 3, including hybrid cosmologies with massive neutrinos and modified gravity. For these extensions, the linear power spectrum deviations from the standard predictions are obtained with MGCAMB¹ ([Zhao et al. 2009](#); [Hojjati et al. 2011](#)). Figure 1 shows a $z = 0$ example of beyond- Λ CDM to Λ CDM linear matter power spectrum ratios.

2.1 The halo model reaction framework

By unlocking the information stored in the statistical distribution of matter on small scales we can substantially increase our sensitivity to physics beyond the vanilla cosmology, be it new particles/fluids, modifications to the theory of gravity, or astrophysics (see, e.g. [Huterer et al. 2015](#); [Copeland et al. 2018](#); [Nori & Baldi 2018](#); [Heymans & Zhao 2018](#); [Schneider et al. 2019](#)). Not surprisingly, then, the non-linear matter power spectrum is found at the core of all cosmological analyses hinged on galaxy survey data ([Parkinson et al. 2012](#); [Kitching et al. 2014](#); [Gil-Marín et al. 2016](#); [Hildebrandt et al. 2017](#); [van Uitert et al. 2018](#); [Joudaki et al. 2018](#); [Abbott et al. 2018](#); [Gil-Marín et al. 2018](#); [Hikage et al. 2018](#)). To take full advantage of the exquisite observations from forthcoming experiments ([Laureijs et al. 2011](#); [LSST Dark Energy Science Collaboration 2012](#); [Levi et al. 2013](#)), theoretical predictions must

¹ <http://aliojjati.github.io/MGCAMB/mgcamb.html>

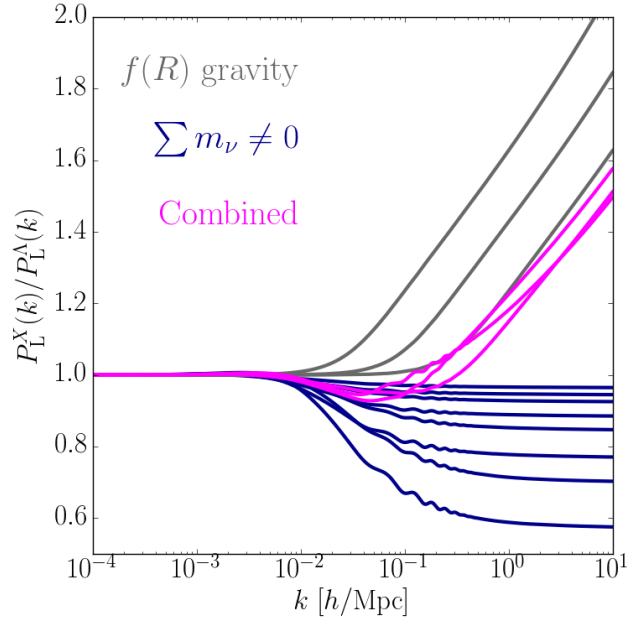


Figure 1. Beyond- Λ CDM (P_L^X) to Λ CDM (P_L^Λ) linear matter power spectrum ratios for the physical models in Table 1 at $z = 0$. Pure $f(R)$ gravity (in grey) and massive neutrino cosmologies (in blue) alter the clustering of matter in similar but opposite ways. This degeneracy is clearly visible for the hybrid models (in magenta), where both extensions are active. These curves will be used as guidelines to design our emulator.

reach percent level accuracy deep in the non-linear regime of structure formation ([Taylor et al. 2018](#)). At present, however, no method is at the same time computationally efficient, accurate and flexible enough to be employed in future analyses aimed at constraining physics beyond the standard paradigm ([Lesgourgues et al. 2009](#); [Bird et al. 2012](#); [Heitmann et al. 2014](#); [Brax & Valageas 2013](#); [Zhao 2014](#); [Blas et al. 2014](#); [Massara et al. 2014](#); [Levi & Vlah 2016](#); [Lawrence et al. 2017](#); [Aviles & Cervantes-Cota 2017](#); [Senatoro & Zaldarriaga 2017](#); [Bose et al. 2018](#); [Euclid Collaboration et al. 2018](#); [Heisenberg & Bartelmann 2019](#); [Winther et al. 2019](#)).

The halo model reaction framework first proposed by C19, and based on the work of [Mead \(2017\)](#), stands out as a promising, practical solution to the long-standing problem of predicting the non-linear matter power spectrum with the required accuracy and speed, while also being readily applicable to a variety of cosmological scenarios. Within this approach, the non-linear power spectrum of a given *real* cosmology takes the form

$$P_{\text{NL}}^{\text{real}}(k, z) = \mathcal{R}(k, z) \times P_{\text{NL}}^{\text{pseudo}}(k, z), \quad (4)$$

where the *pseudo* cosmology is defined as a Λ CDM cosmology sharing the same linear matter power spectrum of the *real* cosmology, that is

$$P_L^{\text{pseudo}}(k, z) = P_L^{\text{real}}(k, z). \quad (5)$$

The *reaction*,

$$\mathcal{R}(k, z) \equiv \left. \frac{P_{\text{NL}}^{\text{real}}(k, z)}{P_{\text{NL}}^{\text{pseudo}}(k, z)} \right|_{\text{HM}}, \quad (6)$$

captures the non-linearities sourced by late-time physics beyond Λ CDM, and is predicted with the halo model (HM) and perturbation theory (see [Cataneo et al. 2019](#), for details). In taking the ratio

of the two halo model predictions, Mead (2017) showed that the impact of the known inaccuracies in the individual power spectra is significantly reduced. On linear scales and for *real* Λ CDM cosmologies one has the trivial relation $\mathcal{R} = 1$. C19 found that the modelling approach summarised by equation 4 is accurate at the percent level for a broad range of general extensions to the standard model.

Here, we formulate a method to provide predictions for $P_{\text{NL}}^{\text{pseudo}}$ that are more accurate than any semi-analytical prescription based on the halo model. It is this accurate estimate that is then used to calibrate the reaction, \mathcal{R} , in equation 4 in order to derive the non-linear matter power spectrum of our target beyond- Λ CDM cosmology, $P_{\text{NL}}^{\text{real}}$. To this end, we develop a novel emulation scheme where in addition to the base Λ CDM parameters we use the shape of the linear matter power spectrum as input. This work is intended as a first feasibility study, and as such we approximate the output of otherwise expensive *pseudo* N -body simulations with the HALOFIT fitting functions (Takahashi et al. 2012) implemented in a modified version of the public Einstein-Boltzmann solver CAMB² (Lewis et al. 2000). In practice, for a model X , CAMB reads in both π_Λ and the externally generated P_L^X/P_L^Λ ratio, evaluates P_L^{pseudo} and gives it to HALOFIT, which finally provides the desired non-linear quantity. The grey curves in the upper panel of Figure 2 show the *pseudo* non-linear power spectra for the physical models in Table 1, where we used the set (A) of base Λ CDM parameters also for the xMNU cosmologies. Note that despite HALOFIT predictions being approximate, this strategy is theoretically consistent, in that *pseudo* cosmologies are simply Λ CDM cosmologies with non-standard initial conditions. Our approach enables a rapid design, construction and performance assessment of the emulator (Heitmann et al. 2009; Euclid Collaboration et al. 2018).

3 METHODOLOGY

In this Section, we describe the basics of Gaussian process emulation of cosmological statistics. We also detail our pipeline, which starts from the linear matter power spectrum of an arbitrary cosmology and maps this onto a point in the multi-dimensional parameter space over which the emulator performs the interpolation. This gives the emulator high versatility, in that it will be able to make predictions for a broad range of non-standard cosmologies absent from the training set.

3.1 Emulation strategy

An in-depth discussion of our emulation strategy, employing Gaussian process (GP) regression, can be found in Appendix A. The accuracy of this method is sensitive to the particular observable being emulated. Our aim is to predict the *pseudo* non-linear matter power spectrum discussed in Section 2.1, a quantity which has a dynamic range of many orders of magnitude and features a change in gradient sign between small and large scales, as shown in the upper panel of Figure 2 for the extensions to Λ CDM listed in Table 1. GP emulators on the other hand perform better at predicting smooth monotonic functions with a narrow dynamic range. Hence,

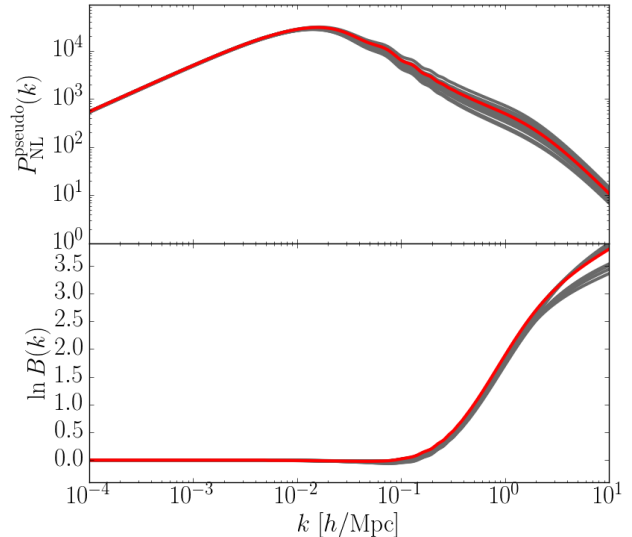


Figure 2. *Upper panel:* in grey, the $z = 0$ pseudo non-linear matter power spectra for the extensions to Λ CDM listed in Table 1, computed with HALOFIT as discussed in Section 2.1. *Lower panel:* the natural logarithm of the boost factor obtained by taking the ratio of pseudo non-linear and linear matter power spectra (see equation 7). For reference, the base Λ CDM cosmology (A), given in Table 1, is shown in red. The monotonicity and narrow dynamic range of the $\ln B$ relative to the $P_{\text{NL}}^{\text{pseudo}}$ help to improve the emulator’s performance significantly.

following Euclid Collaboration et al. (2018), we consider the natural logarithm of the *boost factor*, $B(k, z)$, as our observable of interest, where

$$B(k, z) \equiv \frac{P_{\text{NL}}^{\text{pseudo}}(k, z)}{P_L^{\text{pseudo}}(k, z)}. \quad (7)$$

The lower panel of Figure 2 shows that the natural logarithm of the boost factor indeed matches the desired properties of a GP observable.

An essential step in the emulator design is data compression. For any particular cosmological model, the power spectrum is sampled both in spatial frequencies (wavenumbers) and time (redshifts), for a total of $\sim 10^4$ support points. Training the emulator on each of these individual measurements requires prohibitively large computational resources. To overcome this issue, we perform a principal component analysis (PCA) following previous work (see, e.g. Habib et al. 2007; Schneider et al. 2008; Heitmann et al. 2016). For a given parameter vector, π , and redshift, z , we apply the following linear decomposition

$$\ln B(k, z; \pi) = \mu(k, z) + \sum_{j=1}^{n_\psi} \psi_j(k, z) w_j(\pi) + \epsilon, \quad (8)$$

where μ and $\{\psi_1, \dots, \psi_{n_\psi}\}$ are, respectively, the mean and basis functions of the training set of boosts at z , and the weights w_j are the projections of the mean-subtracted observable onto ψ_i . The error term ϵ has two contributions: one component, arising from the negligible noise on the HALOFIT training predictions (given by $\epsilon_n(\pi)$ in equation A1); and a second component coming from missing relevant basis functions. We find that $n_\psi = 12$ is sufficient to

² <https://github.com/cmbant/CAMB>

satisfactorily reconstruct the training set, and including more basis functions has only a marginal effect on the emulator accuracy. Ultimately, for a given π^* our emulator infers the boost factor by predicting the weights, $w_j(\pi^*)$, conditioned on the training set $w_j(\pi)$.

