# Matter power spectrum emulator for math xmlns="http://www.w3.org/1998/Math/MathML" display="inline">mi>f/mi>mo stretchy="false">(/mo>mi>R/mi>mo stretchy="false">)/mo>/math> modified gravity cosmologies

Nesar Ramachandra, Georgios Valogiannis, Mustapha Ishak, and Katrin Heitmann (LSST Dark Energy Science Collaboration)

Phys. Rev. D **103**, 123525 — Published 10 June 2021

DOI: 10.1103/PhysRevD.103.123525

# Matter Power Spectrum Emulator for $f(R)$ Modified Gravity Cosmologies

Nesar Ramachandra,[1,*] Georgios Valogiannis,[2,3,*] Mustapha Ishak,[4] and Katrin Heitmann[1]
(The LSST Dark Energy Science Collaboration)
[1]*High Energy Physics Division, Argonne National Laboratory, Lemont, IL 60439, USA*
[2]*Department of Astronomy, Cornell University, Ithaca, NY 14853, USA*
[3]*Department of Physics, Harvard University, Cambridge, MA 02138, USA*
[4]*Physics Department, The University of Texas at Dallas, Richardson, Texas 75080, USA*
(Dated: May 5, 2021)

Testing a subset of viable cosmological models beyond General Relativity (GR), with implications for cosmic acceleration and the Dark Energy associated with it, is within the reach of Rubin Observatory Legacy Survey of Space and Time (LSST) and a part of its endeavor. Deviations from GR-$w(z)$CDM models can manifest in the growth rate of structure and lensing, as well as in screening effects on non-linear scales. We explore the constraining power of small-scale deviations predicted by the $f(R)$ Hu-Sawicki Modified Gravity (MG) candidate, by emulating this model with COLA (COmoving Lagrangian Acceleration) simulations. We present the experimental design, data generation, and interpolation schemes in cosmological parameters and across redshifts for the emulation of the boost in the power spectra due to Modified Gravity effects. Three preliminary applications of the emulator highlight the sensitivity to cosmological parameters, Fisher forecasting and Markov Chain Monte Carlo inference for a fiducial cosmology. This emulator will play an important role for future cosmological analysis handling the formidable amount of data expected from Rubin Observatory LSST.

## I. INTRODUCTION

The Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST) [1] [1, 2], together with a wide range of current and future surveys of the large-scale structure (LSS) of the universe, such as DESI [3], Euclid [4], the Nancy Grace Roman Space Telescope [5] and SPHEREx [6], will offer a unique opportunity to test our standard cosmological assumptions at an unprecedented level of accuracy. The widely accepted cosmological model, $\Lambda$ Cold Dark Matter ($\Lambda$CDM), attributes the observed accelerated expansion of the universe [7, 8] to the existence of a positive cosmological constant, $\Lambda$, corresponding to non-zero vacuum energy. This assumption, combined with the existence of pressure-less cold dark matter and gravity described by Einstein's General Relativity (GR), has been very successful at fitting a large spectrum of cosmological observations [9–14].

Despite these remarkable observational accomplishments, $\Lambda$CDM has faced several theoretical challenges, with the cosmological constant problem [15, 16] arguably serving as the primary reason to consider alternative proposals. Furthermore, as our ability to accurately obtain the underlying cosmological parameters from late and early-time observations increases, attention has been drawn to certain tensions between the corresponding extracted values of the Hubble constant, $H_0$ [17–20], and the amplitude of density fluctuations, $\sigma_8$ [14, 21–24], which could however be attributed to unknown systematics. Combined with the long-term interest in exploring deviations from GR in all regimes [25, 26], the above motivate introducing large-scale modifications of gravity as alternative candidates for cosmic acceleration; such theories are called the Modified Gravity (MG) theories [27–30].

In order to be able to evade the existing tight constraints of gravity from observations in the Solar System [25, 26], while at the same time producing detectable large-scale signatures, viable MG candidates typically invoke "screening" mechanisms [31, 32], which suppress deviations in the high-density regime through novel scalar field self-interactions [33–39]. Furthermore, the space of all MG parametrizations that lead to second order equations of motion, the Horndeski class [40–42] has been additionally restricted [43–48] by the simultaneous detection of gravitational waves and electromagnetic counterparts by the LIGO/Virgo collaborations [49–53]. A detailed discussion of viable MG candidates testable by LSST Dark Energy Science Collaboration (DESC) was presented in Ref. [54].

The predicted transition from MG to GR would manifest itself, by means of the dynamical screening mechanism, in the nonlinear regime of structure formation, which will be precisely probed by the Stage-IV surveys of the LSS [55]. As a result, these upcoming observations will offer a unique opportunity to study the large-scale behavior of gravity with unprecedented accuracy. The optimal interpretation of the wealth of upcoming data, however, is conditional upon our ability to produce efficient and reliable theoretical predictions of the expected observable signatures of MG. In the (quasi-)linear regime of structure formation, this can be partially achieved through analytical, perturbation theory approaches, such as Lagrangian Perturbation Theory (LPT) [56–60]. Unfortunately, these approaches break down on nonlinear

* The two lead authors have contributed equally to this work.
[1] `http://www.lsst.org`

scales. Reliable predictions of such small-scale signals from MG theories can only be obtained through performing full N-body simulations (for a comparison of different codes, see Ref. [61]). These are computationally very expensive, particularly in the presence of an additional MG-induced force, due to the inherent nonlinearities introduced by the screening mechanism.

As a consequence of the substantial computational costs associated with performing full N-body simulations for MG models, efficient approaches are essential. To that end, the hybrid COmoving Lagrangian Acceleration (COLA) method was developed in Ref. [62], expanding upon the initial ΛCDM implementation of Ref. [63], for efficiently simulating the MG classes of chameleon and symmetron screening. Utilizing a combination of 2$^{\text{nd}}$ order Lagrangian Perturbation Theory (2LPT) and a pure N-body component, the COLA method provides a great trade-off between accuracy and computational efficiency, that was found to recover the fractional deviation in the nonlinear dark matter power spectra with sufficient accuracy, at only a fraction of the standard computational cost. However, many investigations require even faster prediction capabilities to enable the exploration of parameter space. For example, Markov Chain Monte-Carlo (MCMC) inference relies on tens of thousands of model evaluations and even the fast COLA approach, which takes $\sim 1$ hour on a single core for one MG simulation, would be too slow to enable such an investigation.

To provide faster predictions for, e.g., the power spectrum, fitting functions have been used extensively in the past (see, e.g., Refs. [64, 65] for fits capturing ΛCDM and $w$CDM cosmologies based on the Halofit approach or a halo model-based approach to include baryonic effects [66] or physics beyond ΛCDM [67]). However, fitting functions have several drawbacks. First, the accuracy requirements for ongoing and future surveys of a few to sub-percent are very difficult to obtain by a single functional form and a set of fitting parameters. Second, for models outside the range for which the fit was derived, no error bounds are available and biases can occur. Third, a large range of simulations is needed to enable a good calibration of the parameters that describe the fit. Overall, it has been shown that even for the relatively restricted case of $w$CDM cosmologies, it is very difficult to achieve an accuracy of better than 5–10% as pointed out in one of the recent Halofit papers [65] and subsequent comparisons, e.g., Refs. [68, 69].

Because of the above shortcomings of fitting functions and the need for very accurate and fast predictions, the concept of emulators was introduced to cosmology in Refs. [70, 71]. It was shown that with a relatively small number of high-quality simulations, prediction schemes could be built that provide high-accuracy results for, e.g., the matter power spectrum and $C_\ell$s quickly. The `Coyote Universe` project [72–74] then released a stand-alone prediction scheme for the matter power spectrum, the `CosmicEmu`, based on a set of highly accurate simulations. The work was extended in several ways, including the coverage of a larger $k$-range and redshift and parameter spaces [68, 75]. The emulation concept itself has since then become rather popular and was used in several studies concerning the matter power spectrum, see, e.g. Refs. [69, 76], a range of other summary statistics, e.g., galaxy power spectra [77, 78], concentration mass relation [77], the halo mass function [79] as well as comprehensive simulation efforts that extracted a range of emulators, such as the `AEMULUS` project [80–83] and `DARK QUEST` [84, 85].

In this work we develop a Gaussian Process emulator to estimate the fractional deviation of the nonlinear matter power spectrum $P_{MG}(k)/P_{\Lambda CDM}(k)$ (also referred to as enhancement/boost in the power spectrum or the power spectrum ratio) as a function of cosmological parameters and redshift, based on a set of COLA simulations in the MG scenario. Given that our focus is to efficiently emulate statistics for MG models that will be testable by LSST DESC, our target model needs to be one that exhibits a well-studied phenomenology (including existing full N-body simulations) and will predict detectable deviations in the scales of interest to the survey. For this reason our chosen candidate is the $f(R)$ Hu-Sawicki model [86], which realizes the chameleon screening mechanism and which we have previously identified as one of the prioritized beyond-$w(z)$CDM candidates testable by LSST DESC [54].

In this paper we discuss the construction of the emulator that comprises of an experimental design, training data synthesis and the statistical techniques to perform interpolation across cosmological parameters and redshifts. After training the emulator on the COLA-generated dataset, we validate its ability to recover simulated test cosmologies within our target range, before proceeding to compare its accuracy against results obtained by full N-body simulations for the Hu-Sawicki model. Having quantified its accuracy, we then illustrate the capabilities of our emulator, which delivers a massive speed-up by 6 orders of magnitude over COLA simulations, through three essential applications: a sensitivity analysis, obtaining parameter constraints through Fisher forecasting and MCMC inference. It is worth noting, at this point, that efficient predictions for power spectra in the Hu-Sawicki MG model have also been recently presented in Refs. [87–89], which however relied upon appropriately designed fitting formulas [87] or semi-analytical models [88, 89]. In order to overcome the potential limitations associated with such approaches, as mentioned above, we have developed our emulator based directly on the simulations, expanding upon the established $w$CDM infrastructure of the `CosmicEmu` [90]. In addition to enhancing the variety of available predictive tools in the community, which is an important endeavor by itself, this contribution is novel. While the above techniques probe variations of only one of the two parameters of the Hu-Sawicki model, our emulator spans its full 2-dimensional parameter space, for the first time in the literature.

Our paper is structured as follows: in Sec. II, we describe the approach to generating our training data set by first introducing our target MG model, then discussing our choices for the experimental design, including cosmological parameters and their ranges, and finally describing the efficient COLA approach we use to generate the simulations. Next, we discuss in Sec. III the details of our emulator development. We then proceed in Sec. IV to validate the emulator results, before presenting three applications in Sec. V. Finally, we conclude and discuss future work in Sec. VI. Technical details on Gaussian Processes and the emulator design can be found in Appendices B and C, respectively.

## II. TRAINING DATA AND DESIGN

### A. Modified Gravity Model

Theoretical investigations of potential departures from Einstein's GR, as well as of their consequent observational implications [25], have been an active research topic, particularly in the last two decades, due to their potential implications on resolving the mystery of cosmic acceleration [27, 29, 30]. One of the most commonly considered classes of MG deviates from GR through the addition of a nonlinear function $f(R)$ of the Ricci scalar $R$ to the standard Einstein-Hilbert action. These are the $f(R)$ gravity theories [91, 92], which are described by the following action $S$:

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

with $\mathcal{L}_m$ denoting the matter sector Lagrangian, $G$ the gravitational constant and using units where the speed of light in vacuum, $c$, has been set to unity. The modification of the form (1) activates, in principle, a new additional degree of freedom in the gravitational sector, which could be responsible for driving the cosmic expansion to accelerate, instead of dark energy [93].

The most widely-considered model of this type is the $f(R)$ Hu-Sawicki model [86], with the functional form:

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

where in Equation (2) $m = H_0\sqrt{\Omega_m}$, with $\Omega_m$ the matter fractional energy density and $H_0$ the Hubble Constant, both evaluated today and $n$, $c_1$ and $c_2$ are free parameters of the model.

Any well-motivated MG parametrization should have the flexibility to match the observed expansion history well-described by ΛCDM, a requirement which gives (for sufficiently small values of $|f_{R_0}|$), the following relationship for the background value of the Ricci scalar, $\bar{R}$:

$$\bar{R} = 3\Omega_m H_0^2\left(1 + 4\frac{\Omega_\Lambda}{\Omega_m}\right), \tag{3}$$

and for the derivative $f_R = \frac{df(R)}{dR}$,

$$\bar{f}_{R_0} = -n\frac{c_1}{c_2^2}\left(\frac{\Omega_m}{3(\Omega_m+\Omega_\Lambda)}\right)^{n+1}, \tag{4}$$

where $\Omega_\Lambda$ is the dark energy fractional density evaluated at the present time. Through relationship (4), one can reduce the number of free parameters of the Hu-Sawicki model, which can be fully characterized by the pair $\{|f_{R_0}|, n\}$. We briefly point out here that in the limit of $|f_{R_0}| \to 0$ the background cosmology of ΛCDM is recovered. Through an appropriate conformal transformation, the model can be cast into a scalar-tensor theory (with $f_R$ acting as the MG scalar field) [94] that recovers GR in regions of large Newtonian potential through the chameleon screening mechanism [35, 36].

The Hu-Sawicki model produces novel, distinct signatures which are testable by upcoming cosmological surveys and as a result has been well-explored in the literature through full N-body simulations. It is also known to be free from any physical instabilities [28]. We further note that, despite the increasingly tight observational constraints placed on it over the past decade (see, e.g., [92] for a review), the Hu-Sawicki model remains viable. For all these reasons, it serves as an ideal test-bed for investigating cosmological theories of gravity and is one of the main candidate models to be considered by DESC [54]. The referenced document describes the various beyond-w(z) CDM model prioritized for study by DESC.

### B. Experimental Design

Two main criteria govern the choice of the cosmological parameter space covered by our emulator – first, a consideration of the computational cost incurred in generating a set of COLA simulations, which will determine how many models can be reasonably run, and, second, an estimate for the emulator's target accuracy. The application of these two criteria determines how many parameters we can afford to include and how broad their priors can be.

Our main interest in this paper is the exploration of the two parameters that define the Hu-Sawicki model for $f(R)$ gravity theories, $\{f_{R_0}, n\}$. We aim to vary these parameters over a wide range in order to span a broad array of deviations predicted by the model. We choose

$$-8 \leq \log(f_{R_0}) \leq -4, \tag{5}$$

$$0 \leq \quad n \quad \leq 4, \tag{6}$$

and note that the upper end of the chosen $f_{R_0}$ range corresponds to a large modification case in which screening is absent, whereas for the low end value, modified gravity forces are very strongly suppressed [86, 95]. Most studies in the MG literature, e.g., Refs. [87, 89], restrict their attention on the subspace of fixed $n = 1$ and only vary $f_{R_0}$, given that deviations are more sensitive to the latter parameter. In this work, however, we also consider variations with respect to $n$, thus allowing a more complete theoretical exploration of the Hu-Sawicki model parameter space. Given the wide dynamical range of $f_{R_0}$, over-sampling of large parameter values may occur if the parameter range is sampled linearly. In order to mitigate this problem, we instead use a logarithmic scaling in the experimental design.

