

# Measurement of the Weyl potential evolution from the first three years of dark energy survey data

Received: 19 July 2024

Accepted: 2 October 2024

Published online: 11 November 2024

 Check for updatesIsaac Tutusaus<sup>1</sup>✉, Camille Bonvin<sup>2</sup>✉ & Nastassia Grimm<sup>2</sup>✉

The Weyl potential, which is the sum of the spatial and temporal distortions of the Universe's geometry, provides a direct way of testing the theory of gravity and the validity of the  $\Lambda$ CDM (Lambda Cold Dark Matter) model. Here we present measurement of the Weyl potential at four redshifts bins using data from the first three years of observations of the Dark Energy Survey. We find that the measured Weyl potential is  $2\sigma$ , respectively  $2.8\sigma$ , below the  $\Lambda$ CDM predictions in the two lowest redshift bins. We show that these low values of the Weyl potential are at the origin of the tension between Cosmic Microwave Background measurements and weak lensing measurements, regarding the parameter  $\sigma_8$  which quantifies the clustering of matter. Interestingly, we find that the tension remains if no information from the Cosmic Microwave Background is used. Dark Energy Survey data on their own prefer a high value of the primordial fluctuations, together with a slow evolution of the Weyl potential. An important feature of our method is that the measurements of the Weyl potential are model-independent and can therefore be confronted with any theory of gravity, allowing efficient tests of models beyond General Relativity.

Since the observation of the accelerated expansion of the Universe in 1998<sup>1,2</sup>, extensive studies have been performed to understand if this acceleration is due to a cosmological constant, a dynamical dark energy, or a modification of the theory of gravity. In theories beyond  $\Lambda$ CDM, not only the background evolution of the Universe can be modified, but also the formation of structure. A powerful way of testing these theories is therefore to confront them with measurements from large-scale structure surveys.

The landscape of theories beyond  $\Lambda$ CDM is vast, and testing all models one by one has become infeasible. As a consequence, two complementary approaches have been developed. The first one consists in testing classes of theories, e.g. Horndeski theories<sup>3</sup> that encompass all scalar-tensor theories with second-order equations of motion. The advantage of such an approach is that the functions constrained by the data are directly linked to fundamental ingredients

in the theories. However, the degeneracies between the free functions are large, and current data are not stringent enough to constrain them all well<sup>4</sup>.

The second approach is more phenomenological and consists in parameterising deviations from General Relativity (GR) directly at the level of Einstein's equations by introducing two new functions, often called  $\mu$  and  $\eta$ <sup>5</sup>. The function  $\mu$  encodes changes in Poisson's equation, while  $\eta$  describes the difference between the time distortion and the spatial distortion of the metric, the so-called anisotropic stress. This approach has been successfully used with current data, e.g. combining gravitational lensing measurements from the Dark Energy Survey (DES) and galaxy clustering measurements from BOSS and eBOSS<sup>6</sup>. Even though powerful, this approach suffers from one important complication: the constraints on  $\mu$  and  $\eta$  at a given redshift  $z$  depend on the evolution of the functions at all redshifts above  $z$ . This means

<sup>1</sup>Institut de Recherche en Astrophysique et Planétologie (IRAP), Université de Toulouse, CNRS, UPS, CNES, 14 Av. Edouard Belin, Toulouse, France. <sup>2</sup>Département de Physique Théorique and Center for Astroparticle Physics, Université de Genève, Quai E. Ansermet 24, Geneva, Switzerland.

✉ e-mail: [isaac.tutusaus@irap.omp.eu](mailto:isaac.tutusaus@irap.omp.eu); [camille.bonvin@unige.ch](mailto:camille.bonvin@unige.ch); [nastassia.grimm@unige.ch](mailto:nastassia.grimm@unige.ch)

that either one assumes a fixed time evolution for these functions, as done in ref. 6, and only the values of  $\mu$  and  $\eta$  at  $z=0$  are constrained. Or one needs sophisticated techniques to constrain and reconstruct the time evolution of  $\mu$  and  $\eta$  from a set of chosen redshift nodes, as done for example in ref. 7. In this second case, it is necessary to have measurements over a wide range of redshift to obtain relevant constraints, and any degradation of the data in a given redshift bin will impact the constraints in the other bins. Another limitation of this framework is that the constraints on  $\mu$  and  $\eta$  rely on the validity of Euler's equation for dark matter. Without this additional assumption, these parameters cannot be constrained from redshift-space distortions (RSD), since they are fully degenerated with the parameters encoding changes in Euler's equation<sup>8,9</sup>.

Here we propose an alternative approach, which is fully detached from any theoretical model. We ask ourselves what are the quantities that we can directly measure from the data, in a fully model-independent way. This approach has been extensively used in the past for RSD. Combining the multipoles of the RSD correlation function (or power spectrum) provides direct measurements of the galaxy peculiar velocities in the redshift bins of the surveys. The evolution of the velocity is encoded in the so-called growth rate function,  $f$ , which is measured in combination with  $\sigma_8$  (the amplitude of density perturbations in spheres of 8 Mpc/h)<sup>10–12</sup>. Such measurements are very powerful, since  $f\sigma_8$  can then be compared with the prediction from GR to see if this theory is consistent or not. Moreover, it can be compared with predictions from any other theory of gravity, to put constraints on their parameters, see e.g. ref. 12. To the best of our knowledge, this approach, which is standard in RSD analyses, has never been used for gravitational lensing analyses with real observations. The goal of this paper is to do precisely that, using DES measurements of galaxy-galaxy lensing and galaxy clustering. More precisely, we use the formalism developed in ref. 13 to measure the evolution of the Weyl potential  $\Psi_W = (\Phi + \Psi)/2$  (where  $\Phi$  and  $\Psi$  denote respectively the spatial and temporal perturbations of the metric, see Eq. (1) in methods) in the four tomographic bins of the DES lenses. This provides the first direct and model-independent measurement of the evolution of the perturbed geometry of our Universe. We define a new function,  $\hat{f}$ , that we call the *Weyl evolution*, to encode the evolution of the Weyl potential. The measurement of  $\hat{f}$  from DES data is fully complementary to the growth rate measurement. Modified theories of gravity can indeed change the way structures evolve, the way the geometry evolves, or

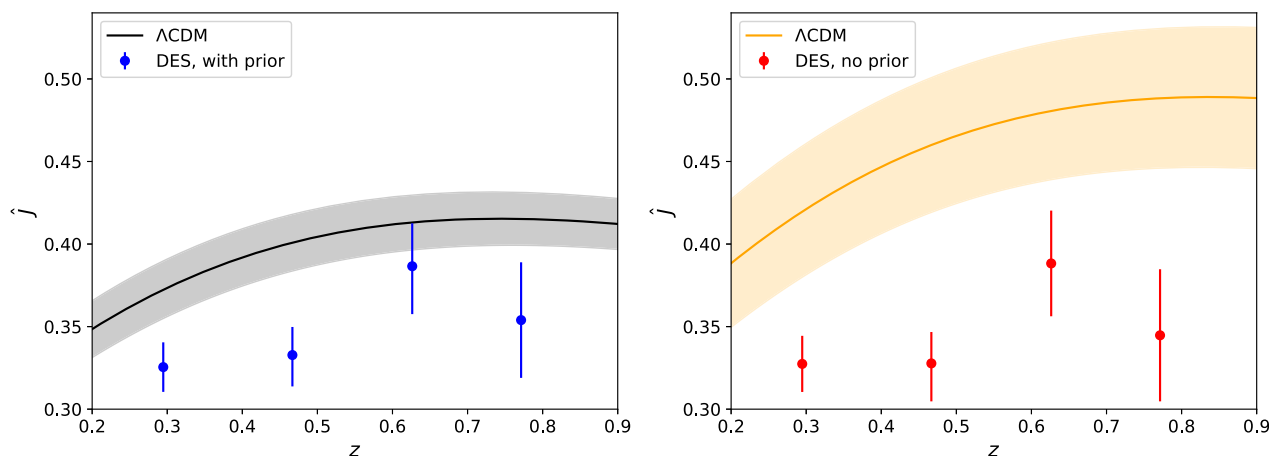
both. As for the measurement of  $f\sigma_8$ , the Weyl evolution  $\hat{f}$  is measured redshift bin per redshift bin, without assuming any time evolution. These model-independent measurements can then be compared with the prediction in  $\Lambda$ CDM, or with any theory of gravity of interest.

## Results

We combine two sets of data from DES: galaxy clustering and galaxy-galaxy lensing. As described in methods, subsection The lensing angular power spectra as function of the Weyl evolution, these two data sets provide constraints on cosmological parameters, nuisance parameters, as well as the galaxy bias,  $b(z_i) = b(z_i)\sigma_8(z_i)$ , and the Weyl evolution,  $\hat{f}(z_i)$ , in each of the redshift bins of the lenses. We focus on the first four tomographic bins of the DES lens sample, with mean redshifts [0.295, 0.467, 0.626, 0.771], since some residual systematic uncertainties were identified in the two highest redshift bins<sup>14</sup>. We consider four different scenarios: our baseline corresponds to the use of  $3\sigma$  Gaussian priors on the cosmological parameters from Planck measurements of the Cosmic Microwave Background (CMB)<sup>15</sup>, while using the fiducial DES angular scale cuts. In this way, the density power spectrum at high redshift is constrained by Planck, while the late time evolution is left free. To test the sensitivity of our treatment on the non-linear modelling, we also consider a “pessimistic” scenario with more stringent angular scale cut, using only scales above 21 Mpc/h for the galaxy-galaxy lensing observable (see “methods”, subsection Measurement of the Weyl potential for more detail). Finally, we also consider two scenarios (with standard and pessimistic scale cuts), removing the priors from Planck. In these scenarios, both the early time power spectrum and the late time evolution of the Weyl potential are constrained by DES data only. For these cases we use the wide fiducial DES priors on the cosmological parameters.

