Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **515,** 4161–4175 (2022) https://doi.org/10.1093/mnras/stac1091

FORGE: the $f(R)$-gravity cosmic emulator project – I. Introduction and matter power spectrum emulator

Christian Arnold [1]★ Baojiu Li [1] Benjamin Giblin [2] Joachim Harnois-Déraps[2,3,4] and Yan-Chuan Cai[2]

[1] *Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*
[2] *Scottish Universities Physics Alliance, Institute for Astronomy, University of Edinburgh, Blackford Hill, Edinburgh, EH9 3JH, UK*
[3] *School of Mathematics, Statistics and Physics, Newcastle University, Herschel Building, Newcastle-upon-Tyne NE1 7RU, UK*
[4] *Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK*

Accepted 2022 March 28. Received 2022 March 28; in original form 2021 September 10

ABSTRACT

We present a large suite of cosmological simulations, the FORGE (F-of-R Gravity Emulator) simulation suite, which is designed to build accurate emulators for cosmological observables in galaxy clustering, weak gravitational lensing, and galaxy clusters for the $f(R)$-gravity model. A total of 200 simulations explore the cosmological parameter space around a standard Planck cosmology with a Latin hypercube, for 50 combinations of $\bar{f}_{R0}$, $\Omega_{\rm m}$, σ_8, and h with all other parameters fixed. For each parameter combination, or node, we ran four independent simulations, one pair using 1024^3 particles in $500\,h^{-1}$ Mpc simulation boxes to cover small scales, and another pair using 512^3 simulation particles in $1.5\,h^{-1}$ Gpc boxes for larger scales. Each pair of initial conditions is selected such that sample variance on large scales is minimized on average. In this work we present an accurate emulator for the matter power spectrum in $f(R)$ gravity trained on FORGE. We have verified, using the cross-validation technique, that the emulator accuracy is better than 2.5 per cent for the majority of nodes, particularly around the centre of the explored parameter space, up to scales of $k = 10\,h\,{\rm Mpc}^{-1}$. We have also checked the power spectrum emulator against simulations that are not part of our training set and found excellent agreement. Due to its high accuracy on small scales, the FORGE matter power spectrum emulator is well suited for weak-lensing analysis and can play a key tool in constraining $f(R)$ gravity using current and future observational data.

Key words: methods: numerical – cosmology: theory.

1 INTRODUCTION

The origin of the accelerated Hubble expansion at late times – the cosmic acceleration – has been one of the most challenging questions in modern cosmology since its discovery over two decades ago (Riess et al. 1998; Perlmutter et al. 1999). A huge body of research works have been conducted ever since, both in the development of theoretical models, such as dark energy (Copeland, Sami & Tsujikawa 2006; Mortonson, Weinberg & White 2013) and modified gravity (Clifton et al. 2012; Joyce et al. 2015; Koyama 2018), that can explain this phenomenon, and in observational efforts to test or constrain these models (see e.g. Albrecht et al. 2006; Weinberg et al. 2013, for general introductions of the various observational probes). While the standard Λ cold dark matter (ΛCDM) model, where the cosmic acceleration is assumed to be driven by a positive cosmological constant, Λ, has been widely accepted as the working model in the field, the profound implications of the discovery – that it could be a signature of new physics in either the particle or the gravity sector – have motivated extensive searches and investigations of other possibilities. Of these, modified gravity models have attracted a lot of attention in recent years, partly thanks to their potential to be used as testbeds to check the validity of general relativity (GR) on length scales relevant for cosmology (e.g. Koyama 2016; Ferreira 2019; Baker et al. 2021).

Evidence of the cosmic acceleration has been supported by various independent cosmological observables/probes, in addition to the luminosity distance measurements with type Ia supernovae. These include the cosmic microwave background (CMB; e.g. Hinshaw et al. 2013; Hou et al. 2014; Planck Collaboration et al. VI 2020; Aiola et al. 2020), baryonic acoustic oscillations (BAO; e.g. Cole et al. 2005; Eisenstein et al. 2005; Beutler et al. 2011; Blake et al. 2011b; Anderson et al. 2012; Alam et al. 2021), weak gravitational lensing (e.g. Heymans et al. 2013; Hamana et al. 2020; Asgari et al. 2021; Secco et al. 2021; Amon et al. 2021), strong gravitational lensing (e.g. Jullo et al. 2010), galaxy clusters (e.g. Vikhlinin et al. 2009; Ade et al. 2014; Mantz et al. 2014, 2015; De Haan et al. 2016; Bocquet et al. 2019; Abbott et al. 2020; Giocoli et al. 2021), and the clustering of galaxies (e.g. Percival et al. 2004; Guzzo et al. 2008; Blake et al. 2011a; Beutler et al. 2012; Alam et al. 2017; Pezzotta et al. 2017; Zarrouk et al. 2018). In the coming years, several much larger galaxy surveys, such as DESI (Aghamousa et al. 2016), *Euclid* (Laureijs et al. 2011) and the Vera C. Rubin Observatory (LSST Science Collaboration et al. 2009) will be able to improve the current observational status by mapping the distribution of matter in the Universe with billions of galaxies.

Given the significant improvement in statistical precision expected from these upcoming observatories, it becomes crucial that the

★ E-mail: christian.arnold@durham.ac.uk

accuracy of the prediction of many large-scale structure (LSS) observables reaches the per cent level. A well established approach to achieving this is by the use of cosmological *N*-body or hydrodynamical simulations, which are able to track the evolution of cosmic structures into the small-scale highly non-linear regime, where the multiple crossings of particle trajectories make it impossible to use linear perturbation theory to obtain reliable predictions. Although this approach is accurate and, with the rapid advancements of supercomputing power and generations of new simulation codes, has led to spectacular achievements in recent years (e.g. Angulo et al. 2012; Schaye et al. 2015; Potter, Stadel & Teyssier 2017; Pillepich et al. 2018; Nelson et al. 2019), it often comes at a heavy computational cost and is therefore not well suited for sampling a large suite of models. More practical approaches often use lower-resolution simulations, sparse cosmology sampling, and approximations to the gravity solver (see e.g. Angulo & White 2010; Tassev, Zaldarriaga & Eisenstein 2013; Monaco et al. 2013; Howlett, Manera & Percival 2015; Feng et al. 2016; Klypin & Prada 2018).

The gravitational force in *N*-body simulations is computed from the Poisson equation, which can be solved using various algorithms such as 'tree codes' as in GADGET (Springel et al. 2005) and AREPO (Springel 2010), multigrid relaxation as in RAMSES (Teyssier 2002) or multigrid fast Fourier transform (FFT) such as in CUBEP^{3}M (Harnois-Déraps et al. 2013). However, in models where the theory of gravity differs from GR, the equation that describes gravity often becomes non-linear in the matter field. This adds to the complexity of the problem and therefore accurate numerical solutions typically come at a significantly higher computational cost. This makes the situation here comparatively worse than that described above for the ΛCDM model. There have been numerous codes developed to simulate the different modified gravity models studied in the literature (e.g. Oyaizu 2008; Schmidt 2009; Chan & Scoccimarro 2009; Li & Zhao 2009; Zhao, Li & Koyama 2011; Li et al. 2012; Puchwein, Baldi & Springel 2013; Li, Zhao & Koyama 2013; Llinares, Mota & Winther 2014; Arnold, Leo & Li 2019a; Hernández-Aguayo et al. 2021), thanks to which we have been able to gain qualitative and initial quantitative insights into the structure formation in these models. However, due to the impeded efficiency of modified gravity simulation codes, so far these simulations mostly suffer from two major limitations: the first is that their sizes (in terms of the box size, particle number, and resolution) are significantly smaller than state-of-the-art simulations of ΛCDM (see e.g. Arnold et al. 2019b, for several of the largest simulations of modified gravity models to date). The second is that the variety of models having been simulated is rather limited – typically one fixes the cosmological parameters and varies only one (or a few) modified gravity parameters to investigate the non-linear effect on the large-scale structure. However, not only are there potential degeneracies between modified gravity and certain cosmological parameters (e.g. Baldi et al. 2014), but also the ability to predict cosmological observables when several – including cosmological – parameters are simultaneously varied is a key to constraining modified gravity models in the absence of prior knowledge of any of the parameters (especially when deviation from GR means that standard cosmological parameters can prefer values different from those in ΛCDM).

In this work, we will address the second limitation mentioned above, leaving the first one to be revisited in the future. In particular, we consider the widely studied $f(R)$-gravity (Buchdahl 1970) model (see e.g. Sotiriou & Faraoni 2010; De Felice & Tsujikawa 2010, for recent reviews) with the $f(R)$ parameter $|\bar{f}_{R0}|$, and study its degeneracy with three cosmological parameters: the matter density $\Omega_{\rm m}$, the normalization of the matter power spectrum σ_8, and the reduced Hubble parameter h. Ideally, one would sample a large number of points inside a wide parameter volume in order to constrain the model from observations; however, the number of sampling points quickly explodes in a high-dimensional parameter space.[1] In fact it is already challenging to simulate $\mathcal{O}(100)$ $f(R)$ models with sufficiently large box size and high resolution with the current generation of codes and supercomputers, making the grid approach not optimal.

A useful technique to overcome this difficulty, which has been widely applied in cosmology research in recent years, is *emulation*. With an optimal sampling of the model parameter space and interpolation of quantities from the sampled points, this makes it possible to make accurate predictions using a relatively small (e.g. $\lesssim \mathcal{O}(100)$) sample of points. The cosmic emulation technique (e.g. Heitmann et al. 2006) was first applied to predict non-linear matter power spectrum $P(k)$ with a per cent accuracy (Heitmann et al. 2009, 2010, 2014), where a Latin hypercube (Heitmann et al. 2006) was used for the efficient sampling, ensuring that the sampled points cover the whole parameter space as uniformly as possible. For effective interpolation, Gaussian processes (Rasmussen & Williams 2008), a non-parametric Bayesian regression method, is most commonly used, though other approaches such as neural networks (e.g. Agarwal et al. 2014) are also employed.

Emulation has since been applied to predict various other physical and observable quantities, such as the galaxy correlation function (Zhai et al. 2019), the galaxy power spectrum and galaxy clustering (Nishimichi et al. 2019; Kobayashi et al. 2020), the halo mass function (McClintock et al. 2019; Bocquet et al. 2020), the lensing shear correlation function (Harnois-Deraps, Giblin & Joachimi 2019), or weak-lensing peak (Harnois-Déraps et al. 2021) and void (Davies et al. 2021) statistics. These are enabled by dedicated suites of numerical simulations such as the AEMULUS (DeRose et al. 2019), *cosmo*-SLICS (Harnois-Deraps et al. 2019) and MassiveNuS (Liu et al. 2018) projects.

More recently, emulators have also been applied in the context of modified gravity models (e.g. Winther et al. 2019; Giblin et al. 2019; Ramachandra et al. 2021), with the objective of providing accurate predictions of the matter power spectrum for *Euclid* (Laureijs et al. 2011) and the Vera C. Rubin Observatory (LSST Science Collaboration et al. 2009) respectively. This work can be considered as a continuation and extension of these previous works. While the variety of models that we simulate and use to construct the $P(k)$ emulator is smaller than that used by Ramachandra et al. (2021), our full $f(R)$ simulations are larger and have a higher resolution: with a total of 100 runs, for 50 cosmologies covering the above 4D parameter space, using a box of $500\,h^{-1}$ Mpc and 1024^3 particles, our simulation program exploited four million core-hours on the COSMA machine, the UK's integrated supercomputing facility for theoretical modelling and HPC-based research in particle physics, astronomy, and cosmology. The simulations were run using the modified gravity AREPO code (Springel 2010; Arnold et al. 2019a), taking advantage of a substantial optimization offered by the algorithm of Bose et al. (2017).