Next, we have to decide which additional parameters we want to vary for the CDM component of the model. We want to focus on the parameters that affect the power spectrum the most and at the same time aim to vary only a small number of parameters. An increase in the number of cosmological parameters varied leads to either a less accurate emulator or a much larger number of simulations needed. It is therefore desirable to keep the number of parameters varied small while keeping the needed accuracy.

Figure 10 in Ref. [73] shows a sensitivity analysis for the five parameters of the $w$CDM "Standard Model" of cosmology. The image shows the variation of the ratio of the power spectrum to a model that is evaluated at the midpoint of the five Standard Model parameters changing one parameter at a time. The baryon density $\Omega_b$ clearly does not have much of an effect on the power spectrum ratio, while $\{\Omega_m h^2, \sigma_8, n_s\}$ all show considerable impact on particularly large scales. While the dark energy equation of state also leads to considerable variations, $w$ is not of interest here as we fix the background of a $\Lambda$CDM model. We note again that the requirement to match a $\Lambda$CDM background expansion imposes additional restrictions on the parameter space of the model, which are reflected in Equations (3) and (4) that we introduced Sec II A. As we briefly discuss at the end of Section IV B, variations of $h$ given current observational bounds do not affect the fractional deviation of the nonlinear matter power spectrum $P_{MG}(k)/P_{\Lambda CDM}(k)$ considerably. We therefore allow variations of the aforementioned additional three parameters for our emulator. In future work, the parameter space will be extended.

We now proceed to set the range for three of the standard CDM parameters $\{\Omega_m h^2, \sigma_8, n_s\}$. We choose the same range used as in the Coyote Universe [73] and the Mira-Titan Universe simulation suites [96]. Both papers provide in-depth discussions about the choices to balance broad parameter coverage and achievable accuracy goals for the emulator, as informed by contemporary and future cosmic microwave background (CMB) and LSS surveys. Choosing the same parameter ranges also has the advantage that results in the future can be compared to the previous work and possibly combined later on. In summary, for the three parameters, we choose:

$$0.12 \leq \Omega_m h^2 \leq 0.15, \tag{7}$$

$$0.85 \leq n_s \leq 1.1, \tag{8}$$

$$0.7 \leq \sigma_8 \leq 0.9. \tag{9}$$

We add that $\Omega_b h^2 = 0.0223$ (while we ignore the effects of massive neutrinos) and that for the dimensionless Hubble constant we have chosen $h = 0.67$. After having decided on the five parameters and their ranges we have to determine the number of simulations needed and pick a sampling scheme that allows us to set up a suitable parameter sampling design. Past experience has shown that roughly 10 simulations per parameter are needed to enable the construction of an accurate emulator (targeting few percent accuracy), leading to a set of $\sim 50$ COLA simulations.

The next choice to be made concerns the employed sampling scheme. For an excellent discussion on different sampling schemes used in computer experiments, the interested reader is referred to Ref. [97]. Ref. [73] provides an extensive discussion about different methods in the cosmology context. In this paper, we use a symmetric Latin hypercube (SLH) design, introduced in Ref. [98]. Latin hypercube (LH) sampling schemes are statistical stratified sampling methods used to generate near-random samples of values from a multi-dimensional distribution, such that there exists only one sample in each sub-division for each parameter range. Compared to a random space-filling scheme, which does not take into account the previous sampled points in a new sample-point generation, an LH sampling strategy guarantees an optimal representation of the variability of parameters. While sampling on a uniform grid also ensures fair representation, the number of required simulations would be prohibitively large. The SLH offers a space-filling design strategy that allows for flexibility with regard to a number of design points and is computationally inexpensive when optimizing the design itself, when compared to other methods such as Orthogonal Array LH implementations. The SLH imposes additional, specific symmetry requirements compared to other LH designs, as described in detail in e.g. Ref. [73]. Ref. [73] provides concrete examples to illustrate these symmetries. The symmetry imposes a space-filling requirement on the designs considered upfront, which carries through to all projections.

The final choice concerns the proposal distribution for the cosmological parameters. We use uniform distributions for the cosmological parameters $\theta = \{\Omega_m h^2, \sigma_8, n_s, \log(f_{R_0}), n\}$ to ensure an unbiased exploration of parameter space. The 50 cosmological models that are chosen using the above prescription are shown in Appendix C. In other efforts, the design has been created based on posteriors from surveys see, e.g., the `AEMULUS` project [99]. This approach reduces the number of required simulations but also restricts the viability of the emulator considerably, because of the limited effective sampling volume and because of potential biases introduced by weighted sampling.

### C. COLA Simulations with Modified Gravity

The signatures introduced by MG models manifest themselves in the nonlinear regime of structure formation, where the screening mechanism is in full effect, as

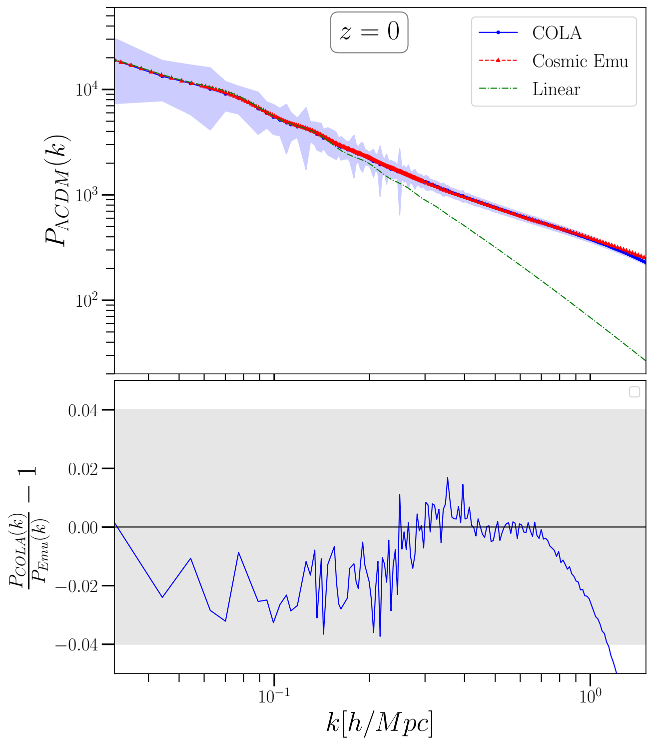


FIG. 1. Top: ΛCDM matter power spectra for the fiducial cosmological parameters at $z=0$, as obtained by the efficient COLA method (blue solid line) and `CosmicEmu` [red dashed line]. The linear matter power spectrum for the same cosmology [green dot-dashed line] is also shown. The shaded blue region represents the standard deviation obtained from the 2000 available $P(k)$ COLA realizations. Bottom: Fractional difference between the COLA and the `CosmicEmu`-generated power spectra of the upper panel. The shaded gray region highlights the 4% level of accuracy.

well as in the intermediate quasi-linear scales. As a result, perturbation theory approaches fail to capture the full picture of structure formation in the presence of a modification to gravity, which can only be performed through N-body simulations. These are particularly computationally expensive, due to the inherent non-linearity of the scalar-field Klein-Gordon equation. An overview and comparison of different full N-body codes in MG can be found in Refs. [61].

Given the substantial computational cost of running multiple full N-body simulations to train our emulator, we instead utilize the Comoving Lagrangian Acceleration (COLA) hybridization scheme for efficient simulations of chameleon and symmetron screening models developed in Ref. [62], expanding upon the initial ΛCDM implementation of Ref. [63]. Through a combination of $2^{\rm nd}$ order Lagrangian PT (2LPT) that evolves the large scales and a pure N-body component for integrating the nonlinear regime, the COLA approach was found to be able to recover the nonlinear dark matter power spectrum (in ΛCDM) using only a few tens of timesteps [63]. The implementation developed for the $f(R)$ Hu-Sawicki

TABLE I. Parameters of the COLA MG simulation suite.

| | |
|---|---|
| Box size, $L$ | 200 $h^{-1}$Mpc |
| Number of particles, $N_p$ | $256^3$ |
| Number of grids, $N_g$ | $512^3$ |
| Initial redshift $z_i$ | 49 |
| Final redshift $z_f$ | 0 |
| Number of time-steps, $N_t$ | 100 |
| Number of realizations, $N_r$ | 1 |
| Dimensionless Hubble constant, $h$ | 0.67 |

model in Ref. [62], which is the one we employ here, was shown to successfully model the fractional deviation in the dark matter power spectrum, $P_{MG}(k)/P_{\Lambda CDM}(k)$, in the nonlinear regime, using an effective "thin-shell" approach for capturing the chameleon screening effect that was presented in Ref. [100]. The reported accuracy in the estimation of this power spectrum ratio was at the level of 1 percent, compared against N-body simulations at $z=0$. Below we summarize the parameters we choose for our COLA runs, while more details about the COLA MG implementation can be found in Ref. [62].

Our simulations are initialized using the 2LPT initial conditions code (`2LPTic`) [101] at an initial redshift of $z_i=49$. Given that the effects of MG are assumed to be negligible at early times within the context of cosmic acceleration, we use the same set of ΛCDM initial conditions for both the ΛCDM and the $f(R)$ runs, as in Ref. [62]. For each of the (50+5) cases in our experimental design (listed in Table III), the linear ΛCDM matter power spectrum is produced with `CAMB` [102], which is used to generate the initial conditions with `2LPTic` at $z_i=49$. The COLA simulations are then run using the parameters of Table I, for both the ΛCDM and the $f(R)$ Hu-Sawicki case with the same initial random seed. At each of the 100 timesteps of the simulations, the matter power spectra are stored, for both ΛCDM and the MG case, with 213 bins equally spaced logarithmically in the $k$-range of $(0.03 - -3.5)h\mathrm{Mpc}^{-1}$. The choice of 100 timesteps is made so that the target redshift range of our emulator is adequately spanned using only a single run for each case, without at the same time making the simulations too time-consuming [2]. Furthermore, the high number of $k$ bins ($n_{bins}=213$) is chosen to guarantee the accuracy of the redshift interpolation. Finally, the exact same specifications are used in the 2000 simulations we perform, each of them for a different randomly chosen initial seed, in order to obtain the covariance matrix of the ratio for the fiducial cosmology of Section V B.

Before we discuss COLA's accuracy in predicting summary statistics in MG (the matter power spectrum ratio here), which will be addressed in detail in Sec. IV B, we start by presenting the ΛCDM benchmark in Fig. 1. We

[2] We should note here that if one wants to make predictions for an individual redshift, the efficient COLA approach was shown to work well with much fewer timesteps [62, 63]

find that the mean of the 2000 COLA-generated $\Lambda$CDM power spectra for the fiducial cosmology remains consistent, within $\sim 3\%$, with the nonlinear `CosmicEmu` prediction for the same cosmology down to $k \sim 1h\text{Mpc}^{-1}$. It is worth noting that the better agreement compared to the initial COLA implementation in Ref. [63], is attributed to the fact that we are using about 3 times as many time-steps as the runs in that original work, for the reasons discussed in the previous paragraph.

Given that the COLA method is known to perform better at recovering the fractional deviation $P_{MG}(k)/P_{\Lambda CDM}(k)$, rather than the power spectrum itself [62], and also because this quantity is much less sensitive to cosmic variance effects, we choose these ratios as our training data in the Gaussian process emulator. Modeling this quantity has also been the target of interest by other studies in the MG literature [89, 103, 104]. The residual noise in the ratio is later smoothed out using a Savitzky-Golay filter, as explained in Sec. III.

## III. EMULATOR DEVELOPMENT

Based on a carefully chosen design strategy to determine a set of training points, a well-matching interpolating strategy can be selected in order to estimate the summary statistics at intermediate cosmologies. Neural networks [76], polynomial chaos expansions [69] and Gaussian processes [70–74] have been successfully employed to construct emulators for the prediction of astrophysical summary statistics. In particular, Gaussian Processes are an attractive way of performing machine learning tasks with small number well sampled data points. This non-parametric Bayesian regression method provides fast, interpretable and high-fidelity estimations with associated uncertainties. For these reasons, we utilize Gaussian processes, along with Principal Component Analysis and linear interpolation schemes to construct our emulator.

### A. Gaussian Process Interpolation across Cosmological Parameters

Our emulation strategy for the cosmological parameters $\theta$ follows a similar routine employed for the `CosmicEmu` [105] construction, using Gaussian processes (GPs) in a representation space [106]. The individual steps, including the data pre-processing are as follows:

1. The individual ratios of the power spectra are noisy since we only perform one realization for each individual COLA simulation. Emulators designed directly based on this data may pick up undesired noise from the data. To avoid this problem, we utilize the Savitzky-Golay filter (or savgol filter, [107, 108]) to obtain a smoothed power spectrum ratio $\chi(k)$, as detailed in Appendix A.

2. A standardization transformation on both smoothed power spectrum ratio and cosmological parameters is performed, to result in a mean of zero and a standard deviation of one for their respective distributions:

$$\theta_i{}' = (\theta_i - \mu_{\theta_i})/\sigma_{\theta_i}, \tag{10}$$

$$\chi'(k) = (\chi(k) - \mu_{\chi(k)})/\sigma_{\chi(k)}. \tag{11}$$

   The standardized power spectrum ratio $\chi'(k)$ is re-scaled using the mean $\mu_{\chi(k)}$ and standard deviation $\sigma_{\chi(k)}$. The mean and standard deviations are computed collectively for all the 50 cosmological models. Similarly the individual cosmological parameters $\theta_i$ are re-scaled to $\theta_i{}'$ by their means $\mu_{\theta_i}$ and standard deviations $\sigma_{\theta_i}$.

3. A Singular Value Decomposition (SVD) is performed on the smoothed and normalized power spectrum enhancement $\chi(k, \theta)$ for dimensionality reduction. This is a generalization of eigenvalue decomposition to any rectangular matrix, whereby a matrix is factorized into a set of orthonormal vectors. Equation (12) below shows the decomposition to the basis $\phi_m(k)$ and weights $w_m(\theta)$ of the representation, truncated at $n_w$ eigenvectors:

$$\chi'(k, \theta') = \sum_{m=1}^{n_w} \phi_m(k) w_m(\theta') + \epsilon. \tag{12}$$

   The excess information that is not captured by this decomposition is represented by $\epsilon$. The Principal Component Analysis (PCA) of the power spectrum enhancement reveals that a total of $n_w = 6$ eigenvectors successfully capture over 99.99 percent of the variance in the data, effectively allowing us to truncate the expansion without significant loss of information. In addition to dealing with a reduced dimension (from $n_{bins} = 213$ to $n_w = 6$) of eigenvectors, this also enables orthogonality in the interpolation space, i.e., the new basis $\phi_m(k)$ that maximizes variance is an uncorrelated representation of $\chi'(k, \theta)$.