### Measurements of $\hat{f}$ using priors from Planck

In Fig. 1, left panel, we show the measured values of  $\hat{f}$  with  $1\sigma$  error bars. We see that  $\hat{f}$  is very well measured, with a precision of 5 to 10% in the four DES redshift bins (see also Table 1). This demonstrates the capacity of the data to significantly constrain the Weyl evolution in a fully model-independent way, i.e. without assuming any theory of gravity. The measured values of  $\hat{f}$  can then be compared with any model of interest. In Fig. 1, we show the prediction for  $\hat{f}$  in  $\Lambda$ CDM, with the cosmological parameters obtained through our cosmological analysis. The shaded region corresponds to the  $1\sigma$  uncertainty on the  $\Lambda$ CDM



**Fig. 1 | Measured values of  $\hat{f}$  with 1 standard deviation error bars at the mean redshifts of the MAGLIM sample.** The left panel shows the measurements assuming a prior from Planck on early time cosmological parameters. The right panel shows the measurements without Planck priors. The black and orange solid lines show the predictions for  $\hat{f}$  in  $\Lambda$ CDM. The shaded regions indicate the  $1\sigma$  uncertainty on those predictions due to uncertainties in the primordial

cosmological parameters. Note that, due to the change in cosmological parameters (see Supplementary Table 1), the  $\Lambda$ CDM prediction of  $\hat{f}$  changes in the case with and without prior, which is particularly noticeable at higher redshifts, where there is a  $1.6\sigma$  difference between the two predictions. Source data are provided as a Source Data file.

**Table 1 | Mean values and  $1\sigma$  error bars for  $\hat{J}(z_i)$  in the four cases considered in our analysis: with and without priors from Planck measurements of the Cosmic Microwave Background (CMB), with standard and pessimistic scale cuts**

	CMB prior		No CMB prior	
	Standard cuts	Pessimistic cuts	Standard cuts	Pessimistic cuts
$\hat{J}(z_1)$	$0.325 \pm 0.015$	$0.329 \pm 0.028$	$0.327 \pm 0.017$	$0.335 \pm 0.030$
$\hat{J}(z_2)$	$0.333^{+0.017}_{-0.019}$	$0.324 \pm 0.035$	$0.328^{+0.019}_{-0.023}$	$0.326 \pm 0.036$
$\hat{J}(z_3)$	$0.387^{+0.026}_{-0.029}$	$0.428 \pm 0.061$	$0.388 \pm 0.032$	$0.433 \pm 0.063$
$\hat{J}(z_4)$	$0.354 \pm 0.035$	$0.368 \pm 0.075$	$0.345 \pm 0.040$	$0.370 \pm 0.080$

We have considered a uniform prior for  $\hat{J}$  between 0.1 and 0.6 for the first two bins, and between 0.15 and 0.65 for the last two bins.

prediction. In the first and second redshift bins, the measurements of  $\hat{J}$  are  $2\sigma$ , respectively  $2.8\sigma$ , below the  $\Lambda$ CDM predictions. In the highest two bins the measurements are compatible with  $\Lambda$ CDM ( $1.6\sigma$  difference in the last bin). The joint constraints on  $\hat{J}$  are plotted in Supplementary Fig. 1.

Even though such a tension is not very significant, it is interesting to see that the deviations from  $\Lambda$ CDM are localised to the two lowest redshift bins. This could hint either at a modification of gravity that may have more impact at low redshift, when the mechanism responsible for the acceleration of the Universe is in full play. Or it could be due to a (partially) incorrect modelling of non-linear effects, which are more important at low redshift. Alternatively, it could also be due to unknown systematic effects in the first two bins of the MAGLIM sample.

Our measurements are in agreement with the results of ref. 16, which used cosmic shear and galaxy clustering data from DES Year 1, KiDS-1000, eBOSS QSOs, DESI Legacy Survey and CMB lensing to reconstruct the redshift evolution of matter density perturbations. Their methodology differs from ours, first because they use GR to link the Weyl evolution to density perturbations, allowing them to combine cosmic shear, CMB lensing and galaxy clustering to measure a single growth function, and second because they reconstruct the growth function using spline interpolation between chosen redshift nodes. As can be seen from Fig. 8 of ref. 16, they find a growth function which is lower than predicted by  $\Lambda$ CDM between  $z = 0.2$  and  $z = 0.5$ , in perfect agreement with our measurements of a low Weyl evolution in the first two redshift bins. Both methods are highly complementary: using GR and spline interpolation allows them to combine different probes, enhancing the precision of their measurement. Our method is more model-independent since it does not rely on GR, which allows us to compare our measurements of  $\hat{J}$  with the predictions of various modified gravity models, see subsection Modified gravity below. Moreover, by measuring  $\hat{J}$  at the redshifts of the lenses, our method can also directly identify if the tension comes from one particular redshift bin (which could indicate a problem with the lens sample in that bin), or from various redshift bins: the two lowest ones in our case. In Supplementary Discussion 2, we also compare our results with the binned  $\sigma_8$  analysis of DES and we show that the two approaches are complementary but different: the binned  $\sigma_8$  analysis is a consistency test of the  $\Lambda$ CDM model, whereas our approach provides a model-independent way of measuring the Weyl evolution. For completeness, in Supplementary Discussion 3, we show and discuss the results for the REDMAGIC lens sample of DES<sup>17</sup>.

Since  $\hat{J}$  has been measured in a fully model-independent way, we can now use these measurements to infer  $\sigma_8(z=0)$  assuming  $\Lambda$ CDM. Combining the four measurements of  $\hat{J}$ , we can fit for  $\sigma_8(z=0)$  at each point of the chain using Eq. (10) in methods, and obtain its posterior distribution which we find to be close to Gaussian (see Supplementary Fig. 2). The mean and  $1\sigma$  errors are  $\sigma_8^{\text{fit}}(z=0) = 0.743 \pm 0.039$ . In

contrast, using the values of  $\Omega_m$ ,  $\Omega_b$ ,  $n_s$  and  $A_s$ , we obtain  $\sigma_8^{\text{cosmo}}(z=0) = 0.849 \pm 0.030$ , which is  $2.2\sigma$  higher than the value obtained from  $\hat{J}$ . With our method, we recover therefore the well-known result from the Year 3 DES data that high redshifts prefer a larger clustering amplitude than low redshifts. Moreover, we are able to pin-point this tension to the behaviour of the Weyl potential in the two lowest redshift bins of DES. We obtain similar results if we use  $\hat{J}$  to infer  $S_8 = \sigma_8 \sqrt{\Omega_m/0.3}$ : the value obtained from  $\hat{J}$  is  $2.1\sigma$  below that obtained from cosmological parameters.

Our results are in agreement with previous analyses, which combined CMB lensing (either from ACT or from Planck) with clustering from DES<sup>18</sup>, BOSS<sup>19</sup>, DESI<sup>20</sup> and high-redshift quasars<sup>21,22</sup>. These analyses all point toward a clustering amplitude which is consistent with  $\Lambda$ CDM above redshift  $0.5^{21-23}$ , but is too small at low redshift<sup>16,18-20</sup>. This shows that the tension is not limited to cosmic shear measured in galaxy surveys, but also appears when using CMB lensing. Combined with the fact that some of those analyses used BOSS and DESI galaxy samples instead of the DES lens sample, we can conclude that the tension is not linked to a particular data set. Interestingly, these results are also in agreement with other probes, e.g. the measurements in ref. 24, which find that high-redshift Lyman- $\alpha$  data prefer a large value of  $\sigma_8$  compared to the number counts of clusters at low redshift pointing towards a lower  $\sigma_8$ .

### Measurements of $\hat{J}$ without Planck prior

We then redo the measurements of  $\hat{J}$ , removing the priors from Planck. From Fig. 1 (right panel), we see that  $\hat{J}$  can still be very well measured in this case: the errors increase by 11–20% only (see Table 1). On the other hand, the errors on the cosmological parameters are 3 to 6 times larger when no priors are used ( $A_s$  and  $h$  being the most degraded). This degradation is not surprising, since in our framework all the information coming from the growth of structure and from the Weyl potential is encoded in the parameters  $\hat{b}_i$  and  $\hat{J}$  and does not contribute to the constraints on the cosmological parameters (see methods, subsection The lensing angular power spectra as function of the Weyl evolution for more detail). The larger uncertainties in cosmological parameters lead to wider errors on the prediction for  $\hat{J}$  in  $\Lambda$ CDM, as is clearly visible from Fig. 1.

We see that removing the priors actually slightly increases the tension between the measured  $\hat{J}$  and the  $\Lambda$ CDM predictions, which is now  $[2.1\sigma, 2.9\sigma, 1.7\sigma, 2.4\sigma]$  in the four bins. This is due to the fact that the  $\Lambda$ CDM predictions for  $\hat{J}$  increase when the prior is removed, due to an increase of the primordial amplitude  $A_s$  from  $(2.13 \pm 0.10) \cdot 10^{-9}$  to  $(2.98^{+0.57}_{-0.69}) \cdot 10^{-9}$ . In this case,  $A_s$  is mainly constrained by the (subdominant) RSD contribution to galaxy clustering that we modelled in GR (see methods, subsection Measurement of the Weyl potential, and Supplementary Discussion 1 for more detail). We note that the value of  $A_s$  is still compatible with Planck value at  $1.2\sigma$ , but its slight increase counterbalances the increase of uncertainty due to the absence of priors, leading to a persisting tension of  $\hat{J}$  with  $\Lambda$ CDM predictions. From this analysis, we see that DES data on their own prefer a high amplitude of perturbations  $A_s$ , together with a growth of the Weyl potential slower than in  $\Lambda$ CDM at low redshifts.

The tension in  $\hat{J}$  leads to a  $2.1\sigma$  tension between  $\sigma_8$  extracted from the cosmological parameters:  $\sigma_8^{\text{cosmo}}(z=0) = 1.028 \pm 0.097$ , and  $\sigma_8$  extracted from  $\hat{J}$  assuming  $\Lambda$ CDM:  $\sigma_8^{\text{fit}}(z=0) = 0.776^{+0.067}_{-0.079}$ . The large value of  $\sigma_8^{\text{cosmo}}$  is directly related to the large value of  $A_s$ . Our analysis shows therefore that the  $\sigma_8$  tension is internally present in DES data. The clustering amplitude extracted from the evolution of the Weyl potential is in tension with the clustering amplitude inferred from the background evolution and the subdominant RSD signal in the distribution of lenses. Our method therefore indicates that the  $\sigma_8$  tension is not only a tension between high redshift and low redshift, but that it is also a tension between the perturbations of the geometry at low

redshift (probed by weak lensing), and the other dynamical fields: expansion rate and density fluctuations (encoded in the RSD contribution). This is in agreement with measurements of the growth rate function  $f$  from spectroscopic RSD surveys, which show no deviations from  $\Lambda$ CDM at low redshift (see e.g. Fig. 1 of ref. 25). What is particularly interesting in our method is that it allows the separation of information from the Weyl potential and from background and density perturbations within DES data, consequently showing that the same trend is present in this single data set. Translating our constraints on  $\sigma_8(z=0)$  to constraints on  $S_8$  we find that the tension is slightly larger, of  $2.6\sigma$ . We summarise the values of  $\sigma_8(z=0)$  obtained in various cases in Table 2.

