Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **525,** 6336–6358 (2023) https://doi.org/10.1093/mnras/stad2700
Advance Access publication 2023 September 7

MGLENS: Modified gravity weak lensing simulations for emulation-based cosmological inference

Joachim Harnois-Déraps ,[1]★ Cesar Hernandez-Aguayo ,[2,3] Carolina Cuesta-Lazaro,[4,5,6,7] Christian Arnold ,[7] Baojiu Li ,[7] Christopher T. Davies[8] and Yan-Chuan Cai[9]

[1] *School of Mathematics, Statistics and Physics, Newcastle University, Herschel Building, NE1 7RU Newcastle-upon-Tyne, UK*
[2] *Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, D-85748 Garching, Germany*
[3] *Excellence Cluster ORIGINS, Boltzmannstrasse 2, D-85748 Garching, Germany*
[4] *Center for Astrophysics, Harvard and Smithsonian, 60 Garden St, Cambridge, MA 02138, USA*
[5] *The NSF AI Institute for Artificial Intelligence and Fundamental Interactions , Cambridge, MA 02139, USA*
[6] *Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[7] *Institute for Computational Cosmology, Department of Physics, Durham University, South Road, DH1 3LE Durham, UK*
[8] *Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, D-81679 Munich, Germany*
[9] *Institute for Astronomy, University of Edinburgh, Blackford Hill, EH9 3HJ Edinburgh, UK*

Accepted 2023 September 5. Received 2023 September 5; in original form 2022 November 11

ABSTRACT

We present MGLENS, a large series of modified gravity lensing simulations tailored for cosmic shear data analyses and forecasts in which cosmological and modified gravity parameters are varied simultaneously. Based on the FORGE and BRIDGE *N*-body simulation suites presented in companion papers, we construct 100 × 5000 deg^2 of mock Stage-IV lensing data from two 4D Latin hypercubes that sample cosmological and gravitational parameters in *f*(*R*) and nDGP gravity, respectively. These are then used to validate our inference analysis pipeline based on the lensing power spectrum, exploiting our implementation of these modified gravity models within the COSMOSIS cosmological inference package. Sampling this new likelihood, we find that cosmic shear can achieve 95 per cent CL constraints on the modified gravity parameters of $\log_{10}[f_{R_0}] < -4.77$ and $\log_{10}[H_0 r_c] > 0.09$, after marginalizing over intrinsic alignments of galaxies and including scales up to $\ell = 5000$. We also investigate the impact of photometric uncertainty, scale cuts, and covariance matrices. We finally explore the consequences of analysing MGLENS data with the wrong gravity model, and report catastrophic biases for a number of possible scenarios. The Stage-IV MGLENS simulations, the FORGE and BRIDGE emulators and the COSMOSIS interface modules will be made publicly available upon journal acceptance.

Key words: gravitational lensing: weak – methods: numerical – dark energy – dark matter – large-scale structure of Universe.

1 INTRODUCTION

Recent measurements from dedicated cosmic shear surveys such as the Kilo Degree Survey[1] (Asgari et al. 2021; van den Busch et al. 2022), the Dark Energy Survey[2] (Amon et al. 2021; Secco et al. 2022), and the HyperSuprime Camera Survey[3] (Hikage et al. 2019; Hamana et al. 2020) have established weak gravitational lensing as one of the most competitive probe of the dark sector of our Universe. In addition to constraining key parameters such as the total matter abundance $\Omega_{\rm m}$, the clustering amplitude σ_8 and the dark-energy equation of state w_0, lensing data have also been used to test the gravitational sector. Indeed, the matter density field could be affected by deviations from the theory of General Relativity (GR) on cosmic scales, where the presence of a fifth force would increase the clustering in a manner detectable by lensing (Schmidt 2008). In most viable models, a screening mechanism is invoked to suppress the impact of modified gravity (MG hereafter) on small scales or high-density regions, as required to satisfy the tight Solar System constraints on GR (Hu & Sawicki 2007). Screening can be achieved in a number of ways, including: 1 - *Chameleon mechanism* (Khoury & Weltman 2004a), in which the range of the fifth force is decreased in regions of high space–time curvature, thus, effectively hiding the additional force; 2 - *Symmetron* (Hinterbichler & Khoury 2010; Hinterbichler et al. 2011), in which the coupling of the scalar field mediating the fifth force is density dependent; 3 - *Vainshtein screening* (Vainshtein 1972), in which the screening effect is sourced by the second derivative of the field value and happens mostly on small scales; 4 - *k-mouflage* screening (Babichev, Deffayet & Ziour 2009). We refer to reader to Koyama (2016) for a review on modified theories of gravitation.

In any case, a clear detection of the resulting excess clustering in galaxy surveys is made difficult by the large uncertainty on the galaxy bias, especially on small non-linear scales. Weak gravitational lensing, however, naturally emerges as a potentially cleaner probe

★ E-mail: joachim.harnois-deraps@ncl.ac.uk
[1] KiDS:kids.strw.leidenuniv.nl
[2] DES:www.darkenergysurvey.org
[3] HSC:www.naoj.org/Projects/HSC

© 2023 The Author(s).
Published by Oxford University Press on behalf of Royal Astronomical Society. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/mnras/article/525/4/6336/7263281 by guest on 30 January 2024

of MG, being unaffected by this severe limitation (Schmidt 2008). While travelling through the foreground large scale structure on its way to our telescopes, the light emitted by distant galaxies acquires coherent distortions, which we measure in cosmic shear surveys. Recently, Harnois-Déraps et al. (2015b) constrained a series of MG models from the cosmic shear analysis of the Canada–France–Hawaii Telescope lensing survey in a pathfinder analysis. Upgraded investigations including a number of systematics inherent to cosmic shear data have since been carried out with the KiDS and DES data (Joudaki et al. 2017; Abbott et al. 2019; Tröster et al. 2021; DES Collaboration 2023), however the constraining power on MG remains weak and model-dependent. As discussed in the above references, exploring multiple MG hypotheses is essential in light of the current $S_8 \equiv \sigma_8\sqrt{\Omega_{\rm m}/0.3}$ tension between low- and high-redshift cosmological data analyses (e.g. Heymans et al. 2021), although it likely will not be the sole solution since MG moves S_8 upwards in both weak lensing and CMB data (Tröster et al. 2021), preserving the tension.

In these previous analyses, the constraints on MG parameters are derived from measurements of lensing two-point statistics, either the two-point correlation functions or the lensing power spectrum. These choices of summary statistics are largely motivated by the simplicity of their modelling, which involves tractable modifications to the matter power spectrum that are well measured from *N*-body simulations. Recent computational efforts led to public power spectrum *emulators*, which predict the enhancement of clustering for a variety of MG models, over a wide range of cosmological parameters[4,5,6,7,8]

It is expected that two-point statistics are not optimal for constraining MG, largely due to the fact that the screening mechanism is typically density-dependent. Instead, statistics that are more sensitive to low-density regions, for example those measuring signals around underdense regions (e.g. Barreira et al. 2017; Davies, Cautun & Li 2019) or upweighting these with marked correlation functions (Armijo et al. 2018; Hernández-Aguayo, Baugh & Li 2018; Peel et al. 2018), have been shown to better constrain the parameters that describe a fifth force. The main difficulty with these alternative measurement methods is the absence of theoretical models to describe this signal, forcing one to rely on emulators trained of a large number of accurate weak lensing simulations to facilitate their interpretation.

Searching for modifications to GR is a complicated enterprise, since different theories predict sometimes radically different effects on the formation of large-scale structures, making this a model-dependent search. Moreover, among all existing MG simulations, only a few have been designed to enable the extraction of weak lensing statistics at the field level, including for example the DUSTGRAIN Pathfinder (Giocoli, Baldi & Moscardini 2018), in which MG models were used to co-evolve dark matter and massive neutrinos. These simulations have shown again that non-Gaussian statistics are better suited to break down the known degeneracy between the increase in structure formation caused by the fifth force, and the decrease caused by neutrino free-streaming. Other simulation efforts studying weak lensing statistics include that of Higuchi & Shirasaki (2016), Barreira et al. (2017), Shirasaki et al. (2017), and Li & Shirasaki (2018), which examine various non-Gaussian statistics in light cones produced by the ECOSMOG modified-gravity *N*-body solver (Li et al. 2012). Fast approximate *N*-body methods such as MG-COLA (Valogiannis & Bean 2017) are generally not accurate enough to model the small scales physics probed by lensing, however speed-up of the MG sector as in Hernández-Aguayo et al. (2022) might prove helpful to reduce the computational cost overhead in the future.

We present in this work the first suite of MG weak lensing simulations designed for the analysis of current cosmic shear surveys. Based on the FORGE (F Of R Gravity Emulator) simulations described in Arnold et al. (2022, hereafter A21) and the BRIDGE (BRaneworld-Inspired DGP Gravity Emulator) simulations presented in Cuesta-Lazaro et al. (in preparation), the Modified Gravity Lensing Simulations (MGLENS) consist of two suites of lensing maps in which three cosmological and one modified gravity parameters are varied on a Latin hypercube over a volume that encloses most of the 2σ posterior allowed by current lensing surveys. The two MG scenarios are modelled separately, and their respective parameters capture the strength of the deviations from GR in the widely studied *f*(*R*) (Hu & Sawicki 2007) and the normal branch of the DGP (nDGP hereafter, see Dvali, Gabadadze & Porrati 2000) gravity models, respectively. With its 2 × 50 nodes, MGLENS has enough sampling points to emulate with better than 2.5 per cent accuracy most lensing statistics. This is timely, as upcoming surveys might be able to place stringent constraints on MG even with two-point statistics when restricted to specific gravity models (Bose et al. 2020), however even stronger constraints can be achieved with non-Gaussian lensing probes, and the latter can also help us to explore the full degeneracy between different gravity models and cosmology (Davies et al. 2019).

As a first application, we use MGLENS to validate a cosmic shear analysis pipeline based on the emulation of the matter power spectrum assuming either *f*(*R*) or nDG gravity models. We next proceed to forecast the constraining power of upcoming Stage-IV lensing surveys on the MG parameters, and investigate their degeneracy with cosmological parameters for a few representative scenarios of interest. In our second application, we deliberately analyse MGLENS simulations with the wrong gravity model and explore the impact on the inferred cosmology and gravity parameters. Upcoming companion papers will use these simulations in forecasts based on higher order statistics, where Gaussian process regression (GPR) or neural network (NN) emulators are used to interpolate the statistics between the nodes and therefore model the likelihood over the full parameter volume for these alternative measurement methods. We emphasize that the MGLENS suite is designed to cover a parameter space that is broad enough to enable the analysis of Stage-III lensing surveys, as done in Harnois-Déraps et al. (2021).

The first part of this paper summarizes the gravitational physics that are captured by the FORGE and BRIDGE simulation suites (Sections 2.1 and 2.2), while Section 2.3 includes a brief overview of their numerical implementation within the high-performance *N*-body code AREPO-MG (Arnold, Leo & Li 2019; Hernández-Aguayo et al. 2021). After reviewing the construction of our matter power spectrum emulator in Section 2.4 and the modelling aspects of weak lensing two-point statistics in Section 2.5, we describe and validate our weak lensing simulations in Section 3. We validate our cosmological inference pipeline in Section 4, then present the results from a series of likelihood analyses where we investigate the detection potential from measurements of the weak lensing power spectrum in a Stage-IV survey such as those to be probed by the Vera

[4] MGEMU:github.com/LSSTDESC/mgemu
[5] MGCAMB:github.com/sfu-cosmo/MGCAMB
[6] FORGE:bitbucket.org/arnoldcn/forge_emulator
[7] HMCODE:github.com/alexander-mead/HMcode
[8] REACT:github.com/nebblu/ReACT

Downloaded from https://academic.oup.com/mnras/article/525/4/6336/7263281 by guest on 30 January 2024

Rubin,[9] *Euclid*,[10] or Nancy Grace[11] telescopes. Finally, we explicitly demonstrate in Section 4.4 the model-dependence of such searches by running cosmological analyses on MG data assuming the wrong gravity model, recording extreme biases both on the gravity and cosmology sectors.

Throughout this paper we assume a flat ΛCDM universe.

2 BACKGROUND

Although GR is well tested on small scales in laboratory experiments and in the Solar system (e.g. Will 2006, 2014), possible deviations are at the moment largely unconstrained on cosmological scales (Mpc and above). To quantify such deviations in a self-consistent way, it is useful to develop an array of simple representative models to be used as templates for making predictions, which can be compared to observational data. There is a large (probably infinite) number of currently viable MG models, and this paper focuses on two of the most widely studied examples, namely the Hu-Sawicky *f*(*R*) and the nDGP gravity models, which we introduce in this section. Note that although these do not support self-acceleration and therefore require dark energy as well, they are two viable, representative MG models that can guide our search.

2.1 *f*(*R*) gravity

The modified Einstein equations in *f*(*R*) gravity can be obtained from a modified Einstein–Hilbert action in which the standard Ricci scalar *R* is supplemented by an algebraic function, *f*(*R*) (hence its name):

$$S = \frac{1}{16\pi G}\int \mathrm{d}^4x\sqrt{-g}(R + f(R)) + S_{\mathrm{m}}(g_{\mu\nu}, \psi_i). \tag{1}$$

In this expression G is the gravitational constant, $g_{\mu\nu}$ is the metric, $g \equiv \det(g_{\mu\nu})$ is its determinant, and S_{m} the action of the matter field, which depends on the metric and the different matter fluids ψ_i. Varying S with respect to $g_{\mu\nu}$, we obtain:

$$G_{\mu\nu} + f_R R_{\mu\nu} - g_{\mu\nu}\left(\frac{1}{2}f(R) - \Box f_R\right) - \nabla_\mu\nabla_\nu f_R = 8\pi G T^{\mathrm{m}}_{\mu\nu}, \tag{2}$$

where $R_{\mu\nu}$ and $G_{\mu\nu}$ are respectively the Ricci and Einstein tensors, ∇_μ is the covariant derivative compatible with the space–time metric $g_{\mu\nu}$ (i.e. $\nabla_\lambda g_{\mu\nu} = 0$), $\Box \equiv \nabla^\mu\nabla_\mu = g^{\mu\nu}\nabla_\mu\nabla_\nu$ is the d'Alembert operator in the 4D space–time, $f_R \equiv \mathrm{d}f(R)/\mathrm{d}R$ and $T^{\mathrm{m}}_{\mu\nu}$ is the energy–momentum tensor for matter.

Despite the small modification to the standard Einstein–Hilbert action, equation (2) differs from the usual Einstein equation in that it contains up to fourth-order, rather than second-order, derivatives of the metric, as a result of the terms $\Box f_R$ and $\nabla_\mu\nabla_\nu f_R$. However, both terms are second derivatives of a scalar quantity f_R, which indicates that the fourth-order differential equation (2) can be written as a second-order Einstein equation if f_R is treated as a (new) scalar degree of freedom (the *scalaron* field), which has its own evolution equation obtained by taking the trace of equation (2). Namely:

$$\Box f_R = \frac{1}{3}[R - f_R R + 2f(R) + 8\pi G\rho_{\mathrm{m}}] \equiv \frac{\mathrm{d}V_{\mathrm{eff}}(f_R)}{\mathrm{d}f_R}, \tag{3}$$

where ρ_{m} is the non-relativistic matter density of the Universe – this terms originates from the trace of the energy momentum tensor, and thus relativistic species do not contribute directly (i.e. through direct coupling) to the dynamics of the scalar field. In the second equality above we have defined an effective potential, $V_{\mathrm{eff}}(f_R)$, of the scalaron field.

Cosmological structure formation can be well described by the quasi-static and weak-field approximations (see e.g. Barrera-Hinojosa et al. 2021, for some quantitative analyses of beyond-Newtonian effects in cosmological settings). The former approximation applies in the limit of slow, non-relativistic, motions of matter, where the time derivatives of the metric potentials can be neglected; the latter assumes that the potentials created by large-scale structure are shallow so that their higher order products can also be ignored. In the presence of a scalar field as in the case of *f*(*R*) gravity, these approximations also apply to the scalaron itself since, as we will show shortly, the latter can be considered as the potential of the modified gravitational force. Note that in general the quasi-static approximation only means that the perturbations of the scalaron have negligible time derivatives compared to spatial derivatives,[12] though in the case of *f*(*R*) models with a viable chameleon screening mechanism, this can apply to the full scalar field f_R. Under these approximations, the modified Einstein's equation (2) and the scalaron equation of motion (3) become:

$$\nabla^2\Phi = \frac{16\pi G}{3}a^2(\rho_{\mathrm{m}} - \bar{\rho}_{\mathrm{m}}) + \frac{1}{6}a^2(R(f_R) - \bar{R}), \tag{4}$$

$$\nabla^2 f_R = -\frac{a^2}{3}[R(f_R) - \bar{R} + 8\pi G(\rho_{\mathrm{m}} - \bar{\rho}_{\mathrm{m}})], \tag{5}$$

where Φ is the gravitational potential, ∇ is the gradient operator in 3D space, and a is the scale factor. Overbars denote the cosmic mean, or background, value of the quantity. Note that the modified Poisson equation (4) can be rewritten as

$$\nabla^2\Phi = 4\pi G a^2\delta\rho_{\mathrm{m}} - \frac{1}{2}\nabla^2 f_R, \tag{6}$$

by using equation (5), with $\delta\rho_{\mathrm{m}} \equiv \rho_{\mathrm{m}} - \bar{\rho}_{\mathrm{m}}$. This shows that $-f_R/2$ can be considered as the potential of the modified gravity force.

In this work we consider the Hu & Sawicki (2007) *f*(*R*) model, for which the functional form of *f*(*R*) is given by

$$f(R) = -m^2\frac{c_1}{c_2}\frac{(-R/m^2)^n}{(-R/m^2)^n + 1}, \tag{7}$$

where $m^2 \equiv \Omega_{\mathrm{m}} H_0^2$ with H_0 and Ω_{m}, respectively, the values of the Hubble parameter and the matter density parameter today, while c_1, c_2, and n are free dimension-less model parameters, with n a non-negative integer. In the limit $|\bar{R}| \gg m^2$ (which holds for the entire cosmic history up to today in the models to be studied), the scalaron field can then be expressed as

$$f_R \simeq -\left|\bar{f}_{R_0}\right|\left(\frac{\bar{R}_0}{R}\right)^{n+1}, \tag{8}$$

where $\bar{R}_0$, $\bar{f}_{R_0}$ are, respectively, the present-day values of the background Ricci scalar and $\bar{f}_R$. We fix the value of the power-law index to $n = 1$ for simplicity (other values of n, such as $n = 0$ and 2, lead to qualitatively similar behaviours of the model, see e.g. Ruan et al. 2022) and we vary $\bar{f}_{R_0}$ in the range $[10^{-4.5}; 10^{-7.0}]$, where larger values lead to larger deviations from GR. See Arnold et al. (2022) and Table A1 below for a complete list of the exact $\bar{f}_{R_0}$ values included in this paper, along with other cosmological

[9] Rubin:www.lsst.org

[10] *Euclid*:euclid-ec.org

[11] Grace:wfirst.gsfc.nasa.gov

[12] See e.g. Oyaizu (2008); Bose, Hellwing & Li (2015) for some results showing the goodness of the quasi-static approximation in *f*(*R*) gravity.