In this work, we introduce this suite of simulations, and present a state-of-the-art $f(R)$ emulator for the fully non-linear matter power spectrum, $P(k)$, that is trained on them. The matter power spectrum is

[1] As an example, assuming that 10 points are sampled in each of the four dimensions investigated here, then a total of 10^4 models need to be simulated; even halving the number of points in each dimension still results in 625 models, which is well beyond the currently available resources.

one of the most powerful statistics since it captures all the statistical properties of an inhomogeneous initially Gaussian density field, which encodes rich information about the structure formation history and its underlying physics. Furthermore, it can be directly related to other two-point statistics, such as galaxy correlation function and weak-lensing power spectrum. These connections will be shown in upcoming papers based on alternative observables constructed from the same suite of simulations. In particular, for each model investigated in this paper, we have saved particle snapshots at pre-selected redshifts to construct past light-cones and therefore can enable the emulation of weak-lensing statistics in $f(R)$ and ΛCDM models. We will present a similar emulator for another widely studied class of models, the Dvali–Gabadadze–Porrati (Dvali, Gabadadze & Porrati 2000) braneworld model, in a companion paper.

This paper is organized as follows. In Section 2 we briefly cover the theory of $f(R)$ gravity and specify our particular model choice. In Section 3 we describe the simulations designed to construct the matter power spectrum emulator in this work. Section 4 will present the results of the emulator, and finally we summarize and conclude in Section 5.

2 $f(R)$ GRAVITY

$f(R)$ gravity (Buchdahl 1970) is an extension of Einstein's general relativity (GR). It adds an additional degree of freedom to gravity, allowing it to vary in space and time. Within a GR framework one can define the action of $f(R)$ gravity as

$$S = \int \mathrm{d}^4x\,\sqrt{-g}\left[\frac{R + f(R)}{16\pi G} + \mathcal{L}_\mathrm{m}\right], \tag{1}$$

where g is the determinant of the metric tensor $g_{\mu\nu}$ of the 4D (GR-like) manifold; R is the Ricci scalar and $f(R)$ is a scalar function of R that gives rise to the additional (scalar) degree of freedom; G is the standard gravitational constant; and $\mathcal{L}_\mathrm{m}$ denotes the matter Lagrangian. If one requires the variation of the action with respect to the metric to vanish, one can derive the field equations for (metric) $f(R)$ gravity,

$$G_{\mu\nu} + f_R R_{\mu\nu} - \left(\frac{f}{2} - \Box f_R\right) g_{\mu\nu} - \nabla_\mu\nabla_\nu f_R = 8\pi G T_{\mu\nu}, \tag{2}$$

again in GR notation and using Einstein's sum convention, where $G_{\mu\nu}$ denotes the Einstein tensor, $R_{\mu\nu}$ is the Ricci tensor, $T_{\mu\nu}$ is the energy–momentum tensor, and ∇_μ the covariant derivative ($\Box \equiv \nabla_\mu\nabla^\mu$). The quantity f_R is the derivative of the scalar function with respect to the Ricci scalar R, $f_R \equiv \mathrm{d}f(R)/\mathrm{d}R$, and is often called the scalar degree of freedom.

In the quasi-static and weak-field limit [often referred to as the Newtonian limit of $f(R)$ gravity], the above equation simplifies to two equations: the modified Poisson equation,

$$\nabla^2\Phi = \frac{16\pi G}{3}\delta\rho - \frac{1}{6}\delta R, \tag{3}$$

and the equation of motion for the scalar degree of freedom f_R, which is obtained by taking the trace of equation (2),

$$\nabla^2 f_R = \frac{1}{3}\left(\delta R - 8\pi G\delta\rho\right), \tag{4}$$

where $\delta\rho$ and δR are respectively the perturbations to the density field and Ricci scalar, while ∇ denotes the gradient operator in three dimensions.

The modified Poisson equation shows that the total gravitational potential Φ and consequently the gravitational forces can be enhanced by a factor of 4/3 in low-curvature environments ($\delta R \approx 0$). If there was no mechanism to counterbalance this enhancement in higher-density environments, this would of course lead to immediate tensions with Solar system constraints on gravity (Will 2014). On the other hand, $f(R)$ gravity is well known to feature the so-called chameleon screening mechanism (Khoury & Weltman 2004; Mota & Shaw 2007; Brax et al. 2008), which could drive $f_R \to 0$ in regions with extensive deep Newtonian potential, therefore bringing the modified Poisson equation back to its behaviour in GR.

A model that is designed to comply with the local constraints on gravity is the one proposed by Hu & Sawicki (2007) using the following functional form for $f(R)$:

$$f(R) = -m^2\frac{c_1\left(\frac{R}{m^2}\right)^n}{c_2\left(\frac{R}{m^2}\right)^n + 1}, \tag{5}$$

where $m^2 \equiv \Omega_\mathrm{m}H_0^2$, with H_0 being the Hubble constant. If one sets its free parameters, c_1 and c_2, to

$$\frac{c_1}{c2} = 6\frac{\Omega_\Lambda}{\Omega_\mathrm{m}} \text{ and } \frac{c_2\,R}{m^2} \gg 1, \tag{6}$$

the theory also features a ΛCDM-like expansion history, fulfilling another important observational requirement (Hu & Sawicki 2007). One can now simplify the differential equation for f_R to

$$f_R \equiv \frac{\mathrm{d}f(R)}{\mathrm{d}R} = -n\frac{c_1\left(\frac{R}{m^2}\right)^{n-1}}{\left[c_2\left(\frac{R}{m^2}\right)^n + 1\right]^2} \approx -n\frac{c_1}{c_2^2}\left(\frac{m^2}{R}\right)^{n+1}. \tag{7}$$

For this paper, we will adopt $n = 1$ as in most of the literature (see e.g. Li & Hu 2011; Ramachandra et al. 2021, for studies of other values of n). The remaining free parameters of the theory can now be re-expressed as $\bar{f}_{R0}$, which is the background value of the scalar degree of freedom at $z = 0$. It controls the potential depth threshold at which chameleon screening becomes active and GR-like forces are recovered.

In our emulator simulations we shall vary $\bar{f}_{R0}$ between $\bar{f}_{R0} = -10^{-4.5}$ (F4.5) and $\bar{f}_{R0} = -10^{-6.2}$ (F6.2). F4.5 corresponds to a very strong modification of gravity that is in obvious conflict with current observational constraints (Terukina et al. 2014). F6.2 on the other hand is a relatively weak modification of gravity where most high-mass haloes are completely screened. To make the emulator a useful tool for observational analysis, it is nevertheless necessary to cover this wide parameter range with our simulations.

3 THE EMULATOR SIMULATION SUITE

In this section, we describe the simulations used for constructing the $f(R)$ matter power spectrum emulators in this work. Our simulation suite consists of a total of 200 collisionless dark-matter-only (DMO) runs covering 50 $f(R)$-gravity models in a ΛCDM background expansion. For each model, or 'node', we ran two independent realizations with initial conditions designed to suppress the sampling variance (as in Harnois-Deraps et al. 2019; see Section 3.2), for two different resolutions. The *high-resolution* simulations were run with 1024^3 DM particles in $500\,h^{-1}$ Mpc side-length simulation boxes at a mass resolution of $m_\mathrm{part} = 9.1 \times 10^9\,h^{-1}\,\mathrm{M}_\odot$. The *low-resolution* simulations use 512^3 DM particles in a $1500\,h^{-1}$ Mpc simulation box with a mass resolution of $m_\mathrm{part} = 1.5 \times 10^{12}\,h^{-1}\,\mathrm{M}_\odot$. The gravitational softening length of the high-resolution runs is $15\,h^{-1}$kpc (the highest resolution of the Adaptive Mesh Refinement (AMR) modified gravity (MG) solver is similar) and the softening of the low-resolution simulations is $75\,h^{-1}$ kpc.

3.1 Cosmological parameters

The simulations presented in this work explore the cosmological parameter space by varying Ω_m, σ_8, h, and $\bar{f}_{R0}$ while keeping the other parameters fixed to $n_\mathrm{s} = 0.9652$, $\Omega_\mathrm{b} = 0.049\,199$, and $\Omega_\Lambda = 1 - \Omega_\mathrm{m}$. The reason for this choice is that Ω_m and σ_8 are well suited to be measured by weak-lensing statistics that we aim to emulate as well. h is varied along with the $f(R)$ parameter due to the tensions between early- and late-time cosmological probes. Because these simulations were partly designed to emulate weak-lensing statistics, we sample directly in the composite structure growth parameter $S_8 \equiv (\Omega_\mathrm{m}/0.3)^{0.5}$ instead of the physical matter fluctuation amplitude parameter σ_8, and therefore better account for the degeneracy between Ω_m and σ_8 in cosmic shear analyses.

The values for the variable parameters are listed in Table 1. For our choice of parameters we take a similar approach as the *cosmo*-SLICS project (Harnois-Deraps et al. 2019), sampling the parameters in a Latin hypercube and ensuring that the parameters are as evenly distributed in all sub-planes of our parameter space as possible. Due to the phenomenology of $f(R)$ gravity, we do not sample evenly in $\bar{f}_{R0}$ but in $\log_{10}|\bar{f}_{R0}|$. While the largest sampled value for the $f(R)$ parameter, $|\bar{f}_{R0}| \approx 10^{-4.5}$, is relatively straightforward to pick considering current constraints on cosmological scales (Terukina et al. 2014), the smallest value is a less obvious choice. As node 0 corresponds to $\log_{10}|\bar{f}_{R0}| = -\infty$ we obviously have to set a lower limit to the emulated parameter range by picking the next largest value for $\log_{10}|\bar{f}_{R0}|$. Selecting a lower value, e.g. $\log_{10}|\bar{f}_{R0}| = -8$, would allow very smooth emulation of the transition from $f(R)$ gravity to ΛCDM but also lead to a poorer emulator performance due to the larger spacing of the $f(R)$ parameters in the Latin hypercube (given that the total number of nodes is limited by computational resources). A larger value leads to better emulator performance but a less smooth transition to ΛCDM. As a compromise, we decided to use $\log_{10}|\bar{f}_{R0}| \approx -6.2$ as our lowest emulated $f(R)$ parameter. The deviations of the matter power spectrum from ΛCDM are expected to be at the level of a few per cent, allowing for a relatively smooth transition to GR without spending too much simulation time on models that lead to extremely small deviations from ΛCDM. The distribution of the variable parameters within the studied parameter space is illustrated in Fig. 1.

3.2 Initial condition generation

The power spectra of initial conditions (ICs) for cosmological simulations can never exactly resemble the theoretical input power spectrum that is used to generate the ICs at high redshift (in our case $z = 127$). While the small-scale differences due to the limited resolution of the ICs become irrelevant as the simulation progresses, the large-scale errors or sample variance, which occur due to the limited box size of the simulations, are carried through while the simulations run and are still apparent at $z = 0$.

To limit these large-scale errors, we run two implementations for a matched pair of ICs with independent phases (i.e. independent random seeds in the '2LPT' initial condition generator) per node (cosmological parameter set) and box. The IC pair are selected such that their average large-scale error is as small as possible. In order to achieve this, we first generate 100 independent ICs for node-0 (fiducial) cosmology and measure their power spectra. The relative difference of the actual IC power spectra to the theoretical input power spectrum is shown in Fig. 2. From the 100 ICs, we selected two using the following criteria:

(i) the mean large-scale error should be small,

Table 1. The simulation parameters varied for the 50 nodes of the cosmic emulator simulations. An ASCII version of this table is available online: https://bitbucket.org/arnoldcn/forge_emulator/.