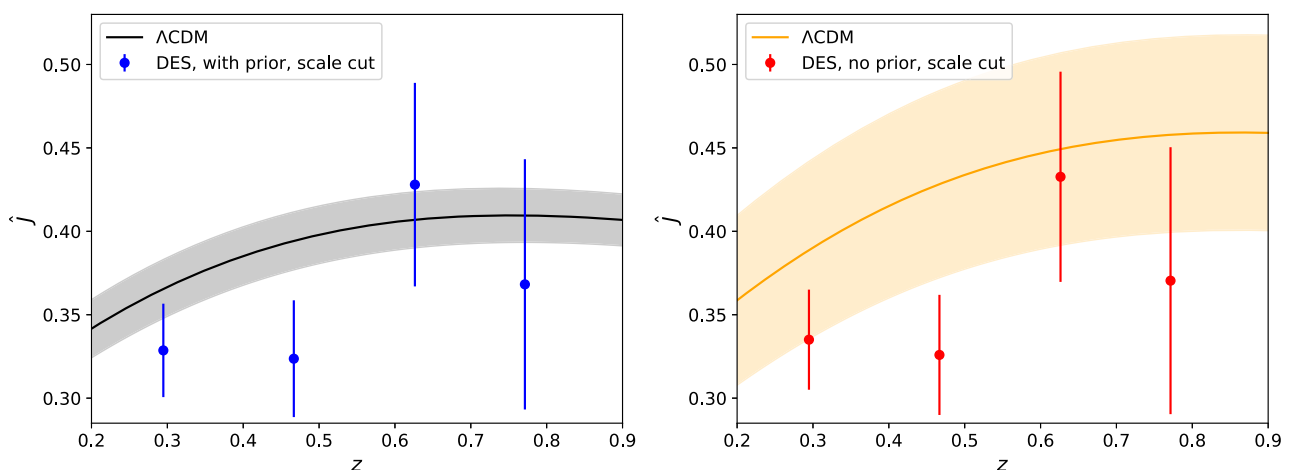
### Measurements of $\hat{f}$ with more stringent non-linear scale cuts

As discussed in methods, subsection Measurement of the Weyl potential, we have implemented more stringent scale cuts to further remove the impact of non-linear scales (see methods, subsection Measurement of the Weyl potential). The results for  $\hat{f}$  are shown in Fig. 2 for the analysis with Planck priors (left panel) and without (right panel). We see that the measurements of  $\hat{f}$  are significantly degraded: the errors increase by roughly a factor 2. This is not surprising since removing scales reduces the available amount of information needed to measure  $\hat{f}$ . The errors on the  $\Lambda$ CDM prediction also increase due to the scale cut, in the case without prior, but only by 14%. Interestingly, the mean values of  $\hat{f}$  are only mildly affected by removing scales and are still below the  $\Lambda$ CDM predictions. However, the increase in the error bars significantly reduces the tension with  $\Lambda$ CDM predictions: only the second bin is still  $1.8\sigma$  away from the  $\Lambda$ CDM prediction in the case with prior ( $1.5\sigma$  without prior). A comparison of the measurements of  $\hat{f}$  with standard and pessimistic scale cuts is provided in Table 1.

**Table 2 | Mean values and  $1\sigma$  error bars for  $\sigma_8(z=0)$ , in the cases with and without CMB priors, using the standard scale cuts**

	CMB Prior		No CMB prior	
	From parameters	From $\hat{f}$	From parameters	From $\hat{f}$
$\sigma_8(z=0)$	$0.849 \pm 0.030$	$0.743 \pm 0.039$	$1.028 \pm 0.097$	$0.776^{+0.067}_{-0.079}$

We compare the value fitted from  $\hat{f}$  with the value obtained from the cosmological parameters.



**Fig. 2 | Measured values of  $\hat{f}$  with 1 standard deviation error bars at the mean redshifts of the MAGLIM sample, using more stringent scale cuts.** The left panel shows the measurements assuming a prior from Planck on early time cosmological parameters. The right panel shows the measurements without Planck priors. The

### Modified gravity

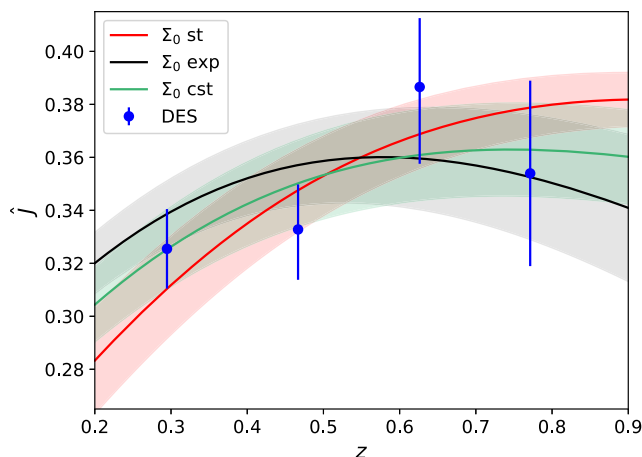
We now compare our measurements of  $\hat{f}$  in the baseline scenario with predictions in modified theories of gravity. We explore two cases: A) the phenomenological  $\mu - \Sigma$  extensions of gravity; and B) Horndeski theories, concentrating on a few specific cases. The relation between  $\hat{f}$  and the parameters of these two types of modified gravity models are given in methods, subsection Modified theories of gravity. In both cases, we use the measured  $\hat{f}$  to constrain the parameters of the theory, and then use those to compute the best fit  $\hat{f}(z)$ .

In case A), we fix  $\mu = 1$ , since as explained in methods,  $\hat{f}$  is mainly affected by  $\Sigma$ . We explore three types of evolution for the function  $\Sigma(z)$ : a standard evolution, where  $\Sigma(z)$  decays proportionally to dark energy; no evolution; and an exponential evolution (see Eq. (13) in methods). The posteriors for  $\Sigma_0 = \Sigma(z=0)$  are very close to Gaussian for the three cases, with mean and  $1\sigma$  errors given by:  $\Sigma_0^{\text{st}} = -0.24 \pm 0.10$  (standard evolution),  $\Sigma_0^{\text{st}} = -0.13 \pm 0.06$  (no evolution) and  $\Sigma_0^{\text{exp}} = -0.027 \pm 0.013$  (exponential evolution). In Fig. 3 we show the predicted  $\hat{f}(z)$  (mean and  $1\sigma$  error bands) for the three models, compared with the measured  $\hat{f}$  in the baseline case. In the three models, a negative value of  $\Sigma_0$  decreases  $\hat{f}$  with respect to  $\Lambda$ CDM, leading to a better agreement with the data (at the cost of one extra parameter though). The reduced chi-squared of  $\Lambda$ CDM is of 2.1 ( $p$ -value 0.078), while the reduced chi-squared of the three models (with one extra free parameter  $\Sigma_0$  each) is very similar: 1.1 for the standard evolution and the constant case ( $p$ -value 0.33), and 0.8 for the exponential evolution ( $p$ -value 0.48). This shows that the data are not able to discriminate between the different time evolutions, as is clear from Fig. 3. The coming generation of surveys has clearly the potential to help, by adding measurements at lower and higher redshifts and reducing the error bars. Here we also see the strong advantage of our method: by first measuring  $\hat{f}$  without assuming any model and then comparing with the predictions in modified gravity, we can easily understand and test the impact of different assumptions about the time evolution.

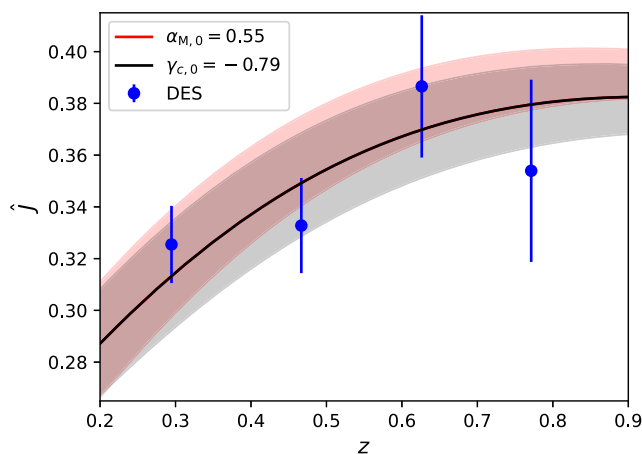
In case B) we consider Horndeski models while additionally allowing for a non-minimal coupling of dark matter, which constitutes a breaking of the weak equivalence principle. We focus on two specific cases: first, we consider the case where  $\alpha_B = \alpha_M/2$  can differ from zero, and dark matter is minimally coupled to the scalar field,  $\gamma_c = 0$  (see e.g. ref. 26 for more detail on these parameters). This special relation between  $\alpha_M$  and  $\alpha_B$  is recovered e.g. in Brans-Dicke theories<sup>27,28</sup> and some  $f(R)$  models<sup>29–32</sup>. Second, we investigate the case with a non-minimal coupling of dark matter,  $\gamma_c \neq 0$ , while all other parameters are

black and orange solid lines show the predictions for  $\hat{f}$  in  $\Lambda$ CDM. The shaded regions indicate the  $1\sigma$  uncertainty on those predictions due to uncertainties in the primordial cosmological parameters. Source data are provided as a Source Data file.





**Fig. 3 | Predictions for  $\tilde{f}$  for modified gravity models with  $\Sigma \neq 1$  and  $\mu = 1$ .** We show the mean value together with the 1 standard deviation error bars for the three choices of time evolution (standard, constant, and exponential) described in methods. The measurements correspond to our baseline, using Planck priors and with standard scale cuts. Source data are provided as a Source Data file.