Downloaded from https://academic.oup.com/mnras/article/525/4/6336/7263281 by guest on 30 January 2024

parameters used in our simulations. Note that hereafter we use f_{R_0} instead of $\bar{f}_{R_0}$ to improve notation.

It is well established that viable *f*(*R*) models for the late-time Universe must invoke the chameleon screening mechanism (Brax et al. 2004; Khoury & Weltman 2004a, b; Mota & Shaw 2006; Brax et al. 2008), an intrinsically non-linear behaviour originating from the functional form of *f*(*R*). The $R(f_R)$ term in the scalaron equation of motion, equation (5), can be considered as a description of the non-linear self-interaction of the scalaron and, along with its interaction with matter, this determines how f_R varies in space. If appropriate parameter values are adopted, for dense spherical objects – such as dark matter haloes in this toy example – inside a homogeneous medium of matter, f_R will transitions from the background value $\bar{f}_R$ far from the object to nearly zero at its centre, and the transition takes place in a thin shell at the boundary of the object, which means that f_R stays constant in all but a thin shell. Because f_R is the potential of the modified gravity force, this means that this force vanishes, or is efficiently 'screened', for most parts inside and outside the object. Another way to see how this screening mechanism works is by looking at equation (6), which shows that inside the object where f_R is nearly identically zero, the modified gravity force vanishes.

Large-scale structures offer a variety of environments, from high-density regions such as the cores of clusters and galaxies, to low-density regions in cosmic voids. As a result, these are ideal places for investigating signatures of chameleon screening and constraining *f*(*R*) gravity. However, it also poses a computational challenge as the non-linear nature of the chameleon mechanism can only be accurately predicted with high-resolution simulations such as FORGE.

2.2 nDGP gravity

In the gravitational model of Dvali, Gabadadze & PorratiDvali et al., all particle species are assumed to be confined to a 4D hypersurface or 'brane', while gravitons can propagate along a fourth spatial dimension and leak into the 5D 'bulk' space–time. The action of this braneworld model is given by

$$S = \int_{\text{brane}} \mathrm{d}^4x\sqrt{-g}\,\frac{R}{16\pi G} + \int_{\text{bulk}} \mathrm{d}^5x\sqrt{-g^{(5)}}\,\frac{R^{(5)}}{16\pi G^{(5)}}, \tag{9}$$

where g, R, G are the values on the brane and have the same meaning as before, while the counterpart bulk quantities are denoted by $g^{(5)}$, $R^{(5)}$, and $G^{(5)}$.

A new parameter can be introduced from the ratio between $G^{(5)}$ and G, known as the *cross-over scale* and denoted by r_c:

$$r_c \equiv \frac{1}{2}\frac{G^{(5)}}{G}. \tag{10}$$

This can be considered as a critical scale above (below) which gravity is well described by the 5D (4D) part of the action. Since r_c is a dimensional quantity, its value is often quoted via H_0r_c/c, which can be considered as the ratio between the cross-over scale and the horizon size c/H_0 (the speed of light c is dropped out hereafter since $c = 1$ in natural units).

The DGP model has two distinct branches of solutions. The first is a self-accelerating branch (sDGP), which supports an accelerated late-time cosmic expansion without the need for exotic dark energy species. The sDGP model, however, is not deemed as a viable alternative to standard ΛCDM due both to theoretical difficulties such as ghost instabilities (e.g. Luty, Porrati & Rattazzi 2003; Charmousis et al. 2006) and to tensions between its predictions and observational data sets (e.g. Fairbairn & Goobar 2006; Maartens & Majerotto 2006; Fang et al. 2008; Lombriser et al. 2009). In this paper, we work with the normal branch (nDGP; Schmidt 2009) model, for which the modified Friedmann equation is given by

$$\frac{H(a)}{H_0} = \sqrt{\Omega_{\rm m}a^{-3} + \Omega_{\rm DE}(a) + \Omega_{\rm rc}} - \sqrt{\Omega_{\rm rc}}, \tag{11}$$

in which $\Omega_{\rm rc} \equiv 1/(4H_0^2r_c^2)$. Similarly to the Hu-Sawicky *f*(*R*) model, the nDGP model does not support self-acceleration, and as a result some additional dark energy component has to be added in order to explain the late-time cosmic acceleration. This naturally makes it less appealing as an alternative to ΛCDM, but it is nevertheless widely used in studies of modified gravity as a representative model featuring the Vainshtein screening mechanism (Vainshtein 1972; Babichev & Deffayet 2013) and other interesting phenomenology. In this study, we take advantage of this flexibility by tuning the additional dark energy component $\Omega_{\rm DE}(a)$ such that it counteracts the effect on the background expansion and gives rise to an expansion history identical to that of ΛCDM: the motivation for this is to enforce expansion histories that are identical between nDGP and ΛCDM, so that the two models only differ in terms of structure formation. Therefore, departures from GR are quantified exclusively by the parameter H_0r_c. As we can see from equation (11), if $H_0r_c \to \infty$ then the expansion of the Universe reduces to ΛCDM, with the additional dark energy, whose density parameter is denoted by $\Omega_{\rm DE}(a)$ in equation (11), closer to a cosmological constant Λ.

Cosmological structure formation in the nDGP model is again governed by a modified Poisson equation:

$$\nabla^2\Phi = 4\pi Ga^2\delta\rho_{\rm m} + \frac{1}{2}\nabla^2\varphi, \tag{12}$$

and an equation of motion for the scalar field (φ) (Koyama & Silva 2007):

$$\nabla^2\varphi + \frac{r_c^2}{3\beta\, a^2c^2}\left[(\nabla^2\varphi)^2 - (\nabla_i\nabla_j\varphi)^2\right] = \frac{8\pi\, G\, a^2}{3\beta}\delta\rho_{\rm m}, \tag{13}$$

where

$$\beta(a) \equiv 1 + 2H\,r_c\left(1 + \frac{\dot{H}}{3H^2}\right) = 1 + \frac{\Omega_{\rm m}a^{-3} + 2\Omega_\Lambda}{2\sqrt{\Omega_{\rm rc}(\Omega_{\rm m}a^{-3} + \Omega_\Lambda)}}, \tag{14}$$

is a time-dependent function, with $\Omega_\Lambda \equiv 1 - \Omega_{\rm m}$. In the nDGP model we consider here, β decreases over time is always positive. The field φ is called the 'brane-bending mode', a scalar quantity describing the position of the 4D brane along the fourth spatial dimension.

Again, from equation (12), we can observe that the brane-bending scalaron field acts as the potential of a fifth force. We can deduce from equation (13) that its solutions have very different behaviours in two opposite limits: (i) low-density regions, where $\nabla^2\varphi$ is small and so the $(\nabla^2\varphi)^2$ and $(\nabla_i\nabla_j\varphi)^2$ terms in equation (13) are subdominant – in this case we have $\nabla^2\varphi \sim 8\pi Ga^2\delta\rho_{\rm m}/(3\beta)$, and so the strength of the fifth force is proportional to that of the standard Newtonian force, leading to an enhancement of Newton's constant from G to $(1 + 1/3\beta)G$; (ii) high-density regions, where $\nabla^2\varphi$ is large, but the quadratic terms in equation (13) become even larger, so that $\nabla^2\varphi \ll 8\pi Ga^2\delta\rho_{\rm m}/(3\beta)$ – in this case the fifth force term in equation (12) is much smaller than the standard Poisson term. This is essentially the Vainshtein screening mechanism at work.

The BRIDGE simulations used in this work cover nDGP models with H_0r_c values between 0.25 and 10 (see Table A1 for further details, and Cuesta-Lazaro et al., in preparation). These simulations share the same cosmological parameter values and initial conditions as the FORGE simulations, and differ only in the gravity model. Moreover,

Downloaded from https://academic.oup.com/mnras/article/525/4/6336/7263281 by guest on 30 January 2024

we matched the order in the strength of the MG parameters, such that models close to GR in FORGE are also close to GR in BRIDGE.

2.3 *N*-body simulations

To date, cosmological simulations are the only known tool for making accurate predictions of physical quantities and observables of the large-scale structure down to the small non-linear scales where perturbation theory fails. The need for simulations in the study of modified gravity models is even stronger because of the additional non-linear behaviours caused by the fifth force. Over the past decade, various simulation techniques and codes have been developed for such models (see e.g. Winther et al. 2015; Li 2018; Llinares 2018, and references therein, for some reviews and code comparison results).

The simulated lensing data described in this paper are based on the FORGE and BRIDGE simulation suites described, respectively, in A21 and Cuesta-Lazaro et al. (in preparation). Four parameters are varied simultaneously, namely the matter density parameter $\Omega_{\rm m}$, the structure growth parameter $S_8 \equiv \sigma_8\sqrt{\Omega_{\rm m}/0.3}$ where σ_8 is the usual root-mean-squared of the density fluctuations smoothed on $8\,h^{-1}$ Mpc scales, the reduced Hubble parameter h and either f_{R_0} or $H_0 r_c$ for the FORGE or the BRIDGE suite, respectively. These two 4D parameter spaces are each sampled at 50 nodes organized in a Latin Hyper cube, as detailed in Table A1. Details of the *N*-body calculations are provided in the references mentioned above, but we provide here a brief summary of the numerical methods used.

For the FORGE simulations, the non-linear evolution of the particle distribution is obtained by the AREPO Poisson solver (Springel 2010; Weinberger, Springel & Pakmor 2020), which is used to compute the standard Newtonian force. This is augmented by a multigrid relaxation solver for equation (5) based on a second-order-accurate finite difference scheme, which computes the fifth force arising from *f*(*R*) gravity on the local grid elements. Adaptive mesh refinement (AMR) is adopted, in which grid elements where the matter density exceeds some threshold are refined (split) into eight child cells with doubled spatial resolution: this ensures that higher resolution is used in regions where a higher accuracy is needed in the scalar field solver. The additional force is then interpolated onto the positions of particles and used to update their velocities using the standard leapfrog scheme, achieving second-order accuracy in the time integral. The relaxation algorithm described in Bose et al. (2017) and extended by Ruan et al. (2022) has been implemented, improving the numerical stability and convergence rate; complete details on AREPO-MG can be found in Arnold et al. (2019).

The BRIDGE simulations are also carried out with AREPO and using multigrid relaxation with the same code structure, except that we are instead solving the differential equation governing the dynamics of the brane-bending mode φ given by equation (13). Since this equation differs in type from equation (5), the algorithm introduced in Li, Zhao & Koyama (2013a); Li et al. (2013b) is used instead to ensure numerical stability. To further improve the efficiency of the code, the scheme described in Barreira, Bose & Li (2015) is used, where, instead of solving the scalar field equation on all levels of mesh refinements (labelled by l), it is only solved on the lowest few levels; in other words, the scalaron solver is truncated at some level $l = l_{\rm trunc}$, and the solutions of φ on level $l_{\rm trunc}$ is interpolated to all higher levels. Barreira et al. (2015) show that this truncation, while an approximation, leads to negligible errors in the quantities of interest in cosmology. This is because the Vainshtein screening mechanism is very efficient at suppressing the fifth force in high-density regions, which happen to be the highly refined regions of the simulation grid; while the truncation and interpolation causes certain errors in the calculated fifth force in such regions, these are small errors on a small quantity, which have a small overall impact on the simulation results. For further details of the implementation in AREPO-MG, see Hernández-Aguayo et al. (2021).

Each of the FORGE and BRIDGE simulation suites consists of a total of 200 collision-less, dark-matter-only runs covering the 50 *f*(*R*) and nDGP models mentioned above. For each node we have run two independent realizations with initial conditions chosen such that the sampling variance is greatly reduced in the mean matter power spectrum (see A21, for more details), at two different resolutions. The *high-resolution* simulations employ 1024^3 particles in a $500\,h^{-1}$ Mpc simulation box, at a mass resolution of $m_{\rm p} \simeq 9.1 \times 10^9 h^{-1} M_{\odot}$[13], while the *low-resolution* simulations evolve 512^3 particles in a simulation box size of $1500\,h^{-1}$ Mpc, with a mass resolution of $m_{\rm p} \simeq 1.5 \times 10^{12} h^{-1} M_{\odot}$ (the values of $m_{\rm p}$ quoted here are for the fiducial ΛCDM node). The gravitational softening lengths are, respectively, 15 and $75\,h^{-1}$kpc for the high- and low-resolution runs. For all simulations, we have fixed the power index of the primordial power spectrum, the present-day baryonic density parameter and the dark energy equation of state to $n_{\rm s} = 0.9652$, $\Omega_{\rm b} = 0.049199$, and $w = -1$. Note that the lensing maps described in this paper only use the high-resolution runs, and that corresponding GR-ΛCDM simulations exist for all 50 nodes.

All simulations start at $z_{\rm ini} = 127$, with initial conditions (ICs) generated using the 2LPTIC (Crocce, Pueblas & Scoccimarro 2006) code, an IC generator based on N-GENIC (Springel et al. 2005) that uses second-order Lagrangian perturbation theory to compute more accurately the initial particle displacements for a given matter power spectrum. The linear matter power spectra for all models are generated with the public Boltzmann code CAMB (Lewis, Challinor & Lasenby 2000), with the cosmological parameters specified in Table A1. Note that for all *f*(*R*) and nDGP models, we assume that the linear power spectra are identical to their ΛCDM *counterparts*, i.e. the ΛCDM models with the same cosmological parameters – this is a good approximation since at the initial time ($z = 127$) any effect of modified gravity is negligible for the models considered here. In other words, they share the same primordial amplitude $A_{\rm s}$. Finally, for each cosmological model, we pre-compute the redshifts z at which particle data[14] have to be written to disc such that the consecutive projections of half-simulation boxes can be used to construct contiguous light cones up to $z = 3.0$. We describe the construction of our mass shells in Section 3.1. We note that the matter power spectrum of the FORGE simulations has been shown in A21 to agree within a few per cent with HALOFIT for node-00 up to $k = 10h^{-1}$ Mpc, and to a slightly lesser level with approximate methods (MG-COLA) and fit functions REACT for non-GR cases. The $P(k)$ emulator itself is calibrated up to $z = 2$, beyond which the departure from GR are highly attenuated. Because of projection effects, the connection between k-scales and angular separations is not clear, however we show in Section 3 that this few per cent level of accuracy generally holds at least up to multipoles of $\ell = 5000$. We have also verified with simulations ran with higher mass resolution that scales up to $k = 8.0h^{-1}$ Mpc are converged to better than two per cent, meaning that resolution limits only affect multipoles larger than $\ell \sim 5000$ (see Appendix B).

[13] This number is for the fiducial ΛCDM model, or Node 0. The actual mass resolution varies in the 50 nodes due to their different cosmological parameter values.

[14] Dark matter haloes are also extracted and will be used in companion papers.

Downloaded from https://academic.oup.com/mnras/article/525/4/6336/7263281 by guest on 30 January 2024

It is important to emphasize here that the σ_8 and S_8 quantities reported in this work correspond to the input truth values at which the GR-ΛCDM N-body simulations are run. When turning on MG however, the non-linear excess structure generated by the fifth force increases the late time σ_8 values by an amount difficult to predict, hence our choice of labelling the simulations by their GR-ΛCDM quantities.[15]

Although this paper focuses on two-point statistics, it serves the additional purpose of presenting the infrastructure necessary for companion papers based on lensing statistics beyond two-point. One of the key ingredients for such measurements is the covariance matrix, for which analytical solutions generally do not exist. We therefore use the public SLICS simulations.[16] For this, a suite of 954 fully independent N-body runs that evolve 1536^3 particles in a box size of 505 h^{-1} Mpc on the side. These are all produced at a fixed cosmology[17] and vary only in their initial conditions, therefore providing an ideal tool for estimating sample covariance. We refer the reader to Harnois-Déraps & van Waerbeke (2015) for full details on the SLICS N-body ensemble. Both of these approaches currently assume GR and are therefore designed to analyse data in which we search for weak MG signature. We decided to keep this matrix fixed even for stronger models, which results in error bars that are likely underestimated. A non-GR analytical covariance matrix could be obtained by using the FORGE $P(k)$ emulator instead of HALOFIT in its calculations, which would likely result in slightly larger error bars (unless it is ran at a cosmology with smaller S_8) which would make MG detection even more difficult. The most accurate posterior is obtained when the covariance matrix is evaluated at the best-fitting cosmology. This does not always make a noticeable difference in the end (see e.g. Burger et al. 2023), hence we leave for the future the study of the dependence covariance matrix on gravitocosmological parameter.

The SLICS, FORGE and BRIDGE simulations are post-processed uniformly, creating mock survey light cones suitable for cosmological inference. Details on the post-processing involved are presented in Section 3.1. Beforehand, we first introduce the basic ingredients that enter our theoretical predictions.

2.4 Modified gravity emulators

The information content of the large-scale structure is largely encapsulated in the matter power spectrum, $P_\delta(k;\, z)$, a two-point statistics that is directly measurable from the matter density fields δ in simulations and that can be inferred from galaxy surveys via clustering or cosmic shear measurements. For example, the N-body simulations described in A21 are used to construct the public P_δ FORGE emulator, obtained by training a Gaussian process regressor (GPR) on the measurements obtained from the 50 FORGE nodes; the emulator provides predictions that are accurate to better than 2.5 per cent up to $k = 10.0\,h\,\mathrm{Mpc}^{-1}$ over the majority of the parameter volume.

As an alternative, we use here the same measurements to train instead fully connected neural networks (FCNN), which are especially powerful at high-dimensional interpolation (as in Cuesta-Lazaro et al., in preparation). We train in this work a neural network with the same characteristics on both FORGE and BRIDGE data, as a function of redshift. Because the number of training simulations is relatively small, we found empirically that larger networks tend to overfit the training data. We ran hyperparameter optimization with OPTUNA,[18] and as long as the number of hidden units was kept small we found no significant benefits of further optimizing the model. In the end, we opted for a neural network defined by an input layer composed of the four cosmological parameters (Ω_m, h, σ_8 and the modified gravity parameter, either $\bar{f}_{R0}$ for $f(R)$, or $H_0 r_c$ for nDGP gravity) plus the redshift z, two hidden layers of 256 units each, and an output layer that returns the power spectrum evaluated at the different k-bins. In between hidden layers, we use a Gaussian error linear unit activation function (Hendrycks & Gimpel 2016) to add a differentiable non-linearity to the relation between inputs and outputs.

To find the optimal parameters that reproduce the statistics measured in the N-body simulations, we minimize the $\mathcal{L}_1$ loss function, defined as:

$$\mathcal{L}_1 = \frac{1}{N}\sum_{i=0}^{N} |y^i_\mathrm{true} - y^i_\mathrm{predicted}| \tag{15}$$

using the Adam optimizer (Kingma & Ba 2014). In the above expression, the y^i are the true and predicted matter power spectra for each of the simulations and each of the snapshots in the simulation suite, and N is the batch size used in the training.