3.2 Model-independent parameterisation of the pseudo cosmologies

Extensions to Λ CDM are described by an arbitrary number of additional physical parameters. Consider, for example, w CDM, $f(R)$ gravity and massive neutrino cosmologies: matter clustering in all these models depends on five standard cosmological parameters, e.g. $\pi_\Lambda = \{\Omega_b h^2, \Omega_m h^2, h, n_s, \ln(10^{10} A_s)\}$, along with one or more parameters describing model-specific features, $\pi_X = \{f_{R0}, \sum m_\nu, \dots\}$. To construct an emulator capable of predicting generic pseudo non-linear matter power spectra, we opt for a model-independent parameterisation mapping an arbitrary model to a certain coordinate, $\pi^* = \{\pi_\Lambda, \Delta\alpha\}$, in our emulation parameter space, where $\Delta\alpha$ is a vector of effective parameters that quantifies smooth deviations from the linear matter power spectrum of the standard cosmology. Thus, for $\Delta\alpha = 0$ we recover the late-time physics of the vanilla Λ CDM model. For this, we model the ‘shape’, namely the ratio of the *linear* pseudo matter power spectrum to that of the Λ CDM cosmology sharing the same standard parameters, as

$$S(k, z; \pi_\Lambda, \Delta\alpha) \equiv \frac{P_L^{\text{pseudo}}(k, z; \pi_\Lambda, \pi_X)}{P_L^\Lambda(k, z; \pi_\Lambda)} \approx 1 + \sum_{i=1}^{n_\Phi} \Phi_i(k, z) \Delta\alpha_i, \quad (9)$$

where $\{\Phi_1, \dots, \Phi_{n_\Phi}\}$ are the principal components of a training set, \mathbb{T} , consisting of smooth curves capturing a range of ‘reasonable’ deviations from Λ CDM³. Use of PCA to parameterise deviations from vanilla Λ CDM cosmology has been previously implemented in a number of studies (see, e.g., Zhao et al. 2009; Hojjati et al. 2012; Eifler et al. 2015).

³ Eq. (9) can be understood in terms of a principal component reconstruction based on the training set \mathbb{T} , that is

$$S(k_m) \approx \nu(k_m) + \sum_{i=1}^{n_\Phi} \Phi_i(k_m) \alpha_i, \quad (10)$$

where for simplicity we dropped the dependence on redshift and cosmology, and made explicit that we sample the shape S at $k_m \in \{k_1, \dots, k_{n_k}\}$. The function ν is the mean across \mathbb{T} , and in general we expect $\nu \neq 1$. Then, for any Λ CDM cosmology we must have

$$1 - \nu(k_m) \approx \sum_{i=1}^{n_\Phi} \Phi_i(k_m) \alpha_i^\Lambda, \quad (11)$$

with α_i^Λ being the weights that exactly compensate the departure of the mean from unity, and upon using the orthonormality of the basis functions they read

$$\alpha_i^\Lambda = \sum_{m=1}^{n_k} \Phi_i(k_m) [1 - \nu(k_m)]. \quad (12)$$

Finally, the combination of Eqs. (10) and (11) gives Eq. (9), with $\Delta\alpha_i \equiv \alpha_i - \alpha_i^\Lambda$. Note that for a given training set, \mathbb{T} , the Λ CDM weights need to be computed only once, and can subsequently be hard-coded in the emulator.

3.2.1 Generation of the basis set

We assemble a set of orthogonal basis functions to reconstruct a broad class of scale-dependent linear departures from the Λ CDM power spectrum. For this purpose, we generate random curves in Fourier space over the range $10^{-4} < k \text{ Mpc}/h < 10$, with the constraint that all curves must converge to unity on large enough scales. This last requirement is motivated by the physics of well-known Λ CDM extensions, such as Generalised Brans-Dicke theories (De Felice & Tsujikawa 2010; Park et al. 2010; Hinterbichler et al. 2011; Pogosian & Silvestri 2016; Quiros et al. 2016; Joyce et al. 2016) and massive neutrino cosmologies (Lesgourgues & Pastor 2006). Both these models possess a characteristic scale, \bar{k} , such that on scales $k \ll \bar{k}$ the linear matter clustering matches that of the vanilla Λ CDM cosmology. For $k \gg \bar{k}$ the growth of structure is either suppressed ($S < 1$) or enhanced ($S > 1$). Note that models with $\bar{k} \rightarrow 0$ exhibit in practice a scale-independent linear growth (see, e.g., Pogosian & Silvestri 2016), and a simple rescaling of the amplitude of the vanilla Λ CDM power spectrum is sufficient to produce the corresponding pseudo power spectrum. Viable models in these two classes of theories above have $\bar{k} \gtrsim 10^{-3} h/\text{Mpc}$ (Brax et al. 2012; Wang et al. 2012; Joyce et al. 2015; Planck Collaboration et al. 2018). We therefore adopt the value $\bar{k} = 10^{-3} h/\text{Mpc}$ for the scale at which the shapes converge to unity, and allow for generous positive smooth deviations in the range $k \in [\bar{k}, 10] h/\text{Mpc}$. Moreover, since the basis functions derived from our generated shapes are sufficiently general to describe the power spectrum ratios at various redshifts, we drop the z -dependence, $\Phi_i(k, z) \rightarrow \Phi_i(k)$, and use one basis set for all redshifts.

In order to reduce the sensitivity of the emulator to the particular method used to generate the set of random shapes, we employ two different and independent generating strategies. The first of these methods, which we refer to as GPCURVES, draws shapes from a 1-dimensional Gaussian process with a squared exponential kernel of the form given in equation A2 (where $d = 1$), conditioned on a ‘training set’ of densely sampled points at $k < 10^{-3} h/\text{Mpc}$, all with values equal to unity, helping to fix the random shapes to this value on large scales. Here, for the kernel amplitude, we use $A = 5$, which corresponds to the maximum value that the random shapes can have on scales $k > 10^{-3} h/\text{Mpc}$. We tune the only length-scale parameter, p , so that the artificial shapes exhibit at most one stationary point for $k > 10^{-3} h/\text{Mpc}$, thus resembling features of the physical shapes in Figure 1. There is no intrinsic restriction in the GP preventing generated shapes from taking unphysical negative values. We therefore generate an initial set of curves from which we sample 1000 viable positive shapes.

As a second method we use SMURVES (Moews et al. 2019), a novel random curve generator. The functionality of the generator is illustrated by considering a particle travelling in Fourier space from low-to-high wavenumbers. Before $k = 10^{-3} h/\text{Mpc}$, the particle experiences no forces and hence travels undisturbed at a vertical coordinate of unity. After this point, forces randomly generated from a uniform distribution perturb the particle in the vertical direction, altering its trajectory. Also randomly generated are the locations of gradient sign changes. In this manner, we produce an additional 1000 smooth curves with a maximum of one stationary point and vertical intervals in the range $[0, 8]$. In order to facilitate extra flexibility in the deviations from Λ CDM, the parameters specifying the curve generation in SMURVES are set purposefully to create a sample of curves which are slightly broader than GPCURVES.

A comparison of 10 curves randomly selected from either

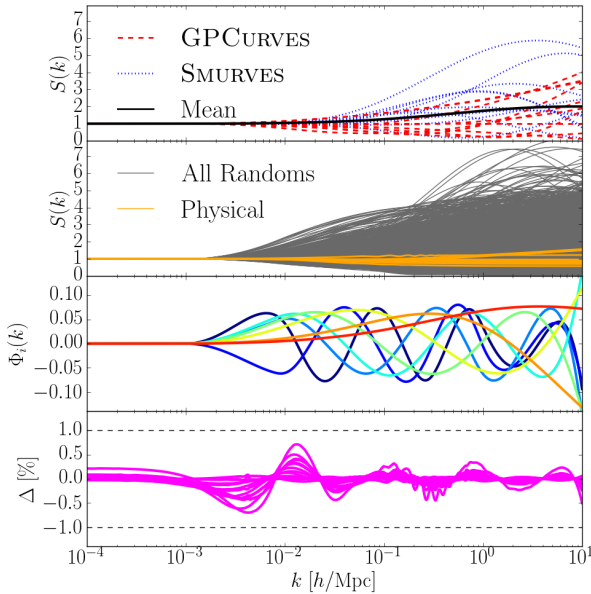


Figure 3. *Upper panel:* A sub-sample of random shapes generated with the GPCURVES (dashed red) and with the SMURVES (dotted blue) methods. The mean across all 2000 random shapes is shown in solid black. *Upper-middle panel:* All 2000 random shapes from GPCURVES and SMURVES (grey) with the 14 shapes for the physical models from Figure 1 shown for reference (orange). *Lower-middle panel:* the 8 orthogonal basis functions, Φ_i , obtained via a PCA of the 2000 random shapes. The colour indicates the ranking of the basis function in the hierarchy of explaining the variance in the shapes, with red indicating the most important, and dark blue indicating the least. *Lower panel:* the accuracy, Δ , in reconstructing the physical model shapes, from Figure 1, using the basis functions, Φ_i .

samples are shown in different colours in the upper panel of Figure 3. The upper-middle panel shows the shapes corresponding to the physical models in Table 1 (orange) relative to the collective sample of 2000 random curves from GPCURVES and SMURVES (grey). The lower-middle panel shows the orthogonal basis functions, Φ_i , obtained via a PCA of the 2000 curves. Note that the resulting basis set allows for a broader range of departures from Λ CDM than our test physical models.

Each basis function, Φ_i , appearing in the shape reconstruction (equation 9) contributes one extra dimension to the parameter space of the emulator, with the additional parameters corresponding to the PCA weights, $\Delta\alpha_i$. Therefore, in order for our emulator to be computationally efficient whilst still achieving accuracies of $\lesssim 1\%$, we keep the smallest number of basis functions that guarantees sub-per-cent reconstruction errors on all physical test shapes (see Figure 1). We find that $n_\Phi = 8$ basis functions are sufficient to reconstruct $f(R)$ gravity shapes with negligible errors. For cosmologies with massive neutrinos, however, changes to the BAO dynamics at early times produce rapid oscillations in their late-time power spectrum shapes. These features require one additional step in the reconstruction, but no extra parameters need be included in the emulation. Further details on this can be found in Appendix B. The resultant shape reconstruction accuracy for all physical models is shown in the lower panel of Figure 3. Overall our emulation volume is therefore carved out of a 13-dimensional space: 5 Λ CDM parameters, π_Λ , and 8 shape parameters, $\Delta\alpha$.

3.3 Building the training and trial sets

Having established a model-independent description for arbitrary cosmological models, we now discuss how we generate training and trial sets for our pseudo matter power spectrum emulator. In Section 3.3.1 we give details on the Latin hypercube (LH) space-filling strategies we adopt to efficiently sample our 13-dimensional space. In Section 3.3.2, we then define suitable ranges for the cosmological and shape parameters, $\{\pi_\Lambda, \Delta\alpha\}$, such that the emulated boost factors are representative of viable extensions to Λ CDM. The nodes of the LHs take values in the range $[0, 1]$, which we then properly rescale to obtain the “physical” coordinates π used to generate the training and trial sets.

3.3.1 Latin hypercube sampling

The use of LH-based distributions of points to evenly sample a parameter space was first introduced by McKay et al. (1979). The LH is a generalisation of the concept of a Latin square in which each row and column of a two-dimensional grid features exactly one sample, taking the form of a chess board with a number of rooks that do not threaten each other. Owing to the desirable space-filling properties, LH sampling of the input parameters for simulations has become a standard practice (Sacks et al. 1989; Currin et al. 1991; Heitmann et al. 2006; Schneider et al. 2008; Agarwal et al. 2012; Liu et al. 2015; Garrison et al. 2018; Harnois-Deraps et al. 2019).