**Fig. 4 | Predictions for  $\tilde{f}$  in Horndeski models.** We show the mean value and 1 standard deviation error bars for a model with  $\alpha_M = 2\alpha_B$  (red) and a model with a non-minimal coupling,  $\gamma_c \neq 0$ , of CDM to the metric (black). In both cases, all other modified gravity parameters are assumed to maintain their GR values. The measurements correspond to our baseline, using Planck priors and with standard scale cuts. Source data are provided as a Source Data file.

kept to their GR value. For the two specific models, we obtain  $\alpha_{M,0} = 2\alpha_{B,0} = 0.55^{+0.22}_{-0.20}$  and  $\gamma_{c,0} = -0.79^{+0.30}_{-0.37}$ . The results for  $\tilde{f}$  are shown in Fig. 4. We see that both  $\alpha_M = 2\alpha_B > 0$  as well as  $\gamma_c < 0$  can decrease the amplitude of the Weyl potential at late time, leading to predicted values of  $\tilde{f}(z)$  in line with the measured data. This shows that while  $\tilde{f}$  provides an excellent test of the  $\Lambda$ CDM model, it cannot, on its own, distinguish between modifications of gravity and a breaking of the weak equivalence principle. A combination with other observables, in particular gravitational redshift, will allow to break degeneracies between different physical effects<sup>26</sup>.

## Discussion

In this paper, we have performed the first measurement of the Weyl potential from galaxy-galaxy lensing. Using data from DES, we were able to measure the function  $\tilde{f}(z)$ , which encodes the evolution of the Weyl potential, with a precision of 5 to 10% in the four bins of the MAGLIM lens sample. We found that the measured values in the two lowest bins are in mild tension with the  $\Lambda$ CDM predictions.

One important feature of our method is that it allows us to separate the information coming from the matter density fluctuations (which we do not use to constrain gravity, since it is fully degenerated with the galaxy bias) from the information coming from the Weyl potential. The latter purely encodes the growth of the spatial and temporal distortions of the geometry, and is therefore independent of the galaxy bias. Moreover, our method does not rely on any assumption about the redshift evolution of  $\tilde{f}$ : the function is directly measured in each bin of the survey.

Our measurements allow us to pin-point the  $\sigma_8$  tension to the fact that the Weyl potential is smaller than predicted in  $\Lambda$ CDM in the two lowest bins of the DES data. Surprisingly, we found that the tension remains even when no priors from the CMB are used. DES data on their own prefer a Universe with a high initial value of the primordial potential (higher than measured by Planck) followed by a slower growth at late time. Further investigation to understand the origin of this behaviour is needed. In particular, it would be interesting to redo the analysis changing the settings to the ones used in the new combined analysis of DES and KiDS<sup>33</sup> (which found a decrease of the  $\sigma_8$  tension) to see how this impacts the measurements of  $\tilde{f}$ . Moreover, having measurements of  $\tilde{f}$  between the second and third bin of DES, where the deviation suddenly appears could help understand the origin of the tension. Surveys like Euclid and LSST will be of crucial importance for this.

More generally, our method allows us to compare the measurements of  $\tilde{f}$  with predictions in any theory of gravity that recovers GR at high redshift. Since the measurements are model-independent, the method allows for various tests of the theory of gravity in an efficient way. One can for example easily change the time evolution of the parameters of the theory, without having to reanalyse the data for each case. Using the  $\mu$ - $\Sigma$  parameterisation of modified gravity, we have shown that current data are not constraining enough to differentiate between different time evolutions. In the future, we will study if correlations of galaxy clustering with CMB lensing may help. Moreover, future surveys like Euclid and LSST are expected to provide sufficiently tight constraints (at the level of 0.1–1% for  $z \in [0.25, 2.1]$  for LSST<sup>13</sup>) over a wider redshift range, to allow for much more stringent tests.

The evolution of the Weyl potential can be combined with measurements of the growth rate of structure, to test the relation between these two fundamental quantities in our Universe. In ref. 25, we have combined these quantities to obtain precise measurements of the  $E_G$  statistic, providing a robust test of GR. Therefore, while  $\tilde{f}$  reveals an interesting tension with  $\Lambda$ CDM on its own, it is particularly powerful in combination with further model-independent observables.

## Methods

The Weyl evolution  $\tilde{f}$  is measured from galaxy-galaxy lensing, combined with galaxy clustering. The formalism to relate gravitational lensing observables to  $\tilde{f}$  has been derived in ref. 13. Here we summarise the main results.

### The lensing angular power spectra as function of the Weyl evolution

We work with the linearly perturbed flat Friedmann-Lemaître-Robertson-Walker (FLRW) metric in the conformal Newtonian gauge, with the line element given by

$$ds^2 = a^2(\tau) [-(1 + 2\Psi)d\tau + (1 + 2\Phi)d\mathbf{x}^2], \quad (1)$$

where  $\tau$  denotes conformal time,  $a$  is the scale factor, and  $\Psi$  and  $\Phi$  are the two metric potentials. Gravitational lensing is directly sensitive to the Weyl potential  $\Psi_W \equiv (\Phi + \Psi)/2$ . We define the Weyl transfer function  $T_{\Psi_W}$ , which relates the value of  $\Psi_W(z, \mathbf{k})$  in Fourier space and at redshift  $z$  to the primordial potential at the end of inflation  $\Psi_{\text{in}}(\mathbf{k})$ . In GR, one can easily show that for wavelengths that are inside the

horizon, the Weyl transfer function simply evolves proportionally to  $D_1(z)\Omega_m(z)$ . Here  $D_1(z)$  is the growth function (governing the evolution of matter density perturbations) and  $\Omega_m(z)$  is the matter density parameter at redshift  $z$ . Hence, in GR, measuring the evolution of density perturbations or measuring the evolution of the Weyl potential provides exactly the same information. In modified theories of gravity the situation is different: Einstein's equations are modified, generically leading to a different evolution of the Weyl potential. To capture this, we introduce a new function,  $J$ , encoding this evolution. More precisely, we write the transfer function of the Weyl potential as (see ref. 13 for more detail)

$$T_{\Psi_w}(k, z) = \frac{\mathcal{H}^2(z)J(k, z)}{\mathcal{H}^2(z^*)D_1(z^*)} \frac{\sqrt{B(k, z)}}{\sqrt{B(k, z^*)}} T_{\Psi_w}(k, z^*), \quad (2)$$

where  $z^*$  is a redshift well in the matter era (before the acceleration of the Universe started),  $\mathcal{H}$  is the Hubble parameter in conformal time, and  $B(k, z)$  is a boost factor, encoding the non-linear evolution of matter density perturbations at small scales. We assume that at  $z^*$  a cold dark matter (CDM) Universe, governed by GR, is recovered. This assumption automatically excludes early dark energy models, where deviations from a CDM Universe take place already at high redshift. The motivation behind this assumption is that, at early times, CMB measurements are highly compatible with a CDM Universe<sup>15</sup>. Theories of modified gravity therefore usually aim at recovering GR at high redshift, while leading to a phase of accelerated expansion at low redshift. In our framework, any deviation in the Weyl potential evolution during the phase of accelerated expansion is encoded in the free function  $J$ . By measuring this free function as a function of redshift, we are able to reconstruct the Weyl evolution and assess its compatibility with GR predictions, or with the predictions of any theory of gravity we may want to test. Note that the function  $J$  has the same information content as the function  $L$  defined in refs. 34,35, which identified for the first time the quantities that can be measured in a model-independent way from large-scale structure surveys.

In full generality,  $J$  is a function of redshift and wavenumber  $k$ . In GR, the evolution is almost scale-independent, except from a small scale-dependence due to massive neutrinos. In our analysis we fix the sum of neutrino masses to 0.06 eV, leading to a scale-dependence in the growth of structure (and therefore in  $J$ ) which is negligible<sup>36</sup>. In models of gravity beyond GR,  $J$  can in principle depend on scales. However, in many models of modified gravity, it turns out that the evolution is scale-independent for sub-horizon modes, where the quasi-static approximation is valid<sup>37,38</sup>. In our analysis, we therefore drop the scale-dependence in  $J$  (as is also usually done for measurements of the growth rate in RSD analyses<sup>10–12</sup>). This simplifies the analysis, but it is not a fundamental limitation of the method, which can be extended to a scale-dependent  $J$ .

As observable we consider the galaxy-galaxy lensing angular power spectrum, i.e. the cross-correlation of cosmic shear of background galaxies with clustering of foreground lenses. As shown in ref. 13, the galaxy-galaxy lensing angular power spectrum can be written in the following way (using the Limber approximation):

$$C_\ell^{\Delta\kappa}(z_i, z_j) = \frac{3}{2} \int dz n_i(z) \mathcal{H}^2(z) \hat{b}_i(z) \hat{J}(z) B(k_\ell, \chi) \frac{P_{\delta\delta}^{\text{lin}}(k_\ell, z^*)}{\sigma_8^2(z^*)} \int dz' n_j(z') \frac{\chi'(z') - \chi(z)}{\chi(z)\chi'(z')}, \quad (3)$$

where  $n_i$  and  $n_j$  denote the galaxy distribution function of the lenses and sources, respectively, and  $k_\ell \equiv (\ell + 1/2)/\chi$ , with  $\chi$  the comoving distance.  $P_{\delta\delta}^{\text{lin}}$  is the linear matter power spectrum, evaluated here at redshift  $z^*$ . We see that the galaxy-galaxy lensing angular power spectrum depends on: 1) the density fluctuations at  $z^*$  where GR is

recovered; 2) the evolution of background quantities,  $\mathcal{H}(z)$  and  $\chi(z)$ ; and 3) the functions  $\hat{J}$  and  $\hat{b}_i$  defined as

$$\hat{J}(z) \equiv \frac{J(z)\sigma_8(z)}{D_1(z)} = \frac{J(z)\sigma_8(z^*)}{D_1(z^*)}, \quad (4)$$

$$\hat{b}_i(z) \equiv b_i(z)\sigma_8(z), \quad (5)$$

where  $b_i$  is the linear galaxy bias in the tomographic redshift bin  $i$ . The second equality in Eq. (4) follows from the fact that  $\sigma_8(z)$  is directly proportional to  $D_1(z)$ . As a consequence  $\hat{J}$  and  $J$  contain the same information about modified gravity, since they are related by the ratio  $\sigma_8(z)/D_1(z)$  which does not depend on the modifications of gravity (since we assume that at  $z^*$  GR is recovered). In the following we will assume the background evolution to be as in  $\Lambda$ CDM. This is an assumption that is often done in large-scale structure analyses, since all current constraints set the background evolution to be close to that of  $\Lambda$ CDM. This assumption can be relaxed, but it requires adding other data sets in the analysis (e.g. supernovae or BAO data) that would constrain the evolution of these functions. This is, however, beyond the scope of this analysis.