Moreover, we avoid fine-tuning the learning rate with a scheduler that reduces the learning rate by a factor of 10 when the validation loss does not improve after 30 epochs. We also stop training the model when the validation loss does not improve after 100 epochs. An in-depth description of the emulator and its validation are presented in Cuesta-Lazaro et al. (in preparation), together with the emulator's code. More precisely, the emulator outputs the modified gravity *enhancement factor*, $B(k,\, z)$, which is defined as:

$$B(k; z) = P_{\delta,\mathrm{MG}}(k; z)/P_{\delta,\mathrm{HaloFIT}}(k; z). \tag{16}$$

Here $P_{\delta,\mathrm{MG}}(k; z)$ is the measurement for a modified gravity model from either the FORGE or BRIDGE simulations, and $P_{\delta,\mathrm{HaloFIT}}(k; z)$ is the prediction by HALOFIT (Takahashi et al. 2012) for the ΛCDM counterpart of that model (we refer the reader to A21 for a more complete description of how this is achieved in practice). The MG enhancement can be as high as 40 per cent depending on the model, for the FORGE nodes. We find that the FCNN slightly outperforms the GPR at modelling the enhancement factor and is therefore our method of choice, for all gravity models.

Finally, we notice that in the weak f_{R_0} limit the emulator does not converge exactly to the GR case: residual deviations of a few per cent are observed at all scales. These same residuals are also present in the power spectrum training set, as reported in fig. 5 of A21. Although generally small, some segments of our analysis require a smooth convergence to GR, hence we linearly interpolate the emulated $B(k)$ in the range $\log_{10}[f_{R_0}] = [-7, -6.0]$, enforcing $B(k) = 1.0$ at the lower end and for any values smaller than -7. The weak nDGP limit does not show such residuals and hence interpolation is not necessary in that case.

2.5 Cosmic shear two-point functions

Two-point functions are the primary cosmic shear measurement methods and exists in different flavours, including two-point cor-

[15] We use σ_8 and S_8 in place of σ_8^GR and S_8^GR to declutter notation.

[16] SLICS: slics.roe.ac.uk

[17] GR-ΛCDM SLICS cosmology: $\Omega_\mathrm{m} = 0.2905$, $\sigma_8 = 0.826$, $h = 0.6898$, $n_\mathrm{s} = 0.969$.

[18] : OPTUNA: optuna.org

Downloaded from https://academic.oup.com/mnras/article/525/4/6336/7263281 by guest on 30 January 2024

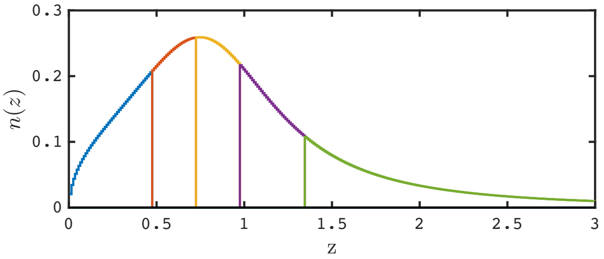

Figure 1. Normalized redshift distribution of the five tomographic bins considered in our mock survey.

Table 1. Properties of our Stage-IV survey. The specifications closely follow those presented in Martinet et al. (2021a), with $n_{\rm eff} = 6.0$ gal arcmin^{-2} per tomographic bin and $\sigma_\epsilon = 0.27$ per component.

tomo	z range	$\langle z \rangle$
bin1	0.0–0.4676	0.286
bin2	0.4676–0.7194	0.600
bin3	0.7194–0.9625	0.841
bin4	0.9625–1.3319	1.134
bin5	1.3319–3.0	1.852

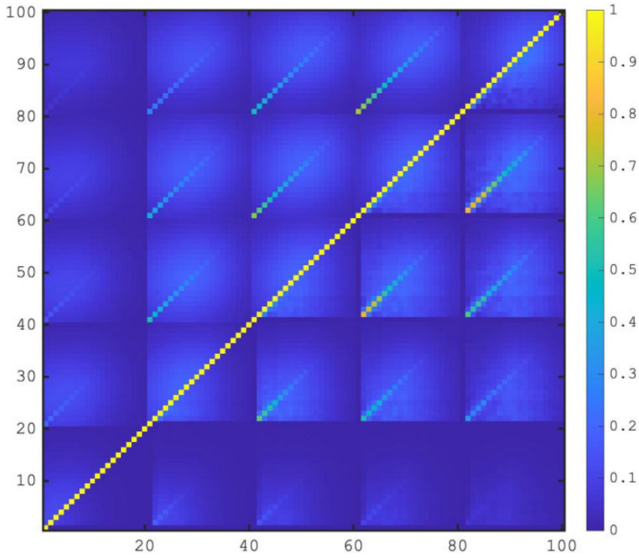

Figure 2. Cross-correlation coefficient matrix of our lensing power spectrum data vector, defined as $r_{ij} = C_{ij}/\sqrt{C_{ii}C_{jj}}$. The upper-left triangle shows the analytical calculations, while the lower right part is estimated from 954 fully independent convergence maps constructed from the SLICS (Section 3.1). The redshift bins increase towards the upper-right corner. We show here the autocorrelations only to enhance the visibility, but cross-redshift correlations show a similar level of agreement.

relation functions, angular power spectra, band powers, or COSEBIs (see Asgari et al. 2021, for a comparison between some of these). One of the key advantage of these cosmic shear statistics is that their modelling can be directly linked to the matter power spectrum, $P_\delta(k; z)$. Thanks to an increased precision in the estimation of the redshift distributions, the lensing catalogues are now routinely split into different redshift bins, allowing for tomographic analyses of the data that better measure those parameters impacting the growth rate of large-scale structure. Specifically, predictions for the cosmic shear power spectrum $C_\ell^{\kappa,ij}$ can be obtained from a Limber integration over the matter power spectrum via (see Kilbinger et al. 2017, for a review):

$$C_\ell^{\kappa,ij} = \int_0^{\chi_{\rm H}} \frac{q^i(\chi)\,q^j(\chi)}{\chi^2}\, P_\delta\left(\frac{\ell+1/2}{\chi}; z(\chi)\right) {\rm d}\chi, \qquad (17)$$

where $\chi_{\rm H}$ is the comoving distance to the horizon, and (i, j) label the different tomographic bins. The lensing kernels $q^i(\chi)$ are computed as:

$$q^i(\chi) = \frac{3}{2}\Omega_{\rm m}\left(\frac{H_0}{c}\right)^2 \frac{\chi}{a(\chi)} \int_\chi^{\chi_{\rm H}} n^i(\chi')\frac{{\rm d}z}{{\rm d}\chi'}\frac{\chi'-\chi}{\chi'}{\rm d}\chi', \qquad (18)$$

where $n^i(z)$ is the redshift distribution of the source galaxies in tomographic bin i.

The matter power spectrum entering equation (17) can be obtained from an array of public codes such as HALOFIT (Takahashi et al. 2012), HMcode (Mead et al. 2021), COSMICEMU (Heitmann et al. 2014), BACCOEMULATOR (Angulo et al. 2021), or the EUCLIDEMULATOR (Euclid Collaboration; Knabenhans et al. 2019). Whereas these codes provide highly accurate predictions tools for many cosmological models, their gravity model is restricted to that of GR only. We therefore generate MG lensing predictions by multiplying the HALOFIT predictions by $B(k; z)$ as in equation (16), and then inserting the results into equation (17). The Limber integral is carried out by COSMOSIS[19] cosmology package (Zuntz et al. 2015), which we also use for parameter inference (see Section 4).

Our mock Stage-IV lensing survey is designed to investigate some of the conditions that would allow MG to be detected by upcoming two-point statistics analyses. We opted for a source redshift distribution described by:

$$n(z) = A\frac{z^a + z^{ab}}{z^b + c}, \qquad (19)$$

with $A = 1.7865$, $a = 0.4710$, $b = 5.1843$, $c = 0.7259$. This $n(z)$ is shown in Fig. 1 and is taken from Martinet et al. (2021a, b) and Harnois-Déraps, Martinet & Reischke (2022). This sample is further split into five tomographic bins, each with a galaxy density of $n_{\rm gal} = 6.0$ gal arcmin^{-2} and shape noise of $\sigma_\epsilon = 0.27$ per component. Our method assumes no overlap between the tomographic bin, a simplifying assumption that does not occur in realistic data distributions but is of no consequence in a cosmic shear forecast. A summary of the mock survey properties is presented in Table 1. We assume a survey area of 5000 deg^2, which corresponds to the total area sampled by our flat-sky simulations at each cosmological nodes (see Section 3).

Section 4 details our likelihood sampling analysis, which takes as input a data vector, a covariance matrix, and a theoretical model in which cosmology, gravity, and nuisance parameters are varied simultaneously. As our baseline we use an analytical covariance matrix that describes the mode correlations, the shape noise, and the sampling covariance expected for the different elements of our data vector. The calculations are fully detailed and validated in Harnois-Déraps, Giblin & Joachimi (2019) and Joachimi et al. (2021a) and we refer the reader to these for more information. In short they include the Gaussian, non-Gaussian, and Super-Sample Covariance terms given a cosmology, a tomographic redshift distribution, a survey area, and binning specifications for the angular multipoles. The non-Gaussian term requires an expensive trispectrum evaluation, while the SSC term assumes a circular survey geometry of 5000 deg^2. We show in Fig. 2 the cross-correlation coefficient matrix obtained with

[19] COSMOSIS: cosmosis.readthedocs.io/en/latest/index.html

Downloaded from https://academic.oup.com/mnras/article/525/4/6336/7263281 by guest on 30 January 2024

our survey specifications,[20] and compare our results to an estimate obtained from the SLICS, which we describe in Section 3.1. Aside some residual noise patterns, both methods completely agree. We will quantify the impact of switching between these two later on, but basically the effect is completely subdominant given our statistical precision. This comparison validates both the theoretical approach and the SLICS maps, which will be used in companion non-Gaussian statistics studies.

3 WEAK LENSING SIMULATIONS

The MGLENS weak lensing simulations are constructed by ray-tracing[21] through series of mass shells obtained by collapsing the particle data either along one of the Cartesian axes (flat-sky method) or along the radial direction (curved-sky). Both methods have their pros and cons; we focus mainly on the flat-sky results in this paper for their ability to probe deeper in the small, non-linear regime, and discuss the curved-sky method in Appendix A. In either case, the mass sheets have a comoving thickness equal to exactly half the simulation box size (i.e. 250 h^{-1} Mpc), and between 15 and 23 shells are needed to continuously fill the light cones up to $z = 3$, depending on cosmology. We finally produce convergence maps for the five tomographic redshift bins shown in Fig. 1. In this paper we do not train our emulator on statistics measured from these maps and instead aim for their validation, however this logical extension will be presented in companion papers.

3.1 Weak lensing maps and power spectra

Our flat-sky method heavily builds from the SIMULLENS algorithm, the multiple-plane technique described in Harnois-Déraps & van Waerbeke (2015): at each pre-selected redshift, the particles from half the simulation volume are projected along the shorter direction and assigned onto a $12\,288^2$ grid. This process is repeated with the other half-volume, and for the other two Cartesian axes, such that six density planes are extracted per snapshot.

Light-cone mass maps, $\delta_{2D}(\boldsymbol{\theta}, z_{\rm lens})$, are extracted from the density planes with an opening angle of 10 deg^2 and 7745^2 pixels. At each redshift, one of the six aforementioned planes is randomly selected and a random origin offset is added. This means that correlations between different mass shells are broken, but it was shown in Zorrilla Matilla, Waterval & Haiman (2020) that this has a subdominant effect on weak lensing statistics due to the line-of-sight projection. Closely following Harnois-Déraps et al. (2019), we repeat this whole ray-tracing procedure in order to create 25 *pseudo*-independent light cones $\delta_{2D}(\boldsymbol{\theta}, z_{\rm lens})$ maps from each N-body run.[22] Periodic boundary conditions are used wherever the area of the light cone becomes larger than the simulation box itself.

In the multiple-plane approximation, each of these mass shells acts as a discrete gravitational lens, distorting the light as it passes through it. Within the Born approximation, the convergence κ experienced by photons propagating along the direction $\boldsymbol{\theta}$ and originating from a source redshift distribution $n(z)$ can be computed as:

$$\kappa^i(\boldsymbol{\theta}) = \sum_{\rm lens} \delta_{2D}(\boldsymbol{\theta}, z_{\rm lens})\; q^i(\chi(z_{\rm lens})), \tag{20}$$

where $q^i(\chi)$ is the tomographic lensing kernel given by equation (18), and the index 'lens' runs over all foreground lens planes in the light cone.

The cosmic shear power spectra are estimated from the square of the Fourier-transformed convergence map, first averaged in annuli of thickness $\Delta\ell = 35$ centred on $\ell \in [35–5000]$:

$$C_\ell^{\kappa,ij} = \langle \kappa^i(\boldsymbol{\ell})\kappa^j(\boldsymbol{\ell})\rangle_{{\rm d}\Omega}, \tag{21}$$

with $\langle ...\rangle_{{\rm d}\Omega}$ denoting an angular average over the solid angle of the annulus. Our measurements are then rebinned into 25 logarithmically spaced bands over the same ℓ-range to further reduced the sampling noise. We refer the reader to Harnois-Déraps & van Waerbeke (2015) for more details on the numerical implementation of our lensing power spectrum estimation method, which includes a mass-assignment de-biasing step; we have also checked that our measurements are consistent with those using the public code LENSTOOLS[23] (Petri 2016). Our fiducial cosmological inference excludes $\ell < 150$ modes as these are not well measured on our 10×10 deg^2 patches, and are affected by the finite lens thickness. The high-ℓ limit is an optimistic scenario, since in the real Universe these multipoles are plagued with systematic effects such as baryonic feedback, which are difficult to model and largely uncertain (Chisari et al. 2018). We therefore consider as well a more conservative scenario that further excludes the $\ell > 3000$ modes. Note that we only extract the auto-angular power spectra in this work, however it is straightforward to extend this to include cross-redshift terms as well.

3.2 Validation

As a first validation test, we examine the fractional error between the C_ℓ measured from the FORGE and BRIDGE simulations and their respective emulator predictions. We can see in Fig. 3 that the agreement is generally at the few per cent level except for the lowest redshift bin, where the deviations are much larger. These are caused by reduced accuracy in the multiple lens approximation, combined with flat-sky projection effects and broken correlations, yielding tilted residuals in the left-most panel. Note that however large this might seem, the precision of lensing surveys is massively reduced at low redshifts, as seen by the black dashed lines, such that these differences should not lead to noticeable biases at the inference stage. On small scales (large ℓ-modes) most of the measurements scatter inside the 2.5 per cent region, consistent with the advertised 2.5 per cent accuracy on the power spectrum emulator reported in A21. The intermediate scales ($300 < \ell < 1000$) exhibit a larger scatter reaching ~ 5 per cent at times, caused by limits in the emulator predictions combined with a small amount of residual sampling variance. For every tomographic bin, we have verified that the mean fractional error over all models and all ℓ-modes is always less than 0.007, which corresponds to $0.5\sigma_{\rm stat}$ in the highest tomographic bin, and much less in all other bins. From this we can expect that emulation of weak lensing statistics from these simulations should also reach 2–3 per cent absolute accuracy. This is also validated at the cosmological inference level presented in Section 4.3.

We next compare the gravitational and cosmological dependence of the signal measured in simulations to that computed by the

[20] We use the SLICS cosmology in the analytical covariance matrix calculations.

[21] Ray-tracing in this paper assumes the Born approximation.

[22] We change the random seeds between the 25 cones at a given cosmology, but use the same 25 seeds for every cosmology node, thereby keeping to a minimum the sampling variance across cosmological models.

[23] lenstools.readthedocs.io/en/latest/

Downloaded from https://academic.oup.com/mnras/article/525/4/6336/7263281 by guest on 30 January 2024

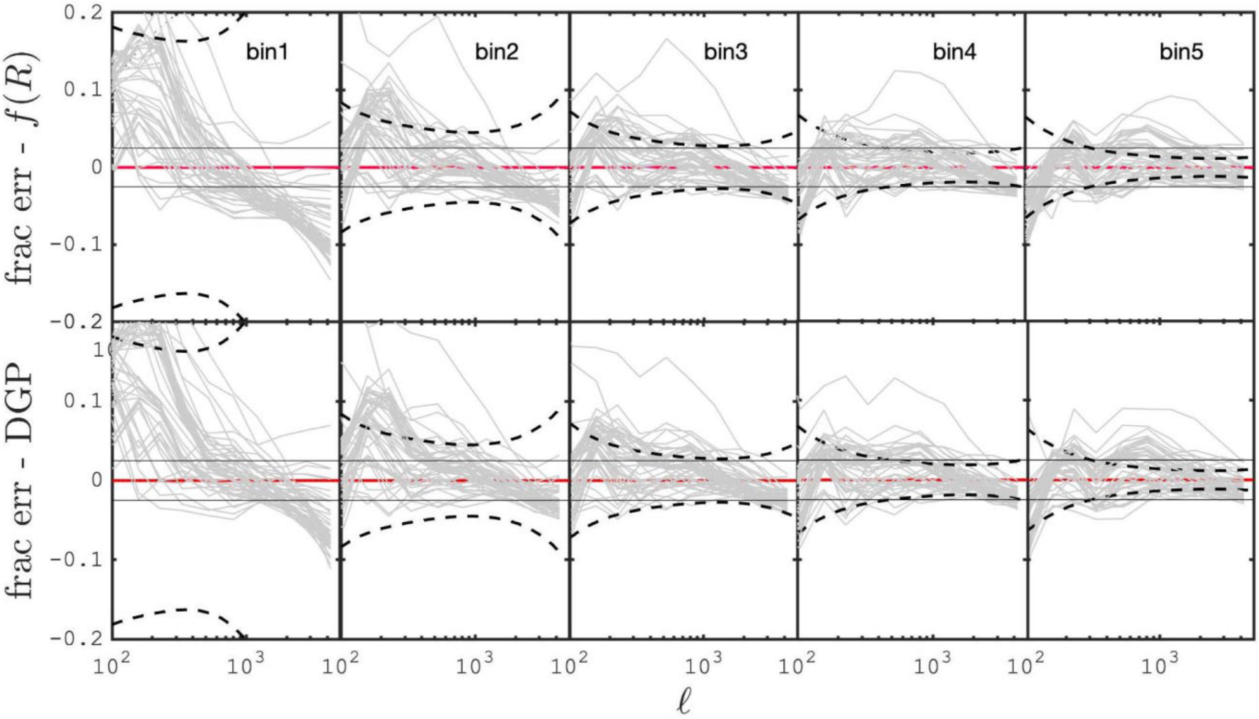

Figure 3. Fractional error between the emulated lensing power spectrum and that measured from the FORGE (upper) and BRIDGE (lower) simulations. The grey lines are obtained for all 50 nodes, each time averaged over the 50 light cones (two per initial conditions). The black dashed lines indicate the 1σ statistical error expected from our mock survey. Redshift increases from left to right, and the thin horizontal lines mark the 2.5 per cent precision target.

emulators, shown in Fig. 4 for a representative sample of FORGE models. These are labelled as *strong* (model-13), *medium* (model-18), and *weak* (model-04), referring to the strength of their departure from GR. The match is excellent here and for most other cases; discrepancies occur only for a handful of nodes with exceptionally low $\Omega_{\rm m}$, which are poorly modelled by HALOFIT and by the FORGE emulator. This is well documented in A21 and is not expected to affect our cosmology and gravity inference results, which are all centred on larger values of the matter density. The emulator predictions (in red solid) is generally within the statistical precision of our mock survey (shown with the dashed black lines) for $\ell < 1000$, beyond which it occasionally deviates by a few per cent. This is caused by residual inaccuracies in the FORGE emulator itself, which was reported in A21 (see their fig. 5) to emulate the simulated matter power spectrum only to a few per cent precision. Similar agreements are found for all other FORGE and BRIDGE models, which validates both the cosmology dependence of the light cones and the COSMOSIS implementation of the two MG emulators.