Node	Ω_m	Ω_Λ	σ_8	h	$\lvert\bar{f}_{R0}\rvert$
0	0.313 15	0.686 85	0.821 72	0.673 70	0 (ΛCDM)
1	0.547 25	0.452 75	0.493 42	0.786 99	$10^{-5.449\,75}$
2	0.539 61	0.460 39	0.637 83	0.683 93	$10^{-5.511\,78}$
3	0.107 21	0.892 79	1.229 74	0.610 90	$10^{-5.480\,08}$
4	0.315 92	0.684 08	0.601 11	0.688 45	$10^{-6.093\,10}$
5	0.157 41	0.842 59	0.911 75	0.710 67	$10^{-4.917\,48}$
6	0.353 39	0.646 61	0.718 86	0.780 52	$10^{-5.283\,68}$
7	0.112 40	0.887 60	1.234 13	0.793 18	$10^{-4.506\,05}$
8	0.393 03	0.606 97	0.721 52	0.752 00	$10^{-6.146\,47}$
9	0.180 96	0.819 04	1.037 76	0.761 32	$10^{-6.038\,17}$
10	0.429 27	0.570 73	0.503 50	0.776 67	$10^{-5.342\,19}$
11	0.402 49	0.597 51	0.555 23	0.691 20	$10^{-5.872\,85}$
12	0.212 86	0.787 14	1.066 87	0.706 61	$10^{-5.147\,80}$
13	0.346 71	0.653 29	0.781 91	0.700 56	$10^{-4.900\,56}$
14	0.154 64	0.845 36	0.933 90	0.772 73	$10^{-5.387\,63}$
15	0.281 72	0.718 28	0.713 67	0.649 68	$10^{-5.303\,26}$
16	0.370 32	0.629 68	0.612 64	0.762 04	$10^{-5.556\,69}$
17	0.416 27	0.583 73	0.742 42	0.634 27	$10^{-4.842\,39}$
18	0.323 31	0.676 69	0.859 87	0.817 49	$10^{-5.434\,73}$
19	0.477 84	0.522 16	0.564 03	0.667 24	$10^{-5.171\,31}$
20	0.205 09	0.794 91	0.756 41	0.644 37	$10^{-5.235\,75}$
21	0.441 03	0.558 97	0.502 37	0.620 46	$10^{-5.205\,64}$
22	0.464 03	0.535 97	0.586 20	0.802 96	$10^{-5.850\,13}$
23	0.136 44	0.863 56	1.258 37	0.624 73	$10^{-5.979\,60}$
24	0.188 32	0.811 68	0.853 96	0.801 74	$10^{-4.777\,81}$
25	0.120 66	0.879 34	1.315 91	0.695 63	$10^{-5.609\,79}$
26	0.288 54	0.711 46	0.653 31	0.739 43	$10^{-5.060\,27}$
27	0.450 16	0.549 84	0.722 41	0.719 54	$10^{-4.662\,74}$
28	0.171 55	0.828 45	1.139 36	0.627 68	$10^{-5.802\,53}$
29	0.519 49	0.480 51	0.595 77	0.744 73	$10^{-5.013\,40}$
30	0.439 09	0.560 91	0.613 27	0.678 56	$10^{-5.750\,21}$
31	0.497 86	0.502 14	0.582 88	0.808 06	$10^{-5.736\,66}$
32	0.409 09	0.590 91	0.541 79	0.737 99	$10^{-5.916\,85}$
33	0.232 27	0.767 73	0.864 33	0.600 28	$10^{-4.720\,39}$
34	0.383 90	0.616 10	0.611 74	0.655 70	$10^{-5.647\,29}$
35	0.262 34	0.737 66	0.886 65	0.769 98	$10^{-5.996\,17}$
36	0.254 53	0.745 47	0.762 12	0.669 18	$10^{-4.749\,84}$
37	0.297 62	0.702 38	0.793 47	0.673 00	$10^{-5.627\,37}$
38	0.224 23	0.775 77	0.889 11	0.646 03	$10^{-4.857\,59}$
39	0.307 99	0.692 01	0.710 46	0.660 01	$10^{-5.930\,63}$
40	0.512 88	0.487 12	0.618 34	0.790 98	$10^{-5.106\,24}$
41	0.140 61	0.859 39	1.171 25	0.731 01	$10^{-4.968\,88}$
42	0.337 82	0.662 18	0.667 02	0.722 56	$10^{-6.097\,96}$
43	0.525 20	0.474 80	0.664 52	0.813 47	$10^{-4.633\,03}$
44	0.194 35	0.805 65	1.017 17	0.639 11	$10^{-4.563\,09}$
45	0.269 63	0.730 37	0.913 66	0.755 11	$10^{-5.022\,80}$
46	0.491 35	0.508 65	0.509 27	0.607 66	$10^{-4.587\,28}$
47	0.472 07	0.527 93	0.580 56	0.615 62	$10^{-5.681\,60}$
48	0.244 24	0.755 76	0.856 76	0.714 36	$10^{-6.174\,88}$
49	0.361 87	0.638 13	0.563 21	0.728 61	$10^{-4.693\,40}$

(ii) their individual large-scale errors should be small,

(iii) and their errors should cross the zero line as often as possible on large scales to avoid a leakage of power from large to small scales.

The two finally selected seeds for the ICs, 74 and 54, are shown in red in Fig. 2 and their mean as the green dashed line. We will refer to these implementations using their seeds, 2080 (74 in Fig. 2) and 4257 (54 in Fig. 2), in the following. Notice that this technique does not produce a matched pair of ICs whose sample variance errors cancel exactly (as in Angulo & Pontzen 2016), but instead follows Harnois-

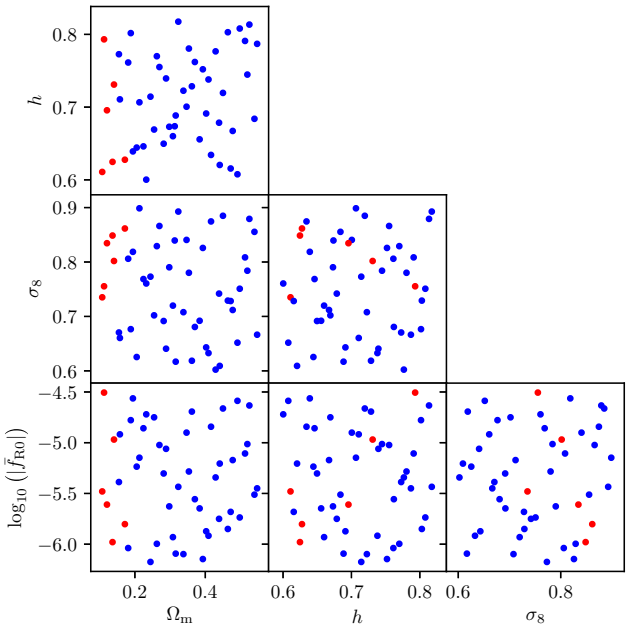

Figure 1. Visualization of the distribution of the parameters in the different sub-planes of the parameter space studied. The fiducial cosmology simulation (node 0) is not shown in the planes involving $\log_{10}|\bar{f}_{R0}|$ since $f_{R0} = 0$ in that case. The nodes represented by red dots are the ones for which the emulator error exceeds 5 per cent at $z = 0$ in a cross-validation test.

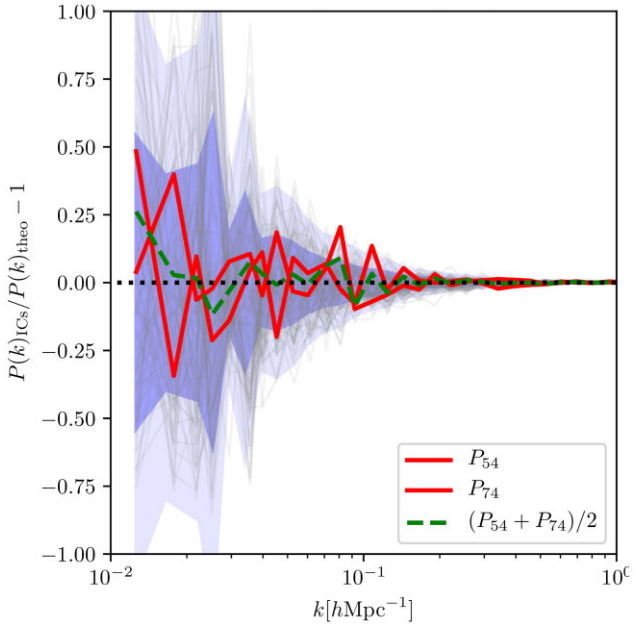

Figure 2. The relative difference of the 100 independent IC power spectra to the theoretical input power spectrum (faint grey lines) as well as the 1σ and 2σ errors of the IC power spectrum distribution (blue shaded regions). The selected seeds are shown in red, and their mean by the green dashed line.

Deraps et al. (2019) to generate a pair of ICs that (on average) roughly converges on the ensemble mean for the 3D power spectrum while each individual Fourier mode is still drawn from a Gaussian distribution. This is likely relevant for measurements of higher-order statistics.

The ICs were generated using the 2LPTIC (Crocce, Pueblas & Scoccimarro 2006) code, an initial condition generator that employs second-order Lagrangian perturbation theory to compute particle displacements for a given initial matter power spectrum, based on N-GENIC (Springel et al. 2005).

3.3 MG simulations with AREPO

The simulations in this work were carried out with the AREPO cosmological simulation code (Springel 2010; Weinberger, Springel & Pakmor 2020), using the modified gravity module presented in Arnold et al. (2019a). In the following, we will only give a very brief overview of the code and refer the reader to the above references for more details. AREPO uses a tree particle-mesh algorithm to calculate the standard gravitational forces. The additional modified gravity force (fifth force) is calculated by employing a multigrid accelerated relaxation solver on an adaptive refining mesh (i.e. an AMR grid).

To obtain the modified gravity forces, the code needs to solve equation (4) in the first place. In order to do so, the density field is binned on to the AMR grid, which is built such that on the finest refinement level each cell can contain at most one particle (except if a pre-set maximum refinement level is reached; the cell size at this level is of the order of the smoothing length of the standard gravity solver). Reformulating equation (4) as (Bose et al. 2017):

$$\nabla^2(u^2) = \frac{1}{3}\left\{\frac{\bar{R}(a)}{\bar{f}_R(a)}\left[\frac{1}{u} - 1\right] - \frac{8\pi G}{\bar{f}_R(a)}\delta\rho\right\}, \tag{8}$$

where $u \equiv \sqrt{f_R/\bar{f}_R(a)}$ and overbars denote background values, the solver now obtains the value of f_R iteratively on the grid. The MG force can then be easily computed from the gradient of the scalar field,

$$\boldsymbol{a}_{\mathrm{MG}} = \frac{c^2}{2}\nabla f_R, \tag{9}$$

where c denotes the speed of light.

As the modified gravity solver can only compute the forces for all particles simultaneously, and the scalar field equation that it solves is highly non-linear, it is computationally more expensive than the standard gravity solver of AREPO. The maximum modified gravity acceleration is, however, smaller than the maximum standard gravity acceleration, largely because the latter occurs in regions with high density (e.g. in massive haloes) that are screened from the fifth-force contribution. This allows us to run the MG solver on larger time steps (see Arnold et al. 2019a, for details on the time-stepping scheme), significantly reducing the computational cost of the simulations.

Together with the efficient MPI parallelization and lean memory footprint of AREPO, these optimized numerical algorithms have made it possible to run the large number of $f(R)$ simulations used in this work. The 100 500 h^{-1} Mpc runs took roughly three million core-hours on the COSMA7 machine hosted at the Institute for Computational Cosmology, Durham University, by using 20 computer nodes, each with 512 GB RAM and 28 cores (Intel Xeon Gold 5120 CPU @ 2.20 GHz). The larger-box runs took much less time due to their lower resolution.