Since  $\hat{b}_i(z)$  and  $\hat{J}(z)$  vary slowly with redshift, we can take them out of the integral in Eq. (3) and evaluate them at the mean redshift of the bin  $z_i$ . We see therefore, that by measuring the galaxy-galaxy lensing angular power spectrum for lenses at redshift  $z_i$ , we can directly measure the functions  $\hat{J}$  and  $\hat{b}_i$  at that redshift, in a fully model-independent way, i.e. without assuming any theory of gravity, nor specifying a redshift evolution for  $\hat{J}$ . The fact that galaxy-galaxy lensing depends on  $\hat{J}$  at the redshift of the lenses follows from the fact that, in the Limber approximation, the signal is fully due to correlations between the Weyl potential and the galaxy density at the position of the lenses. All the other correlations along the photon's trajectory are negligible. The situation for the shear angular power spectrum is different, since it depends on the integral of  $\hat{J}$  from the sources to the observer. For this reason we do not consider it in our analysis. Note however that this could be done by introducing a function  $\hat{J}$  which is piece-wise constant in some chosen redshift bins, with smooth interpolation between them.

Eq. (3) shows that there is a degeneracy between  $\hat{J}$  and  $\hat{b}_i$ , but this degeneracy can be easily broken by adding the clustering angular power spectrum which only depends on  $\hat{b}_i$ . Following<sup>39</sup>, we use the Limber approximation to compute the galaxy clustering spectrum at large  $\ell$

$$C_\ell^{\Delta\Delta}(z_i, z_j) = \int dz n_i(z) n_j(z) \frac{\mathcal{H}(z)(1+z)}{\chi^2(z)} \hat{b}_i(z) \hat{b}_j(z) B(k_\ell, \chi) \frac{P_{\delta\delta}^{\text{lin}}(k_\ell, z^*)}{\sigma_8^2(z^*)}, \quad \text{for } \ell \geq 200. \quad (6)$$

For low  $\ell$ , however, the Limber approximation is not accurate enough and we need to go beyond it. Following ref. 40, we split the non-linear matter power spectrum into its linear contribution and the difference between the non-linear and linear contributions:  $P_{\delta\delta}^{\text{nl}} = P_{\delta\delta}^{\text{lin}} + (P_{\delta\delta}^{\text{nl}} - P_{\delta\delta}^{\text{lin}})$ . For the purely non-linear part,  $P_{\delta\delta}^{\text{nl}} - P_{\delta\delta}^{\text{lin}}$ , the Limber approximation can be used, since non-linearities will only contribute at small scales (large  $k$ ) where the Limber approximation is valid. The linear part on the other hand is computed exactly for  $\ell < 200$ . In our formalism we can write it as (this corresponds to the second term of Eq. (2.10) in ref. 40).

$$C_\ell^{\Delta\Delta}(z_i, z_j)|_{\text{lin}} = \frac{2}{\pi} \int d\chi_1 n_i(\chi_1)(1+z(\chi_1))\mathcal{H}(\chi_1)\hat{b}_i(\chi_1) \int d\chi_2 n_j(\chi_2)(1+z(\chi_2))\mathcal{H}(\chi_2)\hat{b}_j(\chi_2) \times \int_0^\infty dk k^2 \frac{P_{\delta\delta}^{\text{lin}}(k, z^*)}{\sigma_8^2(z^*)} j_\ell(k\chi_1)j_\ell(k\chi_2), \quad \text{for } \ell < 200. \quad (7)$$

Eqs. (6) and (7) are similar to Eq. (3): they depend on the density power spectrum at  $z$ , on the background evolution through  $\mathcal{H}(z)$  and  $\chi(z)$ , and on the bias of the lens sample  $\hat{b}_l$ . Combining galaxy clustering with galaxy-galaxy lensing allows us therefore to measure separately  $\hat{b}_l$  and  $\hat{f}$ . Note that in the case where GR is assumed to be correct, the measurement of  $\hat{f}$  reduces to a measurement of  $\Omega_m(z)\sigma_8(z)$ , as was proposed in ref. 41.

Eqs. (3) and (6) relate the angular power spectra to  $\hat{f}$  and  $\hat{b}_l$ . In practise, however, the baseline DES analysis considered the angular correlation functions as observables. Here we make the same choice, to directly use the data vectors measured by the Collaboration. The angular correlation functions can be related to the angular power spectra via

$$w^{ij}(\theta) = \sum_{\ell} \frac{2\ell+1}{4\pi} P_{\ell}(\cos\theta) C_{\ell}^{\Delta\Delta}(z_i, z_j), \quad (8)$$

$$\gamma_{\ell}^{ij}(\theta) = \sum_{\ell} \frac{2\ell+1}{4\pi\ell(\ell+1)} P_{\ell}^2(\cos\theta) C_{\ell}^{\Delta\kappa}(z_i, z_j), \quad (9)$$

where  $P_{\ell}$  and  $P_{\ell}^2$  are the associated Legendre polynomials of degree  $\ell$  and order 0 and 2, respectively.

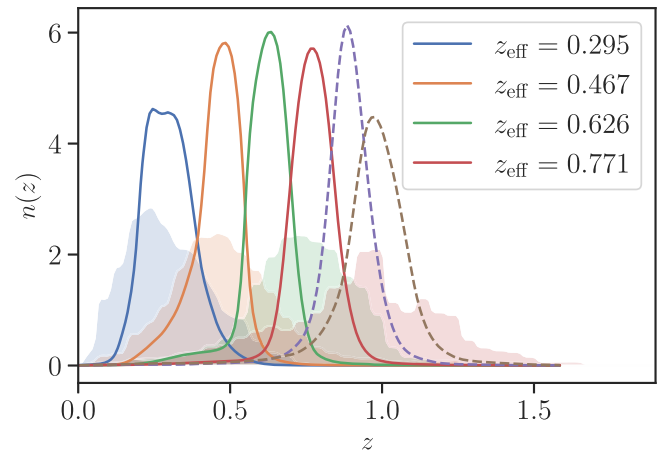
The galaxy-galaxy lensing signal (3) and the galaxy clustering signal (6) depend on the boost  $B(k, z)$  encoding the evolution of matter density fluctuations in the non-linear regime. Without specifying a theory of gravity, we cannot know the form of the boost. Here we follow the strategy presented in ref. 6 for extensions beyond GR, and assume that the boost is the same as in GR. We then restrict the range of angular scales used in the analysis to minimise the impact of the boost. We consider two scenarios: one where we use all scales that were used in the standard DES analysis<sup>14</sup>, and another one where we impose a more stringent scale cut similar to the one used in ref. 6. These two choices are described in detail in the next section. Note that even though choosing a boost as in GR is obviously not a model-independent treatment, it does not invalidate our analysis. Suppose that we observe  $\hat{f}$  different from GR using our formalism. This indicates without ambiguity that GR is invalid (if GR would be valid, the boost would be correct). One can then compare the measured  $\hat{f}$  with modified gravity models and find those that are compatible with the measurements. Within these models, one can then rerun the analysis, using now the correct boost in those models to refine the constraints (which we do not expect to change by much, given that we have chosen stringent angular scale cuts). Of course, if we are in a theory of gravity where the boost is anti-correlated with  $\hat{f}$ , using a GR boost will partially hide the deviations in  $\hat{f}$ , in the regime where non-linear effects are relevant. This is unavoidable and is simply due to the fact that lensing mixes large scales and small scales leading to possible cancellations between opposite effects. Such a situation would however have a very specific signature: the measured values of  $\hat{f}$  would deviate more and more from GR when more stringent scale cuts are used.

Once  $\hat{f}$  has been measured, it can be compared with predictions in any theory of gravity of interest and it can be used to constrain the parameters of these theories. In particular, assuming the validity of the  $\Lambda$ CDM model,  $\hat{f}$  can be used to measure  $\sigma_8(z=0)$ , since in this case it reduces to

$$\hat{f}(z) = \Omega_m(z) \frac{D_1(z)}{D_1(z=0)} \sigma_8(z=0). \quad (10)$$

### Measurement of the Weyl potential

To measure the Weyl potential with the DES observations, we follow closely the baseline analysis for galaxy clustering and galaxy-galaxy lensing presented in ref. 39, which is also called  $2 \times 2$ pt analysis. This



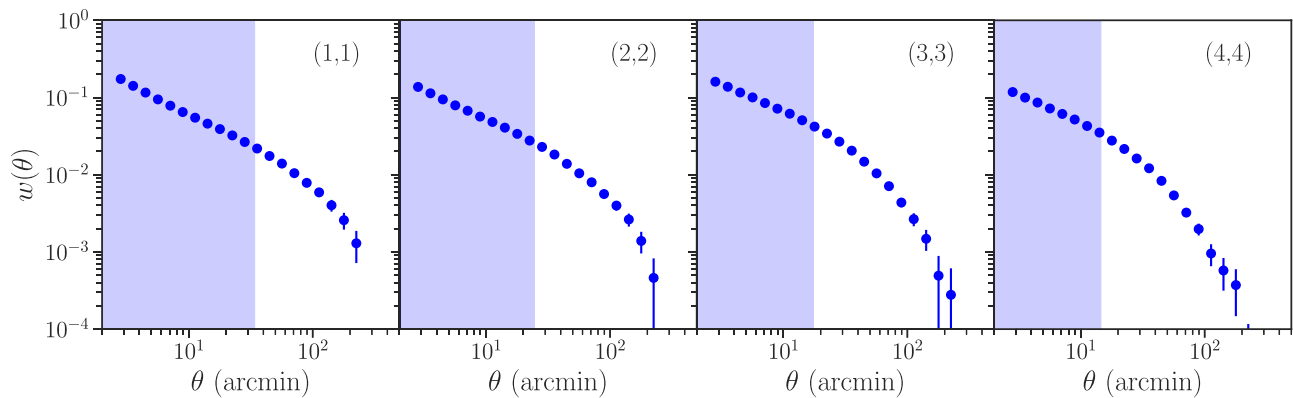
**Fig. 5 | The solid and dashed line show the galaxy (lenses) distribution of the 6 MAGLIM tomographic bins as a function of redshift. Only the first 4 bins (solid lines) have been considered in this analysis. The effective redshift of the lenses, computed as the mean of each distribution, is also provided for information. The shaded regions show the distribution of sources. Source data are provided as a Source Data file.**

includes the data sets presented in that analysis, as well as the modelling of systematic effects and the modelling of the cosmological observables. In the following we describe the main differences with respect to the baseline analysis considered in ref. 39 and publicly-available in the COSMOSIS software<sup>42</sup> (<https://cosmosis.readthedocs.io/en/latest/>).