Assembling 13-dimensional LH designs therefore presents an ideal way to construct predictions to both train the emulator and test its accuracy for different regions of the emulation space. The spatial distribution of the trial cosmologies does not require optimisation. Therefore we sample 300 trial coordinates from a 13-dimensional LH generated with the PYDOE Python package⁴. The training nodes, however, impact significantly on the emulator accuracy, and sampling methods able to cover the high-dimensional volume as uniformly as possible are thus preferable. For this purpose, space-filling criteria based on the distance between samples are commonly used. One of the most prominent approaches is the Maximin LH design, which maximises the minimum distance between samples (Morris & Mitchell 1995). While this approach leads to the sought-after space-filling properties in the bulk of the parameter space, most points cluster in the corners or at the boundaries of the lower-dimensional projection spaces. Since the matter power spectrum is most sensitive to subsets of cosmological and shape parameters, space-filling strategies also optimising projection space sampling would give an obvious advantage.

To this end, Joseph et al. (2015) introduced the maximum projection design (MAXPRO), in which a weighted Euclidean distance, E , is minimised, resulting in the new criterion

$$\min_{\mathbf{D}} E(\mathbf{D}) = \left(\frac{1}{\binom{N}{2}} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{1}{\prod_{l=1}^d (x_{il} - x_{jl})^2} \right)^{\frac{1}{d}}, \quad (13)$$

where $\mathbf{D} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ is an $N \times d$ design matrix containing N samples occupying the d -dimensional parameter space. x_{il} denotes the value of the i -th sample’s coordinate in the l -th dimension. The MAXPRO criterion maximises the bulk and projection space-filling properties, since for any dimension l , if $x_{il} = x_{jl}$ and $i \neq j$,

⁴ <https://pythonhosted.org/pyDOE/>

$E(\mathbf{D}) = \infty$. This ensures that in the design minimising $E(\mathbf{D})$ no two points can be close to each other in any of the projections.

The MaxPro design optimisation implemented in this work is based on the simulated annealing algorithm described in [Morris & Mitchell \(1995\)](#)⁵. This consists of selecting an initial configuration before searching for progressively better design choices by randomly perturbing the current design, keeping the new matrix should it reduce the cost function, $E(\mathbf{D})$, or else do so with a given probability. Finally this implementation arrives at a locally optimal MAXPRO design by ascending the gradient given by

$$\frac{\partial E^d(\mathbf{D})}{\partial x_{rs}} = \frac{2}{\binom{N}{2}} \sum_{i \neq r} \frac{1}{(x_{is} - x_{rs})} \frac{1}{\prod_{l=1}^d (x_{il} - x_{rl})^2}. \quad (14)$$

To investigate the performance of our emulator we generate sets of training nodes with sizes $N = 100, 300, 500$ and 600 , with the additional property of having an optimal space-filling pattern in each projection. Furthermore, to investigate the sensitivity of our results to the specific configuration of nodes in each training set, we produce 10 optimised designs of each size and measure the range of achieved accuracies across them.

3.3.2 Setting parameter ranges

We shall now map our dimensionless training and trial nodes belonging to the unit LHs onto our parameters, $\pi \in \{\pi_\Lambda, \Delta\alpha\}$. To do this, we must identify appropriate ranges for these parameters. We begin with the Λ CDM parameters, and note that our emulator is intended for cosmological analyses employing high-quality data from the next-generation galaxy surveys. Hence, we set the boundaries on π_Λ to be roughly consistent with the 95% marginalised constraints from the combined TT, TE, EE + lowE analysis in [Planck Collaboration et al. \(2018\)](#), namely,

$$\begin{aligned} 0.0215 < \Omega_b h^2 < 0.0235, \\ 0.12 < \Omega_m h^2 < 0.155, \\ 0.6 < h < 0.8, \\ 0.9 < n_s < 1.05, \\ 2.92 < \log(10^{10} A_s) < 3.16. \end{aligned} \quad (15)$$

These ranges are illustrated in Figure 4, which shows the 2D projections of the Λ CDM sub-space, with the grey points corresponding to one of the $N = 500$ MAXPRO distributed training designs.

Turning to the shape parameters, $\Delta\alpha$, we take inspiration from the physical models listed in Table 1, and simply set the range of values in our training set to be slightly broader than those defined by the $\Delta\alpha$'s associated with $f(R)$ gravity, massive neutrino and combined cosmologies. The upper panel of Figure 5 shows the $\Delta\alpha_i$ for the physical models at $z = 0$ in orange, and those related to one of the designs with 500 optimally spaced training nodes (see Section 3.3.1) in grey. The convergence of the $\Delta\alpha_i$ values towards zero as the PCA rank, i , increases, is an indication that we are using a sufficient number of basis functions to reconstruct our test models.

The lower panel of Figure 5 shows the shapes corresponding to the $\Delta\alpha$ vectors shown in the upper panel, both for the 500-node training set (in grey) and for the physical models (in orange), reproduced from Figure 1. We see that the training shapes cover a

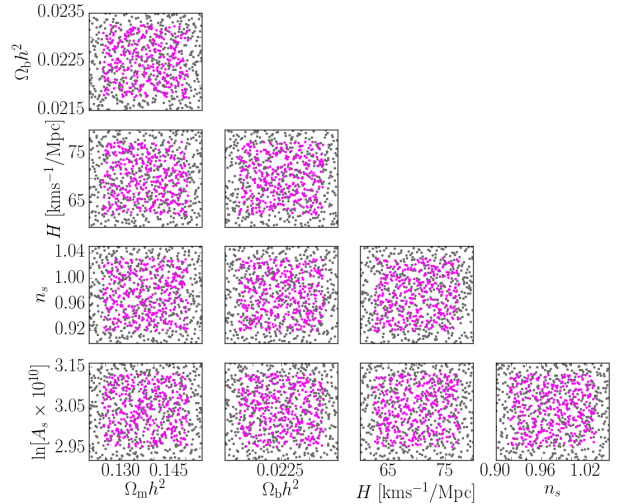


Figure 4. The 2D projections of the 13-dimensional emulation space for the five Λ CDM parameters only. The grey points correspond to a 500-node training set obtained by maximising the distance between nodes in each projection, whereas the magenta points correspond to the trial coordinates, which are simply sampled from a standard Latin hypercube (see Section 3.3.1 for more details). The axes show the full allowed range of the training points. The trial coordinates are confined to a smaller hypercube with sides measuring 75% of the training range, which reduces the impact of boundary effects on the emulator accuracy.

complex range of behaviours, whilst still encompassing the shapes of the physical models.

For the trial set, we opt for narrower parameter ranges, such that in each dimension the trial coordinates have values in the central 75% of the training set intervals. In doing this we test the emulator accuracy only in the bulk of the parameter space, so that our conclusions are less influenced by edge effects, where emulator accuracy degrades due to a shortage of nodes near the boundaries. This is visualised in Figure 4 by the distribution of trial nodes (in magenta). We also test the performance of our emulator for strictly Λ CDM cosmologies, where all shape parameters are set to zero. This Λ CDM trial ensemble is used to test whether the emulator is able to make accurate predictions for the standard model in the non-linear regime, even when Λ CDM cosmologies are completely absent from the training set.

3.3.3 Generating boost factors

With the training and trial nodes now assembled, the final task is to produce the corresponding linear as well as non-linear matter power spectra, and hence the logarithm of the boost factor, $B(k, z)$ (see equation 7), which the emulator is trained on and predicts. Conventionally, the non-linear statistics are measured from cosmological simulations, but for the purposes of this proof-of-concept study HALOFIT can serve as a proxy for a suite of N -body simulations. Note that this approach is self-consistent as long as the growth of structure and background evolution of the ‘simulated’ cosmologies matches those of the Λ CDM model, which is indeed the case for the pseudo cosmologies. To estimate the surrogate F_{NL}^{pseudo} we repeat the process described in Section 2.1 for the physical models: for a given pseudo cosmology, $\pi = \{\pi_\Lambda, \Delta\alpha\}$,

⁵ <https://CRAN.R-project.org/package=MaxPro>

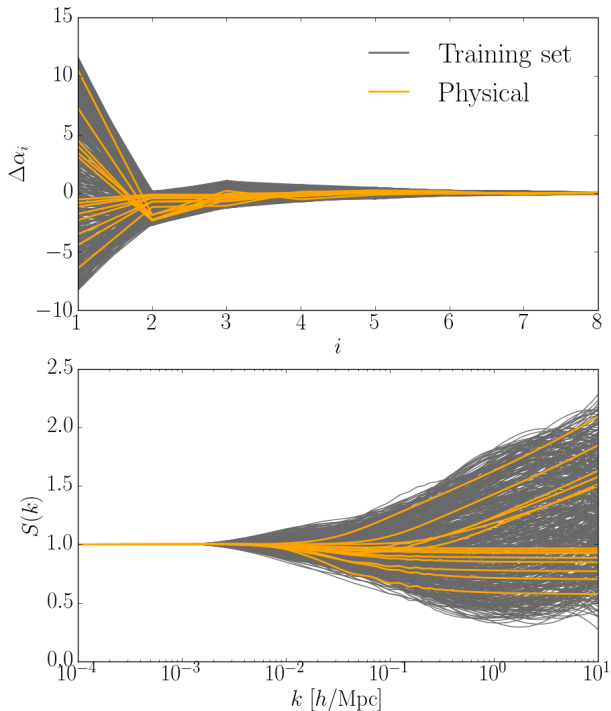


Figure 5. *Upper panel:* the $\Delta\alpha_i$ values, as a function of PCA rank, i , which parameterise deviations from Λ CDM in the $z = 0$ pseudo linear matter power spectra, for a 500 node training set (grey) and physical models (orange). *Lower panel:* synthetic and physical shapes corresponding to the $\Delta\alpha_i$ weights above.

our modified version of CAMB accepts as input the random shape described by $\Delta\alpha$, and internally multiplies this by the Λ CDM linear power spectrum associated with π_Λ . This generates the pseudo linear power spectrum which together with π_Λ serves as input for HALOFIT non-linear predictions. The boost factor is then produced by simply taking the ratio of the non-linear to the linear power spectrum. We generate boost factors for the training and trial sets of our emulator at redshifts 0 and 1. The $z = 0$ pseudo linear and non-linear matter power spectra, along with the corresponding boost factors, are shown in Figure 6 for a 500-node training set (in grey), as well as for the physical models (in orange) also illustrated in Figure 2.

4 RESULTS

The non-linear matter power spectrum is key to the derivation of a number of cosmological observables, therefore we use this rather than the emulated statistic, $\ln B(k, z)$, to evaluate the accuracy of our emulator. We consider wavenumbers in the range $k \in [0.01, 10] h/\text{Mpc}$, excluding larger scales where the boost factor goes to unity in all models, presenting no challenge to the emulator. For each trial cosmology and redshift separately, we quantify the emulator accuracy as

$$\epsilon^{\text{emu}} \equiv \max_{0.01 < k < 10} \left| \frac{P_{\text{NL}}^{\text{emu}}(k) - P_{\text{NL}}^{\text{True}}(k)}{P_{\text{NL}}^{\text{True}}(k)} \right| \quad (16)$$

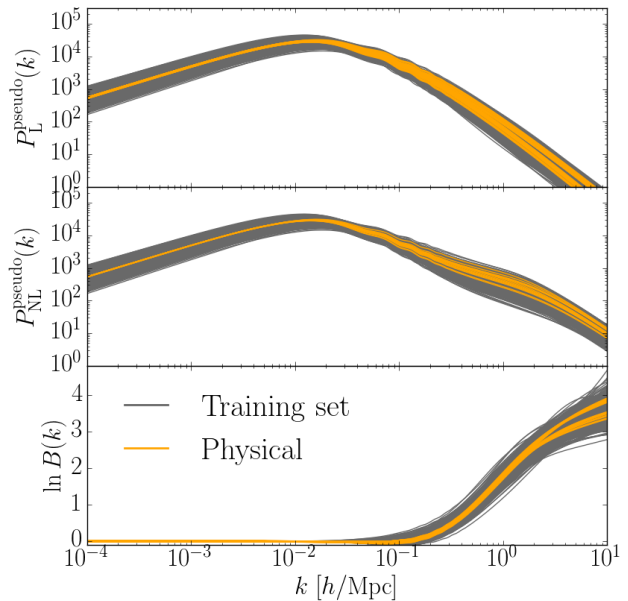


Figure 6. The $z = 0$ linear (upper) and non-linear (middle) matter power spectra for the 500 pseudo cosmologies of a training set realisation (grey) and for the physical models (orange). The lower panel shows the logarithm of the boost factor, defined as the natural logarithm of the ratio of the middle to the upper panels (see equation 7). By design our training set comfortably encompasses the physical model predictions.

where $P_{\text{NL}}^{\text{emu}}$ and $P_{\text{NL}}^{\text{True}}$ refer to the emulated and true pseudo non-linear power spectra respectively, and for the purposes of this study we take HALOFIT non-linear predictions as the truth. Figure 7 presents the fraction of the trial cosmologies with emulation errors smaller than a threshold value, $\bar{\epsilon}$, for various sizes of the training sets at both $z = 0$ (solid lines) and $z = 1$ (dashed). The fraction shown is the average across the 10 realisations per training set size. The left panel shows the results for the pseudo cosmology trials, whilst the right panel corresponds to the pure Λ CDM trial ensemble.