Also shown in Fig. 4 are the predictions for the pure GR case (see the thin red-dashed curve), obtained by setting $B(k, z) = 1.0$ while keeping the cosmology unchanged. The difference with respect to the solid red line is solely due to the absence of the fifth force, and falls well outside the statistical error for most models. In other words, in absence of observational and astrophysical systematics that are not included in this figure, deviations from GR would likely be observed to a high significance in our survey, if the Universe followed either the medium or strong FORGE models. This raises a key question: given our mock survey, how weak could be detectable deviations from GR, if they exist? The first step in answering this is to understand what redshift and angular scales mostly contribute towards such a measurement, an exercise that we carry out next with a Fisher analysis.

3.3 Fisher information

The origin of the constraining potential on f_{R_0} and H_0r_c from measurements of the lensing power spectrum is best understood by first fixing the cosmology in the emulators and varying only the modified gravity parameter. This is shown in Fig. 5 for cosmology otherwise identical to our GR simulation (model-00), where we compare the measurements from the flat-sky GR-ΛCDM simulations (solid black) to the FORGE and BRIDGE predictions with different values of their MG parameters (the thin dotted lines). Also shown are the expected statistical uncertainty. This figure suggests that the information about the f_{R0} parameter mostly comes from the high redshift and high-ℓ modes, where the deviations with respect to GR are amplified and the statistical error bars vastly reduced. In comparison, the constraints on H_0r_c arise from larger scales as well, again with the strongest detection potential coming from the highest redshift bins. This difference is driven by the type of fifth forces and screening mechanisms. In this section we dissect these signals and shine light on the data elements that better contribute to their constraints.

We carry out this investigation with a Fisher analysis (see e.g. Takada & Jain 2009, for a similar Fisher matrix calculation), which is cheaper to run than a full MCMC while providing exactly the information we are seeking. Given measurements of the lensing power spectrum, the Fisher information about a parameter π is obtain from

$$\mathcal{F}_\pi = \left[\frac{{\rm d}C_\ell}{{\rm d}\pi}\right] {\rm Cov}^{-1} \left[\frac{{\rm d}C_\ell}{{\rm d}\pi}\right]^{\rm T}, \tag{22}$$

where Cov is the covariance matrix shown in Fig. 2, which we assume to be cosmology independent in our calculation. A matrix product is taken between the three terms, resulting in a single scalar quantity

Downloaded from https://academic.oup.com/mnras/article/525/4/6336/7263281 by guest on 30 January 2024

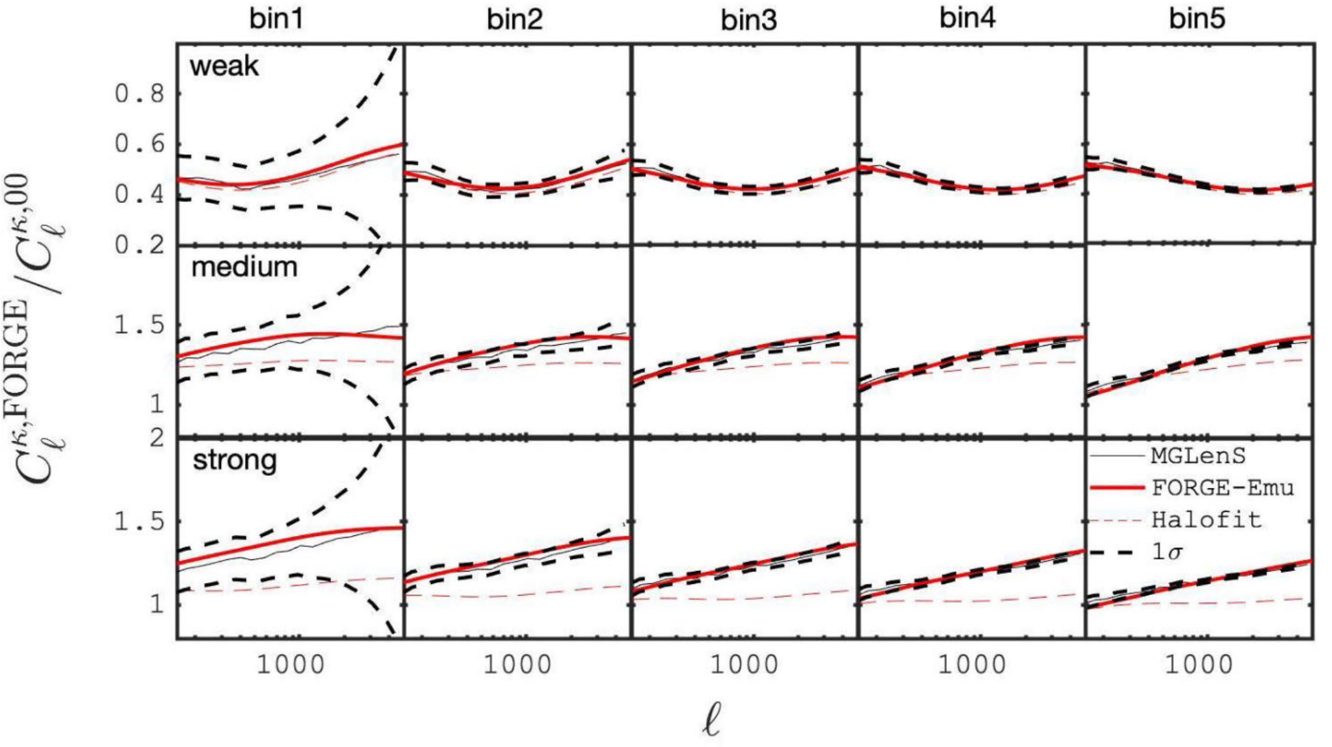

Figure 4. Ratio between the tomographic weak lensing power spectrum of different FORGE models and that of the GR model (model-00). The three rows respectively refer to three chosen models with different strengths of deviation from GR: *weak* (model-04), *medium* (model-18), and *strong* (model-13). Departure from unity is caused by differences in both cosmological and gravitational parameters. The main objective of this figure is to show that measurements from MGLENS maps (shown by the thin black lines) are in excellent agreement with the predictions of C_ℓ^κ using the FORGE matter power spectrum emulator (the solid red lines). The pair of thick dashed lines indicate the $\pm 1\sigma$ statistical uncertainty expected from our mock Stage-IV lensing survey, and redshift increases from left to right, as indicated above the upper panels. As a comparison, we also plot with the thin dashed red lines the GR predictions from HALOFIT at these cosmologies. The BRIDGE simulations and predictions reach a similar level of agreement.

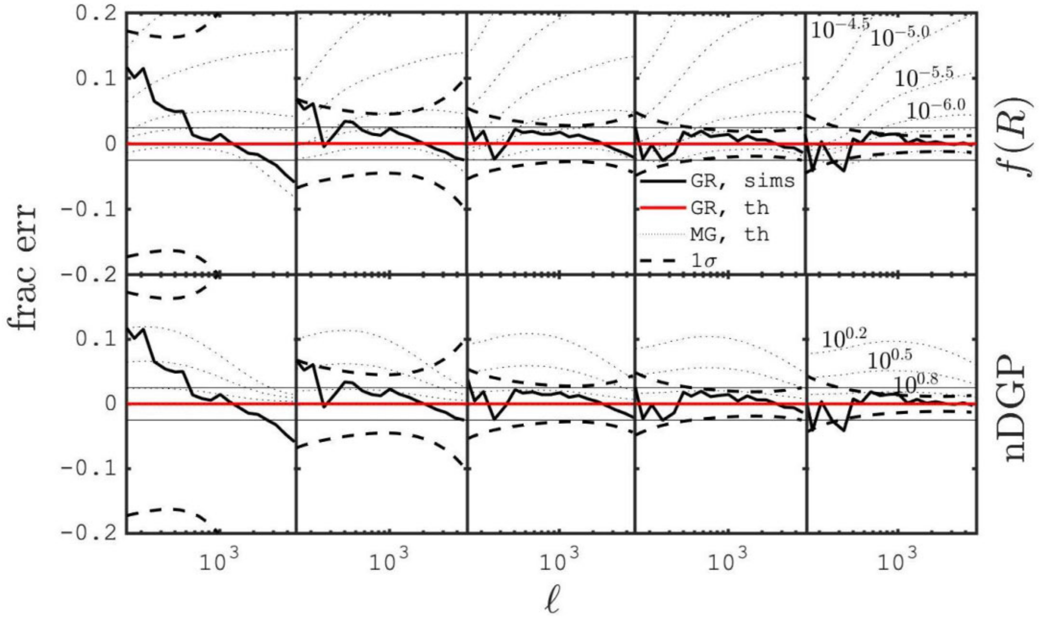

Figure 5. Top: Comparison between the lensing measurements on our GR simulations (model-00, shown with the black solid) relative to GR-theory (obtained from HALOFIT, red solid), along with the expected 1σ error from a 5000 deg^2 tomographic Stage-IV cosmic shear survey (dashed black). Predictions from *f*(*R*) models with respect to GR are shown as thin dotted lines, which can be used as a rough indicator of how well these models can be constrained. As before, redshift bins increases from left to right. Bottom: Same as top panels, but now the dotted lines show nDGP model with different values of $H_0 r_c$.

Downloaded from https://academic.oup.com/mnras/article/525/4/6336/7263281 by guest on 30 January 2024

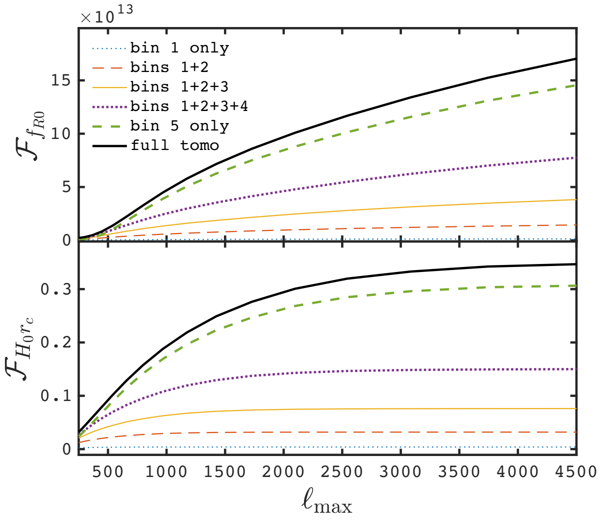

Figure 6. Fisher information as a function of $\ell_{\rm max}$, the highest mode included in the data vector, shown here for different selections of tomographic bins. The top and bottom panels are for f_{R_0} and $H_0 r_c$, respectively.

per parameter π. In short, the contribution to the information is large for elements of the data vector that are highly sensitive to changes in π (i.e. their derivative is large) and for which the covariance is small (the inverse is large). The inverse of $\mathcal{F}$ provides an optimistic estimate of the covariance about π, which in our 1D case gives $\sigma_{f_{R0}} = \sqrt{\mathcal{F}^{-1}_{f_{R0}}}$ and $\sigma_{H_0 r_c} = \sqrt{\mathcal{F}^{-1}_{H_0 r_c}}$.

Starting our dissection, we compute the Fisher information for different selections of the full data vector, first allowing variations in the maximal multipole included in our survey, $\ell_{\rm max}$. The results are shown in Fig. 6 with the solid black line, where for f_{R_0} we observe that the increase in information remains significant for all scales included here. We notice a slight transition past $\ell_{\rm max} = 1500$ where the slope becomes shallower, due to the non-linear coupling between the different Fourier modes (Takada & Jain 2009). The flattening of the slope is more pronounced for $H_0 r_c$, where a full information saturation is observed beyond $\ell = 3000$, similar to that found by Takada & Jain (2009, see their fig. 3) when estimating the information content about the global amplitude of the matter power spectrum. It is generally true that including more angular scales results in an increase of Fisher information about almost any parameter, however the rate at which the Fisher information grows and saturates, and its dependence on redshift, allows us to better understand what parts of the data are most useful.

We next explore the impact of adding each of the tomographic bins one at a time. The second line from the top shows the information contained solely in the highest tomographic bin, while the other lines correspond to different combinations of the lower redshift bins. It is clear from this that most of the information is contained in bin 5, the other four bins providing only a modest additional gain.

Using all scales and all tomographic bins, we could expect a detection of at least 3σ if $f_{R0} > 2.3 \times 10^{-7}$ or if $H_0 r_c < 5.1$, in absence of systematics and assuming that the cosmology is perfectly known from external data. We could include variations with cosmology and marginalization over systematics in an upgraded Fisher calculation, however we choose instead to run full MCMC on mock data, yielding the most accurate picture of the inference capabilities provided by the MGLENS simulations.

Table 2. Priors used in our cosmological inference. Except for δz^i, all parameters are sampled with a uniform (i.e. flat) prior; a Gaussian prior of width 0.01 is applied on the redshift parameters, reflecting a realistic precision we should have on the redshift distributions. The last two parameters are sometimes held fixed, see the main text for more details.

Parameter	Range
$\Omega_{\rm m}$	0.1–0.55
$S_8^{\rm GR}$	0.6–0.9
h	0.6–0.82
$\log_{10}[f_{R_0}]$	−8.0 – −4.5
$\log_{10}[H_0 r_c]$	−0.6–1.0
$A_{\rm IA}$	−5.0–5.0
δz^i	−0.1–0.1

4 COSMOLOGY INFERENCE

This section presents the inference method with which we quantify our ability to distinguish cosmological and gravitational parameters in different scenarios. After validating our inference pipeline on predictions obtained from the FORGE and BRIDGE $P(k)$ emulators, we run a sensitivity test on both MG models, this time varying both cosmological and gravity parameters but first ignoring secondary signals and systematic uncertainties. We next validate the pipeline on measurements from the MGLENS simulations, then investigate the catastrophic impact of analysing mock MG data with the wrong gravity model, thereby demonstrating the strong model-dependence of this approach. We finally study the impact of various systematics effects on these measurements.

In all cases our data vector consists of the auto- and cross-spectra measured from the weak, medium, and strong FORGE/BRIDGE models in all five tomographic bins. Our likelihood assumes a standard multivariate Gaussian functional form with a fixed covariance matrix (see Section 2.5). The predictions are computed at arbitrary cosmologies using equation (17) augmented with the $B(k,\, z)$ emulators, with a flat prior on the four parameters ($\Omega_{\rm m}$, S_8, h, and either $\log_{10}[f_{R_0}]$ or $\log_{10}[H_0 r_c]$) that spans the range for which the emulators are supported, listed in Table 2. One exception to this rule is the lower bound on $\log_{10}[f_{R_0}]$ which we set to -8 in order to reduce prior effects in the weak MG limit. Otherwise the inference pipeline could wrongly reject $\log_{10}[f_{R_0}] \sim -7$ simply because it is poorly sampled. As explained before, we set $B(k,\, z)$ to 1.0 whenever $\log_{10}[f_{R_0}] \in [-8, -7]$. Since the MG parameter range extends over several orders of magnitude, sampling them in log-space reduces prior volume effect that would otherwise artificially upweight the larger values. In theory, one would need to sample MG values up to $\pm\infty$, to recover GR, but in practice $\log_{10}[f_{R_0}] = -8$ and $\log_{10}[H_0 r_c] = 1.0$ are undetectable with the Stage-IV survey under consideration here and therefore serve as our GR limits. As we discuss later, 1D posteriors significantly overlapping with these limits are consistent with GR and only yield one-sided limits on the MG parameters. The other cosmological parameters are held fixed to the values used in the N-body runs. In order to better focus on the gravity/cosmology interplay, the nuisance parameters related to intrinsic alignments and photometric uncertainty are first set to zero. This is relaxed in Section 4.5, at which point they are also varied in the likelihood sampling.

We carry out our cosmology inferences with the likelihood sampler MULTINEST (Feroz, Hobson & Bridges 2009), which is run within COSMOSIS. This sampling method has been used and validated in a number of previous works, notably in the cosmic shear analysis of the KiDS-1000 data (Asgari et al. 2021) and of the DES-Year

Downloaded from https://academic.oup.com/mnras/article/525/4/6336/7263281 by guest on 30 January 2024

3 data (Secco et al. 2022). It has been reported in Lemos et al. (2023) that the projected contours could be slightly overconstraining in some cases compared to alternative samplers, however we opted for MULTINEST as it is much faster and its accuracy is sufficient to support the scientific goals of this paper. The chains all ran in 5000 steps and are analysed with GETDIST.[24]

4.1 Likelihood-based forecasts on f_{R_0} and H_0r_c

Forecasts on weak lensing f_{R_0} and H_0r_c constraints found in the literature need to be revisited, mostly due to recent improvements in modelling the deep non-linear matter power spectrum in presence of a screened fifth force. For example, Pratten et al. (2016) forecast that with a full-sky 3D weak lensing analysis based on spectroscopic data, and assuming that the cosmological background is fixed by CMB data, one could constrain $f_{R_0} < 5 \times 10^{-6}$. Their χ^2 analysis is simpler than our full MCMC approach, they used a hybrid one-loop perturbation theory and halo-model to compute the $P(k)$ in presence of MG, and unlike us they do not include WL systematics. Other examples include the *Euclid* forecast of Thomas, Abdalla & Weller (2009) that predicts from a Fisher analysis that the nDGP signal will be clearly detectable from lensing alone.[25] Martinelli et al. (2011) and Casas et al. (2017) also predicts clear detection of MG signal from *Euclid*, this time using MGCAMB (Hojjati et al. 2011) for the $P(k)$ modelling, including ℓ-modes up to 5000, and assuming the commonly used (μ, Σ) phenomenological parametrization. None of these adequately investigate the sensitivity of modern cosmic shear surveys. Perhaps the most realistic forecast to date is that from Bose et al. (2020), which investigate the constraining power of an LSST-like survey on $f(R)$ and nDGP gravity, but it ignored tomography and secondary signals caused by intrinsic alignments of galaxy. The rest of this paper is therefore a step forward in realism, as we present a series of forecasts based on tomographic cosmic shear, progressively including most of the ingredients that are relevant for lensing. Before bringing on the full machinery, we first start with simplified scenarios in order to gain a better physical and statistical understanding of the measurements at hand.

Fig. 7 (top panel) presents the posterior distributions from three likelihood samplings, in which the data are taken directly from the FORGE emulator predictions, at cosmology-00 and for $\log_{10}[f_{R_0}] = -6.5, -6.0$, and -5.5. We observe a strong degeneracy between f_{R_0} and S_8, expected from the fact that these two parameters both modulate the overall amplitude of the lensing signal. This degrades the constraining performance with respect to our previous Fisher calculation (Section 3.3). If S_8 was fixed, we could indeed detect with high significance these three models (imagine slicing through the $S_8 - f_{R_0}$ contours along the vertical dashed line at the input S_8 value), however the two weakest models are hitting the GR-limit when S_8 becomes large. The $f_{R_0} = 10^{-5.5}$ model, on the other hand, would be detected at the $\sim 3\sigma$ level. This is an order of magnitude less constraining than what was found by our 1D Fisher forecast, but is more realistic as we are now fully including gravity-cosmology degeneracies.