Starting with the data sets, we consider the magnitude-limited MAGLIM sample for the lenses<sup>43</sup>, and the source sample obtained with the METACALIBRATION method<sup>44</sup>. The full lens sample contains 6 tomographic bins, which would in principle allow us to determine  $\hat{f}$  at 6 different effective redshifts. However, some residual systematic uncertainties were identified in the two bins with the highest redshift and this led to the use of only the first 4 bins in the baseline  $3 \times 2$ pt analysis<sup>14</sup>. In this analysis we use a conservative approach and focus only on the first 4 tomographic bins, even if we only consider a  $2 \times 2$ pt analysis. The mean redshift of the first 4 tomographic MAGLIM bins are: [0.295, 0.467, 0.626, 0.771] and their distribution is plotted in Fig. 5.

In terms of modelling the (observational and instrumental) systematic effects, we include nuisance parameters for the width and the position of the distributions of the lenses, as well as nuisance parameters for the shift of the source distributions, see Supplementary Table 1 for a list of all nuisance parameters and Supplementary Discussion 4 for a discussion about degeneracies. We use the officially-provided informative priors on the corresponding nuisance parameters. We further consider the effect of intrinsic alignments in our galaxy-galaxy lensing observables. Although the baseline adopted in the  $3 \times 2$ pt analysis by the DES Collaboration was the tidal alignment and tidal torquing model<sup>45</sup>, we follow the approach used by the Collaboration for extended models and consider the simpler non-linear alignment model<sup>46</sup>. Given the lack of understanding of intrinsic alignments in modified gravity scenarios, we prefer to take a conservative approach and treat these effects as if we were in GR. The main reason behind this choice is to avoid artificially extracting information from modified gravity from a wrong modelling of intrinsic alignments. We further marginalise over the amplitude and redshift dependence of intrinsic alignments to avoid possible biases in our posteriors due to the intrinsic alignment assumptions.

Similar to the treatment of the intrinsic alignment of galaxies, we account for magnification effects in the lens sample assuming GR. In principle, magnification could be written in terms of the function  $\hat{f}$ , thus



**Fig. 6 | Measured galaxy clustering angular two-point correlation function as a function of angle for the first 4 tomographic bins of the MAGLIM sample in the DES Y3 analysis.** The error bars represent 1 standard deviation. The shaded region

corresponds to the measurements removed from the analysis after imposing the baseline scale cut of comoving transverse distance of 8 Mpc/h. Source data are provided as a Source Data file.

providing further constraints on it. However, since this effect is strongly subdominant, we expect the improvement to be negligible and we treat therefore magnification as a nuisance, that can be computed in GR without biasing the results. We fix the magnification bias parameters, as it was done in ref. 39. We also include the effect of RSD in the galaxy clustering observable, and as for magnification we model it in GR, which again is well justified since the effect is subdominant. Following ref. 39 we use a linear galaxy bias model and we consider broad flat priors, given the fact that  $\hat{b}_i$  now contains information relative to the growth, see Eq. (5). Finally, we account for the shear multiplicative bias using the publicly-available informative priors. Concerning the galaxy-galaxy lensing observable, we further marginalise analytically over the mass enclosed below the angular scales considered in the analysis, as was done in ref. 39. Contrary to the baseline analysis, and in order to avoid dealing with the small scales, we prefer not to include any information from shear ratios in the likelihood.

We consider the POLYCHORD<sup>47</sup> sampler (<https://github.com/PolyChord/>) to explore the parameter space with 500 live points, number of repetitions of 60, tolerance of 0.01, fast fraction of 0.1, and boost of the posteriors of 10, consistent with the baseline DES Y3 analysis<sup>14</sup>.

In order to test the dependency of our results on the different assumptions, we consider four different scenarios in this analysis. More precisely, we implement two different scale cuts: a “standard” scale cut, where only separations above 8 Mpc/h for galaxy clustering and 6 Mpc/h for galaxy-galaxy lensing are included. These roughly correspond to angular scales above 34, 24, 17, and 14 arcmin, depending on the tomographic bin, for galaxy clustering, and 25, 18, 13, and 11 arcmin for galaxy-galaxy lensing. Those are the scale cuts that were used in the DES analysis<sup>39</sup>. In addition, we also introduce a more “pessimistic” scale cut, where we further remove all separations below 21 Mpc/h for galaxy-galaxy lensing (which corresponds to angular scales below 89, 64, 46, and 38 arcmin, depending on the tomographic bin). The choice of 21 Mpc/h roughly matches the number of data points considered for galaxy-galaxy lensing in ref. 6 when testing theories of gravity beyond GR. We note that the standard scale cut for galaxy clustering is already enough to not be sensitive to the nonlinear modelling (see Table 1 in ref. 6); therefore, we consider always the standard scale cut for this probe. In Figs. 6–7 we plot the data vectors together with the two scale cuts.

In addition to these two different scale cuts, we consider two different scenarios: one where we use  $3\sigma$  Gaussian Planck priors on the cosmological parameters<sup>15</sup>; and another one where we do not use any information from Planck, i.e. we use the wide fiducial DES priors on the cosmological parameters. In this second case, the cosmological parameters will be constrained through the background evolution, the

shape of the power spectrum at  $z = 10$  (which affects the shape of the  $C_\ell$ 's at lower redshift, see Eqs. (3) and (6)), and the contaminations (RSD, intrinsic alignment and magnification). The priors in both cases are listed in Supplementary Table 1.

To analyse the different chains and visualise posteriors, we use the GETDIST package<sup>48</sup>.

### Modified theories of gravity

Since the measurements of  $\hat{J}$  do not rely on a specific theory of gravity, they can be compared with predictions of any modified theory of gravity. The only limitation is that the theory in consideration must converge to GR at high redshift,  $z = 10$ , since our measurements rely on this assumption. As illustration, we consider two cases: A) the phenomenological  $\mu - \Sigma$  extensions of gravity; and B) Horndeski theories, concentrating on a few specific cases.

**$\mu - \Sigma$  modifications of gravity.** The phenomenological parameters  $\mu$  and  $\Sigma$  encode deviations from GR in the relation between the metric potentials and the matter density fluctuations,

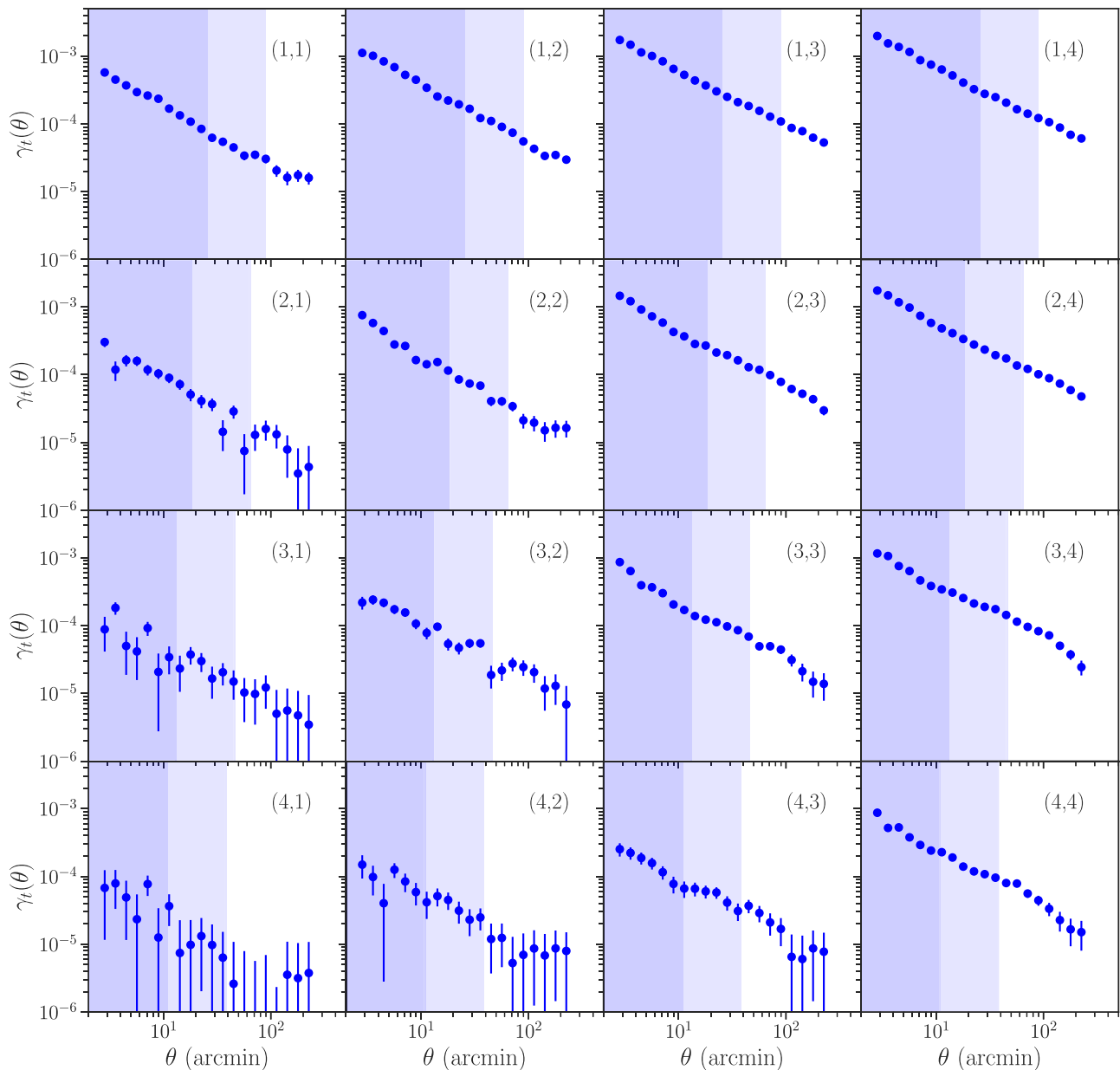
$$k^2\Psi = -4\pi\mu(a,k)Ga^2\rho\delta \quad \text{and} \quad k^2\Psi_w = -4\pi\Sigma(a,k)Ga^2\rho\delta. \quad (11)$$