At $z = 0$, for the pseudo cosmology trials and training sets of size $N = 500$ or 600 , the emulation is accurate to better than 1% for about 85% of the trials, and the full ensemble is recovered to better than 2% accuracy. As expected, the emulator performs noticeably worse when trained on a smaller number of nodes, with only 13% (60%) of the pseudo non-linear spectra emulated to better than 1% when $N = 100$ (300). This is because the emulator has less information on the complex relationship between input parameters, π , and the boost factors. However, we find that the largest inaccuracy across all trial models is still $< 6\%$ ($< 3\%$) for $N = 100$ (300).

At $z = 1$, the fraction of pseudo trials with $\epsilon^{\text{emu}} < 1\%$ are similar to those at $z = 0$. However, the reduced steepness of the cumulative distributions at $z = 1$, compared to those at $z = 0$, suggests an increasing fraction of outliers with redshift. This is caused by the training and trial boost factors displaying a broader dynamic range at the higher redshift, which poses a greater challenge to emulation. Such broadening can be understood in terms of differences in the boost factor evolution induced by the different growth rates

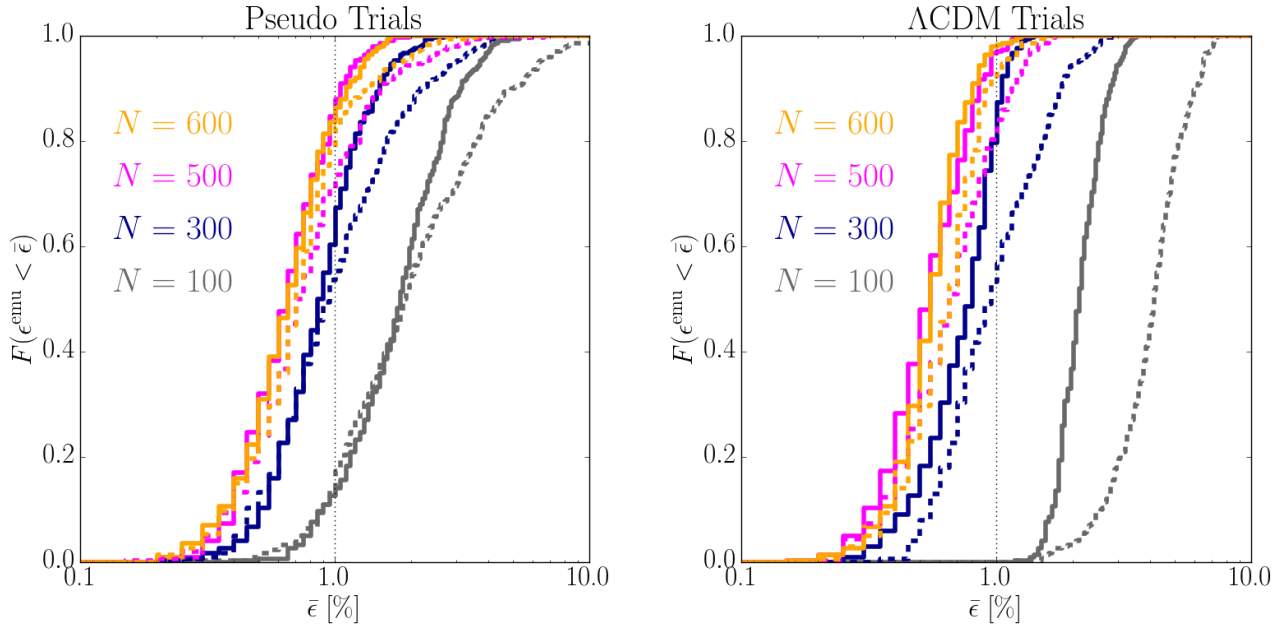


Figure 7. The cumulative distribution, F , of trial predictions for which ϵ^{emmu} , the maximum absolute fractional deviation between the emulated and HALOFIT pseudo non-linear matter power spectra over the range $k \in [0.01, 10] h/\text{Mpc}$ (equation 16), is within a threshold, $\bar{\epsilon}$, for $z = 0$ (solid lines) and $z = 1$ (dashed lines). F is averaged across the 10 realisations per training set size, N , increasing from 100 nodes (grey) to 600 (orange). The left panel corresponds to the 300 pseudo cosmology trials. The right panel illustrates the performance for the 300 pure ΛCDM trials (see Section 3.3.2).

and non-linear mode couplings for the various base ΛCDM parameters and shapes.

The right panel of Figure 7 shows similar trends for the ΛCDM trial ensemble. For the two largest training sets we find that over 90% of the trial cosmologies are reproduced with better than percent accuracy at both redshifts analysed. The larger emulation space dimensionality compared to similar emulator schemes designed for $w\text{CDM}$ cosmologies only (see, e.g. Heitmann et al. 2014) demands a three-fold increase in the number of nodes to achieve equivalent accuracies for ΛCDM cosmologies. In fact, for our smallest training set ($N = 100$) the emulator predictions can have errors as large as 4% (7%) at $z = 0$ ($z = 1$), comparable to those obtained by Heitmann et al. (2014). Although our emulation strategy requires a few hundred training nodes to reach percent accuracy, it is important to realise that it provides non-linear power spectrum predictions for a much broader class of cosmologies, including the ΛCDM model. Moreover, with the aid of *pairing and fixing* techniques (Angulo & Pontzen 2016; Euclid Collaboration et al. 2018) the total number of training simulations required is analogous to that of other GP emulators (Lawrence et al. 2010, 2017).

Finally we assess the accuracy of the emulator for the physical models in Table 1. Motivated by our results for the synthetic trial cosmologies, for this second test we use $N = 500$ training nodes, since a larger training set yields no significant improvement in the emulator performance.

The full range of the fractional errors of the emulated $P_{\text{NL}}^{\text{pseudo}}$ relative to the HALOFIT predictions, across the 10 training set realisations and all physical models, are shown in Figure 8 for $z = 0$ (upper panel) and $z = 1$ (lower panel). The outer (pale) and inner (dark) bands show the ranges including or excluding the xMNU_0.4 cosmology, respectively. Comparison of these reveals that much of

the observed error is associated with this extreme model, which resides outside of the central 75%-per-side hypervolume used in evaluating the accuracies with the synthetic pseudo cosmologies. We find that the remaining models are recovered to better than the target percent accuracy for $k < 10 h/\text{Mpc}$ at $z = 0$, with the majority of the training set realisations. We note that the range plotted, showing the total variation obtained across the training set realisations, presents a conservative estimate of the emulator performance; the 2σ range of *absolute* emulation accuracies, for example, is considerably tighter. We observe a slightly higher outlier rate at $z = 1$, consistent with our observations with the synthetic pseudo cosmologies. Of the models included in the dark band, we find, once again, that the largest inaccuracies occur for cosmologies near the boundary of the sub-volume used to test the synthetic shapes, MNU_0.4 and xMNU_0.2, both lying many standard deviations away from the *Planck* best fit cosmology (Planck Collaboration et al. 2018).

5 DISCUSSION AND CONCLUSIONS

In this paper, we have designed and tested a methodology essential for predicting accurate non-linear matter power spectra in a broad range of extensions to the ΛCDM cosmology in the context of the *halo model reaction* framework (Cataneo et al. 2019, ‘C19’). The effectiveness of the C19 strategy rests on the availability of accurate baseline ΛCDM non-linear power spectra evolved from non-standard initial conditions, the *pseudo* power spectra (see Section 2.1). We showed that such quantities can be readily predicted with a Gaussian process emulator whilst simultaneously satisfying the stringent accuracy requirements. We demonstrated the power of this technique for theoretically motivated models, such as $f(R)$ gravity and massive neutrino cosmologies, as well as for phe-

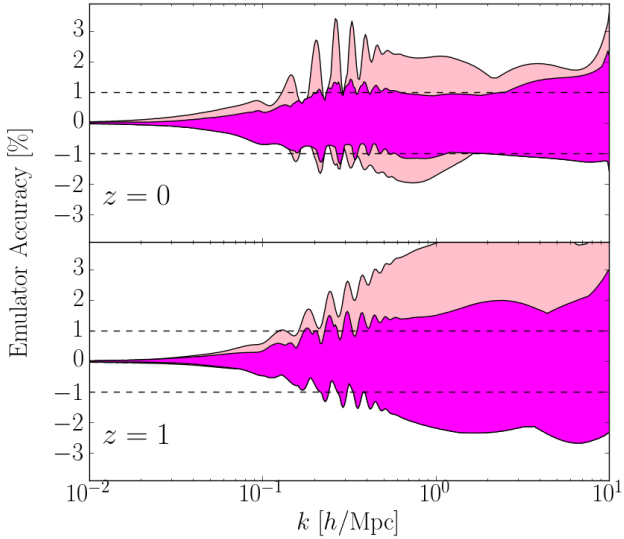


Figure 8. The full range of accuracies of the emulated pseudo matter power spectra for the physical models at $z = 0$ (upper panel) and $z = 1$ (lower panel), obtained with our 10 training sets of 500 nodes. The difference in the outer (pale) and inner (dark) band is whether the xMNU.0.4 cosmology is included in calculating the range of accuracies. Hence, much of the observed error bar can be attributed to this extreme model, which lies close to the edge of our training sets.

nomenological models, where in both cases structures on different scales grow at different rates.

Based on the results of this work, we advocate the following implementation of our methodology in the future:

- (i) A 13-dimensional optimised Latin hypercube is defined in terms of: π_Λ ($\dim = 5$), the cosmological parameters that define the linear Λ CDM power spectrum, and $\Delta\alpha$ ($\dim = 8$), the shape parameters which modify the linear Λ CDM power spectrum to create the linear power spectrum for a given beyond- Λ CDM model.
- (ii) Two standard Λ CDM N -body simulations, with initial conditions *paired* and *fixed* following Angulo & Pontzen (2016), are run for each node in the LH, i.e. with the input cosmology π_Λ , and rescaling of the primordial power spectrum with the shape defined by $\Delta\alpha$.
- (iii) From the output of each N -body simulation, the pseudo non-linear matter power spectrum, $P_{\text{NL}}^{\text{pseudo}}(k, z)$, and hence the natural logarithm of the boost factor, $\ln B(k, z)$ (see equation 7), are measured. An emulator is subsequently trained to predict this quantity for any $\{\pi_\Lambda, \Delta\alpha\}$ combination.
- (iv) A user will provide the analytically computed linear power spectrum in their chosen beyond- Λ CDM model, $P_L^{\text{real}}(k, z; \pi_\Lambda, \pi_X)$, where π_X is a vector of parameters describing the particular extension to the standard cosmology, and the corresponding vector of base cosmological parameters, π_Λ . We then perform a PCA decomposition of the ratio, $P_L^{\text{real}}(k, z; \pi_\Lambda, \pi_X)/P_L^\Lambda(k, z; \pi_\Lambda)$, where the denominator is the standard Λ CDM linear spectrum, to map π_X to the eight PCA weights (see equation 9).
- (v) Given the set of values of the base cosmological parameters and PCA weights for the queried model, the emulator predicts $\ln B(k, z)$.
- (vi) The pseudo non-linear matter power spectrum is then simply obtained as $P_{\text{NL}}^{\text{pseudo}}(k, z) = B(k, z) \times P_L^{\text{real}}(k, z)$.