4. The weights $w_m(\theta)$ corresponding to $n_w = 6$ truncated orthogonal bases $\phi_m(k)$ are then modeled as functions of the input parameters $\theta$. This local interpolation in parameter space is made using multivariate Gaussian Process regression applied to the weights of the Principal Component bases, as explained in Appendix B.

   Configuration of the covariance function and determination of the associated hyperparameters are the key components for learning the correct GP fit. We choose a popular Matern-5/2 kernel [109], and check for robustness of the emulator accuracy with different choices of covariance functions. We search for the best combination of hyper-parameters of the GP using a fast gradient-based optimization called Adagrad [110].

The four steps above are applied to all the 100 snapshots, resulting in a suite of 100 emulators for smoothed power spectrum deviations. For any new cosmological parameters $\overline{\theta}$, the trained GP generated 6 corresponding weights, and these are multiplied by the PCA basis vectors to generate new power spectrum enhancement values with the $k$-range of the emulator.

### B. Redshift Interpolation

The sampling for the redshifts coverage is treated separately from the sampling of the cosmological parameters. The cosmological parameter values are generated using an SLH design to ensure representation across the five dimensional parameter space. In contrast, the COLA snapshots are all created at the same redshifts between $z_i = 49$ and $z_f = 0$ with linear spacing in corresponding scale factors. Due to this difference in the training sampling, we do not employ the same interpolation scheme for all the parameters. Instead a separate interpolation routine is executed between the outputs of independent GP emulators. Equation (13) shows a simple linear interpolation for an intermediate redshift $z$ when emulator results at two nearest redshifts $z_-$ and $z_+$ are calculated from the previous section:

$$\chi_{emu}(k;\theta,z) = \chi(k;\theta,z_-) + \frac{z-z_-}{z_+-z_-}(\chi(k;\theta,z_+) - \chi(k;\theta,z_-)). \quad (13)$$

With our emulator suite for 100 individual redshifts, a simple linear interpolation works within one percent accuracy for the redshift interpolation when the closest two emulators are used for a given cosmology. Although this requires two independent GP evaluations and PCA reconstructions, this final emulator $\chi_{emu}(k,\Omega_m,\sigma_8,n_s,\log(f_{R_0}),n,z)$ is found to be robust and fast.

## IV. EMULATOR ACCURACY AND VERIFICATION

The accuracy of our emulator is determined by two factors. First, limitations of the underlying simulations lead to irreducible errors. We have chosen to use Lagrangian Perturbation Theory for the cosmological simulations over a computationally expensive full N-body alternative. This choice restricts the accuracy of the emulator on small scales. Second, an error arises due to the limited number of training samples and the nature of the sampling and interpolation schemes.

We study these effects by carrying out two types of verifications. For the first test we compare the emulator against a set of six additional COLA simulations that are not part of the design. This allows us to evaluate the accuracy of the emulator construction itself. The second test utilizes three state-of-the-art N-body simulations for beyond $\omega(z)$CDM cosmologies and we compare the emulator performance directly to the measurements from the simulations. This test allows us to evaluate the overall performance of the emulator, including both errors from the limited simulation accuracy and the emulator construction itself.

TABLE II. Cosmological parameters for the six additional COLA test simulations.

| Model | $\Omega_m h^2$ | $n_s$ | $\sigma_8$ | $\log(f_{R_0})$ | $n$ |
|---|---|---|---|---|---|
| T00 | 0.125 | 0.957 | 0.860 | −6.667 | 0.000 |
| T01 | 0.136 | 1.023 | 0.833 | −4.000 | 2.133 |
| T02 | 0.150 | 0.970 | 0.820 | −6.133 | 3.200 |
| T03 | 0.132 | 0.890 | 0.793 | −5.867 | 2.933 |
| T04 | 0.129 | 0.983 | 0.807 | −6.400 | 3.733 |
| T05 | 0.127 | 1.050 | 0.740 | −4.267 | 0.267 |

We restrict the emulator predictions to $k \leq 1h\mathrm{Mpc}^{-1}$. This choice is mainly driven by the accuracy of the COLA simulations. As shown in Fig. 1, the COLA approach is in very good agreement with measurements from high-resolution N-body simulations, represented by the `CosmicEmu` result, out to $k \sim 1h\mathrm{Mpc}^{-1}$. At this point, the COLA power spectrum starts to deviate from the `CosmicEmu` result and underpredicts the power spectrum at the few percent level. Restricting our emulator out to this $k$-range therefore seems well justified. In addition, beyond $k \sim 1h\mathrm{Mpc}^{-1}$ other effects, like baryonic physics become more and more important (see, e.g., Ref. [111] for a recent discussion of the effects of baryonic physics on the power spectrum).

### A. Comparison with COLA Simulations

In this section we show the comparison of the emulator with COLA results. Our trained emulator is tested on parameter values within the limits of our SLH design, but not at the exact cosmologies used to construct the emulator.

In Fig. 2 we show the predictions of the emulator compared to an additional test set of six COLA simulations. The power spectrum ratios for the additional cosmologies ($T00 - T05$, parameters are given in Table II) are randomly chosen within the allowed parameter ranges specified in Equations. (5) – (9). The gray lines in the figure show all power spectrum ratios used to build the training set. The lower panel in the figure shows the relative error of the emulator output to the corresponding COLA results. The relative difference is within 5% for our six test models. Sub-percent level accuracy is observed for models with $f_{R_0} < 10^{-5}$. The variation in accuracy corresponds to the sampling density of power spectrum ratios in the training set. For instance, the

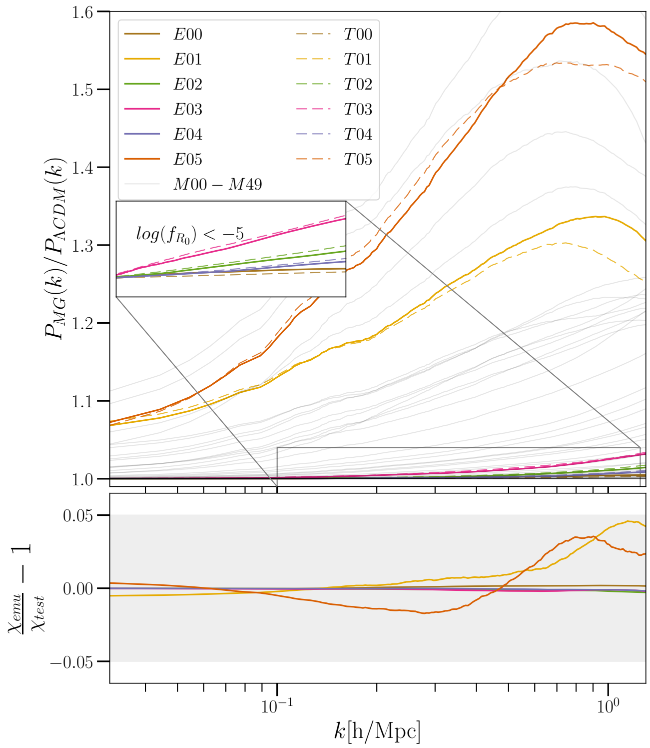


FIG. 2. Test of the emulator accuracy for additional cosmologies. Top: Power spectrum ratios $P_{MG}(k)/P_{\Lambda CDM}(k)$ for the six testing COLA simulations at $z = 0$, $T00-T05$, are compared to the corresponding emulator results, $E00 - E05$. Training cosmologies, $M00 - M49$, are shown in gray. Bottom: Relative error of the emulator prediction compared to the COLA simulations. For the models with large values for $f_{R_0}$ ($T01$ and $T05$) the emulator deviates by up to 5%. The models with $f_{R_0} < 10^{-5}$ are predicted at sub-percent accuracy (corresponding zoom-in panel shown at the top).

test simulation $T01$ with $\log(f_{R_0}) = -4$ is at the very edge of our SLH design used for the training, and the $T05$ model is also close to the edge of the training design, with $\log(f_{R_0}) = -4.267$. By design, we focus more on exploring the modified gravity sector of $f_{R_0} < 10^{-5}$ that corresponds to smaller boosts in the matter power spectra. We stress that the required accuracy in the estimation of the power spectrum is 1-2 percent [112], down to $k \sim 1h\text{Mpc}^{-1}$, which we satisfy for most of the target parameter space.

### B. Comparison with N-body Simulations

Next, we present a comparison of the MG emulator predictions against state-of-the-art N-body simulations for the Hu-Sawicki model. These are the Extended LEnsing PHysics using ANalaytic ray Tracing `ELEPHANT` simulations [113], that were performed with a modified version of the `RAMSES` code, the `ECOSMOG` module [114, 115]. Power spectra at various redshifts have been measured for the Hu-Sawicki case of $n = 1$ and three variations of $|\bar{f}_{R_0}| = \{10^{-6}, 10^{-5}, 10^{-4}\}$. We refer to these as F6, F5 and F4, respectively, in the following discussion. The simulations evolved $1024^3$ dark matter particles, in a simulation volume of $V_{box} = (1024h^{-1}\text{Mpc})^3$ and a $\Lambda$CDM cosmology specified by the following parameters: $\{\Omega_m, \Omega_\Lambda, h, n_s, \sigma_8, \Omega_b\} = \{0.281, 0.719, 0.697, 0.971, 0.82, 0.046\}$. For each model, five different random realizations are available. More details about these simulations can be found in Ref. [113].

In Fig. 3, we compare the predictions from our emulator to the full N-body simulations from Ref. [113] for the F4, F5 and F6 cases at three different redshifts: $z = 0$, $z = 0.397$ and $z = 0.5$, which span the redshift range in which predicted MG signals are more pronounced. Furthermore, in order to independently illustrate the accuracy of our training set, we show the ratio $P_{MG}(k)/P_{\Lambda CDM}(k)$ obtained from COLA simulations that we additionally performed, separately from our design, for these scenarios. COLA is found to recover the simulated ratios at (better than) 1% level of accuracy in the (F6) F5 case for all redshifts, in agreement with previous findings in the literature [62]. For the F4 model, the agreement is still similarly good at $z = 0$, but worsens progressively for higher redshifts, with COLA underestimating the predicted deviation. This behavior is most likely attributed to the approximate MG screening implementation [100] used in COLA, which is known to underestimate the ratio particularly in cases that deviate substantially from GR (such as F4) and at high $z$. We stress that this tendency is not detrimental, since the deviations typically predicted by cases such as F4 are of substantial magnitude [3], and effectively ruled out by observations [92, 116]. We do however choose to include larger values of $f_{R_0}$ in our target range, to enable the exploration of the linearized regime of the Hu-Sawicki MG models. In addition, we also compare a semi-analytical fitting formula provided by Ref. [87] with our emulator. There are two notable differences between our emulator and the fitting formula. First, the fitting formula was trained on a single set of CDM cosmological parameters and a fixed $n = 1$ whereas our emulator is based on the sampling given in Equations (5) – (9). Second, the $k$-range covered by the fitting formula extends to $k \leq 10h\text{Mpc}^{-1}$, whereas with our emulator is restricted to $k \leq 1h\text{Mpc}^{-1}$.

Our emulator is found to successfully recover the target ratios for F5 and F6 at a very similar level of accuracy as the COLA method, for all three redshifts. The predictions are accurate even for the F4 corner case. All of the above findings are consistent with the levels of accuracy previously found by the accuracy tests shown in Sec. IV A.

We end this section by clarifying that the COLA and full N-body simulations shown in Fig. 3 correspond to

[3] The screening mechanism effectively fails for such large values of $f_{R_0}$.

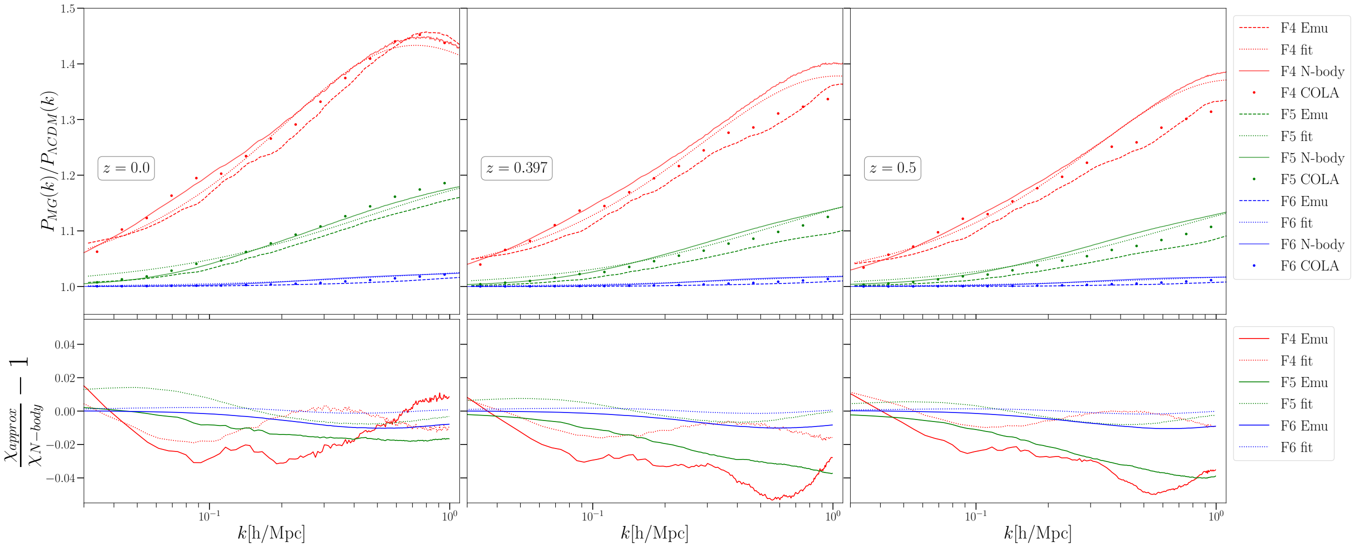


FIG. 3. Comparison of emulator predictions to our corresponding COLA simulations and the N-body simulations of Ref. [113] for the F4 (red), F5 (green) and F6 (blue) models, respectively. Each top panel represents a different redshift: $z = 0$ (left), $z = 0.397$ (middle) and $z = 0.5$ (right). Solid lines are from the three N-body simulations, dashed lines are from the corresponding emulator (with $z$-interpolation) output, the dotted lines are the corresponding outputs from the semi-analytical fitting formula given in Ref. [87]

, and the data points are COLA simulations performed for these models. The COLA and N-body runs correspond to a cosmology of $h = 0.697$ while the emulator was built using $h = 0.67$. The ratios of the power spectra are minimally impacted by these different choices. The bottom panels show the relative differences between the fast approximations (our emulator and the fitting formula) and the N-body simulations.