In GR,  $\mu = \Sigma = 1$ . These parameters affect  $\hat{J}$  in the following way,

$$\hat{J}(z) = \Sigma(z)\Omega_m(z)D_1(z)\frac{\sigma_8(z_*)}{D_1(z_*)}, \quad (12)$$

where as before we have neglected the dependence in  $k$ , as is usually done in the literature, see e.g. ref. 6 (both due to the fact that current data are not constraining enough to probe a possible scale-dependence in  $\mu$  and  $\Sigma$ , and that in many models of modified gravity the scale-dependence is negligible for sub-horizon scales).  $\hat{J}$  depends directly on  $\Sigma$ , but it also depends on  $\mu$  through the evolution of  $D_1(z)$ . Eq. (12) underlines the difference between measuring  $\hat{J}$ , and measuring  $\mu$  and  $\Sigma$ .  $\hat{J}$  can be measured from lensing data on their own. The parameters  $\mu$  and  $\Sigma$ , on the other hand, cannot be measured separately from lensing only, since their impact on  $\hat{J}$  is fully degenerated. To determine if  $\mu$  and  $\Sigma$  are consistent with GR, it is therefore necessary to include other measurements, like RSD. Moreover, in order to constrain  $\mu$  from RSD, one has to assume the validity of Euler's equation for dark matter. The growth of structure is indeed not only affected by  $\mu$ , but also by additional forces or interactions affecting dark matter<sup>8,9</sup>. As a consequence, the constraints on  $\Sigma$  from lensing also rely on the validity of Euler's equation. In contrast,  $\hat{J}$  can be measured without any assumptions on the behaviour of dark matter. Finally, we see from Eq. (12) that  $\hat{J}$  depends directly on the redshift evolution of  $\mu$  and  $\Sigma$ , which





**Fig. 7 | Measured tangential shear (galaxy-galaxy lensing signal) as a function of angle for the first 4 tomographic bins of the MAGLIM lens samples and the 4 tomographic bins of the METACALIBRATION sources in the DES Y3 analysis.** Each tuple  $(i, j)$  represents the combination of the  $i^{\text{th}}$  tomographic bin for the lenses and the  $j^{\text{th}}$  tomographic bin for the sources. The error bars represent 1 standard

deviation. The darker shaded region corresponds to the measurements removed from the analysis after imposing the baseline scale cut of comoving transverse distance of 6 Mpc/h. The lighter shaded region represents the more stringent scale cuts of 21 Mpc/h also considered in this work. Source data are provided as a Source Data file.

is unknown. In particular,  $\hat{J}$  depends on  $\mu(z)$  through a second order evolution equation, which requires knowledge of  $\mu(z)$  not only at the redshift of the analysis, but also at all redshifts above it.

In practice, we find that  $\hat{J}$  is much more sensitive to  $\Sigma$  than to  $\mu$ . This is linked to the fact that  $\hat{J}$  is directly proportional to  $\Sigma$  (see Eq. (12)), whereas it is affected by  $\mu$  only through the evolution of  $D_1(z)$ . Here, we aim to investigate the constraining power of  $\hat{J}$ , without the inclusion of additional RSD data. Therefore, as illustration, we choose  $\mu=1$  (as in GR) and we infer  $\Sigma$  from  $\hat{J}$ . We consider 3 different choices of time evolution, encoded through

$$\Sigma(z) = 1 + \Sigma_0 g(z), \quad (13)$$

with 1) the standard evolution that has been used in the DES analysis<sup>6</sup> and in Planck analysis<sup>15</sup>:  $g(z) = \Omega_A(z)$ ; 2) no evolution:  $g(z)=1$  for  $z \in [0, 1]$  and 0 elsewhere; and 3) an exponential evolution:  $g(z) = \exp(1+z)$  for  $z \in [0, 1]$  and 0 elsewhere. The first case is motivated by the fact that deviations from GR are linked to the accelerated expansion of the Universe, and are therefore expected to decay proportionally to the amount of dark energy. The second and third cases are not physically motivated time evolutions, but they allow us to explore the sensitivity of the data to various behaviours, drastically different from each other.

At each point of the chain, we fit for  $\Sigma_0$  using Eq. (12), combining the four measurements of  $\hat{J}$ . The value  $\sigma_8(z)$  is computed at each point from the cosmological parameters using CAMB<sup>49</sup>.

**Horndeski theories and a non-minimal coupling of CDM.** Besides these phenomenological modifications of GR, we also consider Horndeski models<sup>3</sup>, which constitute the most general class of Lorentz-invariant scalar-tensor theories with second-order equations of motion. In the regime of linear cosmological perturbations, this broad class of theories can be expressed in an effective theory approach by a limited number of parameters<sup>50</sup>, including the running of the Planck mass  $\alpha_M$  and the braiding  $\alpha_B$ . Following the approach of refs. 32,37,51, we additionally allow for dark matter to be non-minimally coupled to the scalar field, since the weak equivalence principle has not been tested for dark matter. While theoretically easier to interpret, since the modifications enter directly at the level of the Lagrangian, Horndeski models with a non-minimal coupling of CDM add more complexity compared to the phenomenological  $\mu - \Sigma$  modifications. In particular, they influence the background evolution of  $\mathcal{H}(z)$  along with the perturbations. As previously mentioned, relaxing the assumption of the background evolution of  $\Lambda$ CDM would necessitate the inclusion of further data sets, which is beyond the scope of this work. Instead, for simplicity we neglect the impact on the background evolution and only constrain the parameters through their impact on  $\hat{J}$ . As such, our constraints should be seen as an illustration of the capability of this model type to explain a  $\hat{J}$  measurement different from the  $\Lambda$ CDM prediction, rather than a full analysis of the models.

We focus on two specific cases: 1) the case where  $\alpha_B = \alpha_M/2$ , and dark matter is minimally coupled to the scalar field, i.e.  $\gamma_c = 0$ ; and 2) the case where  $\gamma_c \neq 0$ , i.e. there is a non-minimal coupling of dark matter, and all other parameters are kept to their GR value. As before, we need a time evolution for the parameters, which we choose to be  $g(z) = \Omega_\Lambda(z)$  as usually considered for Horndeski models. For both cases, we consider the baseline analysis, with Planck priors and standard scale cuts.

To find  $\alpha_{M,0}$  and  $\gamma_{c,0}$  we can use Eq. (12), which is also valid in Horndeski theories. For each value of the parameters, we can compute  $\Sigma(z_i) = \mu(z_i)(1 + \eta(z_i))$ , using the relations between the model parameters ( $\alpha_M, \alpha_B, \gamma_c$ ) and the phenomenological parameters ( $\mu, \eta$ ) stated in Appendix A of ref. 32. We then use the EF-TIGRE code (<https://github.com/Mik3M4n/EF-TIGRE>) developed in ref. 32 to obtain  $\Omega_m(z_i)$  and  $D_1(z_i)/D_1(z_*)$ . In addition to their dependence on  $\alpha_{M,0}$  and  $\gamma_{c,0}$ , these quantities depend also on the primordial cosmological parameters. However, since it is computationally expensive to solve the system of differential equations at each point of the chain, we solve it instead for the mean values of the primordial cosmological parameters obtained from our analysis. Similarly, we compute the mean value of  $\sigma_8(z_*)$  using CAMB. With this, we can compute  $\hat{J}(z_i)$  using Eq. (12). We then fit for  $\alpha_{M,0}$ , respectively for  $\gamma_{c,0}$ , using the mean values for the four measured  $\hat{J}$  and their covariance from our baseline analysis. Since we do not fit for  $\alpha_{M,0}$  and  $\gamma_{c,0}$  at each point of the chain, we do not have posteriors for these quantities. We can, however, compute the  $1\sigma$  error bars by propagating the errors on the primordial parameters. In principle, the way in which the errors on  $\Omega_m$  and  $\Omega_b$  at initial epoch propagate into the errors on  $\Omega_m(z_i)$  and  $D_1(z_i)/D_1(z_*)$  depends on the values of the modified gravity parameters. For simplicity, we assume that they are the same as in  $\Lambda$ CDM.

## Data availability

All the DES data used in this analysis have been produced and publicly released by the DES Collaboration<sup>14</sup>. It is accessible through the COSMOSIS software<sup>42</sup> in <https://github.com/joezuntz/cosmosis-standard-library/tree/main/likelihood/des-y3>. It can also be found in <https://des.ncsa.illinois.edu/releases/y3a2/Y3key-products>. Source data are provided with this paper. The chains are available in [https://github.com/itutaus/cosmosis\\_weyl](https://github.com/itutaus/cosmosis_weyl)<sup>52</sup>. The other data obtained in the analysis are

available from the corresponding author on request. Source data are provided with this paper.

## Code availability

The code used in this analysis is a modified version of COSMOSIS that is accessible in [https://github.com/itutaus/cosmosis\\_weyl](https://github.com/itutaus/cosmosis_weyl)<sup>52</sup>. This analysis also made use of the public codes: CAMB<sup>49</sup>, POLYCHORD<sup>47</sup><https://github.com/PolyChord/>, GETDIST<sup>48</sup>, and EF-TIGRE<sup>32</sup><https://github.com/Mik3M4n/EF-TIGRE>.