- (vii) Finally, a rescaling of the pseudo non-linear power spectrum by the C19 halo model reaction, in equation 4, produces the *real* non-linear power spectrum including the effects of physics for the chosen beyond- Λ CDM model.

The advantage of our emulator scheme over existing approaches (Heitmann et al. 2014; Lawrence et al. 2017; Euclid Collaboration et al. 2018; Winther et al. 2019) is twofold: the model-independent parameterisation enables predictions for beyond-one-parameter extensions to Λ CDM, effectively expanding the domain of applicability to a much broader class of cosmologically interesting models; and the simulations required for the training phase are all based on the standard Λ CDM cosmology, which makes them easier and faster to run.

This work is a first step towards a proper and accurate non-linear matter power spectrum emulator, one with a training set built from the output of N -body simulations. Here, instead, our goal was to perform a feasibility study, in which the approximate semi-analytical predictions of HALOFIT act as surrogate simulations (replacing step (ii) above). Taking inspiration from the linear power spectrum shape of well-studied extensions to the standard cosmology, we computed the training pseudo non-linear power spectra by feeding smooth, random modifications of the Λ CDM linear power spectrum to HALOFIT. Despite its intrinsic 5-10% inaccuracies, previous studies showed that HALOFIT provides a quick and robust way to design, construct and test the emulator (Heitmann et al. 2009; Euclid Collaboration et al. 2018).

With the aid of *pairing and fixing* techniques for the initial conditions (Angulo & Pontzen 2016), we estimate that a total of 1000 Λ CDM-like simulations (two realisations for each of the 500 training nodes) is enough to build a per-cent-level accurate pseudo non-linear matter power spectrum emulator. This setup also allows for predictions of the Λ CDM non-linear matter power spectrum with $\lesssim 1\%$ accuracy, even with the standard cosmology being completely absent from the training set. Despite the remarkable amount of computational resources required, the number of simulations is not too dissimilar from that used for the training of previous, more limited emulators (Lawrence et al. 2010, 2017).

Although not shown here, our methodology can also potentially be extended to include the effects on baryonic physics on non-linear scales (see, e.g., Schneider et al. 2019; Debackere et al. 2019). This correction, which could be incorporated into the halo model reaction, is left for future investigation.

The results presented here should be regarded as conservative, as other emulation schemes, such as *sparse polynomial chaos expansion* (Blatman & Sudret 2011; Euclid Collaboration et al. 2018), and further design optimisation (Rogers et al. 2019; Caron et al. 2019) could help overcome the shortcomings of Gaussian process regression. Furthermore, we have only considered departures from Λ CDM on scales of $k > 10^{-3} h/\text{Mpc}$. This requirement could be relaxed by incorporating only one additional emulation parameter, $\Delta\bar{k}$, describing translations of the scale at which departures occur about this point. Since the scope of this preliminary study was to show the feasibility of pseudo cosmology emulation, these topics are also left to future work.

Our emulation method, coupled to the halo model reactions, enables future cosmic shear analyses by providing fast, accurate, and flexible predictions of the matter power spectrum deep in the non-linear regime, where we are more likely to find imprints of new physics beyond the concordance cosmology, if any (see, e.g., Heymans & Zhao 2018). Moreover, the approach developed in this work can be extended to predict other pseudo cosmology statistics,

for instance galaxy clustering (Zhai et al. 2019), as well as mean halo properties, such as their abundance and concentration (Kwan et al. 2013; McClintock et al. 2019).

6 ACKNOWLEDGEMENTS

BG, MC and CH acknowledge support from the European Research Council under grant number 647112. BM acknowledges the support of a Principal's Career Development Scholarship from the University of Edinburgh. MC thanks Joey Faulkner for stimulating discussions in the early stages of this work.

References

- Abbott T. M. C., et al., 2018, *Phys. Rev. D*, **98**, 043526
- Agarwal S., Abdalla F. B., Feldman H. A., Lahav O., Thomas S. A., 2012, *MNRAS*, **424**, 1409
- Alam S., Ho S., Silvestri A., 2016, *MNRAS*, **456**, 3743
- Angulo R. E., Pontzen A., 2016, *MNRAS*, **462**, L1
- Archidiacono M., Giusarma E., Hannestad S., Mena O., 2013, preprint, ([arXiv:1307.0637](https://arxiv.org/abs/1307.0637))
- Aviles A., Cervantes-Cota J. L., 2017, *Phys. Rev. D*, **96**, 123526
- Baumann D., Green D., Wallisch B., 2018, *J. Cosmology Astropart. Phys.*, **8**, 029
- Bird S., Viel M., Haehnelt M. G., 2012, *MNRAS*, **420**, 2551
- Blas D., Garny M., Konstandin T., Lesgourgues J., 2014, *Journal of Cosmology and Astro-Particle Physics*, 2014, 039
- Blatman G., Sudret B., 2011, *Journal of Computational Physics*, **230**, 2345
- Bose B., Koyama K., Lewandowski M., Vernizzi F., Winther H. A., 2018, *Journal of Cosmology and Astro-Particle Physics*, 2018, 063
- Brax P., Valageas P., 2013, *Phys. Rev. D*, **88**, 023527
- Brax P., Davis A.-C., Li B., Winther H. A., 2012, *Phys. Rev. D*, **86**, 044015
- Caron S., Heskies T., Otten S., Stienen B., 2019, preprint, ([arXiv:1905.08628](https://arxiv.org/abs/1905.08628))
- Cataneo M., et al., 2015, *Phys. Rev. D*, **92**, 044009
- Cataneo M., Lombriser L., Heymans C., Mead A. J., Barreira A., Bose S., Li B., 2019, *MNRAS*, **488**, 2121
- Copeland D., Taylor A., Hall A., 2018, *MNRAS*, **480**, 2247
- Currin C., Mitchell T., Morris M., Ylvisaker D., 1991, *Journal of the American Statistical Association*, **86**, 953
- Cyr-Racine F.-Y., Sigurdson K., Zavala J., Bringmann T., Vogelsberger M., Pfrommer C., 2016, *Phys. Rev. D*, **93**, 123527
- DES Collaboration et al., 2018, preprint, ([arXiv:1810.02499](https://arxiv.org/abs/1810.02499))
- Dakin J., Hannestad S., Tram T., 2019, preprint, ([arXiv:1904.11773](https://arxiv.org/abs/1904.11773))
- De Felice A., Tsujikawa S., 2010, *Journal of Cosmology and Astro-Particle Physics*, 2010, 024
- Debackere S. N. B., Schaye J., Hoekstra H., 2019, preprint, ([arXiv:1908.05765](https://arxiv.org/abs/1908.05765))
- Eifler T., Krause E., Dodelson S., Zentner A. R., Hearin A. P., Gnedin N. Y., 2015, *Monthly Notices of the Royal Astronomical Society*, **454**, 2451
- Eisenstein D. J., Hu W., 1998, *ApJ*, **496**, 605
- Euclid Collaboration et al., 2018, preprint, ([arXiv:1809.04695](https://arxiv.org/abs/1809.04695))
- Fixsen D. J., 2009, *ApJ*, **707**, 916
- Forero D. V., Tórtola M., Valle J. W. F., 2014, *Phys. Rev. D*, **90**, 093006
- Garrison L. H., Eisenstein D. J., Ferrer D., Tinker J. L., Pinto P. A., Weinberg D. H., 2018, *ApJS*, **236**, 43
- Gil-Marín H., et al., 2016, *MNRAS*, **460**, 4188
- Gil-Marín H., et al., 2018, *MNRAS*, **477**, 1604
- Habib S., Heitmann K., Higdon D., Nakhleh C., Williams B., 2007, *Phys. Rev. D*, **76**, 083503
- Harnois-Deraps J., Giblin B., Joachimi B., 2019, preprint, ([arXiv:1905.06454](https://arxiv.org/abs/1905.06454))
- Heisenberg L., Bartelmann M., 2019, preprint, ([arXiv:1901.01041](https://arxiv.org/abs/1901.01041))
- Heitmann K., Higdon D., Nakhleh C., Habib S., 2006, *ApJ*, **646**, L1
- Heitmann K., Higdon D., White M., Habib S., Williams B. J., Lawrence E., Wagner C., 2009, *ApJ*, **705**, 156
- Heitmann K., Lawrence E., Kwan J., Habib S., Higdon D., 2014, *ApJ*, **780**, 111
- Heitmann K., et al., 2016, *ApJ*, **820**, 108
- Heymans C., Zhao G.-B., 2018, *International Journal of Modern Physics D*, **27**, 1848005
- Hikage C., et al., 2018, preprint, ([arXiv:1809.09148](https://arxiv.org/abs/1809.09148))
- Hildebrandt H., et al., 2017, *MNRAS*, **465**, 1454
- Hildebrandt H., et al., 2018, preprint, ([arXiv:1812.06076](https://arxiv.org/abs/1812.06076))
- Hinterbichler K., Khoury J., Levy A., Matas A., 2011, *Phys. Rev. D*, **84**, 103521
- Hložek R., Marsh D. J. E., Grin D., Allison R., Dunkley J., Calabrese E., 2017, *Phys. Rev. D*, **95**, 123511
- Hojjati A., Pogosian L., Zhao G.-B., 2011, *Journal of Cosmology and Astro-Particle Physics*, 2011, 005
- Hojjati A., Zhao G.-B., Pogosian L., Silvestri A., Crittenden R., Koyama K., 2012, *Phys. Rev. D*, **85**, 043508
- Hu W., Sawicki I., 2007, *Phys. Rev. D*, **76**, 064004
- Hu W., Sugiyama N., 1996, *ApJ*, **471**, 542
- Hu W., White M., 1996, *ApJ*, **471**, 30
- Huterer D., et al., 2015, *Astroparticle Physics*, **63**, 23
- Joseph V. R., Gul E., Ba S., 2015, *Biometrika*, **102**, 371
- Joudaki S., et al., 2017, *MNRAS*, **471**, 1259
- Joudaki S., et al., 2018, *MNRAS*, **474**, 4894
- Joyce A., Jain B., Khoury J., Trodden M., 2015, *Phys. Rep.*, **568**, 1
- Joyce A., Lombriser L., Schmidt F., 2016, *Annual Review of Nuclear and Particle Science*, **66**, 95
- Kitching T. D., et al., 2014, *MNRAS*, **442**, 1326
- Kwan J., Bhattacharya S., Heitmann K., Habib S., 2013, *ApJ*, **768**, 123
- LSST Dark Energy Science Collaboration 2012, preprint, ([arXiv:1211.0310](https://arxiv.org/abs/1211.0310))
- Lattanzi M., Gerbino M., 2017, preprint, ([arXiv:1712.07109](https://arxiv.org/abs/1712.07109))
- Laureijs R., et al., 2011, preprint, ([arXiv:1110.3193](https://arxiv.org/abs/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**, 50
- Lesgourgues J., Pastor S., 2006, *Phys. Rep.*, **429**, 307
- Lesgourgues J., Matarrese S., Pietroni M., Riotto A., 2009, *Journal of Cosmology and Astro-Particle Physics*, 2009, 017
- Levi M., Vlah Z., 2016, preprint, ([arXiv:1605.09417](https://arxiv.org/abs/1605.09417))
- Levi M., et al., 2013, preprint, ([arXiv:1308.0847](https://arxiv.org/abs/1308.0847))
- Lewis A., Challinor A., Lasenby A., 2000, *ApJ*, **538**, 473
- Liu J., Petri A., Haiman Z., Hui L., Kratochvil J. M., May M., 2015, *Phys. Rev. D*, **91**, 063507
- Liu X., et al., 2016, *Physical Review Letters*, **117**, 051101
- Llinares C., 2018, *Int. J. Mod. Phys.*, **D27**, 1848003
- Lombriser L., 2014, *Annalen der Physik*, **264**, 259
- Mangano G., Miele G., Pastor S., Pinto T., Pisanti O., Serpico P. D., 2005, *Nuclear Physics B*, **729**, 221
- Marsh D. J. E., 2016, preprint, ([arXiv:1605.05973](https://arxiv.org/abs/1605.05973))
- Marsh D. J. E., Silk J., 2014, *MNRAS*, **437**, 2652
- Massara E., Villaescusa-Navarro F., Viel M., 2014, *Journal of Cosmology and Astro-Particle Physics*, 2014, 053
- McClintock T., et al., 2019, *ApJ*, **872**, 53
- McKay M. D., Beckman R. J., Conover W. J., 1979, *Technometrics*, **21**, 239
- Mead A. J., 2017, *MNRAS*, **464**, 1282
- Mead A. J., Peacock J. A., Heymans C., Joudaki S., Heavens A. F., 2015, *MNRAS*, **454**, 1958
- Moews B., de Souza R. S., Ishida E. E. O., Malz A. I., Heneka C., Vilalta R., Zuntz J., 2019, *Phys. Rev. D*, **99**, 123529
- Morris M.-D., Mitchell T. J., 1995, *J. Stat. Plan. Inference*, **43**, 381
- Mörtsell E., Dhawan S., 2018, *J. Cosmology Astropart. Phys.*, **9**, 025
- Nori M., Baldi M., 2018, *MNRAS*, **478**, 3935
- Park M., Zurek K. M., Watson S., 2010, *Phys. Rev. D*, **81**, 124008
- Parkinson D., et al., 2012, *Phys. Rev. D*, **86**, 103518
- Pedregosa F., et al., 2011, *Journal of Machine Learning Research*, **12**, 2825
- Planck Collaboration et al., 2018, preprint, ([arXiv:1807.06209](https://arxiv.org/abs/1807.06209))