4 AN EMULATOR FOR THE 3D MATTER POWER SPECTRUM

We use the simulations described above to build an emulator for the 3D matter power spectrum. Analyses of gravitational lensing properties, the halo mass function (HMF), lensing, etc., based on these simulations, will be presented in subsequent works. For the

power spectrum emulator we rely on redshift outputs common to all nodes, namely at $z = 0.0$, 0.25, 0.50, 0.75, 1.00, 1.25, 1.50, 1.75, and 2.00. Below we present the measurement details and our emulator verification procedure, first focusing at $z = 0$; however, the measurements at larger z are performed in the same way. A cross-validation of the emulator prediction at $z > 0$ is shown in Appendix A.

The power spectrum emulator is made publicly available online at https://bitbucket.org/arnoldcn/forge_emulator/. A brief user guide can be found in Appendix C.

4.1 Power spectrum measurement

The 3D matter power spectra are measured with the estimator built into AREPO employing the self-folding technique described in Jenkins et al. (1998), Springel et al. (2018). This allows a high resolution for the power spectrum to be obtained while avoiding computationally expensive FFTs on large grids. In this paper, we use an FFT resolution of 2048^3 with a fold-factor $f_{\rm fold} = 4$ and apply the self-folding twice (by using $f_{\rm fold} = 4^2$ for the second folding). This allows for a dynamic range of 32 000 between the largest scales measured and the Nyquist frequency of our measurement. We apply this technique to both our small- and large-box simulations, and both members of our matched IC pair. The resulting $3 \times 2 \times 2$ sections of the power spectra (3 foldings, 2 ICs, 2 resolutions) span different ranges in k that overlap, allowing us to combine them into a single $P(k)$ per cosmological model.

To obtain a single power spectrum per cosmology, we divide the 12 sections by the HALOFIT (Takahashi et al. 2012) non-linear ΛCDM prediction to get the 'response' defined as

$$B(k) \equiv P(k)^{f(R)}_{\rm simulation} \,/\, P(k)^{\Lambda{\rm CDM}}_{\rm HALOFIT}. \tag{10}$$

Note that the HALOFIT prediction does not incorporate any $f(R)$-gravity effects on the power spectrum, i.e. $B(k)$ will be a measure of the discrepancy between HALOFIT and the simulation for our ΛCDM node (node 0), but will include modified gravity effects for all others. The aim of dividing by the theory prediction is to largely flatten out the k-dependence of the power spectrum. This way we can minimize the errors due to bin-centring and limit the dynamic range that has to be emulated, as described below. The 12 $B(k)$ sections (foldings) are finally binned into fine k-bins, from which our estimator extracts the median $B(k)$. Some sections of the large boxes are not actually used as they are dominated by shot noise. This binning process is illustrated for node 0 in Fig. 3. The different coloured symbols indicate the power spectra relative to HALOFIT of the different implementations and boxes before binning, the blue line those after binning. A larger scatter appears towards the low-k end of each folding.

The binned responses and power spectra for all nodes are shown in Fig. 4. To avoid an overcrowded plot, the data for the 50 nodes are split into five rows. The curves for the different nodes are coloured according to the background field of the $f(R)$ model as shown in the colour bar at the top of the figure. Low numbers ($-\log_{10}\left(-\bar{f}_{R0}\right) \approx 4$) mean less efficient chameleon screening and consequently stronger deviation from GR. Higher numbers ($-\log_{10}\left(-\bar{f}_{R0}\right) \approx 6.5$) lead to stronger chameleon screening and a model behaviour more similar to GR.

The left-hand column of Fig. 4 shows the binned power spectra of the simulations (solid lines), as well as the linearly evolved initial

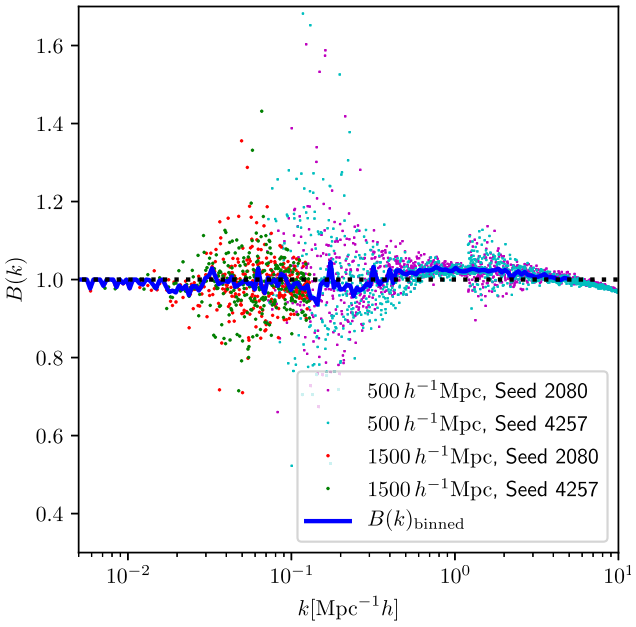

Figure 3. The power spectrum response data [$B(k)$] for the different boxes and foldings of node 0, before (dots) and after (blue line) binning. The black dotted line shows the equality to the HALOFIT prediction.

power spectra[2] (dotted lines). These panels also show the shot-noise level of the simulations (the horizontal grey dashed line) that has been corrected for. Note that we limit our results to $k < 10\,h\,{\rm Mpc}^{-1}$ here, since our main aim is to create a reliable emulator that can be applied to observational data. The noise level would allow us to study the power spectra at even smaller scales, but the simulations may not have a high-enough force resolution in that regime.

The second column of Fig. 4 shows the binned response $B(k)$ described above. As one can see from the amplitude of $B(k)$, $f(R)$ gravity leads to an enhancement in the power spectrum on top of the difference between the theoretical prediction and the simulations. This enhancement is larger for stronger $f(R)$-gravity models, and it depends on the k-scale as expected from previous works (Li et al. 2013; Winther et al. 2015; Arnold et al. 2019a, b).

4.2 Power spectrum smoothing

Gaussian process emulators generally work best for $\mathcal{O}(1)$ quantities with as few minima and maxima as possible. As the power spectrum itself spans several orders of magnitude, we choose to emulate the ratio $B(k)$ to the HALOFIT ΛCDM prediction. As one can see from column 2 of Fig. 4, the (finely) binned $B(k)$ is very noisy, which poses a problem for the emulator. This problem can be overcome by applying a Savitzky–Golay filter (Savitzky & Golay 1964) to $B(k)$, following Ramachandra et al. (2021). A similar technique has been employed in Knabenhans et al. (2019) and Giblin et al. (2019). The filter fits a certain number, N, of data points left and right of the point of interest with a polynomial of pth order. We have applied this technique to $B(k)$ using $N = 51$ and $p = 3$ to obtain the coarsely smoothed $B(k)$ in column 3 of Fig. 4.

[2]For linear evolution we have used the ΛCDM linear growth factor for each node with the cosmological parameters $\Omega_{\rm m}$, h, σ_8 for that node, assuming $f_R = 0$.

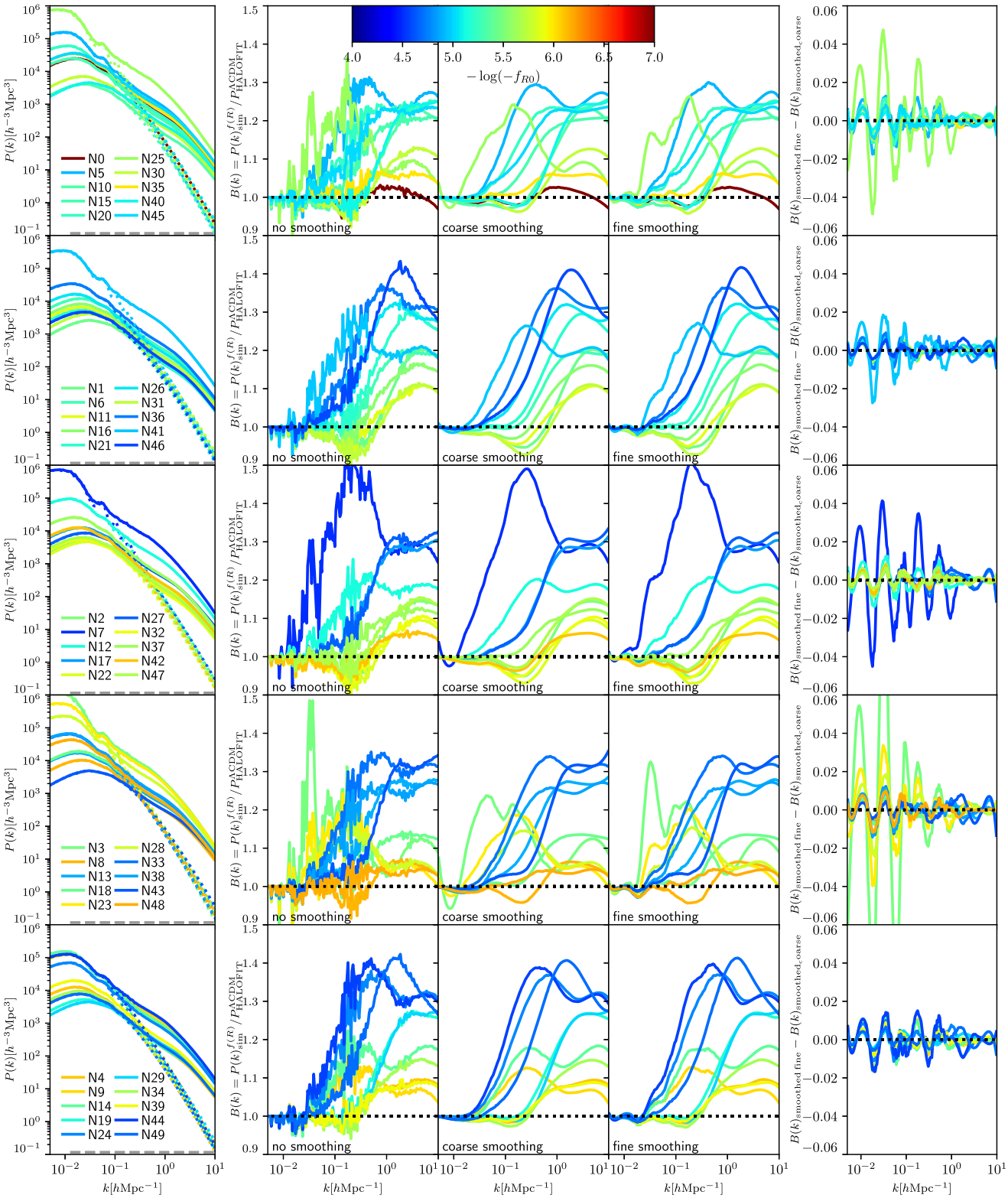

Figure 4. The 3D matter power spectra for the 50 nodes (N0–N49) as measured from the $f(R)$-gravity simulations at $z = 0$. To avoid overcrowded plots, the node data have been distributed over the five rows of the figure. The left-hand panels show the binned power spectra (averaged over all box sizes and implementations, shot noise corrected; see text for details) from the simulations as solid lines. The line colours indicate the background field of the $f(R)$ model for each node as indicated by the colour bar at the top of the figure. Node 0 is ΛCDM, but is shown as $-\log_{10}\left(-\bar{f}_{R0}\right) = 7$. The corresponding linearly evolved initial power spectra are shown as dotted lines in the same colours in the left-hand panels, where the simulation shot-noise level is indicated by the grey dashed lines. The second column shows the binned ratio $B(k)$ between the simulated $f(R)$-gravity power spectrum and the HALOFIT ΛCDM prediction for the cosmology. The third and fourth columns show the same quantity but with a coarse and fine Savitzky–Golay smoothing applied (see text), respectively. The fifth column shows the difference between the fine and the coarse smoothed $B(k)$. Black dotted lines indicate the HALOFIT predictions.