$h = 0.697$, whereas the emulator predictions were generated from a COLA training set that assumed $h = 0.67$, in agreement with the Planck constraints [13, 14]. We have carefully checked and confirmed that this small inconsistency leads to negligible errors in the emulated ratio for all cases and redshifts of our design. As a result, the direct comparison in Fig. 3 is indeed meaningful, despite the different underlying values of $h$ assumed.

## V. EMULATOR PRELIMINARY APPLICATIONS

We present three preliminary applications for the emulator developed in this paper. Using the power spectrum ratio as the summary statistic we first perform a parameter sensitivity analysis. Second, we use the emulator for Fisher forecasting. Finally, results from an MCMC run for a fiducial cosmology are shown. The evaluation time for an emulator prediction is less than 0.001 seconds per computation on an Intel Core i5 Processor, delivering a massive speed-up over numerical calculation using COLA, which typically takes about an hour on similar computational hardware. This is particularly important for our third application, where our GP emulator is implemented in the MCMC likelihood calculation which requires a very large number of accurate predictions for the power spectrum ratio.

### A. Parameter Sensitivity Analysis

In this section we investigate the effect of different cosmological parameters on the power spectrum ratio $P_{MG}(k)/P_{\Lambda CDM}(k)$ using the emulator. For this study, we choose a base model at redshift $z = 0.0$. The base model is evaluated at parameters shown in Equation (14):

$$\begin{aligned} \Omega_m h^2 &= 0.142, \\ n_s &= 0.967, \\ \sigma_8 &= 0.816, \\ \log(f_{R_0}) &= -5, \\ n &= 1.0. \end{aligned} \tag{14}$$

We then measure the sensitivity of the power spectrum ratio to changes in one cosmological parameter at a time while keeping the others at their base values. We stress that the MG parameters span a much broader range than the $\Lambda$CDM parameters and therefore we expect a much larger impact on the power spectrum ratio for varying $\log(f_{R_0})$ and $n$. The results are presented in Fig. 4.

Primarily, and in agreement with the literature [86, 95], we observe that $\log(f_{R_0})$ has the highest impact on the matter power spectrum ratios, showing up to 40 percent variation just around $0.3h\mathrm{Mpc}^{-1} \lessapprox k \lessapprox 1h\mathrm{Mpc}^{-1}$ for the range $-6 \leq \log(f_{R_0}) \leq -4$. Larger $\log(f_{R_0})$ results in enhancement of the power spectrum ratios, due to the progressive weakening of the screening mechanism, and

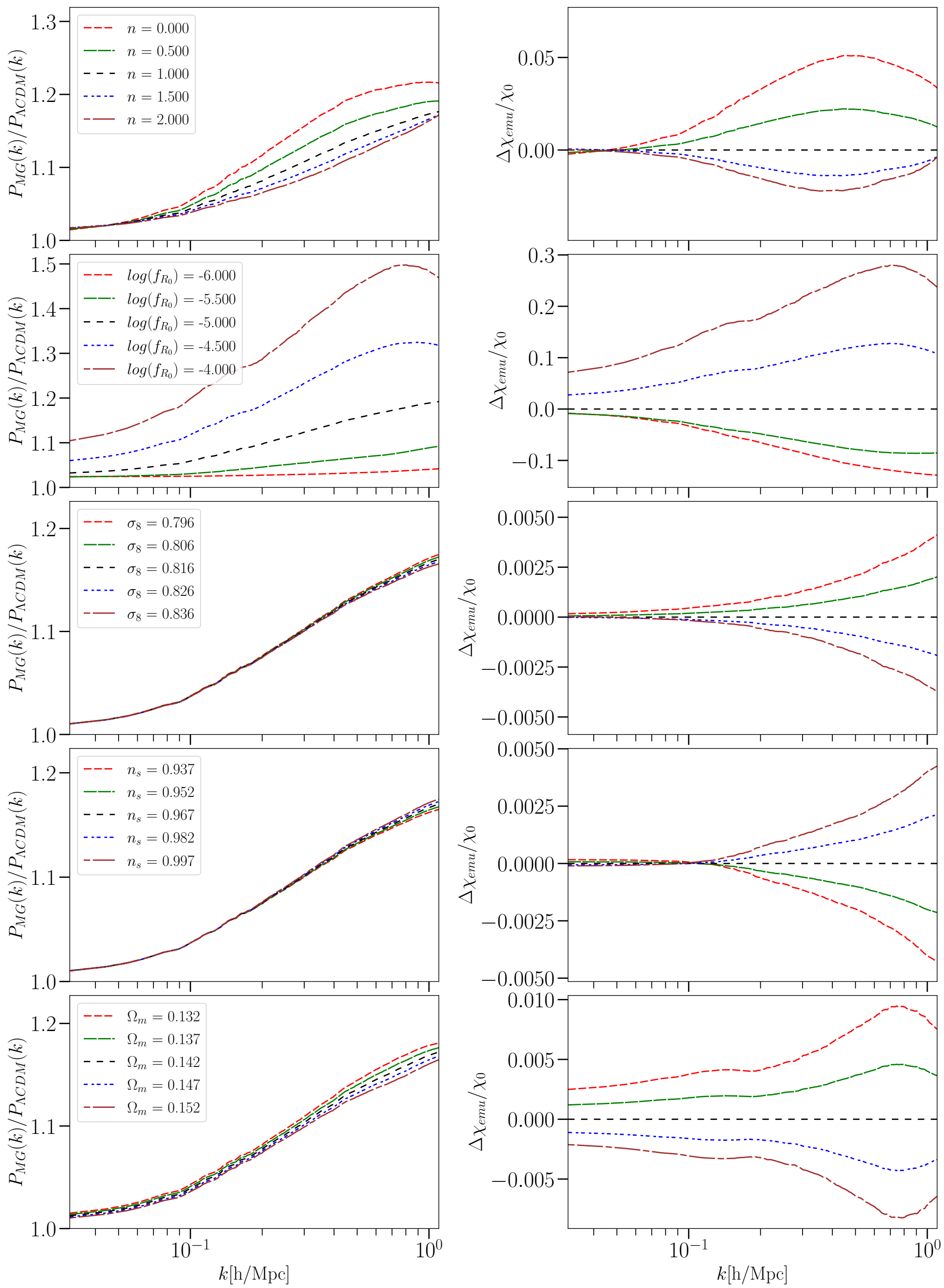


FIG. 4. Results for the sensitivity analysis for the power spectrum ratios using emulator predictions. The impact of five cosmological parameters is investigated. The left panels show the emulator outputs from linearly varying one parameter while keeping the rest at the base cosmology. The right panels show the relative variation of the the power spectrum ratios $\Delta\chi_{emu}/\chi_0$ with respect to the emulator output $\chi_0$ at base cosmology parameters listed in Equation (14). From top to bottom the parameter sensitivity analysis with respect to $n$, $\log(f_{R_0})$, $\sigma_8$, $n_s$ and $\Omega_m h^2$ is shown respectively. Note that the scales on the $y$-axes for the relative variation differ by up to two orders of magnitude. Given the wide range we allowed for $\log(f_{R_0})$ it is not surprising that its impact is by far the largest on the power spectrum ratio.

the increase is monotonous across the full range of testing cosmologies. The analysis also suggests that with $\log(f_{R_0}) > -5$ the modified gravity matter power spectrum $P_{MG}(k)$ is up to 50 percent higher than $P_{\Lambda CDM}(k)$. This shows that models in this range are disfavored by current data, which is consistent with previous studies constraining $f(R)$ models using available data sets, see for example Refs. [92, 116].

The second highest contribution to the changes in the summary statistic is due to the second Hu-Sawicki MG parameter, $n$. The variation is under 10% from the base cosmology for the range $0 \leq n \leq 2$. However, the peak of the departure occurs at slightly larger length scales, between $0.1h\text{Mpc}^{-1} \lessapprox k \lessapprox 0.9h\text{Mpc}^{-1}$. For $k \lessapprox 1h\text{Mpc}^{-1}$, larger values of $n$ cause higher suppression of the matter power spectrum ratios. This trend reverses beyond $k \gtrapprox 1h\text{Mpc}^{-1}$. Unfortunately, there are currently no N-body Hu-Sawicki simulations for $n \neq 1$ available in the literature, and thus a thorough comparison on nonlinear scales is not possible. Nevertheless, our observed large-scale trend seems to be in qualitative agreement with the linear considerations in Ref. [86]. Even though we do not expect the behavior of the model to be substantially different for $n \neq 1$, we defer a detailed study of the emulator accuracy for such values to future work, when corresponding N-body simulations become available.

Finally, the relative change of the emulator output when varying the ΛCDM parameters is more restricted. Both $\sigma_8$ and $n_s$ reveal sub-percent variation within their respective ranges of $0.796 \leq \sigma_8 \leq 0.836$ and $0.937 \leq n_s \leq 0.997$. Moreover, both their peak departures from the base model occur at $k \gtrapprox 1h\text{Mpc}^{-1}$, where the accuracy of the training COLA simulations with respect to full N-body simulations reduces at higher redshifts. Also, the variations of $\sigma_8$ and $n_s$ beyond $k \lessapprox 0.1h\text{Mpc}^{-1}$, are opposite of each other, i.e, increasing $\sigma_8$ reduces the $P_{MG}(k)/P_{\Lambda CDM}(k)$ ratio, whereas $n_s$ has the opposite effect. Variations of the final emulator parameter, $\Omega_m h^2$, lead to a monotonic change in the matter power spectrum ratios, where increasing the values from 0.132 to 0.152 shows a decrease in $P_{MG}(k)/P_{\Lambda CDM}(k)$. This reduced sensitivity of the ratio with respect to variations of the ΛCDM parameters (relative to the response of MG parameters) was also found using the fitting formula approach in Ref. [87].

### B. Fisher Forecasting

The likelihood $\mathcal{L}(\chi|\theta)$ is defined as the probability distribution function of an observed summary statistic $\chi$ for a given model with parameters $\theta$. The emulator output $\chi_{\text{emu}}(k;\theta) = P_{MG,\text{emu}}(k)/P_{\Lambda CDM,\text{emu}}(k)$ at a given redshift itself can be considered as the forward model in the computation of the likelihood. In the case of the observed power spectrum ratio $\chi_{\text{obs}}(k) = P_{MG,\text{obs}}(k)/P_{\Lambda CDM,\text{obs}}(k)$ for a set of cosmological parameters $\theta = \{\Omega_m h^2, \sigma_8, n_s, \log(f_{R_0}), n\}$, the likelihood

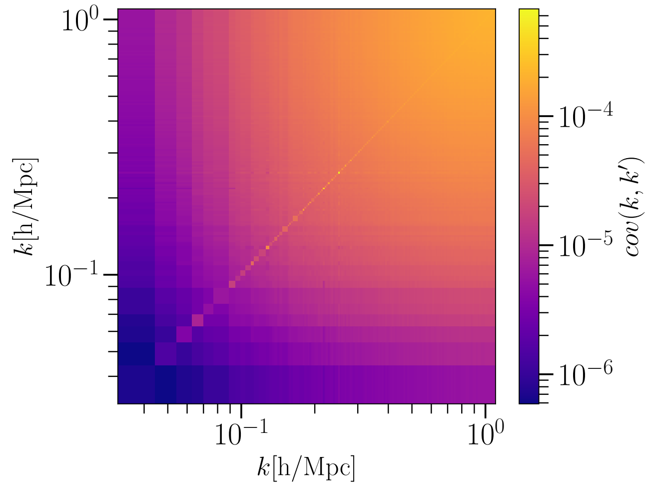


FIG. 5. Covariance matrix for the ratio $P_{MG}(k)/P_{\Lambda CDM}(k)$ obtained from 2000 COLA realizations for our fiducial cosmology parameters $\theta_{fid}$ at redshift $z = 0.02$ as an example. The likelihood obtained using this covariance matrix is used for both Fisher analysis and posterior estimation of the cosmological parameters via MCMC sampling.

is computed using Equation (15) assuming it is of a Gaussian form:

$$\log \mathcal{L}(\chi|\theta) \propto -\frac{1}{2}\left[\chi_{\text{emu}}(k;\theta) - \chi_{\text{obs}}(k)\right] \widehat{C_{ij}^{-1}} \left[\chi_{\text{emu}}(k;\theta) - \chi_{\text{obs}}(k)\right]^{\mathrm{T}}. \quad (15)$$

We construct a mock data set for the power spectrum ratios with an associated mean, $\chi_{\text{obs}}(k)$, and a covariance matrix $C_{ij}$, which captures the effects of cosmic variance. This data set is computed using 2000 COLA simulation realizations run at a fiducial cosmology $\theta_{fid} \equiv \{\Omega_m h^2 = 0.142, n_s = 0.967, \sigma_8 = 0.816, \log(f_{R_0}) = -5, n = 1.0\}$, shown in Figure 5. The unbiased estimator for the inverse covariance matrix $\widehat{C_{ij}^{-1}}$ is then computed using Equation (16) given by Ref. [117]. This correction accounts for the number of COLA simulations used ($N = 2000$) and the size of the data vector $D$, which depends on our range of wavenumbers used in calculating the likelihood:

$$\widehat{C_{ij}^{-1}} = \frac{N - D - 2}{N - 1} C_{ij}^{-1}. \quad (16)$$

The box size, mass resolution and other simulation specifications for these additional COLA simulations are the same as the ones listed in Table I. We choose a single redshift of $z = 0.02$ for this analysis.

The Fisher information matrix assesses how well a cosmological parameter can be inferred from a summary statistic. It is defined in terms of the likelihood $\mathcal{L}$ of the data in the following equation:

$$F_{ij} = -\left\langle \frac{\partial^2 log\mathcal{L}}{\partial\theta_i \partial\theta_j} \right\rangle . \quad (17)$$

The Fisher matrix can be calculated directly using the emulator, either using numerical derivatives, or using the analytical derivatives of Gaussian Processes propagated through the emulator. We choose the former method and calculate the second order partial derivatives numerically. The corresponding confidence ellipses are shown in Figure 6. The numerical derivatives are evaluated with multiple step sizes to ensure consistency.

The Fisher information obtained from the power spectrum ratio analysis reveals correlations between MG parameters and ΛCDM parameters. We present results using 4 different wavenumber thresholds ($k_{max} = 0.15h\mathrm{Mpc}^{-1}$, $0.25h\mathrm{Mpc}^{-1}$, $0.5h\mathrm{Mpc}^{-1}$ and $1.0h\mathrm{Mpc}^{-1}$) corresponding to increasing scales of nonlinear growth of structure. The tightest constraints for forecasts across all parameters are achieved when including nonlinear scales up to $k_{max} = 1.0h\mathrm{Mpc}^{-1}$, and the constraints weaken with decreasing nonlinearity. This is expected, given that including more modes gives us increasingly more information about the underlying cosmological parameters (including the MG parameters), resulting in increased constraining power.