- Pogosian L., Silvestri A., 2016, *Phys. Rev. D*, **94**, 104014
- Poulin V., Serpico P. D., Lesgourgues J., 2016, *Journal of Cosmology and Astro-Particle Physics*, 2016, 036
- Quiros I., García-Salcedo R., Gonzalez T., Horta-Rangel F. A., Saavedra J., 2016, *European Journal of Physics*, **37**, 055605
- Rasmussen C., Williams C., 2006, *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning, MIT Press, Cambridge, MA, USA
- Riess A. G., et al., 2016, *ApJ*, **826**, 56
- Riess A. G., et al., 2018, *ApJ*, **855**, 136
- Riess A. G., Casertano S., Yuan W., Macri L. M., Scolnic D., 2019, *ApJ*, **876**, 85
- Rogers K. K., Peiris H. V., Pontzen A., Bird S., Verde L., Font-Ribera A., 2019, *Journal of Cosmology and Astro-Particle Physics*, 2019, 031
- Sacks J., Welch W. J., Mitchell T. J., Wynn H. P., 1989, *Statist. Sci.*, **4**, 409
- Schneider A., 2015, *MNRAS*, **451**, 3117
- Schneider M. D., Knox L., Habib S., Heitmann K., Higdon D., Nakhleh C., 2008, *Phys. Rev. D*, **78**, 063529
- Schneider A., Teyssier R., Stadel J., Chisari N. E., Le Brun A. M. C., Amara A., Refregier A., 2019, *Journal of Cosmology and Astro-Particle Physics*, 2019, 020
- Senatore L., Zaldarriaga M., 2017, preprint, ([arXiv:1707.04698](https://arxiv.org/abs/1707.04698))
- Smith R. E., et al., 2003, *MNRAS*, **341**, 1311
- Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, *ApJ*, **761**, 152
- Taylor P. L., Kitching T. D., McEwen J. D., 2018, *Phys. Rev. D*, **98**, 043532
- Terukina A., Lombriser L., Yamamoto K., Bacon D., Koyama K., Nichol R. C., 2014, *Journal of Cosmology and Astroparticle Physics*, 2014, 013
- Thomas D. B., Kopp M., Markovič K., 2019, preprint, ([arXiv:1905.02739](https://arxiv.org/abs/1905.02739))
- Troxel M. A., et al., 2018, *Phys. Rev. D*, **98**, 043528
- Verde L., Protopapas P., Jimenez R., 2013, *Physics of the Dark Universe*, **2**, 166
- Wang J., Hui L., Khoury J., 2012, *Phys. Rev. Lett.*, **109**, 241301
- Winther H., Casas S., Baldi M., Koyama K., Li B., Lombriser L., Zhao G.-B., 2019, preprint, ([arXiv:1903.08798](https://arxiv.org/abs/1903.08798))
- Zhai Z., et al., 2019, *ApJ*, **874**, 95
- Zhao G.-B., 2014, *ApJS*, **211**, 23
- Zhao G.-B., Pogosian L., Silvestri A., Zylberberg J., 2009, *Phys. Rev. D*, **79**, 083513
- Zumalacárregui M., Bellini E., Sawicki I., Lesgourgues J., Ferreira P. G., 2017, *Journal of Cosmology and Astro-Particle Physics*, 2017, 019
- van Uitert E., et al., 2018, *MNRAS*, **476**, 4662

APPENDIX A: GAUSSIAN PROCESS REGRESSION METHOD

The mathematics behind Gaussian process (GP) regression emulators have been covered extensively in previous work; we refer the interested reader to [Rasmussen & Williams \(2006\)](#) for a general discussion of GP, and to [Habib et al. \(2007\)](#) and [Schneider et al. \(2008\)](#) for its first applications to cosmology. Here we summarise its key points, which we implement using the publicly available code SCIKIT-LEARN⁶ ([Pedregosa et al. 2011](#)).

GP regression is a non-parametric Bayesian machine learning algorithm for constraining the distribution of functions consistent with observed data. Let us consider a data set, \mathcal{D} , consisting of n realisations of an observable, \mathbf{y} , each corresponding to different values of the d -dimensional input parameters, $\boldsymbol{\pi} \in \mathbb{R}^d$, i.e.

$\mathcal{D} = \{(\boldsymbol{\pi}_i, y_i) | i = 1, \dots, n\}$. The ensemble of matter power spectra measured from a suite of N -body simulations is a typical example of such a data set, and constitutes a training set for the emulator. The task of the GP emulator is to learn the distribution of functions, $f(\boldsymbol{\pi})$, which are consistent with the mapping between the training set input parameters - the ‘nodes’ - and output, via

$$y(\boldsymbol{\pi}) = f(\boldsymbol{\pi}) + \epsilon_n(\boldsymbol{\pi}), \quad (\text{A1})$$

where $\epsilon_n(\boldsymbol{\pi})$ is a noise term sampled from a mean-zero Gaussian distribution with a standard deviation given by the error on the training set observable, $y(\boldsymbol{\pi})$. The prediction, y^* , corresponding to an arbitrary coordinate $\boldsymbol{\pi}^*$, is then sampled from a generalisation of a Gaussian posterior probability distribution over the range of consistent functions⁷. In other words, the GP emulator interpolates the observable from the input coordinates of the training set to the trial coordinates across a d -dimensional parameter space.

A key ingredient of our posterior is the Gaussian prior distribution of functions deemed to reasonably map between input and output. The prior is determined by a mean, conventionally taken to be zero, and a covariance function, known as the ‘kernel’. The kernel can take various functional forms, each described by a vector of hyperparameters, $\mathbf{h} = \{A, p_1, \dots, p_d\}$, governing the kernel’s behaviour. Following [Heitmann et al. \(2009\)](#), in this work we adopt the squared exponential form, which specifies the covariance between the functions $f(\boldsymbol{\pi})$ and $f(\boldsymbol{\pi}^*)$ as

$$K(f, f^*; \mathbf{h}) \equiv \text{cov}(f(\boldsymbol{\pi}), f(\boldsymbol{\pi}^*); \mathbf{h}) = A \prod_{l=1}^d \exp \left[-\frac{(\pi_l - \pi_l^*)^2}{p_l^2} \right]. \quad (\text{A2})$$

This kernel has the following properties: (1) the covariance varies smoothly within the parameter space; (2) it depends only on the Euclidean distance between points, such that $K(f, f^*; \mathbf{h}) = K(f^*, f; \mathbf{h})$; (3) predictions become maximally correlated when $\boldsymbol{\pi} = \boldsymbol{\pi}^*$; (4) the correlation is large for points in relative proximity and small for largely separated points; (5) each p_l corresponds to the functions’ characteristic length-scale of variation in each of the d dimensions, whilst A is the kernel amplitude.

The emulator is trained by finding values for the hyperparameters such that a distribution of functions which are optimally consistent with all realisations in the training set can be defined. In this work, we fit for these using the method built-in to SCIKIT-LEARN, which employs a gradient ascent optimisation of the marginal likelihood conditioned on the training set.

APPENDIX B: MODELLING BAO RESIDUALS

As shown in [Figure 1](#), models with massive neutrinos have shapes characterised by a rapidly oscillating component superimposed over a smooth function. These wiggles are associated with changes in the baryon acoustic oscillations (BAO) produced by differences in the sound horizon at the end of the drag epoch and shallower gravitational potentials before recombination ([Hu & Sugiyama 1996](#); [Hu & White 1996](#); [Eisenstein & Hu 1998](#); [Lattanzi &](#)

⁷ In this work we consider HALOFIT training predictions as a proxy for hypothetical numerical simulations with negligible error. Nevertheless, to prevent unwanted numerical instabilities we use the arbitrarily small constant default ϵ_n in SCIKIT-LEARN for all $y(\boldsymbol{\pi})$.

⁶ <https://scikit-learn.org>

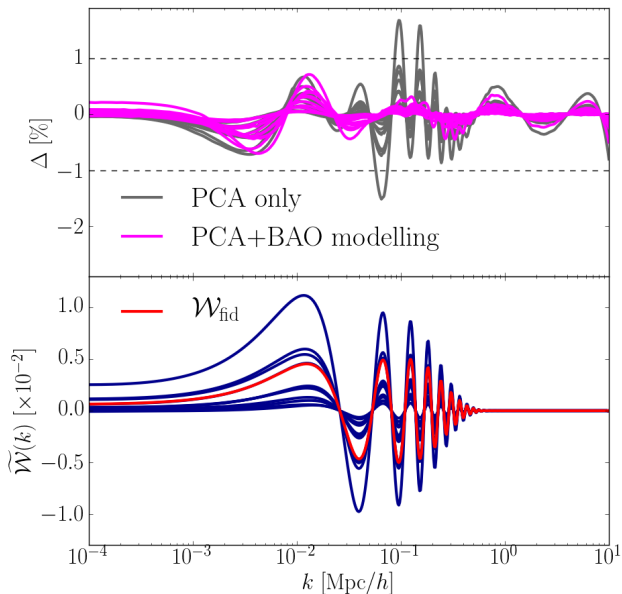


Figure B1. *Upper panel:* the accuracy of the shape reconstruction, Δ , for the physical models featuring massive neutrinos (see Table 1). The grey curves show the results for direct PCA reconstruction using the 8 orthogonal basis functions derived in Section 3.2.1. The magenta curves, reproduced from Figure 3, illustrate the benefit of including the BAO residuals reconstruction via the two-step process in equation B4: the BAO residuals are first extracted with the de-wiggling algorithm presented in Appendix C and then modelled using the template (see equation B3), whilst the PCA reconstruction is performed on the smooth component only. *Lower panel:* the red curve shows the BAO residual associated with the MNU_0.4 cosmology, which we use as the fiducial template. The blue curves show the best fit functions, $\tilde{\mathcal{W}}(k; \bar{a}, \bar{b})$, to the oscillatory component of the shapes for the remaining physical models with massive neutrinos. All curves here correspond to $z = 0$; results are very similar at $z = 1$.