These data are now better suited for a Gaussian process emulator, but lack the enhancement of the BAO fluctuations due to $f(R)$ gravity because of the aggressive smoothing. To capture these enhancements as well, we perform a second, finer, smoothing of the initially binned $B(k)$ but use $N = 41$ and $p = 4$ to obtain the curves in column 4 of the figure. The data in column 4 are less suited for a GP emulator due to the large number of minima and maxima on top of the enhancement. We therefore choose to emulate the BAO 'wiggles' separately after isolating them by taking the difference between the fine and the coarse smoothed $B(k)$ (see column 5 of Fig. 4). The above choice of parameters for the coarse and fine smoothing led to the best overall performance for the emulator during our analysis. In particular we verified that a finer Savitzky–Golay filter for the fine or both smoothing steps does not improve the results.

4.3 Emulating the 3D matter power spectrum

We emulate separately the coarsely smoothed $B(k)$ and the BAO-induced 'wiggles' described above employing a Gaussian process emulator similar to that described in Harnois-Deraps et al. (2019). This emulator uses the publicly available SCIKIT-LEARN Gaussian process regression code.[3] Following numerous previous works (for example, Habib et al. 2007; Heitmann et al. 2009; Kwan et al. 2015), we adopt a radial basis function to model the covariance of the power spectrum response in the Gaussian process. This functional form for the covariance is well motivated, assuming only that the emulated statistic varies smoothly with changes in the cosmological parameters, and that it converges towards the training set measurements at the simulation nodes. Training the emulator corresponds to fitting the five free hyperparameters of this model, one amplitude and four correlation lengths (one per dimension in our cosmological parameter space), by applying a gradient ascent optimization algorithm to a likelihood distribution conditioned on the simulated training data. We refer the interested reader to Rasmussen & Williams (2008) for more details on emulation with Gaussian processes.

On both the coarsely smoothed $B(k)$ and BAO-induced 'wiggles', we perform a principal component analysis (PCA) using four basis functions prior to training, finding this compression of the data beneficial for improving emulation accuracy and reducing training time. We confirmed that this number of basis functions is enough to reliably emulate both the wiggles and the smoothed $B(k)$ and that a larger number does not improve the emulator performance further.

Since the emulator cannot model the ΛCDM response with a value of $|\bar{f}_{R0}| = 0$ for node 0, we set this parameter to a value where no deviation of the $f(R)$-gravity power spectrum from its ΛCDM counterpart is expected. For $z = 0$, this is $|\bar{f}_{R0}| = 10^{-7.5}$; at higher redshift we use a slightly larger value to ensure a smooth transition.

The final power spectrum prediction $P_{\rm prediction}(k)$ is obtained from the emulated coarsely smoothed $B_{\rm coarse}(k)$ and the emulated 'wiggles' $W(k)$:

$$P_{\rm emulated}(k) = P^{\Lambda{\rm CDM}}_{\rm HALOFIT} \times [B_{\rm coarse}(k) + W(k)]\,. \quad (11)$$

The training described above is repeated at all common output redshifts available. In the following section we verify the emulator predictions for $z = 0$, while higher-redshift results are presented in Appendix A.

[3] https://scikit-learn.org/stable/modules/Gaussian_process.html

4.4 Emulator validation

The first step that we take towards validating our emulator predictions is performing a 'leave-one-out' cross-validation test. This method removes one node from the sample at a time, trains the emulator on the remaining 49 nodes, and then predicts the power spectrum response $B_{\rm coarse}(k)$ and the wiggle enhancement $W(k)$ due to $f(R)$ gravity at the excluded node. This process is repeated for all nodes, with results shown in Fig. 5, where the dashed lines show the true data for each node, and solid lines the predictions of the emulator. The two separately emulated quantities, the coarsely smoothed $B(k)$ and the wiggles, are shown in the two left-hand columns of the plot.

As one can see, the emulator performs well except for some of the more extreme cosmologies. This is also reflected in the accuracy of the predicted total $B(k)$ shown in the two right-hand columns, where the rightmost column is the relative difference between the emulated and the true simulated $B(k)$. For most nodes, the latter remains entirely within the ±5 per cent error margin indicated by the light-grey bands in the rightmost column of Fig. 5 for $k < 2\,h\,{\rm Mpc}^{-1}$. The nodes with a > 5 per cent error for $k < 2\,h\,{\rm Mpc}^{-1}$ are 25, 41, 7, 23, 28, and 3. These nodes are highlighted in red in Fig. 1, from which it is obvious that all of them feature an extremely low $\Omega_{\rm m}$ value, resulting in a large baryon fraction. For example, for node 3, approximately 50 per cent of the matter is represented by baryons, leading to much larger than normal BAO wiggles that are further enhanced during cosmic structure formation. The emulator struggles to emulate these very large wiggles for node 3 (see column 2, row 4 of Fig. 5), resulting in a poor prediction accuracy at large scales in that specific case. Similar effects of the large baryon fraction can be observed for the other low-$\Omega_{\rm m}$ nodes. We note that values of $\Omega_{\rm m} \approx 0.1$ are far from the current observational constraints $\Omega_{\rm m} = 0.321 \pm 0.013$ (Planck Collaboration et al. VI 2020, TT+lowE) and the poorer performance of our emulator for these nodes will therefore not affect its applicability in practice. Although this subset of nodes may have extreme baryon fractions, it is necessary to include them in our ensemble of cosmologies to cover the range of $\Omega_{\rm m}$ values consistent with the constraints from current weak-lensing surveys, whilst keeping $\Omega_{\rm b}$ fixed to minimize the dimensionality of our emulation parameter space. It is also apparent from Fig. 5 that the emulated power spectra for the majority of the nodes have an error smaller than 2 per cent for $k < 10\,h\,{\rm Mpc}^{-1}$, as shown by the dark-grey bands in the rightmost column.

As a comparison, we have trained the emulator directly on the HALOFIT non-linear $P(k)$ for the same 50 cosmologies (i.e. $\bar{f}_{R0}$ is set to 0) and found a similar emulator accuracy for the same cross-validation exercise. This means that the accuracy is limited by the number of nodes in general, and is not driven by the modified gravity sector.

As a further step towards the verification of our emulator, we predicted the matter power spectra for two cosmologies that were not part of our training, and compared the emulator predictions to two sets of $f(R)$ simulations, each consisting of eight independent runs with 1024^3 particles in $500\,h^{-1}$ Mpc boxes (the ICs of these eight runs are paired in the same way as for the ones used to train the emulator, to suppress sample variance on large scales). This is thus not only a test of how well the emulator can predict the matter power spectra for different cosmologies, but also a measure of how much cosmic variance affects the results, as the eight simulations per parameter set offer a significantly larger statistical sample. Both simulation sets use the same cosmological base parameters as node 0, but a non-zero modified gravity parameter, $\bar{f}_{R0}$, respectively -10^{-6} (F6) and -10^{-5} (F5). The results are shown in Fig. 6, where the columns are

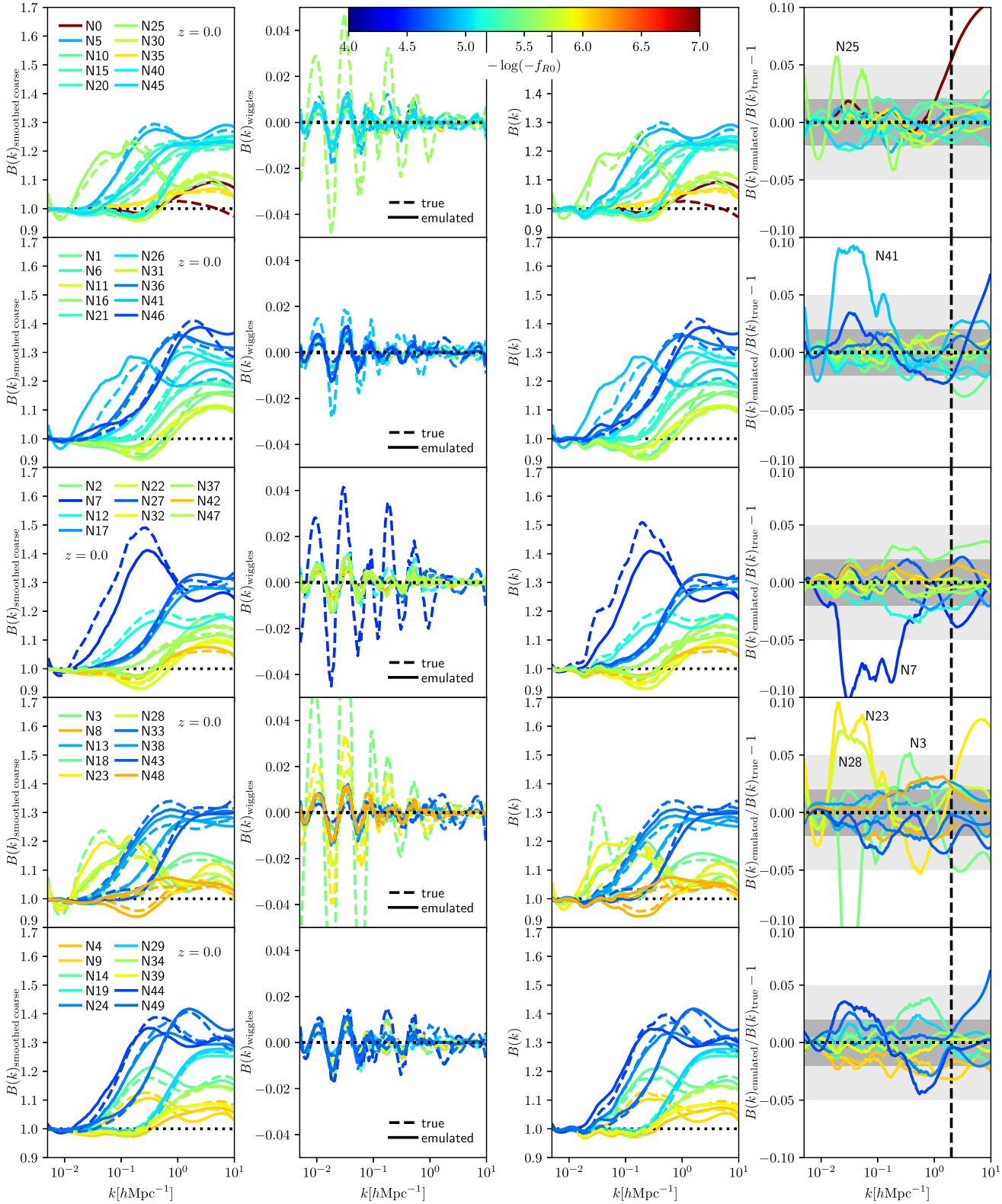

Figure 5. Cross-validation for the matter power spectrum emulator at $z = 0$. The power spectrum for each node is predicted by training the emulator on the 49 other nodes. The results are split into the five different rows of the plot for better readability. Dashed lines show the true data for each node, and solid lines the emulator predictions from the cross-validation. The colours indicate the strength of the $f(R)$ model as shown by the colour bar at the top of the plot. The left-hand column shows the coarsely smoothed $B(k)$. The centre-left-hand column shows the separately emulated BAO 'wiggle' enhancement. The centre-right-hand column shows emulated and true total $B(k)$ (i.e. the sum of the two left-hand columns). The right-hand column shows the relative difference between the emulated and the true $B(k)$. The dark- and light-grey shaded regions in the right-hand panels indicate the ± 2 per cent and ± 5 per cent error margins, respectively. Black dotted lines indicate equality; the dashed black vertical lines are at $k = 2\,h$ Mpc^{-1}.