It should be noted that the aim of this exploratory set-up is not to perform a thorough study of forecasted MG constraints, but rather to highlight the diverse applications of our emulator. As a result, these constraints only provide an approximate view. For a more realistic approach, one would need to combine the emulated ratio with a prediction for $P_{\Lambda CDM}(k)$ (e.g. from the `CosmicEmu`), as well as fold in prescriptions for the galaxy bias and/or redshift-space distortions [60], in order to obtain a prediction for the observed galaxy power spectrum in MG (as for example was done in [87]). This step goes beyond the scope of this paper, but will be studied in detail in future work. As a consequence, our constraints are not directly comparable to the ones obtained by the fitting formula approach in [87].

### C. Parameter Constraints Using MCMC

The speed and precision of our emulator enables quick parameter inference using traditional Bayesian inference schemes like Markov Chain Monte Carlo (MCMC) methods. We illustrate the application of such emulator-based posterior estimation for mock data, the same as shown in Sec. V B. The Gaussian likelihood is computed via Equation (15), using the emulator and the mock covariance matrix shown in Fig. 5 from 2000 COLA realizations for the fiducial cosmology parameters $\theta_{fid}$. Our aim is to demonstrate the recovery of the original parameter values within appropriate error margins. In addition, the exploration of the posterior distribution allows us to study the covariance between the parameters that arises from the data vector and sensitivity of the parameters.

We execute two MCMC sampling schemes. First with fixed values for the ΛCDM parameters at $\{\Omega_m h^2 = 0.142, n_s = 0.967, \sigma_8 = 0.816\}$, we evaluate the posterior distributions of the 2 MG parameters. The logarithmic value for $f_{R_0}$ is chosen to reduce the dynamical range of the MCMC sampling, and a flat prior in the range $-8 \leq \log(f_{R_0}) \leq -4$ is selected. For $n$ we also use a flat prior within the limits of our experimental design: $0 \leq n \leq 4$. While the choice of priors could be more stringent, we select such uninformative priors spanning the entire parameter range of the emulator in order to avoid obtaining prior-dominated constraints for the MG parameters. The likelihood $\mathcal{L}(\chi|\theta)$ of the mock data vector is computed up to $k < 1.0h\mathrm{Mpc}^{-1}$.

We utilize an MCMC approach to approximate the posterior distribution using an affine-invariant ensemble sampling [118] method implemented in the `emcee` [119] sampler. An ensemble of 300 walkers or Markov chains with 3000 evaluations or steps per walker only takes about 4000 seconds on a single processor, hence enabling quick explorations of the posterior space of the parameters of interest. The final constraints obtained were $\theta_{MCMC} \equiv \{\log(f_{R_0}) = -5^{+0.05}_{-0.04}, n = 1.18^{+0.20}_{-0.20}\}$, as shown in the top panel of Fig. 7, and a recovery of parameters within the 2-$\sigma$ margin of error is found. While the 2000 realizations for fiducial covariance matrix have enabled tight parameter constraints via MCMC sampling compared to the broad priors, overcoming the offset within 2-$\sigma$ margin of error from the fiducial targets may require further reduction in noise. The emulator outputs at the $\theta_{fid}$ and $\theta_{MCMC}$ are matched well with the $P_{MG}(k)/P_{\Lambda CDM}(k)$ ratio of the fiducial cosmology simulations. Moreover, the direction of the covariance from the Fisher forecasts in Section V B also matches with the contours of the posteriors. The consistent results are shown in the bottom left panel of Fig. 7.

In addition, we also sample the combined posterior distribution of the ΛCDM parameters $\{\Omega_m h^2, \sigma_8, n_s\}$ and the $f(R)$ Hu-Sawicki model parameters $\{\log(f_{R_0}), n\}$ using an MCMC approach. The fiducial cosmology $\theta_{fid}$, the covariance information and the likelihood function remain the same, with a $k$-range limited to $k < 1.0h\mathrm{Mpc}^{-1}$ for the computation of the likelihood. The priors are again chosen to be flat within the full range covered by the emulator. The corresponding posterior distribution functions are shown in Fig. 8, along with the target cosmological parameters corresponding to the mock data set. The final constraints corresponding to the MCMC sampling are $\theta_{MCMC} \equiv \{\Omega_m h^2 = 0.15^{+0.01}_{-0.01}$, $n_s = 0.91^{+0.06}_{-0.04}, \sigma_8 = 0.79^{+0.05}_{-0.05}, \log(f_{R_0}) = -5.05^{+0.05}_{-0.05}$, $n = 1.22^{+0.25}_{-0.19}\}$.

The MCMC sampling of the parameter space is restricted to the parameter range in our SLH design. Especially for $\Omega_m h^2$, $\sigma_8$ and $n_s$, this results in a partial coverage of the posterior distribution. This is displayed in the individual panels of Fig. 8, where the posteri-

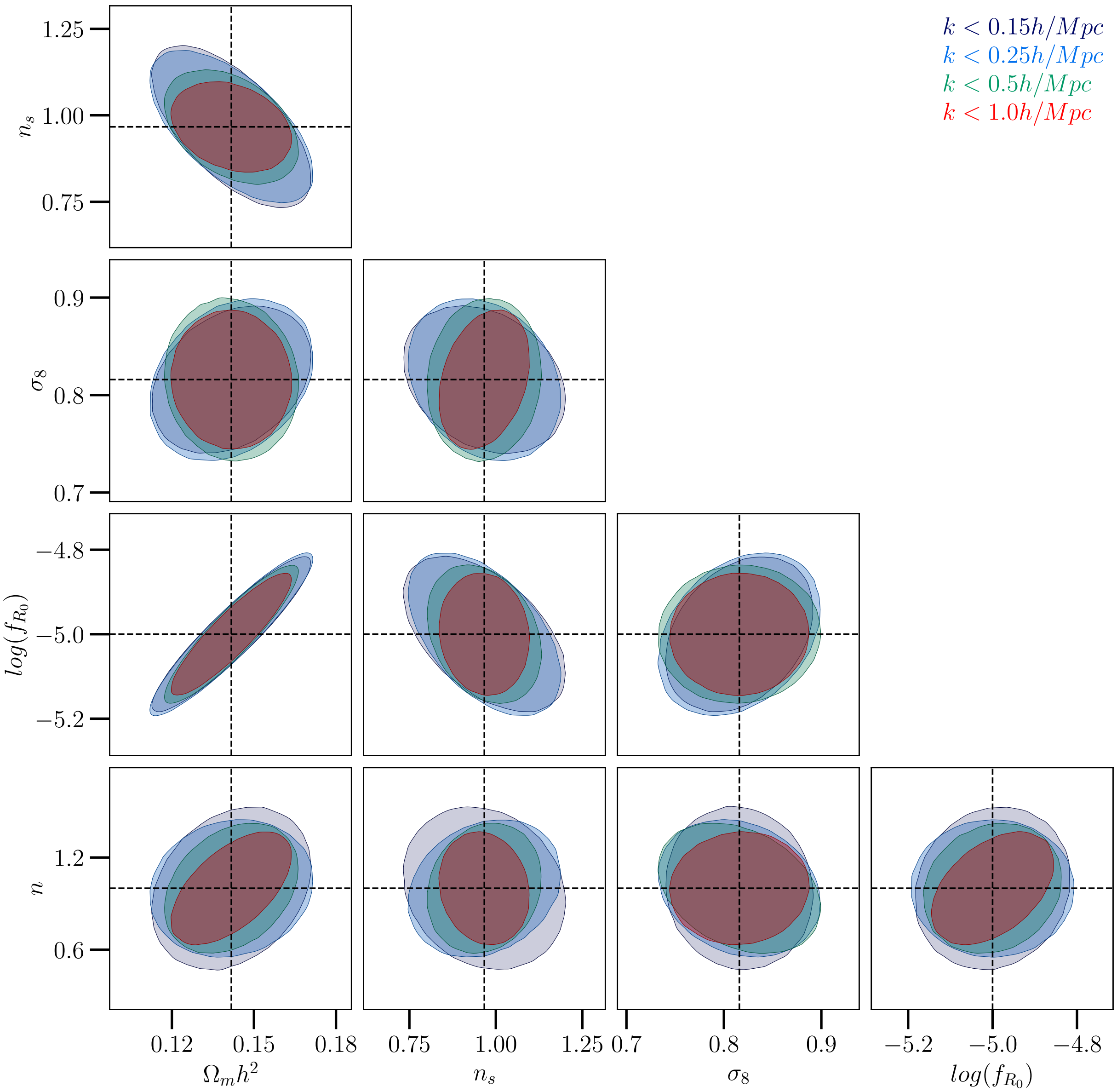


FIG. 6. Confidence ellipses from the Fisher information matrix $F_{ij}$ for the cosmological parameters considered and using the emulator to provide the power spectrum ratio information. Each panel shows the covariance between the ΛCDM and MG parameters. The thresholds for the wave-numbers used in the computation of the likelihood, i.e., $k_{max} = 0.15h\mathrm{Mpc}^{-1}$, $0.25h\mathrm{Mpc}^{-1}$, $0.5h\mathrm{Mpc}^{-1}$ and $1.0h\mathrm{Mpc}^{-1}$ correspond to the four 2-$\sigma$ level confidence ellipses in each panel. The dashed lines in each panel correspond to the fiducial cosmological parameters $\theta_{fid}$.

ors are limited to the parameter range of the emulator. These constraints would be tightened using a joint analysis (with CMB data, for instance). In this study we restrict our attention to explorations of the constraining power of $P_{MG}(k)/P_{\Lambda CDM}(k)$ alone, and reserve combined analyses for future studies. We additionally note that MCMC constraints for the HS model were also presented in [89, 120], but using the power spectrum. For this reason, our results are not directly comparable with the ones in these studies.

The 1, 2 and 3-$\sigma$ contours for the MG parameters $\{\log(f_{R_0}), n\}$ are within the range of our experimental design, as seen in both Fig. 7 and Fig. 8. We note that the constraints obtained from sampling a parameter space of just 2 parameters are tighter than those for 5 parameters. Posterior estimation restricted to fewer parameters removes possible degeneracies, resulting in reduced sampling space.

The original parameters corresponding to the mock data vectors are recovered in both the sample MCMC runs, indicating that the boost in the matter power spectra is a powerful statistic for constraining MG parameters, especially when coupled with robust emulators as presented in this work. In both posterior approximations,

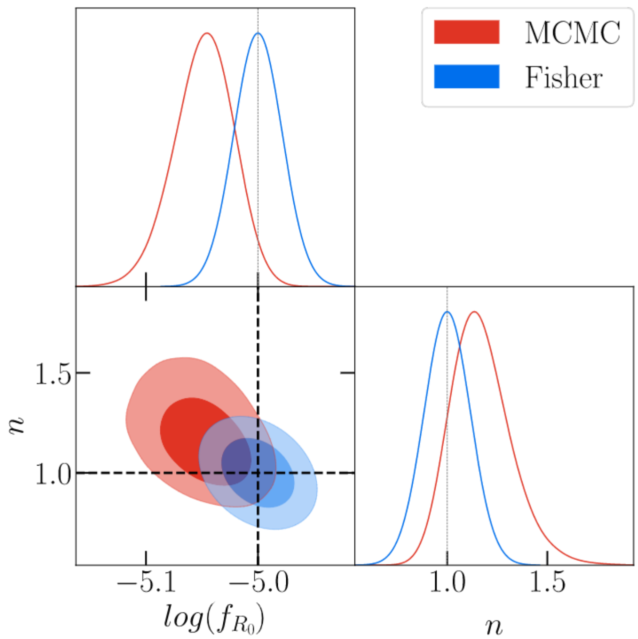


FIG. 7. Bayesian posterior distribution for $\{\log(f_{R_0}), n\}$ obtained using MCMC sampling. The data vector, emulator output and likelihood computation are carried out for wavenumbers $k < 1.0h\mathrm{Mpc}^{-1}$. The two shades of red correspond to 1-, 2-$\sigma$ confidence intervals of the posterior. Dotted lines are the target cosmologies of the mock data, $\log(f_{R_0}) = -5$ and $n = 1$. The constraints are obtained from uniform priors in the ranges $-8 \leq \log(f_{R_0}) \leq -4$ and $0 \leq n \leq 4$. The blue contours are the 1- and 2-$\sigma$ confidence ellipses from the Fisher information constraints for the same $k$-range at the fiducial cosmology, showing the consistency between the two applications in space and direction of correlation.

the chosen prior probability distributions are highly uninformative, i.e., they are uniform with coverage of the entire training limits. Posterior contours in Figure 7 display a shift from the fiducial values, and a partial coverage of posterior distribution is seen in Figure 8. These constraints on the cosmological parameters would be improved either using a tomographic or multi-probe analysis, along with more realistic data vectors and covariance matrices. We reserve these for a future study of modeling cosmological observables and associated forecasting in MG scenario.

## VI. CONCLUSIONS

In this paper, we present an emulator for efficiently predicting the enhancements in the nonlinear matter power spectrum due to beyond-GR effects. This emulator is based on the $f(R)$ Hu-Sawicki model, a MG candidate prioritized for further studies in near-future Stage-IV surveys such as the LSST and Euclid. We aim at simplifying the tasks of model selection and cosmological parameter inference by the use of fast and robust generation of the power spectrum ratio $P_{MG}(k)/P_{\Lambda CDM}(k)$. The emulator provides estimations across the redshift range of $0 < z < 49$, for a combination of cosmological parameters $\{\Omega_m h^2, \sigma_8, n_s, \log(f_{R_0}), n\}$. This way, we enable the full exploration of the parameter space that defines the Hu-Sawicki model.

Our emulator is based on simulations produced by the efficient COLA method, which effectively captures the chameleon screening mechanism through a phenomenological thin-shell factor attached to the scalar field Klein-Gordon equation. The matter power spectrum ratios extracted are computed by running two consistent COLA simulations at each training point – one for MG and the other for the corresponding ΛCDM scenario, reducing the effect of cosmic variance while at the same time highlighting the effects of MG. Our fully trained emulator is validated on additional cosmologies within our target parameter space and is found to achieve sub-percent levels of accuracy for models with $f_{R_0} < 10^{-5}$ and up to 5% agreement when $f_{R_0} > 10^{-5}$. The computation time is less than 0.001 seconds, delivering thus a massive speed-up by 6 orders of magnitude compared to the COLA simulations.

In order to explore and validate the diverse capabilities of our emulator, we further proceed to utilize its predictions for three preliminary applications. First, we perform a sensitivity study of the target summary statistic with respect to the variation of the five cosmological parameters. We find that the power spectrum ratio exhibits the highest sensitivity for the MG parameter $\log(f_{R_0})$ followed by the $n$ parameter, in agreement with previous studies [86, 95]. Next, we produce constraints around a fiducial cosmology using Fisher forecasting as well as MCMC parameter inference. The confidence contours obtained are consistent between the two methods and also consistent with the corresponding analytical expressions. The emulated ratio is thus found to enable accurate constraints for both MG parameters as well as the values of the background ΛCDM cosmological parameters.