Gerbino 2017) compared to a massless neutrino Λ CDM cosmology with the same standard parameters, i.e. $\pi_{\Lambda}^{m_{\nu}=0} = \pi_{\Lambda}^{m_{\nu} \neq 0}$, and $\Omega_c^{m_{\nu}=0} = \Omega_c^{m_{\nu} \neq 0} + \Omega_{\nu}^{m_{\nu} \neq 0}$. Depending on the value of the standard cosmological parameters and neutrino mass, the amplitude of the BAO residual in the shape can be up to $\sim 1\%$ of the slowly varying component. The PCA reconstruction outlined in Section 3.1 requires a large number of components to simultaneously capture long- and short-scale variations. Hence we find errors $\gtrsim 1\%$ in reconstructing the $\sum m_{\nu}$ -induced oscillations with only $n_{\Phi} = 8$ basis functions, as shown by the grey curves in the upper panel of Figure B1.

In order to improve the reconstruction accuracy of the PCA, we isolate and remove the BAO residuals by applying the de-wiggling algorithm presented in Appendix C. In short, the BAO in Fourier space maps to a localised bump in the discrete sine transform of the power spectrum. By removing this bump and performing the inverse transform, we obtain a smooth, “de-wiggled” version of the power spectrum. This is performed on both the model featuring massive neutrinos and on the Λ CDM baseline, before calculating the de-wiggled shape given by

$$S^{\text{dw}}(k, z) = \frac{P_{\text{L}}^{\text{pseudo,dw}}(k, z)}{P_{\text{L}}^{\Lambda, \text{dw}}(k, z)}, \quad (\text{B1})$$

where the superscript ‘dw’ on the quantities on the right-hand side denote power spectra obtained from applying the de-wiggling algorithm. This smoothed component of the shape is captured very accurately by the basis functions. Reconstruction of the *full* shape then rests on modelling the remaining oscillatory component, given by

$$\mathcal{W}(k, z) = S(k, z) - S^{\text{dw}}(k, z). \quad (\text{B2})$$

Notice that all quantities in equations B1 and B2 depend implicitly on the cosmological parameters. We experimented with fitting these oscillations with the function

$$\tilde{\mathcal{W}}(k; a, b) = [a + b \log_{10}(k)] \mathcal{W}_{\text{fid}}(k), \quad (\text{B3})$$

where a and b are free parameters controlling the amplitude and modulation of a fiducial wiggle template, \mathcal{W}_{fid} , which we identify with the wiggle contribution of the MNU_0.4 cosmology in Table 1, represented by the red curve in the lower panel of Figure B1. We checked that our fiducial template choice in equation B3 is robust against changes in redshift, standard cosmological parameters, sum of neutrino masses and neutrino mass eigenstates. This ensures that the parameters $\{a, b\}$ are sufficient to capture BAO variations in the family of pseudo cosmologies treated in this work. The blue curves in the lower panel of Figure B1 show the behaviour of our fitting formula, Eq. (B3), for all of the test cosmologies in Table 1 containing massive neutrinos.

In summary, for each redshift and cosmology separately, we reconstructed the total shapes of the physical models as

$$S(k; \pi_{\Lambda}, \Delta\alpha, a, b) = S^{\text{dw}}(k; \pi_{\Lambda}, \Delta\alpha) + \tilde{\mathcal{W}}(k; a, b), \quad (\text{B4})$$

where only the smooth, de-wiggled term is modelled via the PCA reconstruction given in equation 9. The magenta curves in the upper panel of Figure B1 show the overall accuracy of the two-step reconstruction, originally presented in the lower panel of Figure 3. We find that the inaccuracies in the BAO regime are greatly suppressed compared to performing the PCA directly on the full shape (grey curves).

The fact that the cosmologies with massive neutrinos require two extra parameters to summarise their shapes to sub-per-cent accuracy, might imply that the emulator need be informed of $\{a, b\}$ to achieve this target accuracy in its predictions, thereby increasing the dimensionality from 13 to 15. To test this, we produced MAXPRO training sets optimised in 15 dimensions, with a range of wiggle parameters, a and b , slightly broader than those obtained for the physical models with massive neutrinos. We then compared the emulator’s performance with the physical models, when trained on these 15-dimensional sets, and when trained on the same sets but in 13-dimensions, omitting a and b . We found, in fact, that the emulator is mostly insensitive to these two parameters, which therefore contribute only marginally to improving its accuracy.

Figure B2 holds the key to this somewhat surprising result, where we adopt the xMNU_0.4 cosmology as a worst case scenario among our physical models. Plotted in solid orange is the S/S^{dw} ratio, which is equivalent to the ratio of pseudo linear power spectra with or without changes to the BAO scale compared to Λ CDM (see equation B1). In dashed dark blue is the ratio of their non-linear counterparts obtained with HALOFIT. Solid magenta shows the ratio between the boost factors derived from using the full or de-wiggled shape, which is also equal to the ratio of the dark blue and orange curves. We see that although the BAO residual amounts to $\sim 1\%$ of the non-linear power spectrum, its contribution to the

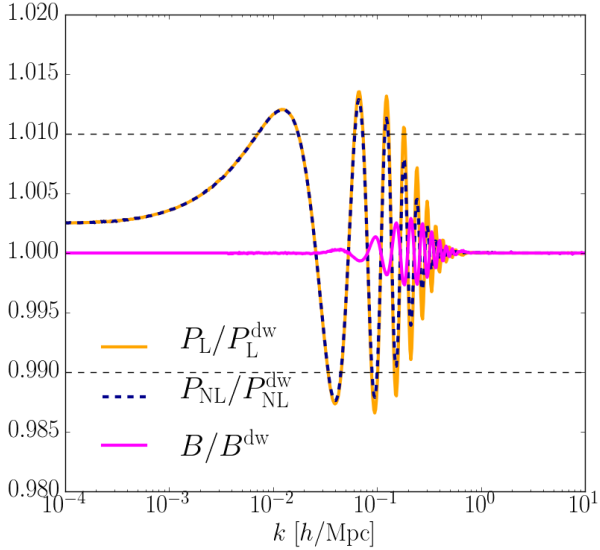


Figure B2. Full to de-wiggled ratios for the boost factor, linear and non-linear matter power spectra of xMNU_0.4 in Table 1. Quantities in the legend are defined as follows: $P_L = S \times P_L^\Lambda$; $P_L^{\text{dw}} = S^{\text{dw}} \times P_L^\Lambda$; $P_{\text{NL}} = \text{HF}[P_L]$; $P_{\text{NL}}^{\text{dw}} = \text{HF}[P_L^{\text{dw}}]$. The de-wiggled shape, S^{dw} , is obtained as in equation B1, while HF is shorthand for HALOFIT. The solid orange and dashed blue lines represent the linear and non-linear BAO residuals, respectively, whose ratio is equivalent to the ratio of the boost factors, $B = P_{\text{NL}}/P_L$ and $B^{\text{dw}} = P_{\text{NL}}^{\text{dw}}/P_L^{\text{dw}}$ (solid magenta). The contribution of the BAO residual to the non-linear power spectrum exceeds 1%, but is at most $\simeq 0.3\%$ in the boost, which implies we can safely ignore the wiggle parameters, $\{a, b\}$, in emulation.

boost factor is much smaller. In turn, this means that the emulator can reconstruct the boost factor with the required accuracy irrespective of whether the information on the BAO residual is provided or not. In other words, the boost factor effectively damps down the correction associated with changes to the BAO caused by massive neutrinos, eliminating the need for the additional wiggle modelling (equation B3)⁸.

The fact that early-time modifications to the BAO physics have such a weak effect on the boost factor is a strength of our emulation scheme; an emulator designed to directly predict the pseudo non-linear power spectra is likely to incur $\gtrsim 1\%$ errors unless detailed modelling of the BAO residual is employed. The unimportance of the wiggle parameters means that the 13-dimensional emulator is sufficient to achieve our target accuracy. In Appendix D, we investigate the performance of the emulator with cosmologies featuring extra relativistic degrees of freedom, for which changes in the acoustic oscillations prior to the epoch of last scattering leave especially prominent wiggles in the late-time shape ratios. We find the emulator achieves sub-per-cent accuracy at $z = 0$, but less consistent results at higher redshift for this model.

Although modelling of the BAO residuals proves to be unnecessary, their extraction through the de-wiggling algorithm remains important. By acting as a low-pass filter, the PCA reconstruction can in principle remove rapid oscillations and capture the broad-

⁸ Although we verified this statement with HALOFIT we expect it to remain valid in simulations as well, assuming that HALOFIT provides accurate predictions for the ratio $P_{\text{NL}}/P_{\text{NL}}^{\text{dw}}$.

band shape. However, the PC weights thus derived are somewhat different from those obtained after de-wiggling, generating an error that ultimately propagates to the boost factor. We compare the emulator performance with and without de-wiggling in Appendix E.

APPENDIX C: DE-WIGGLING METHODOLOGY

Here we outline in more detail the method, discussed briefly in Appendix B, for de-wiggling linear matter power spectra following Baumann et al. (2018). This procedure is performed on the pseudo and Λ CDM spectra separately, and therefore we denote both with $P_L(k)$ in the following:

- (i) We firstly interpolate the linear matter power spectra featuring BAO wiggles, $P_L(k)$, from the logarithmically-spaced k values onto a linear-spaced array, k_n , sampled at 2^n points, where n is an integer. This is so as to improve the computational efficiency of the Fast Discrete Sine Transform (DST), which is then performed on $\log_{10}[k_n P_L(k_n)]$ using the orthonormalised type-II sine transform. The resulting array, which has indices denoted by i_{dst} and length 2^n , is split into two separate arrays, one with the even i_{dst} indices and the other with odd i_{dst} indices.
- (ii) The DST of the even and odd arrays each feature a bump, which is a localised manifestation of the BAO in the k -space of the matter power spectrum. This is shown by the solid lines for one of the physical models in the upper panel of Figure C1. In order to de-wiggle the power spectrum, we must remove these bumps. To this end, we take the second derivative of the DST, converting the observed bump in each array into a prominent single-period oscillation, as shown by the lower panel of Figure C1. We then identify $i_{\text{dst}}^{\text{min}}$ and $i_{\text{dst}}^{\text{max}}$ values bracketing the oscillation, which are close to the local minimum on the left-hand side of the approximate centre of this feature and the local maximum on the right-hand side respectively.
- (iii) We remove the corresponding i_{dst} -range from the even and odd DST arrays, multiply each by $(i_{\text{dst}} + 1)^2$ and interpolate across the gap between $i_{\text{dst}}^{\text{min}}$ and $i_{\text{dst}}^{\text{max}}$ with cubic splines. The $(i_{\text{dst}} + 1)^2$ factors are then divided out, generating smooth “de-bumped” even and odd DST arrays, where the information corresponding to the BAO has been removed. This is shown by the dashed lines in the upper panel of Figure C1.
- (iv) We recombine the even and odd DST arrays into a single array. Performing an inverse type-II DST produces the logarithmic de-wiggled linear matter power spectra weighted by the linear-spaced k_n array, $\log_{10}[k_n P_L^{\text{dw}}(k_n)]$. From this we obtain the de-wiggled linear spectra itself, $P_L^{\text{dw}}(k_n)$.
- (v) After performing this on both the pseudo and Λ CDM power spectra, we take the ratio of the respective $P_L^{\text{dw}}(k_n)$ to obtain $S^{\text{dw}}(k_n)$, the de-wiggled shape sampled at the linearly-spaced wavenumbers. Finally we interpolate the de-wiggled shape back to the logarithmically-spaced k sampling, obtaining $S^{\text{dw}}(k)$. We find that there is less numerical noise at the upper and lower wavenumber bounds of the de-wiggled shape when we take the ratio of the linearly-sampled de-wiggled power spectra before interpolating, rather than the other way around.