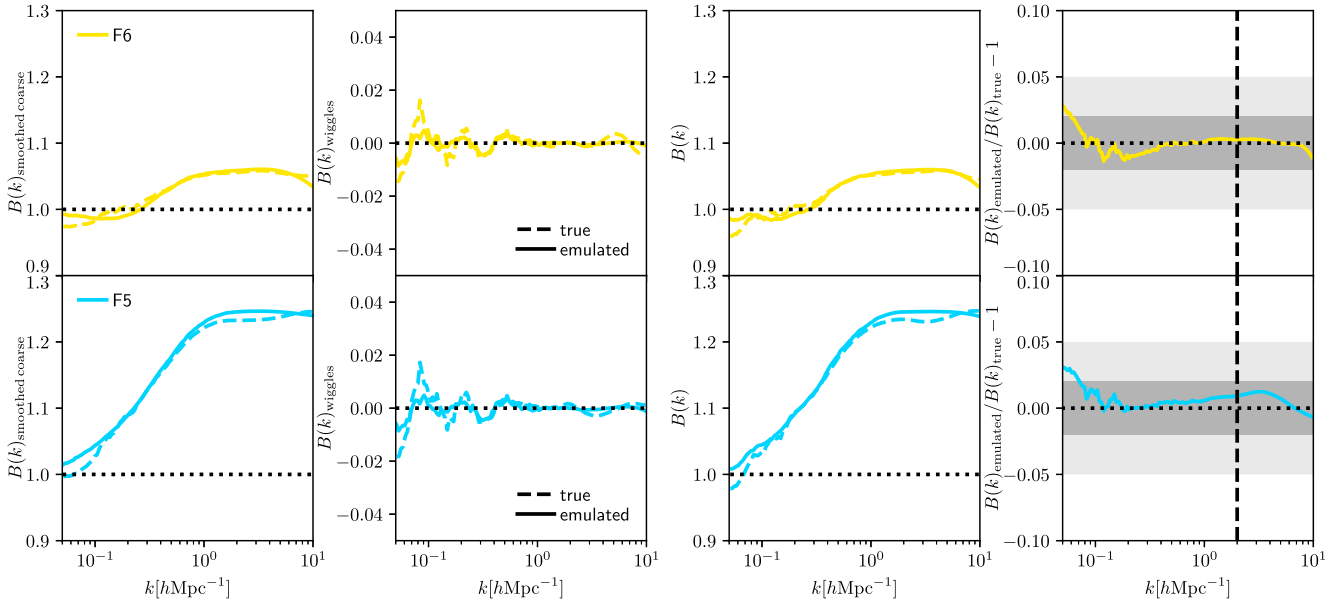

Figure 6. The power spectra responses for two additional cosmologies using fiducial base parameters (like node 0) but $|f_{R0}| = 10^{-6}$ (top) and $|f_{R0}| = 10^{-5}$ (bottom), at $z = 0$. The dashed lines show the measured spectra from the sets of eight simulations per parameter set, while the solid lines show the emulator predictions. As in Fig. 5, the left-hand column shows the coarsely smoothed response, $B(k)$, the centre-left-hand column the BAO wiggle enhancement, the centre-right-hand column the total $B(k)$, and the right-hand column the relative error of the emulated $B(k)$. The dotted horizontal lines indicate equality, and the dashed vertical line in the right-hand column shows $k = 2\,h\,{\rm Mpc}^{-1}$. The dark- and light-grey regions in the right-hand column indicate ± 2 per cent and ± 5 per cent error intervals, respectively.

the same as in Fig. 5. We do not have large-box implementations for these cosmologies, and therefore limit the k-range of the simulation power spectra to $k > 5 \times 10^{-2}\,h\,{\rm Mpc}^{-1}$. As one can see from Fig. 6, the emulator works very well over the whole k-range shown, with a maximum relative error of roughly 2.5 per cent for F6 and 3 per cent for F5.

The largest errors occur at the largest scales for $k < 0.07\,h\,{\rm Mpc}^{-1}$. These are partially caused by cosmic variance, which has the strongest effect at large scales, and partially by the fact that we emulate the relative difference to HALOFIT, which introduces additional errors. Both points could be addressed by emulating the relative difference with respect to a ΛCDM simulation using the same cosmological parameters but we can currently not afford the extra computational effort that this would imply.

The relative errors for $0.07 < k/(\,h\,{\rm Mpc}^{-1}) < 10$ stay completely within the 2 per cent error margin for both parameter sets, showing that our emulator can robustly predict the power spectrum response relative to HALOFIT. The high accuracy achieved at such non-linear scales is critical for weak-lensing science, which is mostly sensitive to these physical scales. We plan to use these $f(R)$ simulations in upcoming cosmic shear data analyses.

5 CONCLUSIONS

We have presented a new suite of 200 cosmological simulations in $f(R)$ gravity that explore a wide cosmological parameter space for 50 different combinations of $\Omega_{\rm m}$, h, σ_8, and $\bar{f}_{R0}$. These simulations are designed to aid the building of emulators for weak lensing, the halo mass function (which will both be presented in future works), and a matter power spectrum emulator, which we present in this work.

The simulations combine two different resolutions per parameter set, allowing us to probe a wide range of cosmological scales and dynamical range. For each resolution, two implementations with independent initial conditions are performed. The ICs for these are selected such that the large-scale cosmic variance errors of the two implementations approximately cancel each other. We then combine the data from the four individual runs per cosmological parameter set using a fine-binning technique. The noise is subsequently removed from the combined power spectrum using a Savitzky–Golay filter. In order to provide a smooth order-unity quantity as our training data, we choose to emulate the ratio of the simulation power spectrum to the HALOFIT ΛCDM predicted power spectrum, in other words a proxy for the relative difference between $f(R)$ gravity and GR, using an easily accessible GR baseline. We also choose to emulate the enhancement of the BAO wiggles separately as this gives a better overall performance. The emulator is run at different redshifts between $z = 0$ and $z = 2$ independently and is made publicly available online.

We have tested the performance of the power spectrum emulator using cross-validation, finding that it has an accuracy better than 5 per cent for all nodes with $\Omega_{\rm m} > 0.18$ at all emulated redshifts. Nodes with even lower $\Omega_{\rm m}$ have a large baryonic-to-total matter fraction, as $\Omega_{\rm b}$ remains unchanged, which leads to strong BAO wiggles that our emulator struggles to predict. This will nevertheless not limit the practical applicability of our emulator, as it performs very well (better than 5 per cent) for reasonable choices of $\Omega_{\rm m}$. We have also tested the emulator predictions against simulations for parameter sets that are not part of the training set (fiducial cosmology but with different values of $\bar{f}_{R0}$), and found that the emulator can predict those to better than 2 per cent accuracy. As an additional check, we have also trained the emulator using non-linear matter power spectra for 50 ΛCDM models with the same cosmological parameters (but $f_R = 0$), generated by HALOFIT, and obtained the same accuracy in the emulated $P(k)$. This implies that the 5 per cent accuracy in the cross-validation test is not due to the noisy simulation data and our smoothing of them, or the effect of $f(R)$ enhancement, but is more likely caused by the small size of the training data set,

especially considering that our 50 nodes covers very wide parameter ranges. The 5 per cent maximum error should be considered as the worst-case scenario, not only since the other test suggests a 2 per cent emulator accuracy, but also because even in the cross-validation case a 2 per cent accuracy is achieved for the majority of nodes at $k < 10\,h\,\mathrm{Mpc}^{-1}$.

We conclude that our simulation suite is suitable for building emulators for different cosmological observables in $f(R)$ gravity. These will be presented in subsequent works. The power spectrum emulator, in particular, can be used to make quick and accurate predictions for the fully non-linear matter power spectrum in $f(R)$ gravity, e.g. to analyse observational data using an MCMC pipeline over a wide k-range. This emulator does not include baryonic effects, but as shown by Arnold et al. (2019a), baryonic and $f(R)$-gravity effects on the matter power spectrum can be estimated, to a good precision for reasonable $f(R)$ models, from separate baryonic ΛCDM simulations and $f(R)$-gravity dark-matter-only simulations. This opens the possibility of combining our emulator with a baryonic physics power spectrum emulator for future analysis. Finally, in this paper we have focused on a particular $f(R)$ model; in subsequent works we will run and present simulations and emulators for other modified gravity models.

ACKNOWLEDGEMENTS

The authors thank Ben Bose for useful discussions and for providing matter power spectrum predictions generated with the REACT code. CA and BL are supported by the European Research Council (ERC) through a starting grant (ERC-StG-716532 PUNCA). BL is further supported by the UK Science and Technology Funding Council (STFC) Consolidated Grant No. ST/I00162X/1 and ST/P000541/1. JHD acknowledges support from an STFC Ernest Rutherford Fellowship (project reference ST/S004858/1). BG acknowledges the support of the Royal Society through an Enhancement Award (RGF/EA/181006). YC acknowledges the support of the Royal Society through a University Research Fellowship and an Enhancement Award. This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The equipment was funded by BEIS via STFC capital grants ST/K00042X/1, ST/P002293/1, ST/R002371/1, and ST/S002502/1, Durham University, and an STFC operation grant ST/R000832/1. DiRAC is part of the UK National e-Infrastructure.

DATA AVAILABILITY

The matter power spectrum training data, as well as the power spectrum emulator, are publicly available from the following git repository: https://bitbucket.org/arnoldcn/forge_emulator/. The simulation outputs can be made available upon reasonable request to the authors.

REFERENCES

Abbott T. M. C. et al., 2020, Phys. Rev. D, 102, 023509
Ade P. A. R. et al., 2014, A&A, 571, A20
Agarwal S., Abdalla F. B., Feldman H. A., Lahav O., Thomas S. A., 2014, MNRAS, 439, 2102
Aghamousa A. et al., 2016, preprint (arxiv:1611.00036)
Aiola S. et al., 2020, J. Cosmol. Astropart. Phys., 2020, 047
Alam S. et al., 2017, MNRAS, 470, 2617
Alam S. et al., 2021, Phys. Rev. D, 103, 083533
Albrecht A. et al., 2006, preprint (astro-ph/0609591)
Amon A. et al., 2021, preprint (arXiv:2105.13543)
Anderson L. et al., 2012, MNRAS, 427, 3435
Angulo R. E., Pontzen A., 2016, MNRAS, 462, L1
Angulo R. E., White S. D. M., 2010, MNRAS, 405, 143
Angulo R. E., Springel V., White S. D. M., Jenkins A., Baugh C. M., Frenk C. S., 2012, MNRAS, 426, 2046
Arnold C., Leo M., Li B., 2019a, Nat. Astron., 3 945–954
Arnold C., Fosalba P., Springel V., Puchwein E., Blot L., 2019b, MNRAS, 483, 790
Asgari M. et al., 2021, A&A, 645, A104
Baker T. et al., 2021, Rev. Modern Phys., 93, 015003
Baldi M., Villaescusa-Navarro F., Viel M., Puchwein E., Springel V., Moscardini L., 2014, MNRAS, 440, 75
Beutler F. et al., 2011, MNRAS, 416, 3017
Beutler F. et al., 2012, MNRAS, 423, 3430
Blake C. et al., 2011a, MNRAS, 415, 2876
Blake C. et al., 2011b, MNRAS, 418, 1707
Bocquet S., Heitmann K., Habib S., Lawrence E. Uram T., Frontiere N., Pope A., 2019, ApJ, 878, 55
Bocquet S., Heitmann K., Habib S., Lawrence E., Uram T., Frontiere N., Pope A., Finkel H., 2020, ApJ, 901, 5
Bose S., Li B., Barreira A., He J.-h., Hellwing W. A., Koyama K., Llinares C., Zhao G.-B., 2017, J. Cosmol. Astropart. Phys., 2017, 050
Bose B., Cataneo M., Tröster T., Xia Q., Heymans C., Lombriser L., 2020, MNRAS, 498, 4650
Bose B. et al., 2021, preprint (arXiv:2105.12114)
Brax P., van de Bruck C., Davis A.-C., Shaw D. J., 2008, Phys. Rev. D, 78, 104021
Buchdahl H. A., 1970, MNRAS, 150, 1
Cataneo M., Lombriser L., Heymans C., Mead A., Barreira A., Bose S., Li B., 2019, MNRAS, 488, 2121
Chan K. C., Scoccimarro R., 2009, Phys. Rev. D, 80, 104005
Clifton T., Ferreira P. G., Padilla A., Skordis C., 2012, Phys. Rep., 513, 1
Cole S. et al., 2005, MNRAS, 362, 505
Copeland E. J., Sami M., Tsujikawa S., 2006, Int. J. Modern Phys. D, 15, 1753
Crocce M., Pueblas S., Scoccimarro R., 2006, MNRAS, 373, 369
Davies C. T., Cautun M., Giblin B., Li B., Harnois-Déraps J., Cai Y.-C., 2021, MNRAS, 507, 2267
De Felice A., Tsujikawa S., 2010, Living Rev. Relativ., 13, 3
De Haan T. et al., 2016, ApJ, 832, 95
DeRose J. et al., 2019, ApJ, 875, 69
Dvali G., Gabadadze G., Porrati M., 2000, Phys. Lett. B, 485, 208
Eisenstein D. J. et al., 2005, ApJ, 633, 560
Feng Y., Chu M.-Y., Seljak U., McDonald P., 2016, MNRAS, 463, 2273
Ferreira P. G., 2019, ARA&A, 57, 335
Giblin B., Cataneo M., Moews B., Heymans C., 2019, MNRAS, 490, 4826
Giocoli C. et al., 2021, A&A 653 A19
Guzzo L. et al., 2008, Nature, 451, 541
Habib S., Heitmann K., Higdon D., Nakhleh C., Williams B., 2007, Phys. Rev. D, 76, 083503
Hamana T. et al., 2020, PASJ, 72, 16
Harnois-Déraps J., Pen U.-L., Iliev I. T., Merz H., Emberson J. D., Desjacques V., 2013, MNRAS, 436, 540
Harnois-Deraps J., Giblin B., Joachimi B., 2019, A&A, 631, A160
Harnois-Déraps J., Martinet N., Castro T., Dolag K., Giblin B., Heymans C., Hildebrandt H., Xia Q., 2021, MNRAS, 506, 1623
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., White M., Wagner C., Habib S., Higdon D., 2010, ApJ, 715, 104
Heitmann K., Lawrence E., Kwan J., Habib S., Higdon D., 2014, ApJ, 780, 111
Hernández-Aguayo C., Arnold C., Li B., Baugh C. M., 2021, MNRAS, 503, 3867
Heymans C. et al., 2013, MNRAS, 432, 2433