We conservatively limit the use of the emulator to $k \leq 1.0h\mathrm{Mpc}^{-1}$ throughout the analysis of this paper. Our tests shows that the COLA prescription agrees with available N-body simulations within a relative error of 5 % up to $k \sim 1.0h\mathrm{Mpc}^{-1}$, and hence we advocate the use of this emulator only up to this limit.

We also advise the reader to exercise caution in extrapolating the emulator beyond the limits of cosmological parameters and redshifts used in the experimental design. Gaussian Processes, like any interpolation schemes, may give estimates with large extrapolation uncertainties beyond the training range. We note again that our tests on additional COLA simulations show that models with $\log(f_{R_0}) < -5$ agree within 1%, while the estimation error rises up to 5% for larger values of $\log(f_{R_0})$. We also note that the emulator is trained on one realization per cosmology, with just 50 training points. With a large training sample with better sampling and larger

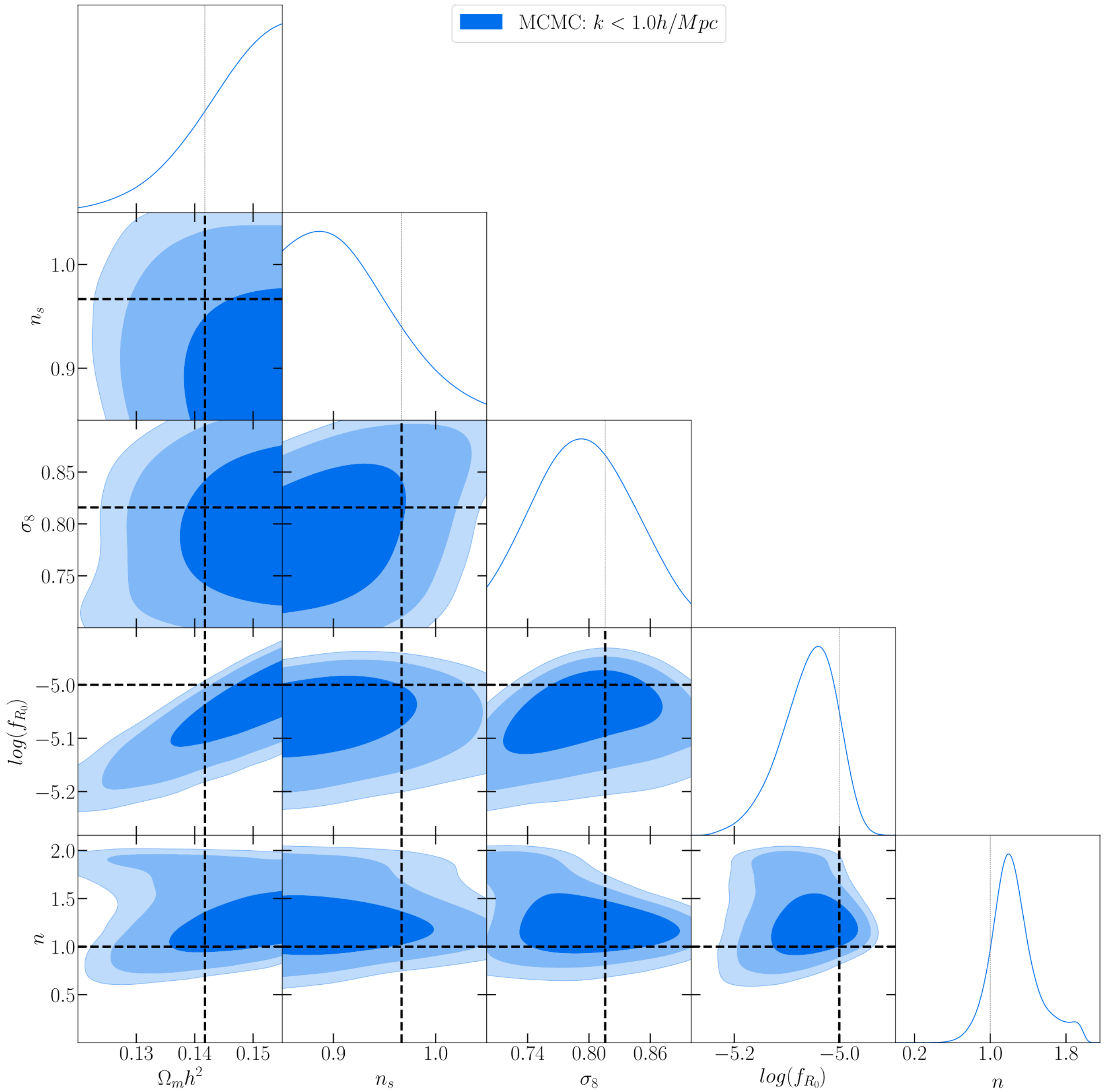


FIG. 8. Posterior distribution evaluated from an MCMC sampling for five cosmological parameters. The data vector, emulator output and likelihood function are computed for wave-numbers $k < 1.0h\mathrm{Mpc}^{-1}$. The three shades of blue correspond to 1-, 2- and 3-$\sigma$ confidence intervals. Dotted lines are the target cosmological parameters of the mock data set $\theta_{fid}$. It is very satisfactory to observe that the target cosmological parameters fit well within the 2-$\sigma$ confidence intervals, despite the relatively limited constraining power of the ratio of power spectra when used with no additional cosmological information.

number of realizations per cosmology, the relative error with respect to COLA simulations is naturally expected to reduce.

In addition to enabling an efficient exploration of the deviations predicted by MG, our emulator also serves as a stepping stone to allow a broad portfolio of future applications. Through a simple multiplication by the $\Lambda CDM$ power spectrum (from emulators like `CosmicEmu`), the emulated power spectrum ratio can straightforwardly provide a prediction for the MG nonlinear matter power spectrum itself. The latter can then also be utilized to incorporate the effects of galaxy bias and Redshift Space Distortions (RSD) in MG (for example as in Ref. [60]), which are crucial for comparing with observations, or to enable obtaining MG constraints from weak lensing cosmic shear measurements. Such predictions should however take into account the accuracy and sensitivity of the estimators at corresponding length scales, and we reserve this for future studies. Furthermore, the emulator will be a useful standardized routine to be included in the Core

Cosmology Library (CCL [4], [121]) in order to calculate basic cosmological observables, specifically for Rubin Observatory LSST analyses. We also intend in a follow-up work to employ the emulator in providing forecasts using a detailed joint probes cosmological parameters analysis. Last but not least, we plan to expand our emulator's capabilities in order to incorporate the effects of massive neutrinos, and also to support the more general class of Horndeski MG models prioritized by our collaboration. The necessary modifications to achieve these steps with the efficient COLA approach are already underway. The combined outcome of these efforts will be a diverse set of tools that will allow efficient and reliable explorations of the broad spectrum of beyond-$w(z)$CDM candidates.

It is worth emphasizing the multitude of ways in which our work expands upon the previous ones in the literature. While the matter power spectrum ratio in the Hu-Sawicki scenario has already been modeled using fitting formulas or semi-analytical approaches [87–89], our emulator is the first predictive tool purely based on numerical simulations. Not having to rely on a single functional form of the modeled observable, emulators are guaranteed to maintain consistent levels of accuracy in their predictions across the full target parameter space. Fitting formulas, on the other hand, may have reduced accuracy outside the particular parameter choices used for their calibration, introducing unknown systematic biases to their predictions, even in the less complicated extension of $w$CDM cosmologies [65, 68, 69]. Furthermore, while previous approaches focused their attention on variations of only one the two MG parameters of the Hu-Sawicki scenario, our emulator spans the full 2-dimensional parameter space of the model, being the first one of its kind in that regard. Last but not least, given the ever increasing interest in testing theories of gravity with cosmological surveys, a wide range of complementary predictive tools available in the community is very desirable. In order to contrast different methods, we compared our predictions against the ones made by the fitting formula approach of Ref. [87] , finding good agreement over the parameter and $k$-range that can be compared.

In this decade, precise cosmological observations will offer a unique opportunity to constrain beyond-$w(z)$CDM models at an unprecedented level of accuracy. Emulators like the one developed in this work will be essential and necessary to perform cosmological analyses with accurate and fast theoretical predictions in the nonlinear regime, and to take full advantage of the formidable data expected from Rubin Observatory LSST, DESI, Euclid, SPHEREx and Roman Space Telescope.

[4] `https://github.com/LSSTDESC/CCL`

## Appendix A: Savitzky-Golay smoothing

The Savitzky-Golay smoothing filter [107, 108] performs convolution operations on adjacent data points with a polynomial function, which gives us the effect of smoothing the input dataset. The window size and the order of the polynomial are the two parameters that specify the smoothing operation. Equation (A1) shows the value of $j$-th bin of the smoothed power spectrum ratio $\chi(k)$:

$$\chi(k_j) = \sum_{i=\frac{1-m}{2}}^{\frac{m-1}{2}} C_i \, \frac{P_{MG}(k_{j+i})}{P_{\Lambda CDM}(k_{j+i})}. \tag{A1}$$

The individual convolution coefficients $C_i$ are defined by the analytical expression given in Ref. [107]. Within every window, a polynomial (of order $p$) is fitted, which provides a smoothing effect of the input dataset. The window size $m$ is the number of data points chosen for individual regression.

The tuning of the two free parameters $m$ and $p$ depends on the largely on the type of data and the desired level of smoothing. These parameters are hand-tuned for smoothing the power spectrum ratios, and checked for consistency with various choices of smoothing filters and window sizes. As the ratio of window width to polynomial order, $m/p$ increases the amount of smoothing increases. First, the window width was appropriately tuned to be effective against the noise and is set to be $m = 51$ points. For this window width, a third order polynomial ($p = 3$) is fitted. A polynomial of order $p > 3$ would closely follow the undesired noise resulting from the single realizations of the COLA simulations. For a given polynomial order, decreasing the window length has a similar effect, i.e., while the bias decreases (the smoothing function closely follows the raw data power spectrum ratios from the COLA simulations), the estimation variance increases, resulting in an over-fitted smoothing function.

Smoothing near the boundaries requires additional considerations. The data points at the edges cannot be placed at the center of a symmetric window, hence the Equation (A1) is applied for $\frac{m-1}{2} \leq j \leq n_{bins} - \frac{m-1}{2}$ only. We use a separate treatment for smoothing at the boundaries, where the polynomial fitted to the windows near the edges of the data is used to evaluate the first and the last $m/2$ smoothed outputs of $\chi(k_j)$. The effect due to the edge effects is less significant at low-$k$ boundary (i.e., $k < 0.17hMpc^{-1}$) since the raw $P_{MG}(k)/P_{\Lambda CDM}(k)$ is less noisy. On the higher end, this near-boundary effects smoothing at $k > 2.35hMpc^{-1}$. Since our emulator outputs are restricted to $k \leq 1hMpc^{-1}$, this does not effect our estimations directly. However, the consistency of the smoothing results across all the 213 bins (up-to $k = 3.5hMpc^{-1}$) is considered whilst tuning the free-parameters.

## Appendix B: Gaussian Processes

A parametric regression task [122] involves an estimation of finite number model parameters that fit the data. For $n$ training targets $\{y_1, \ldots, y_n\}$ at training locations $\{x_1, \ldots, x_n\}$, one may define a polynomial regression relationship and estimate the finite number of polynomial coefficients. With a frequentist approach, point estimates of these parameters can be estimated. Alternatively, a Bayesian approach treats these model parameters as probability distributions which are to be inferred using the training points. With these either a point estimation or distribution of a test target $y_*$ can be made at a new location $x_*$.

In contrast, Gaussian process [109] regression is a non-parametric approach i.e., one finds a distribution of possible functions $f(x)$ that are consistent with observed data. GPs are remarkably good Bayesian tools that perform regression tasks with associated uncertainties. Although the inference is fast, computational cost of GP regression can be very expensive and grows cubicly with the training set size and dimension. Due to this constraint, dimensionality reduction is usually performed on the training data, as is the case with our approach of building the emulator. We employ a data reduction technique called Principal Component Analysis (PCA) to reduce computational expenses during GP training. With a PCA decomposition (in Equation 12), the bases $\phi_i(k)$ are independent of the cosmological parameters, and only the weights $w_i(\theta)$ are used in GP interpolation. Thus the training locations are the set of parameters $\theta$ (tabulated in Appendix C) where 50 COLA simulations are computed, and the training targets are the corresponding weights $w_i(\theta)$.

With GP, we first assume that the joint probability distribution $p\left(f(x_1), \ldots, f(x_n)\right)$ are jointly Gaussian with mean $\mu(x)$ and covariance $\mathbf{K}$, where the elements $\mathbf{K}_{ij} = k(x_i, x_j)$ with $k(x_i, x_j)$ being the covariance kernel. For simplicity, the GP prior can be defined using a zero mean and covariance as $p\left(f(x)\right) = GP(0, \mathbf{K})$. In our emulator construction, the weights are essentially sampled from this distribution, i.e., $w_i(\theta) \sim GP(0, \mathbf{K})$, where the covariance $\mathbf{K} = k(\theta, \theta')$.

The kernel function $k$ is usually selected depending on how smooth the function is expected to be. One popular choice is the squared exponential or the Radial basis function kernel: $k_{RBF}(x, x') = \sigma^2 \exp\left(-\frac{(x-x')^2}{2\theta^2}\right)$ with hyperparameters $\left(\sigma^2, \theta\right)$ corresponding to the process variance and lengthscale, respectively. These hyperparameters can be inferred based on the maximum likelihood or other more Bayesian techniques. Once optimized, the GP model has learned a distribution of functions that fits the training data.

Using the GP assumption that our data can be represented as a sample from a multivariate Gaussian distribution, the above definition can also be extended to a hold-out target $y_*$ at a new location $x_*$ in terms of the trained covariance $\mathbf{K}$. The joint probability of training targets $y$ and test targets $y_*$ shown on Equation (B1) is also a Gaussian Process:

$$p(y, y_*) = \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{K} & \mathbf{K}_*^T \\ \mathbf{K}_* & \mathbf{K}_{**} \end{bmatrix}\right). \tag{B1}$$

The covariance $\mathbf{K}$ is obtained by $\mathbf{K}_{ij} = k(x_i, x_j)$ is the matrix we get by applying the trained kernel function to our training values, i.e., the similarity of each observed $x$ to each other observed $x$. $\mathbf{K}_* = k(x_i, x_*)$ shows the similarity of the training values to the test value whose output values were trying to estimate. $\mathbf{K}_{**} = k(x_*, x_*)$ gives the similarity of the test values to each other.