The lower panel of Fig. 7 shows a similar exercise carried out on nDGP data taken directly from the BRIDGE emulator. We observe that in all cases the three parameters are correctly inferred, and that the $[S_8 - H_0r_c]$ degeneracy direction is inverted compared to f_{R_0} due to the fact that in this model strongest deviations occur for smaller H_0r_c values. Finally, whereas the posterior from weakest nDGP model in this figure (grey contours, corresponding to $\log_{10}[H_0r_c] = 0.2$) is prior-dominated towards the higher H_0r_c bound, the other two models are not: $H_0r_c < 1.0$ could be detected beyond 3σ in this forecast. Once again this error is less constraining than our Fisher forecast, as expected from the added realism. Fixing cosmology would significantly help in this measurement as well, as the posteriors are narrow along a fixed S_8 value.

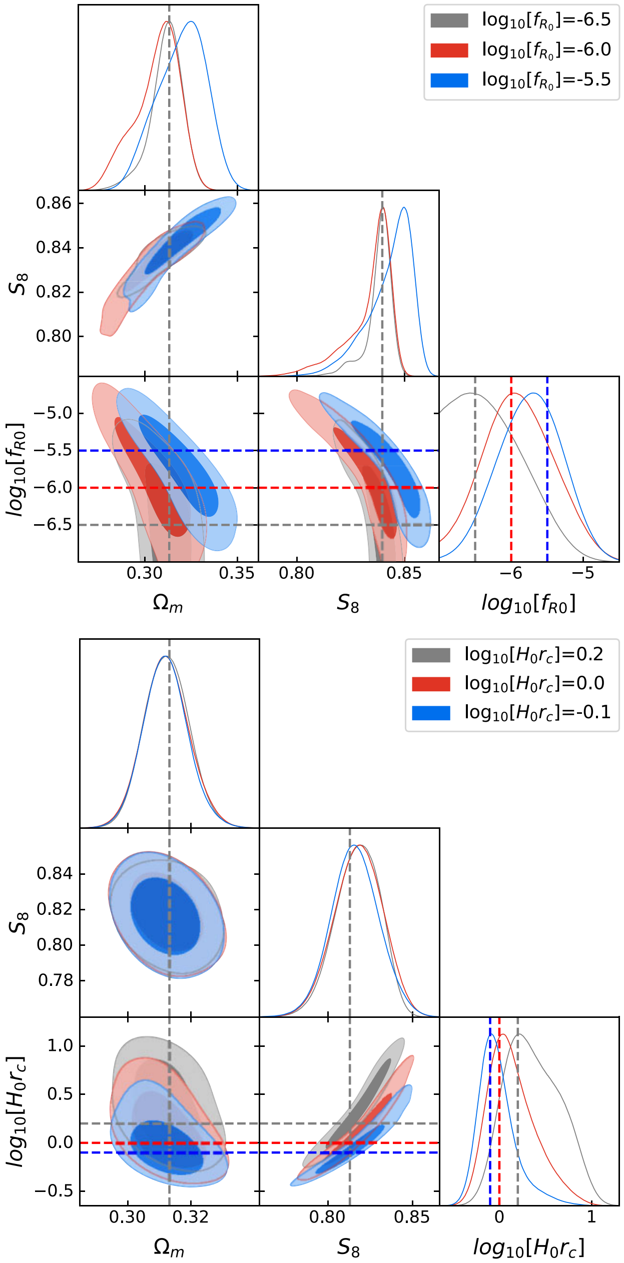

Figure 7. Marginalized constraints on the FORGE (upper panel) and BRIDGE (lower) parameters when analysing data taken directly from our $f(R)$ and nDGP $P(k)$ emulators, for different input values of $\log_{10}[f_{R_0}]$ and $\log_{10}[H_0r_c]$ indicated in the legend. Values of Ω_m, S_8, and h are otherwise matching the GR-model.

[24] GETDIST:getdist.readthedocs.io/en/latest/.

[25] In their work, Thomas et al. (2009) use a different DGP parametrization, replacing H_0r_c by a derived α parameter.

Downloaded from https://academic.oup.com/mnras/article/525/4/6336/7263281 by guest on 30 January 2024

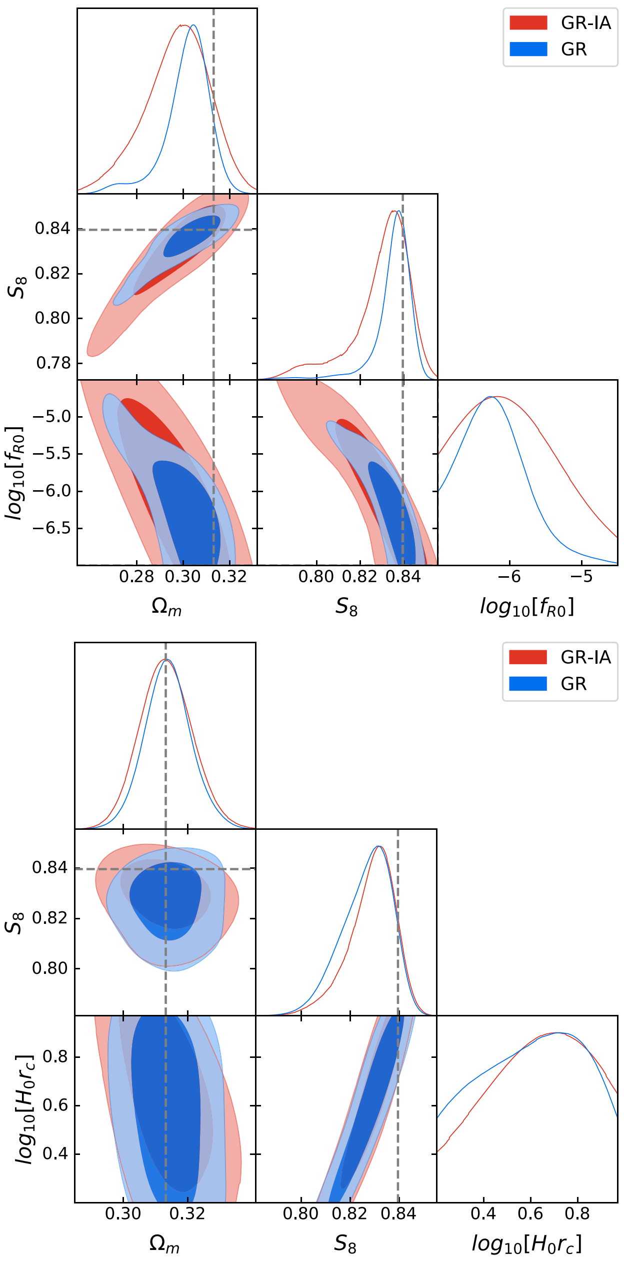

Figure 8. Marginalized constraints on the FORGE (upper) and BRIDGE (lower) parameters when analysing lensing maps from the GR simulations. Given our prior limits and the important degeneracy between S_8 and the MG parameters, we recover the expectation that the input truth is well inside the 1σ contours, but not necessarily at the centre.

4.2 Recovering the GR-ΛCDM simulation

Fig. 8 presents our first inference validation test on the MGLENS simulations, where we run our analysis pipeline on the GR-only simulations, assuming consecutively a FORGE and BRIDGE gravity model (top and bottom panels, respectively). It is important to note here that our noise-free data have been measured from 5000 deg^2, and our analytical covariance matrix assumes the same area and includes shape noise. We therefore expect the input truth to lie close to the centre of the 1σ regions, but offset can be caused by residual sampling variance in the mocks and interpolation errors from the emulators. This is indeed consistent with what we observe in Fig. 8, establishing that we correctly infer the input cosmological parameters, and prefer modified gravity models that are beyond detection, with:

$\log_{10}[f_{R0}] < -5.42,$

and

$\log_{10}[H_0r_c] > 0.140,$

in absence of systematics (both upper limits are reported with 95 per cent CL). Note that these one-sided limits depend on the prior range we adopt: larger sampled volumes (on the weak MG side) down-weight the tails and hence artificially increase the constraining power. For example, truncating the MCMC chains at $\log_{10}[f_{R0}] = [-7.0, -7.5, -8.0]$ yield upper limits of $[-5.36, -5.42,$ and $-5.48]$, respectively. We selected the middle value in this work, but care must be taken when comparing these results with others found in the literature. Similarly, we truncate the nDGP chains at $\log_{10}[H_0r_c] = 0.8$ to avoid false two-sided constraints coming from hitting the prior edge. Note that the results obtained here seem at first to contradict Fig. 5, in which models with $f_{R_0} > 10^{-6.0}$ are more that 3σ away from GR at high-redshift (see the right-most panel), but this observation ignores the $[f_{R_0} - S_8]$ degeneracy, which hinder possible MG detections.

An important feature of this figure is that the degeneracy between f_{R_0} and S_8 vanishes when sampling lower f_{R_0} values, as seen in the lower part of the contours which are close to vertical; this is also seen in Fig. 7. That is likely due to the fact that a small f_{R_0} tends to have little modification to the clustering in the linear regime on large scales, where the amplitude of clustering is influenced by S_8 more directly; instead, it tends to cause stronger deviations to its GR counterpart only at the very small scales, where there is also a stronger non-linearity, thus a weaker connection to the amplitude parameter S_8. Put together, these two factors, the relatively stronger effect of f_{R_0} on small scales and stronger non-linearity, naturally break the degeneracy between f_{R_0} and S_8 when f_{R_0} is small. This is not the case for other FORGE models with a stronger MG sector, as we will see in the following section.

For nDGP, shown on the bottom panel of Fig. 8, the degeneracy with S_8 is present at every value of H_0r_c, even for weak deviations from GR, but the input cosmology is well recovered, even though this model is at the edge of the latin hypercube.

4.3 Recovering the FORGE and BRIDGE simulations

We now turn our attention to other MGLENS nodes, with Fig. 9 showing the inferred parameters when analysing a series of FORGE and BRIDGE data vectors (left and right panels, respectively), specifying the correct gravity framework (*f*(*R*) or DGP) at the moment; we investigate later the result of specifying the wrong framework. We present, from top to bottom, models with increasing deviations from GR. Once again the input cosmologies are recovered within 1σ, which validates both the MGLENS simulations and the COSMOSIS implementation of the FORGE and BRIDGE emulators in our end-to-end cosmological inference. One of the most important features seen here is the strong degeneracy between the MG parameters (f_{R_0}, H_0r_c) and S_8. Looking now at the posteriors, according to these results, if the gravitational physics of our Universe matched the medium or strong models in these survey conditions, we could strongly rule out GR and constrain the MG sector with our survey. The marginalized posteriors on the parameters of interests are summarized in Table 3, where, for example, our measurement for the weak FORGE yields $\log_{10}[f_{R0}] = -6.62^{+0.79}_{-0.79}$, which is fully consistent

Downloaded from https://academic.oup.com/mnras/article/525/4/6336/7263281 by guest on 30 January 2024

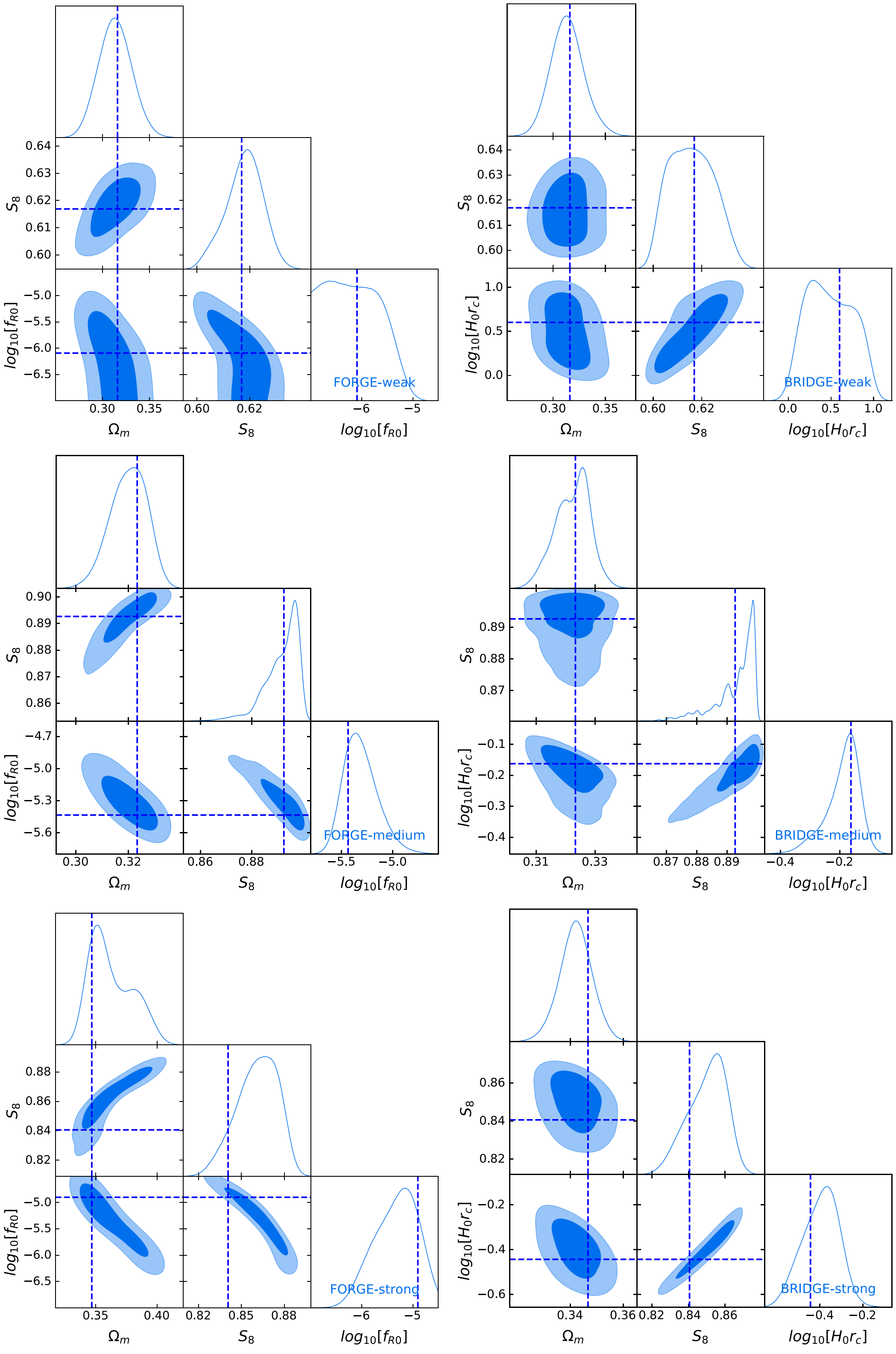

Figure 9. Marginalized constraints on the FORGE (left) and BRIDGE (right) parameters, for models-04 (upper panels, weak MG), −18 (middle panels, medium MG), and −13 (lower panels, strong MG), when analysing lensing maps from the MGLENS simulations. No systematics are included here.

with the input truth (−6.09). Similar results can be seen for the nDGP inference analyses, where the large values of H_0r_c are heavily disfavoured, while successfully recovering the input simulation values.

The observed $[f_{R_0}, S_8]$ degeneracy limits the precision we can achieve on these two parameters separately, which incites us to define a combination that is better measured. Inspired by the $\Sigma_8 \equiv \sigma_8(\Omega_{\rm m}/0.3)^\alpha$ composite lensing parameter, we introduce a new

Downloaded from https://academic.oup.com/mnras/article/525/4/6336/7263281 by guest on 30 January 2024

Table 3. Measurements of the modified gravity parameters inferred from the tomographic weak lensing power spectrum analysis of the FORGE and BRIDGE simulations. We show the results for a selection of models (top to bottom show GR, *f*(*R*), and nDGP gravity). The last column shows the impact of marginalizing over the $A_{\rm IA}$ nuisance parameter. In our FORGE and BRIDGE emulators, the GR node is taken at $\log_{10}[f_{R0}] = -7.0$ and $\log_{10}[H_0 r_c] = 1.0$, respectively. Upper and lower limits are reported at 95 per cent CL.

Model	Parameter	Truth	No-syst	IA
GR	$\Omega_{\rm m}$	0.313	$0.302^{+0.010}_{-0.006}$	$0.298^{+0.015}_{-0.012}$
	$S_8^{\rm GR}$	0.840	$0.835^{+0.009}_{-0.004}$	$0.830^{+0.015}_{-0.006}$
	$\log_{10}[f_{R0}]$	$-\infty$	<-5.42	<-4.77
	$\log_{10}[H_0 r_c]$	∞	>0.140	>0.090
FORGE-weak	$\Omega_{\rm m}$	0.316	$0.313^{+0.017}_{-0.017}$	$0.313^{+0.017}_{-0.019}$
	$S_8^{\rm GR}$	0.617	$0.618^{+0.008}_{-0.006}$	$0.618^{+0.008}_{-0.007}$
	$\log_{10}[f_{R0}]$	-6.09	$-6.62^{+0.79}_{-0.79}$	$-6.63^{+0.78}_{-0.78}$
	ζ_{R_0}	-25.3	$-27.2^{+3.9}_{-2.4}$	$-27.3^{+3.7}_{-2.4}$
FORGE-medium	$\Omega_{\rm m}$	0.323	$0.320^{+0.008}_{-0.006}$	$0.319^{+0.012}_{-0.0098}$
	$S_8^{\rm GR}$	0.893	$0.892^{+0.008}_{-0.003}$	$0.886^{+0.014}_{-0.0053}$
	$\log_{10}[f_{R0}]$	-5.43	$-5.32^{+0.13}_{-0.19}$	$-5.19^{+0.22}_{-0.26}$
	ζ_{R_0}	-3.55	$-3.49^{+0.06}_{-0.05}$	$-3.51^{+0.10}_{-0.075}$
FORGE-strong	$\Omega_{\rm m}$	0.347	$0.362^{+0.018}_{-0.018}$	$0.365^{+0.016}_{-0.016}$
	$S_8^{\rm GR}$	0.841	$0.861^{+0.017}_{-0.012}$	$0.864^{+0.016}_{-0.014}$
	$\log_{10}[f_{R0}]$	-4.90	$-5.32^{+0.50}_{-0.39}$	$-5.34^{+0.49}_{-0.44}$
	ζ_{R_0}	-4.33	$-4.15^{+0.12}_{-0.09}$	$-4.10^{+0.12}_{-0.12}$
BRIDGE-weak	$\Omega_{\rm m}$	0.316	$0.313^{+0.015}_{-0.017}$	0.314 ± 0.018
	$S_8^{\rm GR}$	0.617	$0.616^{+0.009}_{-0.011}$	0.6162 ± 0.0084
	$\log_{10}[H_0 r_c]$	0.602	$0.49^{+0.26}_{-0.33}$	0.51 ± 0.26
	$\zeta_{r_c} \times 10^3$	0.36	$0.36^{+0.12}_{-0.35}$	$0.36^{+0.16}_{-0.34}$
BRIDGE-medium	$\Omega_{\rm m}$	0.323	$0.322^{+0.006}_{-0.005}$	$0.3245^{+0.0080}_{-0.0063}$
	$S_8^{\rm GR}$	0.893	$0.893^{+0.007}_{-0.004}$	$0.8886^{+0.0082}_{-0.0057}$
	$\log_{10}[H_0 r_c]$	-0.163	$-0.189^{+0.065}_{-0.035}$	$-0.209^{+0.071}_{-0.064}$
	ζ_{r_c}	-1.478	$-1.68^{+0.22}_{-0.26}$	$-1.67^{+0.36}_{-0.31}$
BRIDGE-strong	$\Omega_{\rm m}$	0.347	$0.342^{+0.006}_{-0.006}$	$0.340^{+0.0098}_{-0.012}$
	$S_8^{\rm GR}$	0.841	$0.850^{+0.013}_{-0.008}$	$0.855^{+0.011}_{-0.0095}$
	$\log_{10}[H_0 r_c]$	-0.443	$-0.395^{+0.095}_{-0.076}$	$-0.355^{+0.091}_{-0.079}$
	ζ_{r_c}	-0.845	$-1.00^{+0.11}_{-0.15}$	-1.047 ± 0.089

variable which runs across the minor axis of the degeneracy ellipse:

$$\zeta^\alpha_{R_0} \equiv \log_{10}[f_{R_0}] \left(\frac{S_8^{\rm GR}}{0.82}\right)^\alpha, \tag{23}$$

where α is a free parameter to be optimized. For small values of f_{R_0}, $\alpha = 5.0$ returns a $\zeta^\alpha_{R_0}$ that is mostly orthogonal to both S_8 and $\Omega_{\rm m}$, making this an attractive target measurement for future cosmic shear experiments. We post-pone to future work the impact of letting α free in a likelihood analysis.