Hinshaw G. et al., 2013, ApJS, 208, 19
Hou Z. et al., 2014, ApJ, 782, 74
Howlett C., Manera M., Percival W. J., 2015, Astron. Comput., 12, 109
Hu W., Sawicki I., 2007, Phys. Rev. D, 76, 064004
Jenkins A. et al., 1998, ApJ, 499, 20
Joyce A., Jain B., Khoury J., Trodden M., 2015, Phys. Rep., 568, 1
Jullo E., Natarajan P., Kneib J. P., D'Aloisio A., Limousin M., Richard J., Schimd C., 2010, Science, 329, 924
Khoury J., Weltman A., 2004, Phys. Rev. D, 69, 044026
Klypin A., Prada F., 2018, MNRAS, 478, 4602
Knabenhans M. et al., 2019, MNRAS, 484, 5509
Kobayashi Y., Nishimichi T., Takada M., Takahashi R., Osato K., 2020, Phys. Rev. D, 102, 063504
Koyama K., 2016, Rep. Progress Phys., 79, 046902
Koyama K., 2018, Int. J. Modern Phys. D, 27, 1848001
Kwan J., Heitmann K., Habib S., Padmanabhan N., Lawrence E., Finkel H., Frontiere N., Pope A., 2015, ApJ, 810, 35
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Li Y., Hu W., 2011, Phys. Rev. D, 84, 084033
Li B., Zhao H., 2009, Phys. Rev. D, 80, 044027
Li B., Zhao G.-B., Teyssier R., Koyama K., 2012, J. Cosmol. Astropart. Phys., 1, 51
Li B., Hellwing W. A., Koyama K., Zhao G.-B., Jennings E., Baugh C. M., 2013, MNRAS, 428, 743
Li B., Zhao G.-B., Koyama K., 2013, J. Cosmol. Astropart. Phys., 05, 023
Liu J., Bird S., Zorrilla Matilla J. M., Hill J. C., Haiman Z., Madhavacheril M. S., Petri A., Spergel D. N., 2018, J. Cosmol. Astropart. Phys., 2018, 049
Llinares C., Mota D. F., Winther H. A., 2014, A&A, 562, A78
LSST Science Collaboration et al., 2009, preprint (arXiv:0912.0201)
Mantz A. B., Allen S. W., Morris R. G., Rapetti D. A., Applegate D. E., Kelly P. L., von der Linden A., Schmidt R. W., 2014, MNRAS, 440, 2077
Mantz A. B. et al., 2015, MNRAS, 446, 2205
McClintock T. et al., 2019, ApJ, 872, 53
Monaco P., Sefusatti E., Borgani S., Crocce M., Fosalba P., Sheth R. K., Theuns T., 2013, MNRAS, 433, 2389
Mortonson M. J., Weinberg D. H., White M., 2013, preprint (arXiv:1401.0046)
Mota D. F., Shaw D. J., 2007, Phys. Rev. D, 75, 063501
Nelson D. et al., 2019, MNRAS, 490, 3234
Nishimichi T. et al., 2019, ApJ, 884, 29
Oyaizu H., 2008, Phys. Rev. D, 78, 123523
Percival W. J. et al., 2004, MNRAS, 353, 1201
Perlmutter S. et al., 1999, ApJ, 517, 565
Pezzotta A. et al., 2017, A&A, 604, A33
Pillepich A. et al., 2018, MNRAS, 475, 648
Planck Collaboration et al., 2020, A&A, 641, A6
Potter D., Stadel J., Teyssier R., 2017, Comput. Astrophys. Cosmology, 4, 2
Puchwein E., Baldi M., Springel V., 2013, MNRAS, 436, 348
Ramachandra N., Valogiannis G., Ishak M., Heitmann K., 2021, Phys. Rev. D, 103, 123525
Rasmussen C. E., Williams C. K., 2008, Gaussian Processes for Machine Learning. MIT Press, Cambridge, MA
Riess A. G. et al., 1998, AJ, 116, 1009
Savitzky A., Golay M. J. E., 1964, Analytical Chemistry, 36, 1627
Schaye J. et al., 2015, MNRAS, 446, 521
Schmidt F., 2009, Phys. Rev. D, 80, 123003
Secco L. F. et al., 2021, PRD, 105, 2
Sotiriou T. P., Faraoni V., 2010, Rev. Modern Phys., 82, 451
Springel V., 2010, MNRAS, 401, 791
Springel V. et al., 2005, Nature, 435, 629
Springel V. et al., 2018, MNRAS, 475, 676
Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, ApJ, 761, 152
Tassev S., Zaldarriaga M., Eisenstein D., 2013, J. Cosmol. Astropart. Phys., 06, 036
Terukina A., Lombriser L., Yamamoto K., Bacon D., Koyama K., Nichol R. C., 2014, J. Cosmol. Astropart. Phys., 4, 013
Teyssier R., 2002, A&A, 385, 337
Vikhlinin A. et al., 2009, ApJ, 692, 1060
Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, Phys. Rep., 530, 87
Weinberger R., Springel V., Pakmor R., 2020, ApJS, 248, 32
Will C. M., 2014, Living Rev. Relativ., 17, 4
Winther H. A. et al., 2015, MNRAS, 454, 4208
Winther H., Casas S., Baldi M., Koyama K., Li B., Lombriser L., Zhao G.-B., 2019, Phys. Rev. D, 100, 123540
Zarrouk P. et al., 2018, MNRAS, 477, 1639
Zhai Z. et al., 2019, ApJ, 874, 95
Zhao G.-B., Li B., Koyama K., 2011, Phys. Rev. D, 83, 044007

APPENDIX A: EMULATOR PERFORMANCE AT HIGHER REDSHIFTS

As mentioned in the main part of the paper, the 3D power spectrum emulator is available not only at $z = 0$ but also for higher redshifts. In this appendix, we show the cross-verification results for $z = 1$ and $z = 2$. At even higher redshifts, the $f(R)$-gravity effect on the power spectrum is expected to be negligible on the scales predicted by our emulator. The ratio between the power spectrum measured from our simulations and HALOFIT is therefore dominated by the inaccuracies of both and cannot reliably be predicted by an emulator. Fortunately though, the need for an emulator at those higher redshifts is not as strong as for low z anyway.

As one can see from Figs A1 and A2 the emulator produces very accurate results for both redshifts. At $z = 1$ all nodes except nodes 3 and 7 – which have a very low value of $\Omega_{\rm m}$ (see the discussion for $z = 0$ in Section 4.4) – are within the 5 per cent error margin and most nodes are completely within the 2 per cent error band. For $z = 2$ we find a similar accuracy from the cross-validation tests. As one can see from Fig. A2, some nodes show a large emulator error on very small scales ($k > 2\,h\,{\rm Mpc}^{-1}$). A closer look at the power spectrum behaviour for these nodes shows that they either take a very low value for $B(k)$, which ultimately leads to large relative errors even if the deviation is small, or experience a very wiggly behaviour of $B(k)$ for $k > 2\,h\,{\rm Mpc}^{-1}$. We therefore re-emphasize that the marked region ($k < 2\,h\,{\rm Mpc}^{-1}$) is the parameter range for which we trust the emulator for reasonable cosmologies.

To keep this section brief, we do not show the results for the intermediate redshifts for which the emulator is also available here. We did nevertheless check that the emulator performance is better than that at $z = 0$ for all available redshifts.