The desired posterior of our prediction is the conditional probability distribution the test target $p(y_*|y)$. This is finally derived using the joint probability distribution in Equation (B1) using marginalization:

$$p(y_*|y) = \frac{p(y, y_*)}{p(y)} = \mathcal{N}\left(\mathbf{K}_*\mathbf{K}^{-1}y, \mathbf{K}_{**} - \mathbf{K}_*\mathbf{K}^{-1}\mathbf{K}_*^T\right). \tag{B2}$$

Hence the mean of our estimation at new test location is simply given by $\mu(y_*) = \left(\mathbf{K}_*\mathbf{K}^{-1}y\right)$, and the uncertainty of this prediction is $\sigma(y_*) = \left(\mathbf{K}_{**} - \mathbf{K}_*\mathbf{K}^{-1}\mathbf{K}_*^T\right)$. In the prediction phase of our emulator, this mean and variance for the predictive weights $w_{i_*}(\theta_*)$ are calculated for any new set of cosmological parameters $\theta_*$. Since the form of $\mathbf{K}$ is already determined from training points using hyperparameter optimization, the GP prediction is simply matrix operations. This makes GP inference extremely fast and easily parallelizable.

## Appendix C: Parameters of COLA simulation

Table of all parameters of each COLA simulation.

## ACKNOWLEDGMENTS

This paper has undergone internal review by the LSST Dark Energy Science Collaboration. The internal reviewers were Tessa Baker, Francois Lanusse and Danielle Leonard. The authors would like to thank them for useful feedbacks.

The two lead authors, NR and GV have contributed equally to this work. NR was involved in development and applications of the emulator, and contributed to the text of the paper. GV was involved in providing the idea for this paper, producing the COLA and N-body MG simulations and contributed to the writing of the text. MI was involved in developing the idea for the paper, the MG part of the design, contributed to the analysis of Fisher and MCMC results, mentored the project, and contributed to the text of the paper. KH was involved in developing the idea for the paper, provided the design, and contributed to the text of paper.

LSST DESC acknowledges ongoing support from the Institut National de Physique Nucléaire et de Physique des Particules in France; the Science & Technology Facilities Council in the United Kingdom; and the Department of Energy, the National Science Foundation, and the LSST Corporation in the United States. LSST DESC uses the resources of the IN2P3 / CNRS Computing Center (CC-IN2P3–Lyon/Villeurbanne - France) funded by the Centre National de la Recherche Scientifique; the Univ. Savoie Mont Blanc - CNRS/IN2P3 MUST computing center; the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under contract No. DE-AC02-05CH11231; STFC DiRAC HPC Facilities, funded by UK BIS National E-infrastructure capital grants; and the UK particle physics grid, supported by the GridPP Collaboration. This work was performed in part under DOE contract DE-AC02-76SF00515.

Argonne National Laboratory's work was supported under the U.S. Department of Energy contract DE-AC02-06CH11357. GV recognizes financial support by DoE grant DE-SC0011838, NASA ATP grants NNX14AH53G and 80NSSC18K0695, NASA ROSES grant 12-EUCLID12-0004 and funding related to the WFIRST Science Investigation Team. MI acknowledges that this material is based upon work supported in part by the Department of Energy, Office of Science, under Award Number DE-SC0019206.

We would like to thank Salman Habib and Eske Pedersen for useful discussions related to this project. We also would like to thank Baojiu Li, for kindly providing us with the MG N-body simulation data. GV would also like to thank Rachel Bean and Hans Winther for useful discussions related to this project. NR would like to thank Andrew Hearin for assistance with Python packaging.

The emulator is built using the following Python packages: `GPflow` [123], `TensorFlow` [124] and `Scikit-learn` [125]. The analyses performed in this paper utilize the following: `Numpy` and `Scipy` [126], `Matplotlib` [127], `emcee` [119] and `GetDist` [128]. The final emulator is provided at `https://github.com/LSSTDESC/mgemu` [129].

TABLE III. Cosmological parameters of the all the COLA simulations in the suite. 50 training (M00-M49) cosmologies are tabulated.

| Model | $\Omega_m h^2$ | $n_s$ | $\sigma_8$ | $\log(f_{R_0})$ | $n$ |
|---|---|---|---|---|---|
| M00 | 0.142 | 1.038 | 0.818 | $-6.367$ | 1.878 |
| M01 | 0.152 | 0.911 | 0.884 | $-7.102$ | 2.286 |
| M02 | 0.154 | 1.034 | 0.778 | $-6.612$ | 0.816 |
| M03 | 0.135 | 0.891 | 0.827 | $-4.245$ | 1.061 |
| M04 | 0.141 | 0.952 | 0.724 | $-4.980$ | 1.224 |
| M05 | 0.149 | 0.964 | 0.896 | $-4.490$ | 2.041 |
| M06 | 0.146 | 0.928 | 0.769 | $-7.184$ | 1.633 |
| M07 | 0.139 | 0.977 | 0.851 | $-7.265$ | 0.327 |
| M08 | 0.124 | 0.960 | 0.765 | $-4.327$ | 3.429 |
| M09 | 0.151 | 0.874 | 0.757 | $-6.939$ | 1.388 |
| M10 | 0.138 | 1.030 | 0.892 | $-5.878$ | 3.347 |
| M11 | 0.146 | 0.956 | 0.814 | $-4.408$ | 0.082 |
| M12 | 0.143 | 0.981 | 0.871 | $-4.163$ | 3.592 |
| M13 | 0.144 | 0.854 | 0.741 | $-6.694$ | 0.898 |
| M14 | 0.141 | 0.887 | 0.847 | $-6.857$ | 4.000 |
| M15 | 0.139 | 0.850 | 0.806 | $-6.286$ | 2.449 |
| M16 | 0.150 | 1.042 | 0.720 | $-5.551$ | 2.531 |
| M17 | 0.154 | 0.895 | 0.863 | $-7.347$ | 1.796 |
| M18 | 0.147 | 1.021 | 0.712 | $-7.918$ | 3.837 |
| M19 | 0.131 | 0.903 | 0.700 | $-5.224$ | 3.510 |
| M20 | 0.127 | 0.985 | 0.761 | $-5.959$ | 0.735 |
| M21 | 0.130 | 1.017 | 0.810 | $-7.429$ | 1.306 |
| M22 | 0.120 | 0.932 | 0.867 | $-6.531$ | 0.980 |
| M23 | 0.122 | 0.907 | 0.798 | $-4.000$ | 0.245 |
| M24 | 0.149 | 1.001 | 0.855 | $-6.204$ | 1.143 |
| M25 | 0.126 | 0.899 | 0.745 | $-5.796$ | 2.857 |
| M26 | 0.153 | 0.993 | 0.802 | $-8.000$ | 3.755 |
| M27 | 0.155 | 0.968 | 0.733 | $-5.469$ | 3.020 |
| M28 | 0.145 | 0.883 | 0.790 | $-4.571$ | 2.694 |
| M29 | 0.148 | 0.915 | 0.839 | $-6.041$ | 3.265 |
| M30 | 0.144 | 0.997 | 0.900 | $-6.776$ | 0.490 |
| M31 | 0.128 | 0.879 | 0.888 | $-4.082$ | 0.163 |
| M32 | 0.121 | 1.005 | 0.737 | $-4.653$ | 2.204 |
| M33 | 0.125 | 0.858 | 0.880 | $-6.449$ | 1.469 |
| M34 | 0.136 | 1.050 | 0.794 | $-5.714$ | 1.551 |
| M35 | 0.134 | 1.013 | 0.753 | $-5.143$ | 0.000 |
| M36 | 0.131 | 1.046 | 0.859 | $-5.306$ | 3.102 |
| M37 | 0.132 | 0.919 | 0.729 | $-7.837$ | 0.408 |
| M38 | 0.129 | 0.944 | 0.786 | $-7.592$ | 3.918 |
| M39 | 0.137 | 0.870 | 0.708 | $-6.122$ | 0.653 |
| M40 | 0.124 | 1.026 | 0.843 | $-5.061$ | 2.612 |
| M41 | 0.151 | 0.940 | 0.835 | $-7.673$ | 0.571 |
| M42 | 0.136 | 0.923 | 0.749 | $-4.735$ | 3.673 |
| M43 | 0.129 | 0.972 | 0.831 | $-4.816$ | 2.367 |
| M44 | 0.126 | 0.936 | 0.704 | $-7.510$ | 1.959 |
| M45 | 0.134 | 0.948 | 0.876 | $-7.020$ | 2.776 |
| M46 | 0.140 | 1.009 | 0.773 | $-7.755$ | 2.939 |
| M47 | 0.121 | 0.866 | 0.822 | $-5.388$ | 3.184 |
| M48 | 0.123 | 0.989 | 0.716 | $-4.898$ | 1.714 |
| M49 | 0.133 | 0.862 | 0.782 | $-5.633$ | 2.122 |

[1] P. A. Abell et al. (LSST Science Collaborations, LSST Project), (2009), arXiv:0912.0201 [astro-ph.IM].
[2] A. Abate et al. (LSST Dark Energy Science), (2012), arXiv:1211.0310 [astro-ph.CO].
[3] M. Levi et al. (DESI collaboration), (2013), arXiv:1308.0847 [astro-ph.CO].
[4] R. Laureijs et al. (EUCLID Collaboration), (2011), arXiv:1110.3193 [astro-ph.CO].
[5] D. Spergel, N. Gehrels, J. Breckinridge, M. Donahue, A. Dressler, et al., (2013), arXiv:1305.5422 [astro-ph.IM].
[6] O. Dor et al., (2014), arXiv:1412.4872 [astro-ph.CO].
[7] S. Perlmutter et al. (Supernova Cosmology Project), Astrophys. J. **517**, 565 (1999), arXiv:astro-ph/9812133 [astro-ph].
[8] A. G. Riess et al. (Supernova Search Team), Astrophys. J. **607**, 665 (2004), arXiv:astro-ph/0402512 [astro-ph].
[9] D. J. Eisenstein et al. (SDSS Collaboration), Astrophys.J. **633**, 560 (2005), arXiv:astro-ph/0501171 [astro-ph].
[10] W. J. Percival et al., Monthly Notices of the Royal Astronomical Society **381**, 1053 (2007), arXiv:0705.3323 [astro-ph].
[11] W. J. Percival et al., (2009), arXiv:0907.1660 [astro-ph.CO].
[12] E. A. Kazin, J. Koda, C. Blake, and N. Padmanabhan, (2014), arXiv:1401.0358 [astro-ph.CO].
[13] P. Ade et al. (Planck Collaboration), (2013), arXiv:1303.5076 [astro-ph.CO].
[14] P. A. R. Ade et al. (Planck Collaboration), Astron. Astrophys. **594**, A13 (2016), arXiv:1502.01589 [astro-ph.CO].
[15] S. Weinberg, Rev. Mod. Phys. **61**, 1 (1989).
[16] S. M. Carroll, Living Reviews in Relativity **4**, 1 (2001), arXiv:astro-ph/0004075 [astro-ph].
[17] G. Efstathiou, Monthly Notices of the Royal Astronomical Society **440**, 1138 (2014), https://academic.oup.com/mnras/article-pdf/440/2/1138/18498204/stu278.pdf.
[18] A. G. Riess, S. Casertano, W. Yuan, L. M. Macri, and D. Scolnic, Astrophys. J. **876**, 85 (2019), arXiv:1903.07603 [astro-ph.CO].
[19] B. R. Zhang, M. J. Childress, T. M. Davis, N. V. Karpenka, C. Lidman, B. P. Schmidt, and M. Smith, Monthly Notices of the Royal Astronomical Society **471**, 2254 (2017), https://academic.oup.com/mnras/article-pdf/471/2/2254/19491062/stx1600.pdf.
[20] L. Lombriser, Physics Letters B **803**, 135303 (2020).
[21] C. Heymans et al., Monthly Notices of the Royal Astronomical Society **432**, 2433 (2013), https://academic.oup.com/mnras/article-pdf/432/3/2433/12627095/stt601.pdf.
[22] H. Hildebrandt et al., Monthly Notices of the Royal Astronomical Society **465**, 1454 (2016), https://academic.oup.com/mnras/article-pdf/465/2/1454/24243465/stw2805.pdf.
[23] T. Abbott et al. (DES), Phys. Rev. D **98**, 043526 (2018), arXiv:1708.01530 [astro-ph.CO].
[24] W. Lin and M. Ishak, Phys. Rev. D **96**, 083532 (2017).
[25] C. M. Will, Living Rev. Rel. **9**, 3 (2006), arXiv:gr-qc/0510072 [gr-qc].
[26] C. M. Will, Living Rev. Rel. **17**, 4 (2014), arXiv:1403.7377 [gr-qc].
[27] K. Koyama, Rept. Prog. Phys. **79**, 046902 (2016), arXiv:1504.04623 [astro-ph.CO].
[28] T. Clifton, P. G. Ferreira, A. Padilla, and C. Skordis, Phys. Rept. **513**, 1 (2012), arXiv:1106.2476 [astro-ph.CO].
[29] M. Ishak, Living Rev. Rel. **22**, 1 (2019), arXiv:1806.10122 [astro-ph.CO].
[30] P. G. Ferreira, (2019), arXiv:1902.10503 [astro-ph.CO].
[31] J. Khoury, arXiv e-prints , arXiv:1011.5909 (2010), arXiv:1011.5909 [astro-ph.CO].
[32] J. Khoury, (2013), arXiv:1312.2006 [astro-ph.CO].
[33] A. Vainshtein, Physics Letters B **39**, 393 (1972).
[34] E. Babichev and C. Deffayet, Class. Quant. Grav. **30**, 184001 (2013), arXiv:1304.7240 [gr-qc].
[35] J. Khoury and A. Weltman, Phys. Rev. D **69**, 044026 (2004).
[36] J. Khoury and A. Weltman, Phys. Rev. Lett. **93**, 171104 (2004).
[37] K. Hinterbichler and J. Khoury, Phys. Rev. Lett. **104**, 231301 (2010).
[38] K. A. Olive and M. Pospelov, Phys. Rev. **D77**, 043524 (2008), arXiv:0709.3825 [hep-ph].
[39] G. Dvali, G. F. Giudice, C. Gomez, and A. Kehagias, JHEP **08**, 108 (2011), arXiv:1010.1415 [hep-ph].
[40] G. W. Horndeski, International Journal of Theoretical Physics **10**, 363 (1974).
[41] C. Deffayet, X. Gao, D. A. Steer, and G. Zahariade, Phys. Rev. D **84**, 064039 (2011).
[42] T. Kobayashi, M. Yamaguchi, and J. Yokoyama, Prog. Theor. Phys. **126**, 511 (2011), arXiv:1105.5723 [hep-th].
[43] L. Lombriser and A. Taylor, JCAP **1603**, 031 (2016), arXiv:1509.08458 [astro-ph.CO].
[44] L. Lombriser and N. A. Lima, Phys. Lett. **B765**, 382 (2017), arXiv:1602.07670 [astro-ph.CO].
[45] J. Sakstein and B. Jain, Phys. Rev. Lett. **119**, 251303 (2017), arXiv:1710.05893 [astro-ph.CO].
[46] J. M. Ezquiaga and M. Zumalacarregui, Phys. Rev. Lett. **119**, 251304 (2017), arXiv:1710.05901 [astro-ph.CO].
[47] P. Creminelli and F. Vernizzi, Phys. Rev. Lett. **119**, 251302 (2017), arXiv:1710.05877 [astro-ph.CO].
[48] T. Baker, E. Bellini, P. G. Ferreira, M. Lagos, J. Noller, and I. Sawicki, Phys. Rev. Lett. **119**, 251301 (2017), arXiv:1710.06394 [astro-ph.CO].
[49] B. A. Abbott et al. (Virgo, LIGO Scientific), Phys. Rev. Lett. **119**, 161101 (2017), arXiv:1710.05832 [gr-qc].
[50] A. Goldstein et al., Astrophys. J. **848**, L14 (2017), arXiv:1710.05446 [astro-ph.HE].
[51] V. Savchenko et al., Astrophys. J. **848**, L15 (2017), arXiv:1710.05449 [astro-ph.HE].
[52] B. P. Abbott et al. (Virgo, Fermi-GBM, INTEGRAL, LIGO Scientific), Astrophys. J. **848**, L13 (2017), arXiv:1710.05834 [astro-ph.HE].
[53] B. P. Abbot et al., ApJLett **848**, L12 (2017), arXiv:1710.05833 [astro-ph.HE].
[54] M. Ishak et al., (2019), arXiv:1905.09687 [astro-ph.CO].
[55] A. Albrecht et al., (2006), arXiv:astro-ph/0609591.
[56] A. Aviles and J. L. Cervantes-Cota, Phys. Rev. **D96**, 123526 (2017), arXiv:1705.10719 [astro-ph.CO].