The equivalent degeneracy-breaking parameter for nDGP models can be constructed as

$$\zeta^\alpha_{r_c} \equiv \log_{10}[H_0 r_c] \left(\frac{S_8^{\rm GR}}{0.82}\right)^\alpha, \tag{24}$$

where $\alpha = 26$ works better for the nDGP models. We show in Fig. 10 the marginalized constraints on these two new parameters, $\zeta^\alpha_{R_0}$ and $\zeta^\alpha_{r_c}$, where the degeneracy with respect to S_8 and $\Omega_{\rm m}$ is highly suppressed. The accuracy on these composite parameters is increased, where for example a 8 per cent measurement[26] of $\log_{10}[f_{R_0}]$ results in a 3 per cent precision on $\zeta^\alpha_{R_0}$ in the strong FORGE model. Similar improvements are seen on nDGP parameters, where a 22 per cent measurement of $\log_{10}[H_0 r_c]$ becomes a 13 per cent measurement of $\zeta^\alpha_{r_c}$ in the strong BRIDGE model. The measurements reported in Table 3 indicate a net gain in precision for all models.

By construction, the variables $\zeta^\alpha_{R_0}$ and $\zeta^\alpha_{r_c}$ down-weight parameter regions of weak modified gravity, which therefore interacts with prior limits. These parameters are therefore mostly useful for medium and strong modified gravity models, but we advise against using them for one-sided limits.

4.4 Degeneracies between gravity models and cosmology

One of the main difficulties in detecting deviations from GR comes from the abundance of models to be tested, which each affect the growth of structures in different ways. A key question to be answered is whether one can confuse a clear detection of gravity model 'A' at some cosmology with a different gravity model 'B' at a different cosmology. The first part of the answer is already provided in the GR-only validation test, where both the FORGE and BRIDGE emulators recognize negligible deviations from GR in the GR-only model, both inferring the right cosmology. This is encouraging since it suggests that GR can be recognized as such.

Complications arise when analysing truly non-GR data with the wrong gravity model. The lower panels of Fig. 11 shows such examples, where three FORGE data vectors are analysed with the BRIDGE emulator. For the weak model (left), this results in a minor bias in $\Omega_{\rm m}$ and S_8, and a wide posterior on $H_0 r_c$ that hits the upper edge prior, leading to inconclusive detection of MG. The central and right panels, however, reveal catastrophic biases on the cosmological parameters for the medium and strong models. The two cosmological parameters are shifted towards higher values, while the posteriors indicate an apparent $H_0 r_c$ detection. We report these shifts in Table 4, in units of statistical precision σ. Similarly catastrophic results are observed when, on the contrary, we analyse nDGP data with the FORGE emulator (see the upper panels of Fig. 11); in this case most inferred cosmological parameters are also far from the truth, and the f_{R_0} parameter is falsely detected with high significance for the medium nDGP model. Biases also occur if data from a modified gravity universe is analysed within GR, in which case the additional structure formation caused by the fifth force is interpreted as a higher S_8 value, as expected from the degeneracy between these quantities. We see again that the weak models has almost no impact on the inferred cosmology (shift $\sim 1\sigma$), whereas the stronger models can offset $\Omega_{\rm m}$ and S_8 by tens of σ. For example, with sub-per cent statistical precision on S_8, a bias of $\Delta S_8 = 0.05$ is almost a 8σ shift.

This inevitably raises the question of whether we could discover that we are analysing the data with the wrong gravity model. One of the approaches commonly used is to examine the *goodness-of-fit*, which informs us on the quality of the data-model match. This can be computed with the *p*-value measured at the best-fitting parameters for different gravity models, from which one can test different hypotheses.[27] A *p*-value below 0.01 generally indicates that

[26] The precision is defined here as the ratio between the error and the best-fitting value for a given parameter.

[27] The *p*-value is computed from the χ^2 conditional distribution function and the number of degrees of freedom; it is routinely used for rejection of null-hypotheses.

Downloaded from https://academic.oup.com/mnras/article/525/4/6336/7263281 by guest on 30 January 2024

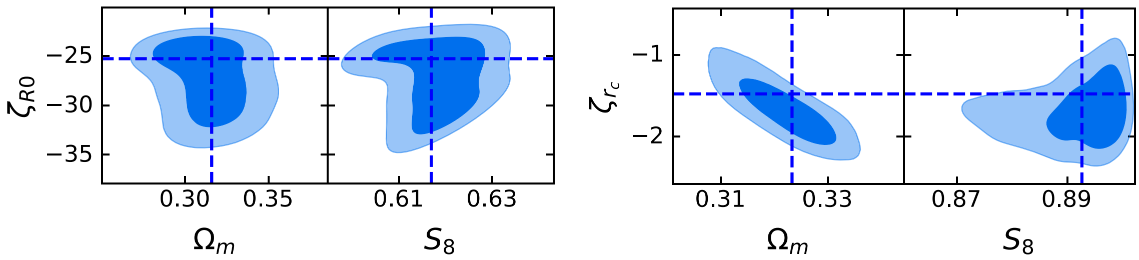

Figure 10. Marginalized constraints on the two composite parameters introduced in this paper, $\zeta^{\alpha}_{R_0}$ and $\zeta^{\alpha}_{r_c}$ (see equations 23 and 24), which are best measured by cosmic shear data when cosmology and modified gravity parameters are jointly varied. These are extracted from the MGLENS simulated data at the weak *f*(*R*) (left) and medium nDGP (right) gravity models.

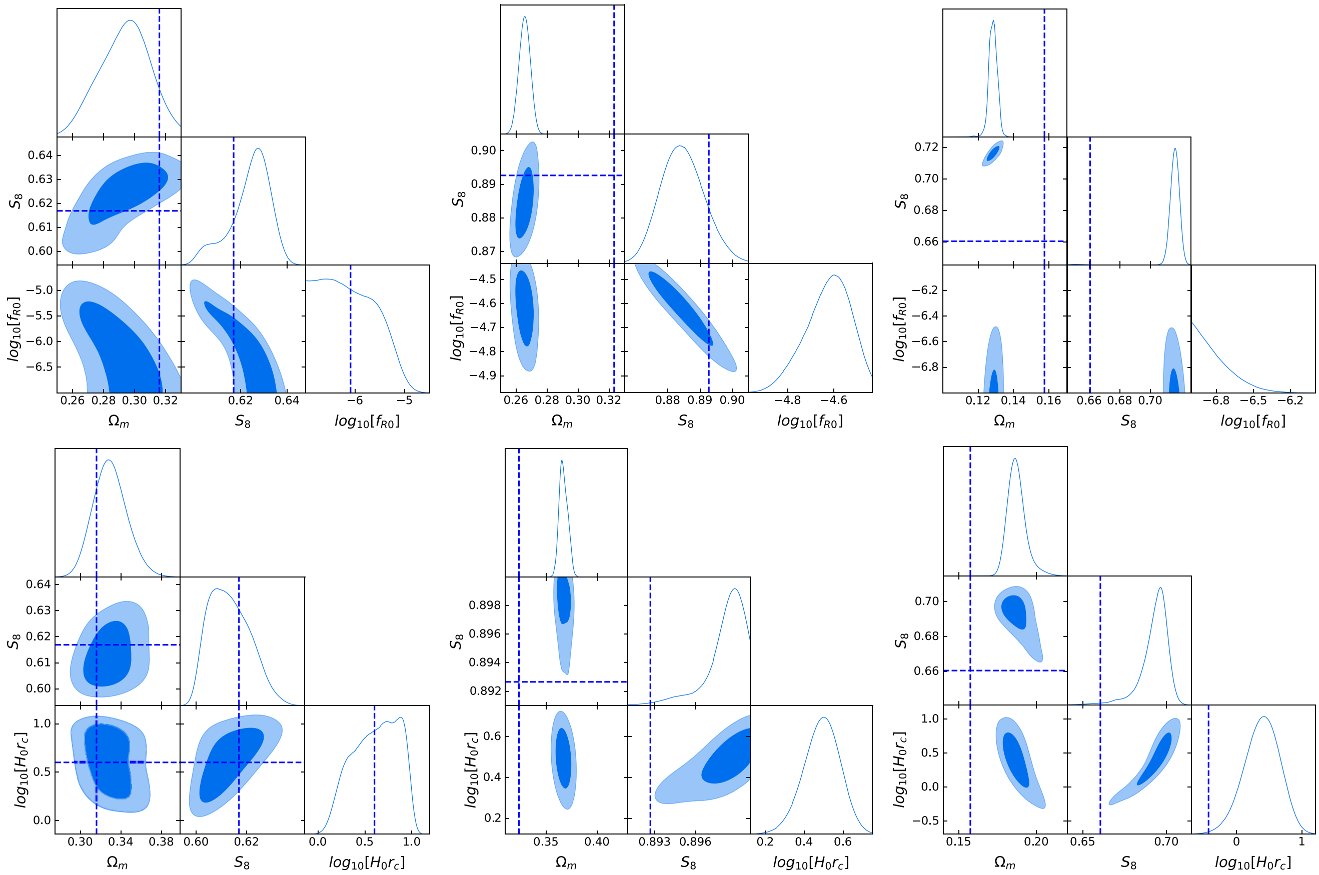

Figure 11. Catastrophic impact of mixing the gravity model. (Upper:) Marginalized parameter constraints when analysing BRIDGE simulations (left is the weak model, centre is the medium, right is the strong model-05) with the FORGE emulator, yielding to catastrophic biases. (Lower:) Counterpart of the upper panels, now analysing FORGE simulations with the BRIDGE emulator.

the hypothesis should be rejected. In our case this test is done with noise-free data, so the *p*-values can approach 1.0 in case of excellent fits. The ideal case here would be to obtain low *p*-values whenever the wrong gravity model is being used. Table 4 presents the measured *p*-values for different plausible analysis scenarios. It turns out that some simulated data (e.g. FORGE-weak and BRIDGE-weak) can be well fitted by all three gravity models (i.e. their *p*-values are high), due to weakness of the departure from GR. *f*(*R*) gravity can also provide a good fit to the BRIDGE-medium data, which is achieved at the cost of significantly lowering $\Omega_{\rm m}$. This bias is clearly seen in the up per central panel of Fig. 11. One would have problems, in such a case, to distinguish between gravity models from the sole goodness-of-fit results. Other test cases are easier to reject based on their bad goodness-of-fit, such as FORGE-strong and BRIDGE-strong, which can only be well fit with the correct gravity model. We also observe that analysing some models within GR pushes the likelihood outside of our already wide prior, which is in itself an indication that something is off with the modelling, even though the solution (to switch gravity model) might not be obvious at first.

Other metrics are better suited for model-selection, notably the *Bayesian Evidence ratio* (Hobson, Bridle & Lahav 2002; Marshall,

Downloaded from https://academic.oup.com/mnras/article/525/4/6336/7263281 by guest on 30 January 2024

Table 4. Impact on the cosmological parameters $\Omega_{\rm m}$ and S_8 when analysing MG simulated data with the wrong gravity model. Column ΛCDM+GR shows the results of analysing MGLENS simulations with a GR model (i.e. HALOFIT), while the 'Wrong MG' columns consider FORGE data analysed with the BRIDGE emulator and vice versa. The parameter shifts are computed as |bestfit – true|/σ, and the p-values assume four free parameters. Posteriors overlapping with prior edges are flagged as such. We also show the evidence ratio $\mathcal{R}$, defined in the main text, which is often used in model selection. It can be interpreted here as the odds of the true model describing the data compared to the alternative model (GR or wrong MG), and $\mathcal{R}\sim\mathcal{O}(1)$ means that both models are equally likely.

True		True model		ΛCDM + GR			Wrong MG		
Gravity model	Param	Shift	p-value	Shift	p-value	$\mathcal{R}$	Shift	p-value	$\mathcal{R}$
FORGE-weak	$\Omega_{\rm m}$	0.2σ	1.0	0.25σ	1.0	1.08	0.8σ	1.0	1.17
	S_8	0.1σ		<0.1σ			0.5σ		
FORGE-medium	$\Omega_{\rm m}$	0.4σ	1.0	prior limited	–	–	prior limited	–	–
	S_8	0.1σ							
FORGE-strong	$\Omega_{\rm m}$	0.8σ	1.0	9.2σ	0.68	3.42e5	29.6σ	0.0	2.18
	S_8	1.4σ		11.0σ			11.7σ		
BRIDGE-weak	$\Omega_{\rm m}$	0.2σ	1.0	1.1σ	1.0	2.44	1.3σ	1.0	2.46
	S_8	0.1σ		2.2σ			0.9σ		
BRIDGE-medium	$\Omega_{\rm m}$	0.1σ	1.0	prior limited	–	–	17.0σ	1.0	836
	S_8	0.0σ					1.2σ		
BRIDGE-strong	$\Omega_{\rm m}$	0.8σ	1.0	prior limited	–	–	12.9σ	0.0	2.18
	S_8	0.8σ					19.0σ		

Rajguru & Slosar 2006), which relies on computing the prior-marginalized likelihood, a quantity directly available from the output of our MULTINEST chains. Specifically, the ratio between the Bayesian evidences, $\mathcal{R}[1,2]\equiv\mathcal{Z}_1/\mathcal{Z}_2$, which are respectively computed by integrating over the full posterior volumes obtained from analysing the same data with models 1 and 2, provides the Bayesian probability that model 1 better describes the data over model 2. Both models are plausible when $\mathcal{R}$ is of the order of unity, while model 1 would be strongly favoured over model 2 for $\mathcal{R}\gg 1.0$.

We therefore compute the evidence ratios between $f(R)$, nDGP, and GR given the weak, medium, and strong FORGE and BRIDGE data. For example, the evidences obtained from analysing FORGE-strong simulation with the three gravity models are $\log[\mathcal{Z}_{f(R)}]=-16.6$, $\log[\mathcal{Z}_{\rm DGP}]=-17.4$, and $\log[\mathcal{Z}_{\rm GR}]=-29.4$, from which we obtain $\mathcal{R}[f(R),{\rm DGP}]=2.18$ and $\mathcal{R}[f(R),{\rm GR}]=3.42\times10^5$. In this case, GR can be safely ruled out, but none of the two MG theories can be rejected based on the evidence ratio, albeit a only minor preference for the $f(R)$ model. The p-value is more informative here, being close to 0.0 when using the wrong model. All results are reported in Table 4.

Interestingly, the two weak MG cases provide evidence ratios of the order of unity when analysed with all gravity models, and all have p-values of the order of unity as well. This means that given the current summary statistics, the data are not precise enough for us to recover with certainty the true gravity model. Possible solutions to overcome this are to augment the analysis with prior knowledge of the cosmological parameters from e.g. the CMB, or analyse the data with higher order statistics to further break degeneracies, which will be the subject of future work. In any case, having a variety of MG simulations is critical to properly understand how gravity models are degenerate with cosmology and propose meaningful mitigation strategies.

It is worth mentioning that the evidence metric is dependent on the prior volume, and for this reason the *Suspiciousness* statistics (Lemos et al. 2020) is often viewed as superior, being more robust to prior-effects, although computationally expensive (see Joachimi et al. 2021b, for a recent discussion on the application of such metrics to real cosmological data).

4.5 Impact of systematics

The results from the beginning of Section 4 are obtained in unrealistically clean conditions; as discussed previously, cosmic shear surveys are in fact affected by poorly constrained intrinsic alignments (IA), by uncertainty on the photometric redshift (photo-z) distributions and shape calibration, as well as by largely unconstrained baryonic feedback. Additionally the weak lensing signal is mildly sensitive to some of the other cosmological parameters such as the baryon density $\Omega_{\rm b}$, the sum of neutrino masses Σm_ν or the tilt in the primordial power spectrum, $n_{\rm s}$, such that our constraints are likely slightly overprecise. Here we focus on two of these, namely the photo-z and the IA, leaving a more comprehensive study of the others for future work. To some extent the impact of baryon can be reduced by removing some of the non-linear scales, which we also touch upon below.

Using COSMOSIS for the calculation of the theoretical cosmic shear predictions has key advantages when it comes to modelling and marginalizing over the known weak lensing systematics. First, the public version includes an implementation of the widely used non-linear alignment model (Bridle & King 2007), which describes the IA contamination from a linear coupling between the intrinsic galaxy orientations and the local tidal field. This results in a two-component secondary signal that can be computed from the matter power spectrum as (Hirata & Seljak 2004; Bridle & King 2007):

$$P_{II}(k,z)=\left(\frac{A_{\rm IA}\bar{C}_1\bar{\rho}(z)}{\overline{D}(z)}\right)^2 a^4(z)P_\delta(k,z) \qquad (25)$$

and

$$P_{GI}(k,z)=-\frac{A_{\rm IA}\bar{C}_1\bar{\rho}(z)}{\overline{D}(z)}a^2(z)P_\delta(k,z). \qquad (26)$$

In the above expressions, $P_\delta(k,z)$ is the matter power spectrum including the MG enhancement, $\bar{\rho}(z)$ is the background matter density, $\overline{D}(z)$ is the 'rescaled linear growth factor' defined as $\overline{D}\equiv D(1+z)$, and $\bar{C}_1$ is a constant calibrated in Brown et al. (2002), set to $5\times10^{-14}{\rm M}_\odot^{-1}h^{-2}\,{\rm Mpc}^3$. These are then inserted in the Limber integral (equation 17), where now the lensing kernels $q^i(\chi)q^j(\chi)$ are replaced by $q^i(\chi)n^j(\chi)$ and $n^i(\chi)n^j(\chi)$ for the *GI* and *II* terms,

Downloaded from https://academic.oup.com/mnras/article/525/4/6336/7263281 by guest on 30 January 2024

respectively. A further redshift dependence can be implemented with a multiplicative term of the form

$$\left(\frac{1+z}{1+z_{\rm pivot}}\right)^{\eta_{\rm IA}}$$

with $z_{\rm pivot}$ and $\eta_{\rm IA}$ two additional free parameters. This model has been shown to accurately capture the IA signal in many cosmic shear analyses (see e.g. Troxel et al. 2018; Asgari et al. 2021), with weak signs of potential limitations in the most recent DES-Y3 analysis by Secco et al. (2022). In all cases, IA significantly biases the inferred cosmology if left unmodelled, however the redshift evolution is only weakly constrained in these surveys. In fact, assuming no IA redshift evolution affects the inferred cosmology by less than 0.3σ, which is significantly less than the shift caused by switching to the more physical model that includes tidal torquing (Blazek et al. 2019). Choosing the right IA model given the data is still an open issue (Campos, Samuroff & Mandelbaum 2023), and in light of this uncertainty we opted to ignore the poorly constrained redshift evolution of the IA signal in our forecasts. We therefore model IA with a single scaling parameter, $A_{\rm IA}$, which we vary over the range [−5.0, 5.0] in line with these previous analyses, and set $\eta_{\rm IA} = 0$.