Whereas Baumann et al. (2018) chose fixed values of the $i_{\text{dst}}^{\text{min}}$ and $i_{\text{dst}}^{\text{max}}$ relative to the oscillation in the second derivative of the DST arrays, we optimise for these parameters in our analysis with the physical models. Specifically, we cycle through a grid

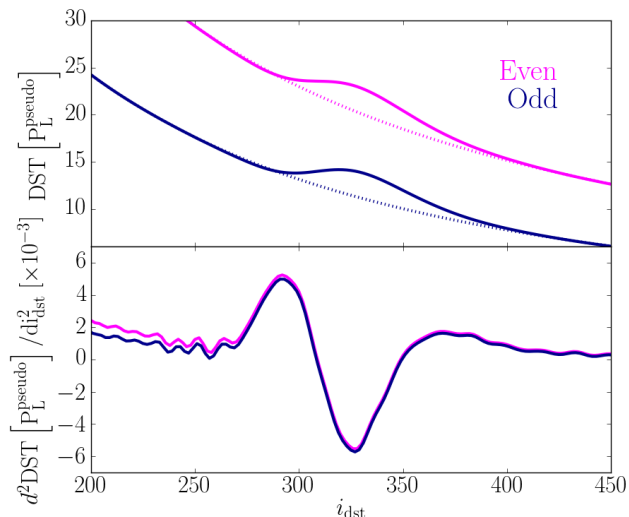


Figure C1. *Upper panel:* The discrete sine transform (DST) of the pseudo linear matter power spectrum for the F4-MNU.0.4 model, split into two arrays with even (magenta) and odd (dark blue) indices, i_{dst} , before (solid lines) and after (dashed lines) the bump is removed. The bump is a localised manifestation of the BAO wiggles present in the power spectrum. *Lower panel:* The second derivative of the DST used to identify the $i_{\text{dst}}^{\text{min}}$ and $i_{\text{dst}}^{\text{max}}$ values which bracket the bump in the DST. We adjust the values of these for each model, optimising the smoothness of the returned de-wiggled shape. The final $i_{\text{dst}}^{\text{min}}$ and $i_{\text{dst}}^{\text{max}}$ values are close to, respectively, the local minimum on the left-hand side of the centre of the prominent oscillation in the second-derivative, and the local maximum on the right-hand side.

of $(i_{\text{dst}}^{\text{min}}, i_{\text{dst}}^{\text{max}})$ combinations which are centred on the local minimum and maximum on either side of the oscillation, and identify those which return the smoothest de-wiggled shape. The smoothness is measured by integrating the second derivative of $S^{\text{dw}}(k)$, convolved with a Gaussian filter, in the range of the BAO wiggles. The optimal $(i_{\text{dst}}^{\text{min}}, i_{\text{dst}}^{\text{max}})$ combination is that which minimises this integral. We find the de-wiggled shapes are insensitive to the width of the Gaussian filter, which is employed so as to ensure numerical noise does not bias identification of the optimal $(i_{\text{dst}}^{\text{min}}, i_{\text{dst}}^{\text{max}})$ range.

APPENDIX D: EXTENSION TO EXTRA RELATIVISTIC DEGREES OF FREEDOM

Appendix B demonstrated the unimportance of modelling the BAO residuals in the shapes of the physical cosmologies, justifying the 13-dimensional emulation scheme. In order to further verify the validity of this, we test our emulator scheme against a theoretically motivated cosmology with an especially prominent BAO residuals in the shape ratio.

The existence of a relic sea of neutrinos with effective relativistic degrees of freedom, $N_{\text{eff}} = 3.046$, is a general prediction of the standard model (Mangano et al. 2005). However, physics beyond this paradigm includes scenarios with additional relativistic particles (or ‘dark radiation’) at the epoch of decoupling, typically quantified by the variation ΔN_{eff} , such that

$$N_{\text{eff}} = 3.046 + \Delta N_{\text{eff}}. \quad (\text{D1})$$

This has the effect of reducing the expansion rate and increasing the acoustic scale (Archidiacono et al. 2013), thereby acting to

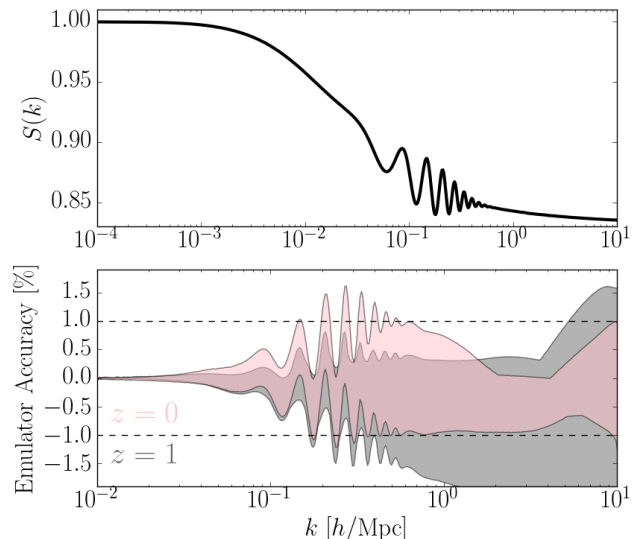


Figure D1. *Upper:* the $z = 0$ shape ratio for the model with $\Delta N_{\text{eff}} = 1$ extra relativistic degrees of freedom. The BAO residual is at the level of $\approx 2\%$, about twice as large as the residuals for physical models with massive neutrinos, shown in Figure 1. *Lower:* the full range of emulation accuracies achieved with the 10 realisations of the $N = 500$ training set at $z = 0$ (pink) and $z = 1$ (grey). Despite the larger prominence of the BAO residual to the shape, we find its contribution to the boost factor remains at the sub-per-cent level, facilitating reconstruction of the $z = 0$ pseudo non-linear power spectrum to better than the target accuracy with almost all of the training set realisations. The larger inaccuracies observed at $z = 1$ suggest that this model cannot reliably be constructed at higher redshifts with the current emulation scheme.

ease the tension between the early- and late-time measurements of the Hubble constant (Riess et al. 2019). We consider a model with $\Delta N_{\text{eff}} = 1$ extra relativistic degrees of freedom, which lies several standard deviations away from the best-fit result of the Planck Collaboration et al. (2018) analysis. The BAO residuals in the shape of this cosmology at $z = 0$ are shown in the upper panel of Figure D1.

Following the same procedure employed for the physical models in Table 1, we obtain a smoothed version of this shape using the de-wiggling algorithm (see Appendix C), perform a PCA decomposition to obtain the PCA weights, $\Delta\alpha$, but refrain from modelling the BAO residual component as described in Appendix B. We then infer the $P_{\text{NL}}^{\text{pseudo}}$ for this model using the emulator in the fiducial 13-dimensional setup and compare to the prediction from HALOFIT.

The results using the 500-node training sets are shown in the lower panel of Figure D1. We find that despite the BAO residual in the shape being as large as 2% for this cosmology⁹, the emulator reproduces the $P_{\text{NL}}^{\text{pseudo}}$ to better than 1% accuracy at $z = 0$ (pink) in almost all of the training set realisations. The reason that these accuracies are achievable, as with the physical models, is due to the boost factor suppressing the early-time modifications of the BAO to sub-per-cent levels. We also note that changes in the linear matter power spectrum for models with $\Delta N_{\text{eff}} \neq 0$ (see Figure D1) are

⁹ For comparison, this feature never exceeds 1% in our other test cosmologies.

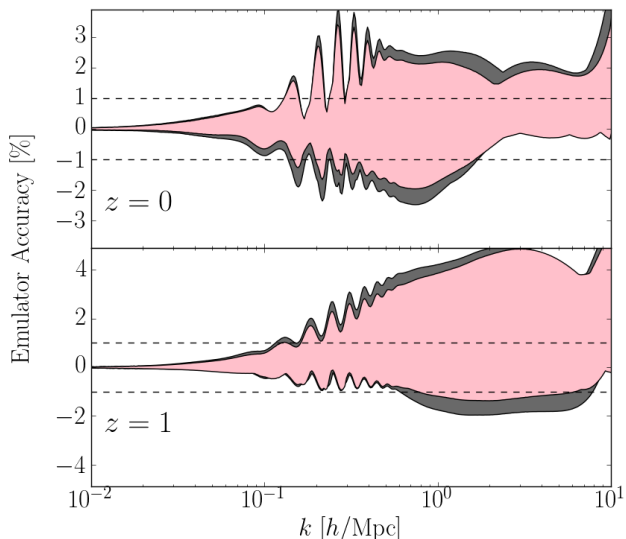


Figure E1. The range of accuracies of the emulated pseudo matter power spectra for the xMNU.0.4 cosmology (see Table 1) at $z = 0$ (upper panel) and $z = 1$ (lower panel), obtained with our 10 training sets of 500 nodes. The inner (pink) band corresponds to our fiducial emulation method, reproduced from Figure 8. The outer (grey) band corresponds to the case with no de-wiggling and PCA decomposition performed directly on the full shape.

caused by processes happening well before the matter dominated epoch. At redshifts $\gtrsim 100$ the background and growth evolution in these extensions are indistinguishable from those in Λ CDM. Therefore, the *full* non-linear matter power is entirely captured by the pseudo cosmology (i.e. the reaction is unity for all scales, see Section 2.1). At $z = 1$, however, we find that this model cannot be reliably predicted to within the target accuracy across the training set realisations. This is perhaps an indication that models with such marked BAO residuals are beyond the capabilities of our emulator as specified in this work. This may be improved upon given an alternative emulation method, such as *sparse polynomial chaos expansion* (Blatman & Sudret 2011; Euclid Collaboration et al. 2018), or with careful optimisation of the training set configuration (Rogers et al. 2019; Caron et al. 2019).

APPENDIX E: EMULATION WITHOUT DE-WIGGLING

Although modelling of the BAO residuals in the shapes to obtain the wiggle parameters, a and b , has negligible effect on the results (see Appendix B), we find that shape de-wiggling remains important for improving emulation accuracy. This is because the PCA is a low-pass filter, effective at capturing only the low-frequency component of the shapes. This property is imparted by the basis set, Φ_i , derived from smooth curves (see Section 3.2.1), such that the corresponding weights, $\Delta\alpha_i$, are designed to contain smooth component information. Consequently, performing the PCA directly on the full shape leads to misestimation of the $\Delta\alpha_i$ values, and bias in the emulation.

Figure E1 presents the effect of this for the xMNU.0.4 cosmology (see Table 1) at $z = 0$ (upper panel) and $z = 1$ (lower panel). The pink band shows the full range of emulation accuracies achieved with the 10 realisations of the 500-node training set for

our fiducial emulation method incorporating the de-wiggling, reproduced from Figure 8. The grey band shows this quantity for the case with no de-wiggling, where the PCA decomposition is performed directly on the full shape. Indeed, excess inaccuracies of $\sim 0.5\%$ are observed in the absence of de-wiggling, for the offset PCA weights map to a boost factor related to a slightly different shape compared to the HALOFIT predictions. Although currently subdominant, these errors will become more relevant for future emulators designed to have target performance of $\sim 0.5\%$, a requirement imposed by irreducible inaccuracies in the halo model reactions.