## References

- Supernova Search Team. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *Astron. J.* **116**, 1009 (1998).
- Supernova Cosmology Project. Measurements of  $\Omega$  and  $\Lambda$  from 42 high redshift supernovae. *Astrophys. J.* **517**, 565 (1999).
- Horndeski, G. W. Second-order scalar-tensor field equations in a four-dimensional space. *Int. J. Theor. Phys.* **10**, 363 (1974).
- Noller, J. & Nicola, A. Cosmological parameter constraints for Horndeski scalar-tensor gravity. *Phys. Rev. D.* **99**, 103502 (2019).
- Koyama, K. Cosmological tests of modified gravity. *Rept. Prog. Phys.* **79**, 046902 (2016).
- DES collaboration. Dark energy survey year 3 results: constraints on extensions to  $\Lambda$ CDM with weak lensing and galaxy clustering. *Phys. Rev. D.* **107**, 083504 (2023).
- Pogosian, L. et al. Imprints of cosmological tensions in reconstructed gravity. *Nat. Astron.* **6**, 1484 (2022).
- Castello, S., Grimm, N. & Bonvin, C. Rescuing constraints on modified gravity using gravitational redshift in large-scale structure. *Phys. Rev. D.* **106**, 083511 (2022).
- Bonvin, C. & Pogosian, L. Modified Einstein versus modified Euler for dark matter. *Nat. Astron.* **7**, 1127 (2023).
- Song, Y.-S. & Percival, W. J. Reconstructing the history of structure formation using redshift distortions. *JCAP* **10**, 004 (2009).
- Blake, C. et al. The WiggleZ dark energy survey: the growth rate of cosmic structure since redshift  $z=0.9$ . *Mon. Not. R. Astron. Soc.* **415**, 2876 (2011).
- eBOSS collaboration. Completed SDSS-IV extended Baryon oscillation spectroscopic survey: cosmological implications from two decades of spectroscopic surveys at the apache point observatory. *Phys. Rev. D.* **103**, 083533 (2021).
- Tutusa, I., Sobral-Blanco, D. & Bonvin, C. Combining gravitational lensing and gravitational redshift to measure the anisotropic stress with future galaxy surveys. *Phys. Rev. D* **107** <https://doi.org/10.1103/physrevd.107.083526> (2023).
- DES collaboration. Dark Energy Survey Year 3 results: cosmological constraints from galaxy clustering and weak lensing. *Phys. Rev. D.* **105**, 023520 (2022).
- Planck, N. et al. Planck 2018 results. VI. Cosmological parameters. *Astron. Astrophys.* **641**, A6 (2020).
- García-García, C. et al. The growth of density perturbations in the last -10 billion years from tomographic large-scale structure data. *JCAP* **10**, 030 (2021).
- Rozo, D. E. et al. redMaGiC: selecting luminous red galaxies from the DES science verification data. *Mon. Not. Roy. Astron. Soc.* **461**, 1431 (2016).
- Marques, G. A. et al. Cosmological constraints from the tomography of DES-Y3 galaxies with CMB lensing from ACT DR4. *JCAP* **01**, 033 (2024).
- Chen, S.-F., White, M., DeRose, J. & Kokron, N. Cosmological analysis of three-dimensional BOSS galaxy clustering and Planck CMB lensing cross correlations via Lagrangian perturbation theory. *JCAP* **07**, 041 (2022).

20. White, M. et al. Cosmological constraints from the tomographic cross-correlation of DESI luminous red galaxies and planck CMB lensing. *JCAP* **02**, 007 (2022).
21. Alonso, D. et al. Constraining cosmology with the Gaia-unWISE quasar catalog and CMB lensing: structure growth. *JCAP* **11**, 043 (2023).
22. Piccirilli, G. et al. Growth history and quasar bias evolution at  $z < 3$  from Quaia, <http://arxiv.org/abs/2402.05761>
23. Madhavacheril, M. S. et al. The atacama cosmology telescope: DR6 gravitational lensing map and cosmological parameters. *Astrophys. J.* **962**, 113 (2024).
24. Esposito, M. et al. Weighing cosmic structures with clusters of galaxies and the intergalactic medium. *Mon. Not. Roy. Astron. Soc.* **515**, 857 (2022).
25. Grimm, N., Bonvin, C. & Tutusaus, I. New measurements of  $E_G$ : testing general relativity with the Weyl potential and galaxy velocities, <http://arxiv.org/abs/2403.13709>
26. Castello, S. et al. Gravitational redshift constraints on the effective theory of interacting dark energy. *JCAP* **05**, 003 (2024).
27. Brans, C. & Dicke, R. H. Mach's principle and a relativistic theory of gravitation. *Phys. Rev.* **124**, 925 (1961).
28. Sawicki, I., Saltas, I. D., Amendola, L. & Kunz, M. Consistent perturbations in an imperfect fluid. *JCAP* **01**, 004 (2013).
29. Carroll, S. M., Duvvuri, V., Trodden, M. & Turner, M. S. Is cosmic speed - up due to new gravitational physics? *Phys. Rev. D.* **70**, 043528 (2004).
30. Song, Y.-S., Hu, W. & Sawicki, I. The Large Scale Structure of  $f(R)$  Gravity. *Phys. Rev. D.* **75**, 044004 (2007).
31. Carroll, S. M., Sawicki, I., Silvestri, A. & Trodden, M. Modified-Source Gravity and Cosmological Structure Formation. *N. J. Phys.* **8**, 323 (2006).
32. Vollick, D. N.  $1/R$  Curvature corrections as the source of the cosmological acceleration. *Phys. Rev. D.* **68**, 063510 (2003).
33. Kilo-Degree Survey. Dark Energy Survey DES Y3 + KiDS-1000: consistent cosmology combining cosmic shear surveys. *Open J. Astrophys.* **6**, 2305.17173 (2023).
34. Amendola, L., Kunz, M., Motta, M., Saltas, I. D. & Sawicki, I. Observables and unobservables in dark energy cosmologies. *Phys. Rev. D.* **87**, 023501 (2013).
35. Amendola, L., Fogli, S., Guarnizo, A., Kunz, M. & Vollmer, A. Model-independent constraints on the cosmological anisotropic stress. *Phys. Rev. D.* **89**, 063538 (2014).
36. Euclid. Euclid preparation: VII. forecast validation for Euclid cosmological probes. *Astron. Astrophys.* **642**, A191 (2020).
37. Gleyzes, J., Langlois, D., Mancarella, M. & Vernizzi, F. Effective theory of dark energy at redshift survey scales. *J. Cosmol. Astropart. Phys.* **02**, 056 (2016).
38. Raveri, M. et al. Principal reconstructed modes of dark energy and gravity. *J. Cosmol. Astropart. Phys.* **02**, 061 (2023).
39. Porredon, A. et al. Dark Energy Survey Year 3 results: cosmological constraints from galaxy clustering and galaxy-galaxy lensing using the MagLim lens sample. *Phys. Rev. D.* **106**, 103530 (2022).
40. Fang, X., Krause, E., Eifler, T. & MacCrann, N. Beyond limber: efficient computation of angular power spectra for galaxy clustering and weak lensing. *JCAP* **05**, 010 (2020).
41. Camera, S. A novel method for unbiased measurements of growth with cosmic shear, <http://arxiv.org/abs/2206.03499>
42. Zuntz, J. et al. CosmoSIS: modular cosmological parameter estimation. *Astron. Comput.* **12**, 45 (2015).
43. Porredon, A. et al. Dark Energy Survey Year 3 results: optimizing the lens sample in a combined galaxy clustering and galaxy-galaxy lensing analysis. *Phys. Rev. D.* **103**, 043503 (2021).
44. Sheldon, E. S. & Huff, E. M. Practical weak lensing shear measurement with metacalibration. *Astrophys. J.* **841**, 24 (2017).
45. Blazek, J., MacCrann, N., Troxel, M. A. & Fang, X. Beyond linear galaxy alignments. *Phys. Rev. D.* **100**, 103506 (2019).
46. Bridle, S. & King, L. Dark energy constraints from cosmic shear power spectra: impact of intrinsic alignments on photometric redshift requirements. *N. J. Phys.* **9**, 444 (2007).
47. Handley, W. J., Hobson, M. P. & Lasenby, A. N. PolyChord: nested sampling for cosmology. *Mon. Not. Roy. Astron. Soc.* **450**, L61 (2015).
48. Lewis, A. GetDist: a Python package for analysing Monte Carlo samples, <http://arxiv.org/abs/1910.13970>
49. Lewis, A., Challinor, A. & Lasenby, A. Efficient computation of CMB anisotropies in closed FRW models. *Astrophys. J.* **538**, 473 (2000).
50. Bellini, E. & Sawicki, I. Maximal freedom at minimum cost: linear large-scale structure in general modifications of gravity. *J. Cosmol. Astropart. Phys.* **07**, 050 (2014).
51. Gleyzes, J., Langlois, D., Mancarella, M. & Vernizzi, F. Effective theory of interacting dark energy. *J. Cosmol. Astropart. Phys.* **08**, 054 (2015).
52. Tutusaus, I., Bonvin, C. & Grimm, N. Measurement of the Weyl potential evolution from the first three years of dark energy survey data, <https://doi.org/10.5281/zenodo.13734927> (2024).

## Acknowledgements

We thank Ruth Durrer, Agnès Ferté and Pablo Fosalba for comments and discussions. We acknowledge the use of the HPC cluster Yggdrasil at the University of Geneva for conducting our analysis. C.B. and N.G. acknowledge support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 863929; project title "Testing the law of gravity with novel large-scale structure observables"). C.B. acknowledges support from the Swiss National Science Foundation.

## Author contributions

C.B. conceived the method, developed the parameterization of the Weyl potential, and led the project. I.T. modified the public code COSMOSIS to measure the Weyl potential and ran different analyses on DES data. I.T., C.B., and N.G. analyzed and interpreted the results. C.B. used the measurement to constrain phenomenological modifications of gravity and N.G. to constrain Horndeski's theories. I.T., C.B., and N.G. wrote the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-53363-6>.

**Correspondence** and requests for materials should be addressed to Isaac Tutusaus, Camille Bonvin or Nastassia Grimm.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024