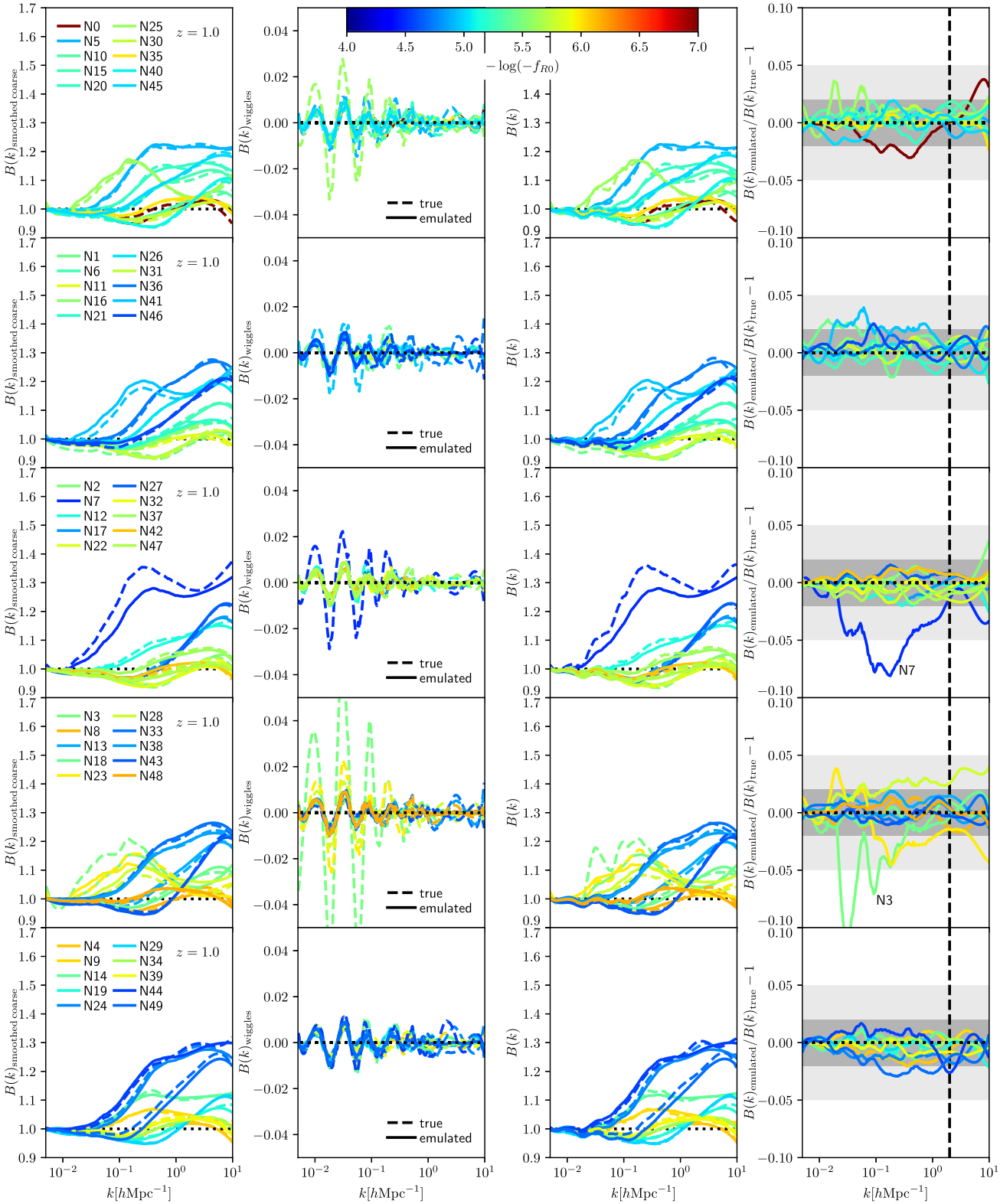

Figure A1. Same as Fig. 5 but at $z = 1$.

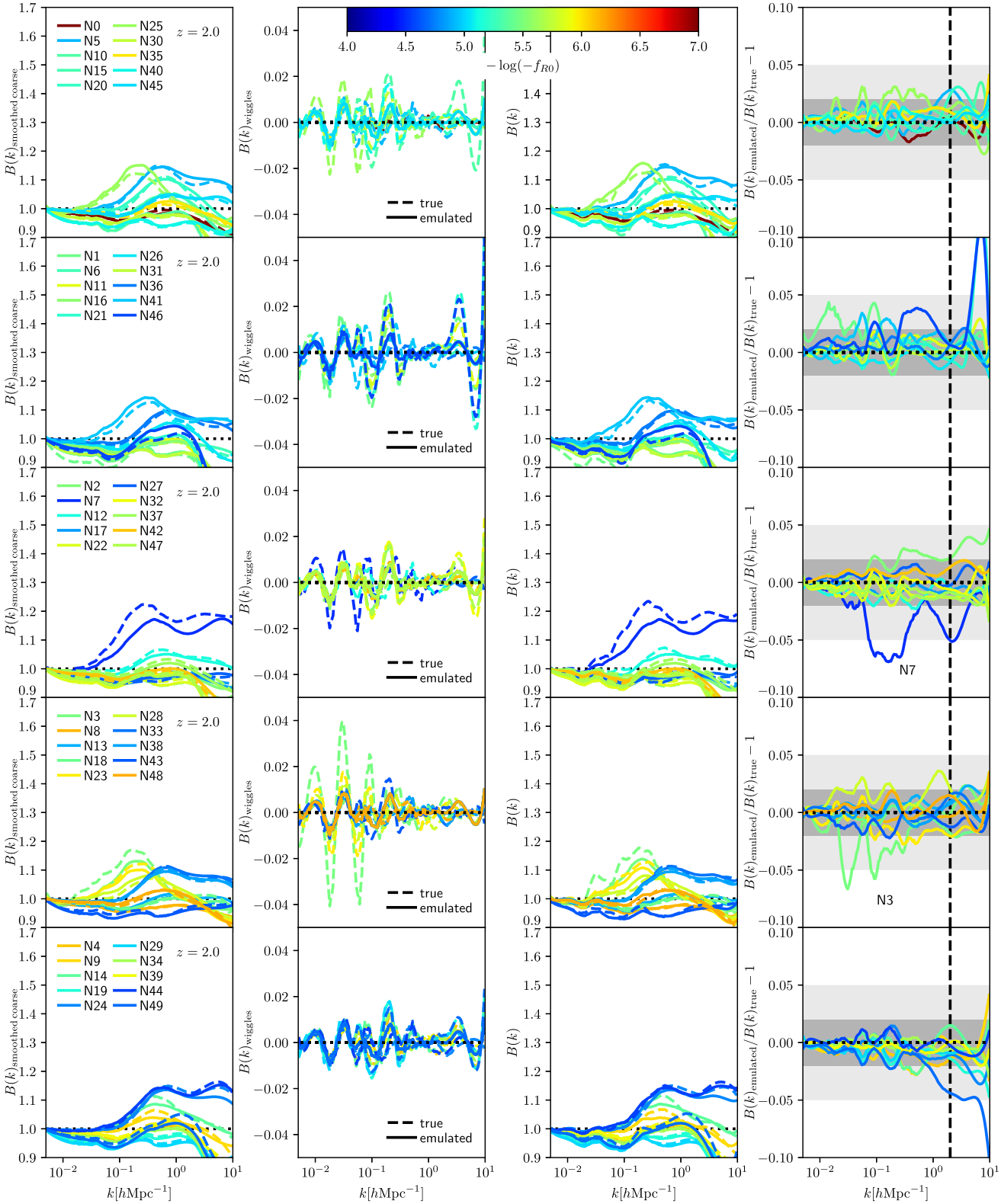

Figure A2. Same as Fig. 5 but at $z = 2$.

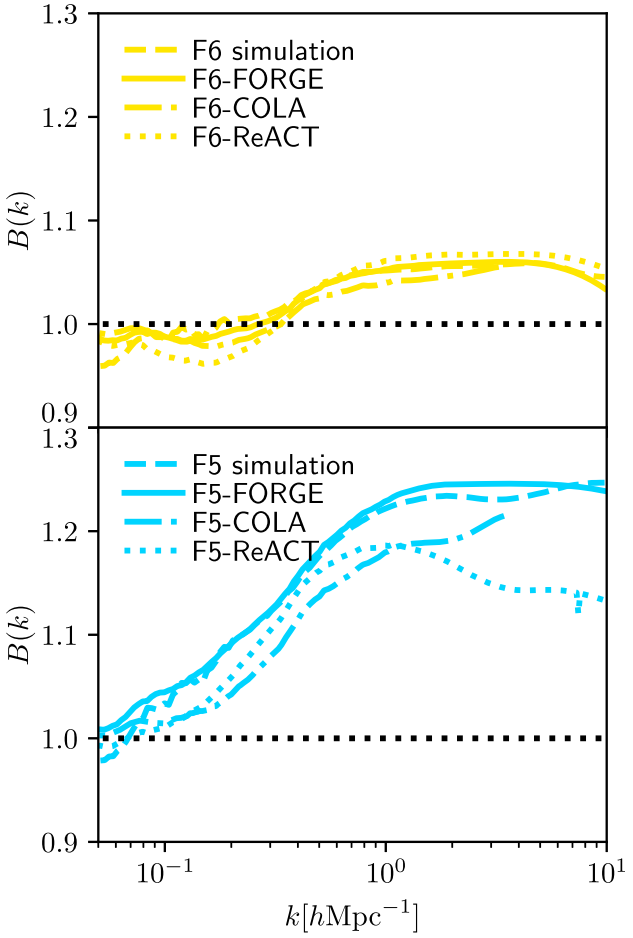

Figure B1. The average power spectrum response $B(k)$ measured from the eight node-0 cosmology simulations (dashed lines) for F6 (top panel, yellow) and F5 (bottom panel, cyan) compared to several different predictions. The responses from the FORGE emulator presented in this paper are shown as solid lines, predictions made with the emulator based on MG-COLA simulations (Ramachandra et al. 2021) as dash–dotted lines, and the $B(k)$ given by REACT (Cataneo et al. 2019; Bose et al. 2020, 2021) as dotted lines. Black dotted lines indicate unity. The REACT and MG-COLA predictions are corrected for the difference between ΛCDM simulations and HALOFIT in this plot (see the text for more details).

APPENDIX B: COMPARISON WITH OTHER *f*(*R*) MATTER POWER SPECTRUM PREDICTIONS

To compare the performance of the FORGE power spectrum emulator with other theoretical predictions for the matter power spectrum in $f(R)$ gravity, we plot the predictions of different approaches for $B(k)$, as well as our simulation results for F5 and F6 and node-0 background cosmology, in Fig. B1. As we have already seen from Fig. 6, the FORGE emulator matches the result of the eight F6 and F5 simulations (reproduced here as dashed lines) on a 1–2 per cent level for scales up to $k = 10\,h\,\mathrm{Mpc}^{-1}$. Further, we have shown the predictions by an $f(R)$ power spectrum emulator constructed using MG-COLA simulations[4] (Ramachandra et al. 2021, dash–dotted lines), and by the reaction method REACT[5] (Cataneo et al. 2019; Bose et al. 2020, 2021, coloured dotted lines). As the two additional approaches give $P(k)$ or its enhancement with respect to ΛCDM, rather than $B(k)$, to get the latter we used the definition (10) with the numerator replaced by the predictions of these approaches.

For F6, the MG-COLA emulator prediction agrees equally well with the simulations within its range of validity (the authors do not recommend using it for scales of $k > 1\,h\,\mathrm{Mpc}^{-1}$), but it differs from the simulation result by about 5–7 per cent for F5. REACT also agrees very well with the simulation result for F6, while for F5 it agrees within about 5 per cent for $k < 0.7\,h\,\mathrm{Mpc}^{-1}$ but then starts to deviate significantly on smaller scales. The difference between REACT and our F6/F5 simulations (and therefore FORGE) is likely because in the theoretical prediction by the former the fitting functions for ΛCDM [as opposed to the $f(R)$ model] halo mass functions and concentration–mass relations are used (these will be completed in upcoming public versions of REACT).

To summarize the performance of the different power spectrum predictions, one can say that all these methods perform well for F6, but there are differences for stronger $f(R)$ models. The good agreement in F6 is not surprising, since the power spectrum in this model differs from that in ΛCDM by at most $\simeq 5$ per cent for the whole k-range. While the FORGE emulator reproduces the simulation result for F5 to great accuracy as well, the other two methods show stronger deviations. This result has to be taken with caution though, for the reason mentioned above, and also because REACT and MG-COLA have been calibrated or trained with simulations produced with other simulation codes, which might lead to differences of up to a few per cent in the training data already.

APPENDIX C: HOW TO USE THE POWER SPECTRUM EMULATOR

The FORGE matter power spectrum emulator and the power spectrum response data are publicly available from a git repository at https://bitbucket.org/arnoldcn/forge_emulator/. The repository contains all the training data used to train the emulator, as well as the emulator package in the form of a PYTHON module (`GPR_Emulator.py`). This module can be imported into any PYTHON application. The repository also contains an example for a cross-validation application of the emulator using the power spectrum data, as well as an example script for training the emulator on the power spectrum data and making a prediction for a certain parameter combination. Note that when the emulator has been trained on the power spectrum data once, its internal state is saved and will be reloaded for predictions. It is hence not necessary to retrain the emulator before making predictions enabling MCMC applications. A more detailed breakdown of the package functionalities is given in a README included in the repository.

[4]https://github.com/LSSTDESC/mgemu.
[5]https://github.com/nebblu/ReACT/tree/react_with_neutrinos.

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