A second advantage of using COSMOSIS is that it deals with the uncertainty on the redshift distribution by shifting the tomographic $n^i(z)$ by a constant quantity δ_z^i, which we treat independently for each tomographic bin i: $n^i(z) = n^i(z + \delta_z^i)$. It has been shown that in some cases these shift parameters are correlated (Wright et al. 2020), however we ignore this here. Our five δ_z^i parameters are sampled assuming a Gaussian prior of width 0.01, similar to the accuracy achieved by current weak lensing surveys (for example, an accuracy between 0.0084 and 0.0116 on these δ_z parameters is achieved with the KiDS-1000 data, see Hildebrandt et al. 2021). We do not include the uncertainty on shape calibration (i.e. the m-bias, see Giblin et al. 2021) as it is currently subdominant compared to the effect of IA and photometric redshift (Asgari et al. 2021; Secco et al. 2022). Importantly, we neglect the impact of baryon feedback, which is arguably the largest approximation in our analysis. Indeed, baryons significantly redistribute the matter distribution and suppress the lensing signal by tens of per cent depending on the scales and baryonic physics (Semboloni et al. 2011; Harnois-Déraps et al. 2015a). We could extend our results by using for instance the matter power spectrum provided by HMCode (Mead et al. 2021) in which the impact of baryons is modelled, but we leave this for future work. We finally assume a constant total neutrino mass set to $\Sigma m_\nu = 0.0$ eV, in order to be consistent with the FORGE and BRIDGE simulations. All of these analysis choices have an impact on the inference and will need to be revisited in order to make robust constraints on the MG parameters from cosmic shear data, however our simplified likelihood evaluations represent an important first step in this direction.

We show in Fig. 12 (and summarize the results in Table 3) the impact of IA on the marginalized constraints for some of the FORGE and BRIDGE models. As expected, the presence of IA degrades the constraints on most parameters, where for example the 1.4 per cent measurement of S_8 value in the FORGE medium model becomes a 1.9 per cent measurement. The same model sees the constraints on $\log_{10}[f_{R_0}]$ degrade from a 5.4 per cent to a 6.7 per cent measurement. We also note that for some models (e.g. FORGE medium, BRIDGE strong), the IA contamination acts mostly along the $[f_{R_0} - S_8]$ or $[H_0 r_c - S_8]$ degeneracy directions, whereas for other models the posterior is inflated in all dimensions (e.g. FORGE-strong). Finally low-S_8 models appear to be less affected (e.g. FORGE weak), which is expected since the IA signal also scales with S_8, causing them to be harder to distinguish from the cosmological signal given our fixed covariance matrix. Also worth repeating here is that our data vector includes the cross-tomographic terms, which are more affected by IA as they are highly sensitive to the ‘GI’ alignment term, i.e. the coupling between the background shearing and the intrinsic alignment of foreground galaxies (Hirata & Seljak 2004). These increase the contamination, but at the same time further help in constraining the IA sector and therefore self-calibrate. Indeed, $A_{\rm IA}$ is one of parameters that is best measured by cosmic shear data (Asgari et al. 2021; Secco et al. 2022; Heydenreich et al. 2022), even though it is an ‘effective’ model that depends on a number of physical selection effects such as galaxy types, colours, and bias (Blazek et al. 2019). Interestingly, there is a mild degeneracy between the $A_{\rm IA}$ and the MG parameters, such that using the wrong gravity model can lead to an apparent IA signal. The effect is generally small, but can lead to a false detection larger than 1σ, as it is the case for the GR analysis of the strong BRIDGE model.

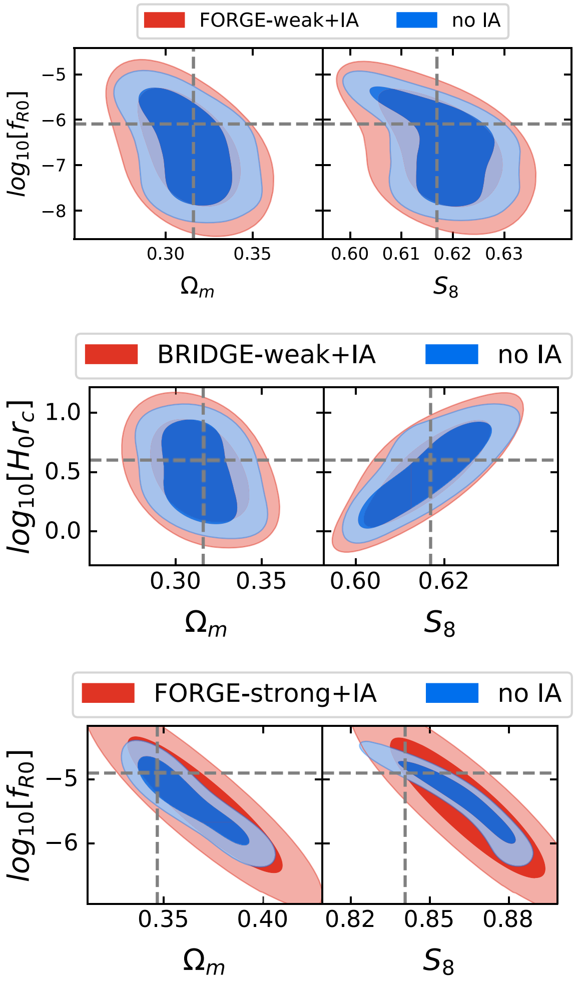

Figure 12. Marginalized constraints on the parameters best probed by lensing, with and without including contamination from intrinsic alignment in the modelling, inferred from the MGLENS simulated lensing data.

The redshift error are in comparison very small due to the narrow informative Gaussian prior that we are able to use. We have tested a few chains with the photo-z nuisance turned on and found almost no visible effect on the marginalized contours. Since this is the case for all models analysed we conclude that under these circumstances

Downloaded from https://academic.oup.com/mnras/article/525/4/6336/7263281 by guest on 30 January 2024

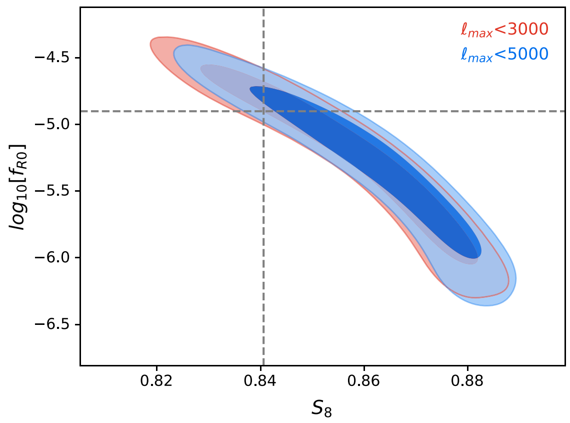

Figure 13. Impact of scale cuts on the marginalized constraints obtained from the analysis of MGLENS simulated lensing data in the strong FORGE model.

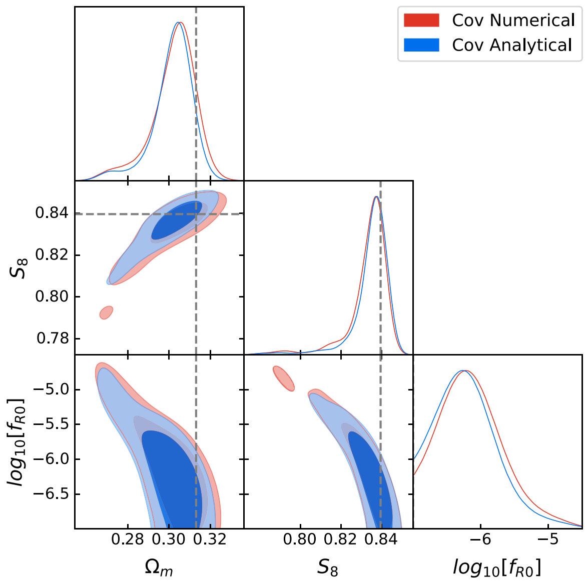

Figure 14. Comparison between the cosmological inference resulting from using the analytical or the numerical covariance matrix when analysing the GR simulated data.

photo-z errors are completely subdominant to IA and we do not investigate this any further.

Regarding baryons, a common approach to protect analyses against their uncertain impact consists in excluding the deeply non-linear scales from the data vector (as in, e.g. Troxel et al. 2018; Amon et al. 2021), which in our case are the high-ℓ modes. Lowering the highest ℓ from 5000 to 3000 typically results in a degraded constraint on the modified gravity parameters, largely due to an increased degeneracy with S_8, but this degradation is not catastrophic, as shown in Fig. 13. This is consistent with our Fisher calculations, according to which the information partly saturates by $\ell = 3000$. Therefore, while we expect the impact of varying $\ell_{\rm max}$ to lower the precision, the amount by which it does is not easily predictable due to the highly non-trivial degeneracies that exists in the high-dimensional likelihood space.

Finally, as mentioned earlier, an ingredient central to cosmological inference is the covariance matrix, which in the case of two-point statistics can be either modelled analytically or estimated numerically. This choice is not guaranteed to exists for all probes, and in fact many other weak lensing statistics must rely on an ensemble of mock data such as the SLICS to estimate the matrix. The validation process of these multipurpose mocks generally includes a comparison with the analytical predictions for covariance matrix about two-point statistics. A first step of this comparison is shown already in Fig. 2, which visually demonstrate that the cross-correlation coefficient matrices are consistent with one another. A full quantitative validation must go beyond this, and we show in Fig. 14 the cosmological inference resulting from using the two matrices. We observe that both posteriors fully overlap, providing identical best-fitting values on $\Omega_{\rm m}$, and differences on S_8 that vary by less than 0.2σ. The upper limits of $\log_{10}[f_{R_0}]$ shift by under 4 per cent, from -5.42 to -5.18. Note that the differences observed here are not exclusively caused by inaccuracies in the mocks, as many other factors can source important deviations, such as choices in the implementation of shape noise or masking (Joachimi et al. 2021a). In particular, the total survey areas match in both cases, however the analytical calculations assume a spherical survey whereas the mocks are square-shaped. Thus the small observed shifts in the cosmological inferences should be viewed as systematic uncertainties, not as biases, which thereby establishes the precision on the covariance one can expect from these SLICS mocks for any alternative weak lensing probes.

Also note that in an actual data analysis, the accuracy of the $B_\delta(k, z)$ emulator itself should be propagated into the covariance matrix in order to capture the modelling uncertainty.

5 DISCUSSIONS AND CONCLUSIONS

This paper introduces the MGLENS simulations, a large set of lensing maps sampling five cosmological and MG parameters within a volume that is wide and dense enough to analyse current Stage-III cosmic shear surveys. We demonstrate that the lensing power spectra measured from these simulations match well with the theoretical predictions obtained by the BRIDGE and FORGE emulators, validating at the same time both the simulation suite and our gravitocosmological inference pipeline implemented within COSMOSIS.

We next carry out a series of investigations using MGLENS and our analysis pipeline. Notably, we find that next-generation lensing surveys will be powerful at constraining the gravity sector: in our simplified systematics-free analysis, we forecast that 5000 deg^2 of upcoming data could lead to 3σ detection of a value of f_{R_0} as weak as 5.5×10^{-5}, and $H_0 r_c$ as low as 1.0. We acknowledge a number of caveats, including the absence of marginalization over baryon feedback, or fixing the values of other cosmological parameters that have a secondary impact on the cosmic shear signal. These will inevitably translate into a slightly larger uncertainty budget in an more complete data analysis, however the statistical power displayed in our survey should remain relatively unchanged. Moreover, these forecasts are for cosmic shear data alone; adding clustering, galaxy–galaxy lensing and/or CMB data could improve the constraints further. An additional gain of precision could be achieved by analysing the data with non-Gaussian statistics.

When inferring cosmology from different input model vectors, we identify in many cases a strong degeneracy between the input S_8 value (related to the primordial power spectrum amplitude $A_{\rm s}$) and the modified gravity parameters; we propose new composite parameters that are better measured by lensing, namely $\zeta^{\alpha}_{R_0}$ and $\zeta^{\alpha}_{r_c}$, on which the precision is increased by up to a factor of two.

Downloaded from https://academic.oup.com/mnras/article/525/4/6336/7263281 by guest on 30 January 2024

We lastly explored the impact of analysing data with the wrong gravity model, typically finding a catastrophic impact on the inferred cosmology with biases exceeding at times 20σ in some cases, as well as an unphysical detection of MG features. The goodness-of-fit is generally best when using the correct gravity models, but not always: some data are well fitted by more than one model and the Bayesian evidence ratio is unable to tell them apart. This means that other analysis methods will need to be developed in order to better differentiate the gravity sector, such as the *Suspiciousness* metric, the recent empirical approach of Campos et al. (2023), or by looking at probes different from the lensing power spectrum.

The MGLENS simulations are organized as a series of flat-sky and curved-sky convergence maps, which can be analysed with any weak lensing statistics. Combined with the large SLICS ensemble produced for the evaluation of covariance matrix, the MGLENS suite are ideally suited to explore the sensitivity of novel statistics to cosmological and gravitational parameters. To validate this approach, we test our inference framework with either an analytical or simulation-based covariance matrix, finding an excellent recovery of the input data vector in both cases. Cosmic shear analyses beyond two-point statistics will be presented in companion papers. Our goal is to provide the community with some of the best tools with which to search for MG in current and upcoming lensing surveys.

ACKNOWLEDGEMENTS

We thank Ben Giblin for useful discussions, and the anonymous referees for their constructive comments on earlier versions of this manuscript. JHD acknowledges support from an STFC Ernest Rutherford Fellowship (project reference ST/S004858/1). CHA acknowledges support from the Excellence Cluster ORIGINS which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2094–390783311. CA and BL are supported by the European Research Council (ERC) through a starting Grant (ERC-StG-716532 PUNCA). BL is also supported by the UK Science and Technology Funding Council (STFC) Consolidated Grant No. ST/I00162X/1 and ST/P000541/1. CTD is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2094–390783311. YC acknowledges the support of the Royal Society through a University Research Fellowship and an Enhancement Award. This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The equipment was funded by BEIS capital funding via STFC capital grants ST/K00042X/1, ST/P002293/1, ST/R002371/1, and ST/S002502/1, Durham University and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure. The SLICS simulations were enabled by the Digital Research Alliance of Canada (alliancecan.ca).

All authors contributed to the development and writing of this paper. JHD created the MGLENS suites, the COSMOSIS interface module and led the inference analysis; CA and CHA respectively led the *N*-body computations for the FORGE and BRIDGE simulations; CC developed the ML emulators, BL assisted in the design and in the data analysis while CTD and YC helped in the interpretation of the data.

DATA AVAILABILITY

The MGLENS numerical simulations, the $B_\delta(k, z)$ CNN emulators, and the COSMOSIS interface module will be made available upon acceptance to the Journal, while the SLICS simulations are already public.

REFERENCES

Abbott T. M. C. et al., 2019, Phys. Rev. D, 99, 123505
Alonso D., Sanchez J., Slosar A., LSST Dark Energy Science Collaboration, 2019, MNRAS, 484, 4127
Amon A. et al., 2022, Phys. Rev. D., 105, 023514
Angulo R. E., Zennaro M., Contreras S., Aricò G., Pellejero-Ibañez M., Stücker J., 2021, MNRAS, 507, 5869
Armijo J., Cai Y.-C., Padilla N., Li B., Peacock J. A., 2018, MNRAS, 478, 3627
Arnold C., Leo M., Li B., 2019, Nat. Astron., 3, 945
Arnold C., Li B., Giblin B., Harnois-Déraps J., Cai Y.-C., 2022, MNRAS, 515, 4161 (A21)
Asgari M. et al., 2021, A&A, 645, A104
Babichev E., Deffayet C., 2013, Class. Quantum Gravity, 30, 184001
Babichev E., Deffayet C., Ziour R., 2009, Int. J. Mod. Phys. D, 18, 2147
Barreira A., Bose S., Li B., 2015, J. Cosmol. Astropart. Phys., 2015, 059
Barreira A., Bose S., Li B., Llinares C., 2017, J. Cosmol. Astropart. Phys., 2017, 031
Barrera-Hinojosa C., Li B., Bruni M., He J.-h., 2021, MNRAS, 501, 5697
Blazek J. A., MacCrann N., Troxel M. A., Fang X., 2019, Phys. Rev. D, 100, 103506
Bose S., Hellwing W. A., Li B., 2015, J. Cosmol. Astropart. Phys., 2015, 034
Bose S., Li B., Barreira A., He J.-h., Hellwing W. A., Koyama K., Llinares C., Zhao G.-B., 2017, J. Cosmol. Astropart. Phys., 2017, 050
Bose B., Cataneo M., Tröster T., Xia Q., Heymans C., Lombriser L., 2020, MNRAS, 498, 4650
Brax P., van de Bruck C., Davis A. C., Khoury J., Weltman A., 2004, AIP Conf. Proc. Vol. 736, PHI IN THE SKY: The Quest for Cosmological Scalar Fields. Am. Inst. Phys., New York, p. 105
Brax P., van de Bruck C., Davis A.-C., Shaw D. J., 2008, Phys. Rev. D, 78, 104021
Bridle S., King L., 2007, New J. Phys., 9, 444
Brown M. L., Taylor A. N., Hambly N. C., Dye S., 2002, MNRAS, 333, 501
Burger P. A. et al., 2023, A&A, 669, A69
Campos A., Samuroff S., Mandelbaum R., 2023, MNRAS, 525, 1885
Casas S., Kunz M., Martinelli M., Pettorino V., 2017, Phys. Dark Univ., 18, 73
Charmousis C., Gregory R., Kaloper N., Padilla A., 2006, J. High Energy Phys., 2006, 066
Chisari N. E. et al., 2018, MNRAS, 480, 3962
Crocce M., Pueblas S., Scoccimarro R., 2006, MNRAS, 373, 369
DES Collaboration, 2023, Phys. Rev. D, 107, 083504
Davies C. T., Cautun M., Li B., 2019, MNRAS, 490, 4907
Dvali G. R., Gabadadze G., Porrati M., 2000, Phys. Lett., 485, 208
Euclid Collaboration, 2019, MNRAS, 484, 5509
Fairbairn M., Goobar A., 2006, Phys. Lett. B, 642, 432
Fang W., Wang S., Hu W., Haiman Z., Hui L., May M., 2008, Phys. Rev. D, 78, 103509
Feroz F., Hobson M. P., Bridges M., 2009, MNRAS, 398, 1601
Giblin B. et al., 2021, A&A, 645, A105
Giocoli C., Baldi M., Moscardini L., 2018, MNRAS, 481, 2813
Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Hamana T. et al., 2020, PASJ, 72, 16
Harnois-Déraps J., van Waerbeke L., 2015, MNRAS, 450, 2857
Harnois-Déraps J., van Waerbeke L., Viola M., Heymans C., 2015a, MNRAS, 450, 1212
Harnois-Déraps J., Munshi D., Valageas P., van Waerbeke L., Brax P., Coles P., Rizzo L., 2015b, MNRAS, 454, 2722
Harnois-Déraps J., Giblin B., Joachimi B., 2019, A&A, 631, A160
Harnois-Déraps J., Martinet N., Castro T., Dolag K., Giblin B., Heymans C., Hildebrandt H., Xia Q., 2021, MNRAS, 506, 1623
Harnois-Déraps J., Martinet N., Reischke R., 2022, MNRAS, 509, 3868