[57] A. Aviles, M. A. Rodriguez-Meza, J. De-Santiago, and J. L. Cervantes-Cota, JCAP **1811**, 013 (2018), arXiv:1809.07713 [astro-ph.CO].

[58] A. Aviles, J. L. Cervantes-Cota, and D. F. Mota, Astron. Astrophys. **622**, A62 (2019), arXiv:1810.02652 [astro-ph.CO].

[59] G. Valogiannis and R. Bean, Phys. Rev. **D99**, 063526 (2019), arXiv:1901.03763 [astro-ph.CO].

[60] G. Valogiannis, R. Bean, and A. Aviles, JCAP **01**, 055 (2020), arXiv:1909.05261 [astro-ph.CO].

[61] H. A. Winther et al., Monthly Notices of the Royal Astronomical Society **454**, 4208 (2015), arXiv:1506.06384 [astro-ph.CO].

[62] G. Valogiannis and R. Bean, Phys. Rev. **D95**, 103515 (2017), arXiv:1612.06469 [astro-ph.CO].

[63] S. Tassev, M. Zaldarriaga, and D. Eisenstein, JCAP **1306**, 036 (2013), arXiv:1301.0322 [astro-ph.CO].

[64] R. E. Smith, J. A. Peacock, A. Jenkins, S. D. M. White, C. S. Frenk, F. R. Pearce, P. A. Thomas, G. Efstathiou, and H. M. P. Couchman, MNRAS **341**, 1311 (2003), arXiv:astro-ph/0207664 [astro-ph].

[65] R. Takahashi, M. Sato, T. Nishimichi, A. Taruya, and M. Oguri, Astrophys. J. **761**, 152 (2012), arXiv:1208.2701 [astro-ph.CO].

[66] A. J. Mead, J. A. Peacock, C. Heymans, S. Joudaki, and A. F. Heavens, Monthly Notices of the Royal Astronomical Society **454**, 19581975 (2015).

[67] A. J. Mead, C. Heymans, L. Lombriser, J. A. Peacock, O. I. Steele, and H. A. Winther, Monthly Notices of the Royal Astronomical Society **459**, 14681488 (2016).

[68] E. Lawrence, K. Heitmann, J. Kwan, A. Upadhye, D. Bingham, S. Habib, D. Higdon, A. Pope, H. Finkel, and N. Frontiere, Astrophys. J. **847**, 50 (2017), arXiv:1705.03388 [astro-ph.CO].

[69] M. Knabenhans et al. (EUCLID Collaboration), MNRAS **484**, 5509 (2019), arXiv:1809.04695 [astro-ph.CO].

[70] K. Heitmann, D. Higdon, C. Nakhleh, and S. Habib, ApJLett **646**, L1 (2006), arXiv:astro-ph/0606154 [astro-ph].

[71] S. Habib, K. Heitmann, D. Higdon, C. Nakhleh, and B. Williams, Phys. Rev. D **76**, 083503 (2007), arXiv:astro-ph/0702348 [astro-ph].

[72] K. Heitmann, M. White, C. Wagner, S. Habib, and D. Higdon, Astrophys. J. **715**, 104 (2010), arXiv:0812.1052 [astro-ph].

[73] K. Heitmann, D. Higdon, M. White, S. Habib, B. J. Williams, E. Lawrence, and C. Wagner, Astrophys. J. **705**, 156 (2009), arXiv:0902.0429 [astro-ph.CO].

[74] E. Lawrence, K. Heitmann, M. White, D. Higdon, C. Wagner, S. Habib, and B. Williams, Astrophys. J. **713**, 1322 (2010), arXiv:0912.4490 [astro-ph.CO].

[75] K. Heitmann, E. Lawrence, J. Kwan, S. Habib, and D. Higdon, Astrophys. J. **780**, 111 (2014), arXiv:1304.7849 [astro-ph.CO].

[76] S. Agarwal, F. B. Abdalla, H. A. Feldman, O. Lahav, and S. A. Thomas, Monthly Notices of the Royal Astronomical Society **439**, 21022121 (2014).

[77] J. Kwan, S. Bhattacharya, K. Heitmann, and S. Habib, The Astrophysical Journal **768**, 123 (2013).

[78] B. D. Wibking, D. H. Weinberg, A. N. Salcedo, H.-Y. Wu, S. Singh, S. Rodrguez-Torres, L. H. Garrison, and D. J. Eisenstein, Monthly Notices of the Royal Astronomical Society **492**, 28722896 (2019).

[79] S. Bocquet, K. Heitmann, S. Habib, E. Lawrence, T. Uram, N. Frontiere, A. Pope, and H. Finkel, arXiv e-prints (2020), arXiv:2003.12116 [astro-ph.CO].

[80] J. DeRose, R. H. Wechsler, J. L. Tinker, M. R. Becker, Y.-Y. Mao, T. McClintock, S. McLaughlin, E. Rozo, and Z. Zhai, The Astrophysical Journal **875**, 69 (2019).

[81] T. McClintock, E. Rozo, M. R. Becker, J. DeRose, Y.-Y. Mao, S. McLaughlin, J. L. Tinker, R. H. Wechsler, and Z. Zhai, The Astrophysical Journal **872**, 53 (2019).

[82] Z. Zhai, J. L. Tinker, M. R. Becker, J. DeRose, Y.-Y. Mao, T. McClintock, S. McLaughlin, E. Rozo, and R. H. Wechsler, The Astrophysical Journal **874**, 95 (2019).

[83] T. McClintock, E. Rozo, A. Banerjee, M. R. Becker, J. DeRose, S. McLaughlin, J. L. Tinker, R. H. Wechsler, and Z. Zhai, arXiv e-prints (2019), arXiv:1907.13167 [astro-ph.CO].

[84] T. Nishimichi et al., The Astrophysical Journal **884**, 29 (2019).

[85] Y. Kobayashi, T. Nishimichi, M. Takada, R. Takahashi, and K. Osato, arXiv e-prints , arXiv:2005.06122 (2020), arXiv:2005.06122 [astro-ph.CO].

[86] W. Hu and I. Sawicki, Phys. Rev. **D76**, 064004 (2007), arXiv:0705.1158 [astro-ph].

[87] H. A. Winther, S. Casas, M. Baldi, K. Koyama, B. Li, L. Lombriser, and G.-B. Zhao, Phys. Rev. D **100**, 123540 (2019), arXiv:1903.08798 [astro-ph.CO].

[88] B. Giblin, M. Cataneo, B. Moews, and C. Heymans, MNRAS **490**, 4826 (2019), arXiv:1906.02742 [astro-ph.CO].

[89] B. Bose, M. Cataneo, T. Trster, Q. Xia, C. Heymans, and L. Lombriser, (2020), arXiv:2005.12184 [astro-ph.CO].

[90] K. Heitmann, E. Lawrence, J. Kwan, S. Habib, and D. Higdon, Astrophys. J. **780**, 111 (2014), arXiv:1304.7849 [astro-ph.CO].

[91] A. De Felice and S. Tsujikawa, Living Rev. Rel. **13**, 3 (2010), arXiv:1002.4928 [gr-qc].

[92] C. Burrage and J. Sakstein, Living Reviews in Relativity **21**, 1 (2018), arXiv:1709.09071 [astro-ph.CO].

[93] S. M. Carroll, V. Duvvuri, M. Trodden, and M. S. Turner, Phys. Rev. **D70**, 043528 (2004), arXiv:astro-ph/0306438 [astro-ph].

[94] P. Brax, C. van de Bruck, A.-C. Davis, and D. J. Shaw, Phys. Rev. **D78**, 104021 (2008), arXiv:0806.3415 [astro-ph].

[95] G.-B. Zhao, B. Li, and K. Koyama, Phys. Rev. D **83**, 044007 (2011), arXiv:1011.1257 [astro-ph.CO].

[96] K. Heitmann et al., Astrophys. J. **820**, 108 (2016), arXiv:1508.02654 [astro-ph.CO].

[97] T. Santer, B. Williams, and W. Notz, Analysis of Computer Experiments (Springer, New York, 2003).

[98] K. Q. Ye, W. Li, and A. Sudjianto, Journal of Statistical Planning and Inference **90**, 145 (2000).

[99] J. DeRose, R. H. Wechsler, J. L. Tinker, M. R. Becker, Y.-Y. Mao, T. McClintock, S. McLaughlin, E. Rozo, and Z. Zhai, Astrophys. J. **875**, 69 (2019), arXiv:1804.05865 [astro-ph.CO].

[100] H. A. Winther and P. G. Ferreira, Phys. Rev. **D91**, 123507 (2015), arXiv:1403.6492 [astro-ph.CO].

[101] R. Scoccimarro, Monthly Notices of the Royal Astronomical Society **299**, 1097 (1998), arXiv:astro-ph/9711187 [astro-ph].

[102] A. Lewis, A. Challinor, and A. Lasenby, Astrophys. J.

**538**, 473 (2000), arXiv:astro-ph/9911177 [astro-ph].

[103] M. Cataneo, L. Lombriser, C. Heymans, A. J. Mead, A. Barreira, S. Bose, and B. Li, Monthly Notices of the Royal Astronomical Society **488**, 2121 (2019), https://academic.oup.com/mnras/article-pdf/488/2/2121/28967640/stz1836.pdf.

[104] H. A. Winther, S. Casas, M. Baldi, K. Koyama, B. Li, L. Lombriser, and G.-B. Zhao, Phys. Rev. D **100**, 123540 (2019).

[105] K. Heitmann, D. Higdon, C. Nakhleh, and S. Habib, ApJLett **646**, L1 (2006), astro-ph/0606154.

[106] D. Higdon, J. Gattiker, B. Williams, and M. Rightley, Journal of the American Statistical Association **103**, 570 (2008), https://doi.org/10.1198/016214507000000888.

[107] A. Savitzky and M. J. E. Golay, Analytical Chemistry **36**, 1627 (1964).

[108] W. H. Press and S. A. Teukolsky, Computers in Physics **4**, 669 (1990).

[109] C. E. Rasmussen and C. Williams, Gaussian Processes for Machine Learning (MIT Press, 2006).

[110] J. Duchi, E. Hazan, and Y. Singer, Journal of Machine Learning Research **12**, 2121 (2011).

[111] V. Springel et al., MNRAS **475**, 676 (2018), arXiv:1707.03397 [astro-ph.GA].

[112] D. Huterer and M. Takada, Astropart. Phys. **23**, 369 (2005), arXiv:astro-ph/0412142.

[113] M. Cautun, E. Paillas, Y.-C. Cai, S. Bose, J. Armijo, B. Li, and N. Padilla, Monthly Notices of the Royal Astronomical Society **476**, 3195 (2018), arXiv:1710.01730 [astro-ph.CO].

[114] B. Li, G.-B. Zhao, R. Teyssier, and K. Koyama, Journal of Cosmology and Astroparticle Physics **2012**, 051 (2012).

[115] S. Bose, B. Li, A. Barreira, J.-h. He, W. A. Hellwing, K. Koyama, C. Llinares, and G.-B. Zhao, JCAP **1702**, 050 (2017), arXiv:1611.09375 [astro-ph.CO].

[116] L. Lombriser, Annalen der Physik **264**, 259 (2014), arXiv:1403.4268 [astro-ph.CO].

[117] J. Hartlap, P. Simon, and P. Schneider, A&A **464**, 399 (2007), arXiv:astro-ph/0608064 [astro-ph].

[118] J. Goodman and J. Weare, Communications in Applied Mathematics and Computational Science, Vol. 5, No. 1, p. 65-80, 2010 **5**, 65 (2010).

[119] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, Publications of the Astronomical Society of the Pacific **125**, 306 (2013), arXiv:1202.3665 [astro-ph.IM].

[120] A. Schneider, A. Refregier, S. Grandis, D. Eckert, N. Stoira, T. Kacprzak, M. Knabenhans, J. Stadel, and R. Teyssier, Journal of Cosmology and Astroparticle Physics **2020**, 020 (2020).

[121] N. E. Chisari et al., ApJS **242**, 2 (2019), arXiv:1812.05995 [astro-ph.CO].

[122] R. A. Fisher, in Breakthroughs in statistics (Springer, 1992) pp. 66–70.

[123] A. G. d. G. Matthews, M. van der Wilk, T. Nickson, K. Fujii, A. Boukouvalas, P. León-Villagrá, Z. Ghahramani, and J. Hensman, Journal of Machine Learning Research **18**, 1 (2017).

[124] M. Abadi et al., arXiv preprint arXiv:1603.04467 (2016).

[125] F. Pedregosa et al., Journal of Machine Learning Research **12**, 2825 (2011).

[126] T. Oliphant, "Python for Scientific Computing," (2007).

[127] J. D. Hunter, "Matplotlib: A 2D Graphics Environment," (2007).

[128] A. Lewis, (2019), arXiv:1910.13970 [astro-ph.IM].

[129] G. Valogiannis, N. Ramachandra, M. Ishak, and K. Heitmann, "MGemu: An emulator for cosmological models beyond general relativity," (2020).