Downloaded from https://academic.oup.com/mnras/article/525/4/6336/7263281 by guest on 30 January 2024

Heitmann K., Lawrence E., Kwan J., Habib S., Higdon D., 2014, ApJ, 780, 111

Hendrycks D., Gimpel K., 2016, preprint (arXiv:1606.08415)

Hernández-Aguayo C., Baugh C. M., Li B., 2018, MNRAS, 479, 4824

Hernández-Aguayo C., Arnold C., Li B., Baugh C. M., 2021, MNRAS, 503, 3867

Hernández-Aguayo C., Ruan C.-Z., Li B., Arnold C., Baugh C. M., Klypin A., Prada F., 2022, J. Cosmol. Astropart. Phys., 2022, 048

Heydenreich S., Brück B., Burger P., Harnois-Déraps J., Unruh S., Castro T., Dolag K., Martinet N., 2022, A&A, 667, A125

Heymans C. et al., 2021, A&A, 646, A140

Higuchi Y., Shirasaki M., 2016, MNRAS, 459, 2762

Hikage C. et al., 2019, PASJ, 71, 43

Hildebrandt H. et al., 2021, A&A, 647, A124

Hinterbichler K., Khoury J., 2010, Phys. Rev. Lett., 104, 231301

Hinterbichler K., Khoury J., Levy A., Matas A., 2011, Phys. Rev. D, 84, 103521

Hirata C. M., Seljak U., 2004, Phys. Rev. D, 70, 063526

Hobson M. P., Bridle S. L., Lahav O., 2002, MNRAS, 335, 377

Hojjati A., Zhao G. B., Pogosian L., Silvestri A., 2011, Astrophysics Source Code Library, record ascl:1106.013

Hu W., Sawicki I., 2007, Phys. Rev., D76, 064004

Joachimi B. et al., 2021a, A&A, 646, A129

Joachimi B., Köhlinger F., Handley W., Lemos P., 2021b, A&A, 647, L5

Joudaki S. et al., 2017, MNRAS, 471, 1259

Khoury J., Weltman A., 2004a, Phys. Rev. D, 69, 044026

Khoury J., Weltman A., 2004b, Phys. Rev. Lett., 93, 171104

Kilbinger M. et al., 2017, MNRAS, 472, 2126

Kingma D. P., Ba J., 2014, preprint (arXiv:1412.6980)

Koyama K., 2016, Rep. Prog. Phys., 79, 046902

Koyama K., Silva F. P., 2007, Phys. Rev. D, 75, 084040

Lemos P., Köhlinger F., Handley W., Joachimi B., Whiteway L., Lahav O., 2020, MNRAS, 496, 4647

Lemos P. et al., 2023, MNRAS, 521, 1184

Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473

Li B., 2018, Simulating Large-Scale Structure for Models of Cosmic Acceleration. IOP Publishing, Bristol, UK, p. 2514

Li B., Shirasaki M., 2018, MNRAS, 474, 3599

Li B., Zhao G.-B., Teyssier R., Koyama K., 2012, J. Cosmol. Astropart. Phys., 2012, 051

Li B., Zhao G.-B., Koyama K., 2013a, J. Cosmol. Astropart. Phys., 2013, 023

Li B., Barreira A., Baugh C. M., Hellwing W. A., Koyama K., Pascoli S., Zhao G.-B., 2013b, J. Cosmol. Astropart. Phys., 2013, 012

Llinares C., 2018, Int. J. Mod. Phys. D, 27, 1848003

Lombriser L., Hu W., Fang W., Seljak U., 2009, Phys. Rev. D, 80, 063536

Luty M. A., Porrati M., Rattazzi R., 2003, J. High Energy Phys., 2003, 029

Maartens R., Majerotto E., 2006, Phys. Rev. D, 74, 023004

Marshall P., Rajguru N., Slosar A., 2006, Phys. Rev. D, 73, 067302

Martinelli M., Calabrese E., de Bernardis F., Melchiorri A., Pagano L., Scaramella R., 2011, Phys. Rev. D, 83, 023012

Martinet N., Harnois-Déraps J., Jullo E., Schneider P., 2021a, A&A, 646, A62

Martinet N., Castro T., Harnois-Déraps J., Jullo E., Giocoli C., Dolag K., 2021b, A&A, 648, A115

Mead A. J., Brieden S., Tröster T., Heymans C., 2021, MNRAS, 502, 1401

Mota D. F., Shaw D. J., 2006, Phys. Rev. Lett., 97, 151102

Oyaizu H., 2008, Phys. Rev. D, 78, 123523

Peel A., Pettorino V., Giocoli C., Starck J.-L., Baldi M., 2018, A&A, 619, A38

Petri A., 2016, Astron. Comput., 17, 73

Pratten G., Munshi D., Valageas P., Brax P., 2016, Phys. Rev. D, 93, 103524

Ruan C.-Z., Hernández-Aguayo C., Li B., Arnold C., Baugh C. M., Klypin A., Prada F., 2022, J. Cosmol. Astropart. Phys., 2022, 018

Schmidt F., 2008, Phys. Rev. D, 78, 043002

Schmidt F., 2009, Phys. Rev. D, 80, 123003

Secco L. F. et al., 2022, Phys. Rev. D, 105, 023515

Semboloni E., Hoekstra H., Schaye J., van Daalen M. P., McCarthy I. G., 2011, MNRAS, 417, 1461

Shirasaki M., Nishimichi T., Li B., Higuchi Y., 2017, MNRAS, 466, 2402

Springel V., 2010, MNRAS, 401, 791

Springel V. et al., 2005, Nature, 435, 629

Takada M., Jain B., 2009, MNRAS, 395, 2065

Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, ApJ, 761, 152

Thomas S. A., Abdalla F. B., Weller J., 2009, MNRAS, 395, 197

Tröster T. et al., 2021, A&A, 649, A88

Troxel M. A. et al., 2018, Phys. Rev. D, 98, 043528

Vainshtein A. I., 1972, Phys. Lett. B, 39, 393

Valogiannis G., Bean R., 2017, Phys. Rev. D, 95, 103515

van den Busch J. L. et al., 2022, A&A, 664, A170

Weinberger R., Springel V., Pakmor R., 2020, ApJS, 248, 32

Will C. M., 2006, Living Rev. Relat., 9, 3

Will C. M., 2014, Living Rev. Relat., 17, 4

Winther H. A. et al., 2015, MNRAS, 454, 4208

Wright A. H., Hildebrandt H., van den Busch J. L., Heymans C., Joachimi B., Kannawadi A., Kuijken K., 2020, A&A, 640, L14

Zorrilla Matilla J. M., Waterval S., Haiman Z., 2020, AJ, 159, 284

Zuntz J. et al., 2015, Astron. Comput., 12, 45

Zürcher D., Fluri J., Sgier R., Kacprzak T., Refregier A., 2021, J. Cosmol. Astropart. Phys., 2021, 028

APPENDIX A: CURVED-SKY WEAK LENSING LIGHT CONES

We develop a curved-sky ray-tracing algorithm adapted from UFALCON[28] (Zürcher et al. 2021), in which the particle data falling into spherical mass shells are assigned onto a HEALPIX (Górski et al. 2005) maps with NSIDE = 4096, instead of the Cartesian grids used in this paper. We again use periodic boundary conditions to fill the light-cone volume whenever it exits the simulation box, and repeat the procedure for 24 different observer's positions. We modified the original UFALCON full-sky map making algorithm to implement instead a pencil-beam method, significantly reducing the memory load required to fill the high-redshift shells. This is achieved by stacking the simulation boxes along the [RA-Dec] = [0,0] direction only, and masking any pixel with RA/Dec > 12 deg. *Pseudo*-independent light cones are then extracted by selecting at random one of the 24 shells for each redshift, repeating the procedure 24 times per *N*-body simulation. The curved-sky angular power spectrum measurements are obtained from the standard HEALPY[29] routine MAP2ALM, which performs Legendre transforms on the sphere and provides measurements for $\ell \in [1-12\,288]$, which we rebin to match the flat-sky measurements for an improved comparison.

We show that both flat- and curved-sky lensing simulations produce similar C_ℓ^κ measurements. Fig. A1 presents the ratio between the lensing spectra from two models (the *f*(*R*) model-49 and the GR model-00). The thin black lines present the mean over all flat-sky measurements while the thin blue lines show the curved-sky equivalent. The agreement between these two methods is excellent in the first four tomographic bins, whereas the last tomographic bin exhibits strong discrepancies on large scales. This is caused by the mixing between the maps and the mask, and can be removed with *pseudo*-C_ℓ estimators such as NAMASTER (Alonso et al. 2019).

[28] cosmo-docs.phys.ethz.ch/UFalcon

[29] healpy.readthedocs.io/en/latest/

Downloaded from https://academic.oup.com/mnras/article/525/4/6336/7263281 by guest on 30 January 2024

Table A1. Cosmological and gravity parameters of the FORGE and BRIDGE simulations. The listed values of the structure growth parameters σ_8 and S_8 correspond to the input truth in the corresponding GR+ΛCDM simulations; the actual values in MGLENS are larger than these. Note that the emulators are specifically trained on $\Omega_{\rm m}$, S_8, h, and either $\log_{10}[f_{R_0}]$ or $\log_{10}[H_0 r_c]$. In this paper we focus on weak, medium, and strong models, which are respectively models-04, -18, and -13.

Model	$\Omega_{\rm m}$	σ_8	S_8	h	f_{R_0}	$H_0 r_c$
00	0.31315	0.82172	0.83954	0.6737	0	Inf
01	0.54725	0.49342	0.66642	0.78699	3.5502e-06	0.72533
02	0.53961	0.63783	0.85542	0.68393	3.0776e-06	0.81161
03	0.10721	1.2297	0.73513	0.6109	3.3107e-06	0.76647
04	0.31592	0.60111	0.61685	0.68845	8.0706e-07	3.9962
05	0.15741	0.91175	0.66044	0.71067	1.2093e-05	0.37375
06	0.35339	0.71886	0.78021	0.78052	5.2037e-06	0.56467
07	0.1124	1.2341	0.75539	0.79318	3.1185e-05	0.25000
08	0.39303	0.72152	0.82585	0.752	7.1372e-07	6.7113
09	0.18096	1.0378	0.80599	0.76132	9.1585e-07	3.3057
10	0.42927	0.5035	0.60228	0.77667	4.5479e-06	0.62132
11	0.40249	0.55523	0.64312	0.6912	1.3401e-06	1.7208
12	0.21286	1.0669	0.89867	0.70661	7.1154e-06	0.47331
13	0.34671	0.78191	0.84059	0.70056	1.2573e-05	0.36029
14	0.15464	0.9339	0.6705	0.77273	4.0961e-06	0.65314
15	0.28172	0.71367	0.69158	0.64968	4.9744e-06	0.59191
16	0.37032	0.61264	0.68066	0.76204	2.7753e-06	0.86134
17	0.41627	0.74242	0.87454	0.63427	1.4375e-05	0.33547
18	0.32331	0.85987	0.89266	0.81749	3.6751e-06	0.6877
19	0.47784	0.56403	0.71183	0.66724	6.7404e-06	0.49385
20	0.20509	0.75641	0.62541	0.64437	5.8109e-06	0.53938
21	0.44103	0.50237	0.60912	0.62046	6.2281e-06	0.51583
22	0.46403	0.5862	0.72906	0.80296	1.4121e-06	1.5615
23	0.13644	1.2584	0.84862	0.62473	1.0481e-06	2.4364
24	0.18832	0.85396	0.67659	0.80174	1.668e-05	0.32401
25	0.12066	1.3159	0.83454	0.69563	2.4559e-06	0.91639
26	0.28854	0.65331	0.6407	0.73943	8.7041e-06	0.43601
27	0.45016	0.72241	0.88492	0.71954	2.174e-05	0.2835
28	0.17155	1.1394	0.86159	0.62768	1.5757e-06	1.4266
29	0.51949	0.59577	0.78399	0.74473	9.6963e-06	0.40305
30	0.43909	0.61327	0.74195	0.67856	1.7774e-06	1.3111
31	0.49786	0.58288	0.75088	0.80806	1.8337e-06	1.2109
32	0.40909	0.54179	0.63268	0.73799	1.211e-06	1.9119
33	0.23227	0.86433	0.76052	0.60028	1.9037e-05	0.30276
34	0.3839	0.61174	0.69201	0.6557	2.2527e-06	1.0462
35	0.26234	0.88665	0.82914	0.76998	1.0089e-06	2.8097
36	0.25453	0.76212	0.702	0.66918	1.7789e-05	0.31312
37	0.29762	0.79347	0.79031	0.673	2.3584e-06	0.97764
38	0.22423	0.88911	0.76866	0.64603	1.3881e-05	0.34755
39	0.30799	0.71046	0.71985	0.66001	1.1732e-06	2.1452
40	0.51288	0.61834	0.80849	0.79098	7.8299e-06	0.45407
41	0.14061	1.1712	0.80186	0.73101	1.0743e-05	0.38798
42	0.33782	0.66702	0.70781	0.72256	7.9806e-07	5.0232
43	0.5252	0.66452	0.87924	0.81347	2.3279e-05	0.27454
44	0.19435	1.0172	0.8187	0.63911	2.7347e-05	0.25781
45	0.26963	0.91366	0.86618	0.75511	9.4886e-06	0.41903
46	0.49135	0.50927	0.65176	0.60766	2.5865e-05	0.26599
47	0.47207	0.58056	0.72827	0.61562	2.0816e-06	1.1234
48	0.24424	0.85676	0.77304	0.71436	6.6853e-07	10.0000
49	0.36187	0.56321	0.61856	0.72861	2.0258e-05	0.2929

Downloaded from https://academic.oup.com/mnras/article/525/4/6336/7263281 by guest on 30 January 2024

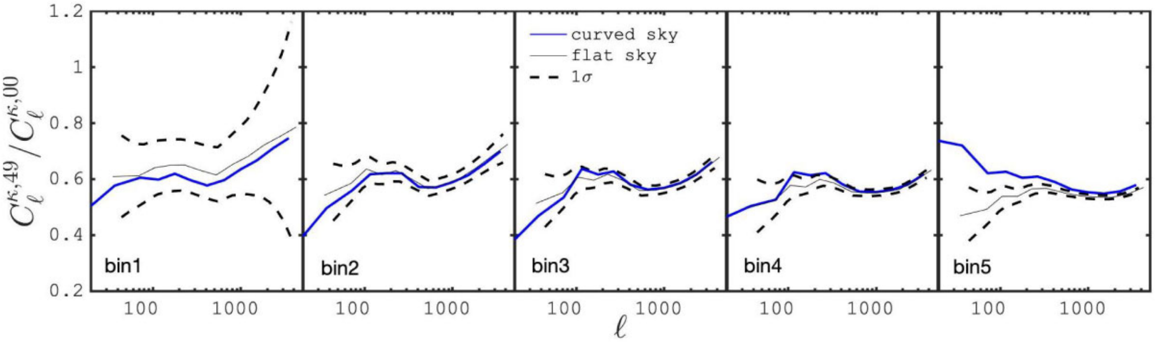

Figure A1. Comparison between the curved- and flat-sky lensing power spectra. Plotted is the ratio between the measurements from the nodes 49 and 00, for all five tomographic redshift bins. The right-most plot exhibits large-scales systematics due to masking, which are increasingly important towards higher redshifts. Our flat-sky methods are mostly immune to this.

APPENDIX B: P(K) VALIDATION

We present in this section the matter power spectra $P_m(k)$ measured from dedicated ΛCDM+GR N-body runs in which the box size is varied between 1000, 500, and 200h^{-1} Mpc. The upper panel of Fig. B1 presents the three measurements at $z = 0$, while the bottom panel shows the ratio with respect to the L200 case – given that the particle count is fixed to 1024^3, the latter has the highest resolution.

Small fluctuations in the ratio are observed at low k modes are due to residual sapling variance. While the L1000 measurements shows a 5 per cent difference in power at most scales, the L500 case shows an excellent match up to $k = 8.0$ Mpc h^{-1}. Equivalent measurements carried out at $z = 1$ reach the same conclusion, thereby establishing that our N-body runs are converged to a few per cent over the scales relevant for lensing ($k < 5$–8.0 Mpc h^{-1}, depending on the redshift of the sources).

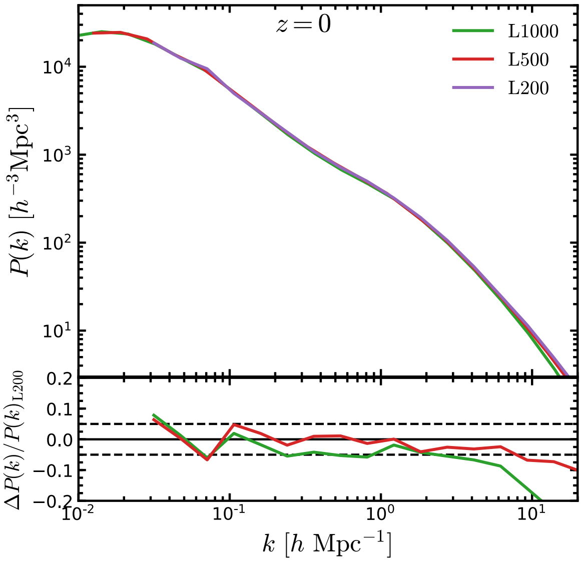

Figure B1. (upper:) Power spectra measured at $z = 0$ from GR-only N-body simulations in which the box size is varied, keeping the particle count fixed. (lower:) Ratio between the three curves shown in the upper panel curves and the L200 case.

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

Downloaded from https://academic.oup.com/mnras/article/525/4/6336/7263281 by guest on 30 January 2024

© 2023 The Author(s).
Published by Oxford University Press on behalf of Royal Astronomical Society. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.