



HAL
open science

Higher-order statistics of shear field via a machine learning approach

Carolina Parroni, Édouard Tollet, Vincenzo F. Cardone, Roberto Maoli,
Roberto Scaramella

► **To cite this version:**

Carolina Parroni, Édouard Tollet, Vincenzo F. Cardone, Roberto Maoli, Roberto Scaramella. Higher-order statistics of shear field via a machine learning approach. *Astronomy and Astrophysics - A&A*, 2021, 645, pp.A123. 10.1051/0004-6361/202038715 . hal-03086774

HAL Id: hal-03086774

<https://hal.science/hal-03086774v1>

Submitted on 10 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Higher-order statistics of shear field via a machine learning approach

Carolina Parroni¹, Édouard Tollet^{2,3,4}, Vincenzo F. Cardone^{1,5}, Roberto Maoli⁶, and Roberto Scaramella^{1,5}

¹ I.N.A.F. – Osservatorio Astronomico di Roma, via Frascati 33, 00040 Monte Porzio Catone (Roma), Italy
e-mail: carolina.parroni@inaf.it

² Observatoire de Paris, GEPI and LERMA, PSL University, 61 Avenue de l'Observatoire, 75014 Paris, France

³ Université Paris Diderot, Sorbonne Paris Cité, 75013 Paris, France

⁴ Centre de Recherche Astrophysique de Lyon UMR5574, ENS de Lyon, Univ. Lyon1, CNRS, Université de Lyon, 69007 Lyon, France

⁵ Istituto Nazionale di Fisica Nucleare – Sezione di Roma 1, Piazzale Aldo Moro, 00185 Roma, Italy

⁶ Dipartimento di Fisica, Università di Roma “La Sapienza”, Piazzale Aldo Moro, 00185 Roma, Italy

Received 22 June 2020 / Accepted 11 November 2020

ABSTRACT

Context. The unprecedented amount and the excellent quality of lensing data expected from upcoming ground and space-based surveys present a great opportunity for shedding light on questions that remain unanswered with regard to our universe and the validity of the standard Λ CDM cosmological model. The development of new techniques that are capable of exploiting the vast quantity of data provided by future observations, in the most effective way possible, is of great importance.

Aims. This is the reason we chose to investigate the development of a new method for treating weak-lensing higher-order statistics, which are known to break the degeneracy among cosmological parameters thanks to their capacity to probe non-Gaussian properties of the shear field. In particular, the proposed method applies directly to the observed quantity, namely, the noisy galaxy ellipticity.

Methods. We produced simulated lensing maps with different sets of cosmological parameters and used them to measure higher-order moments, Minkowski functionals, Betti numbers, and other statistics related to graph theory. This allowed us to construct datasets with a range of sizes, levels of precision, and smoothing. We then applied several machine learning algorithms to determine which method best predicts the actual cosmological parameters associated with each simulation.

Results. The most optimal model turned out to be a simple multidimensional linear regression. We use this model to compare the results coming from the different datasets and find that we can measure, with a good level of accuracy, the majority of the parameters considered in this study. We also investigated the relation between each higher-order estimator and the different cosmological parameters for several signal-to-noise thresholds and redshifts bins.

Conclusions. Given the promising results we obtained, we consider this approach a valuable resource that is worthy of further development.

Key words. gravitational lensing: weak – cosmology: theory – methods: statistical

1. Introduction

Over the past several decades, the availability of multi-band astronomical data of increasingly greater quality has led to impressive progress in the field of observational cosmology and resulted in the establishment of the concordance Λ cold dark matter (Λ CDM) model. Its parameters (which specify the contribution of matter and the cosmological constant to the energy budget, its expansion rate, and growth of structures) have been measured with an unprecedented level of precision through the joint application of various cosmological probes, such as the angular anisotropy of the cosmic microwave background (CMB), the baryon acoustic oscillation (BAO), galaxy clustering (GC), and weak lensing (WL) studies (e.g., [Dunkley et al. 2009](#); [Planck Collaboration XVI 2014](#); [Planck Collaboration XIII 2016](#); [Planck Collaboration VI 2020](#); [Alam et al. 2017](#); [DES Collaboration 2018](#)).

In particular, the [Planck Collaboration VI \(2020, hereafter PC18\)](#) combined measurements of CMB polarization, temperature, and lensing with the BAO and type Ia supernova (SN) data to obtain the tightest possible constraints on the cosmological parameters. The constraints on PC18 are in good agreement with

different BAO, SN, and some galaxy lensing observations, but they show a slight tension with the [DES Collaboration \(2018\)](#) results, which were obtained with GC and WL data. These constraints also present a more significant tension with the local measurements of the Hubble constant ([Riess et al. 2018](#)). Considering the precision of these measurements and the accuracy with which those studies were performed, we could speculate that the observed tensions may be linked to new physics or to phenomena that are not accounted for in the standard cosmological model – more so than to systematic errors. In fact, despite the successful results that were obtained, it is important to stress that the nature of the main energy contents of the universe predicted by the Λ CDM model (i.e., dark energy that drives cosmic speed up and dark matter that is responsible for the formation of large-scale structures) is still unknown.

These open questions can be explored by ongoing surveys, such as the [DES Collaboration \(2005\)](#), the Hyper Suprime-Cam (HSC, [Aihara et al. 2017](#)), and the Kilo-Degree Survey (KiDS, [de Jong et al. 2012](#)), as well as next-generation surveys from the ground, such as the Legacy Survey of Space and Time (LSST, [Abell 2009](#)), or space-based, such as ESA's *Euclid* ([Laureijs et al. 2011](#)) and NASA's Wide Field Infrared Survey

Telescope (WFIRST, Green et al. 2012) mission. The *Euclid* survey, notably, will collect both imaging and spectroscopic datasets using WL and galaxy clustering as its primary probes to constrain, with unprecedented precision, the dark energy equation of state, measure the rate of cosmic structure growth to discriminate between general relativity against and modified gravity and to look for deviations from Gaussianity of initial density perturbations to test inflationary scenarios. In particular, *Euclid* will obtain high-quality data on sub-arcsec scale of galaxy shape measurements for galaxies up to $z \geq 2$, covering 15 000 deg² of the extragalactic sky.

Although the density field on large scales is well-approximated by a Gaussian distribution, the information brought on by the measurement of non-Gaussianity on small scales can help to break degeneracies and further constrain the cosmological parameters. With regard to accessing this information, WL is considered to be one of the best tools. As predicted by general relativity (and any metric theory of gravity), the matter distribution along the line of sight deflects the light rays because they propagate along the geodesic lines, causing a distortion of the image of the emitting sources. In the WL regime, this effect is too small to be detected on single galaxies and a statistical approach is needed to access the information contained in the cosmic shear field.

Given its sensitivity to the background expansion and to the growth of structures, second-order lensing statistics have been employed with remarkable success in the past, via the analysis of the two-point correlation function and its Fourier counterpart, namely, the power spectrum (see, e.g., Munshi et al. 2008; Kilbinger 2015; Bartelmann & Maturi 2017; Köhlinger et al. 2017; Hildebrandt et al. 2017; Troxel et al. 2018; Hikage et al. 2019; Hamana et al. 2020, and references therein). In order to access the non-Gaussian information originating from the non-linear collapse of the primordial density fluctuations, however, it is necessary to go to a higher order in the statistical description of the shear field. Along with the more traditional three- and four-point correlation functions, and the corresponding bi- and tri-spectra in Fourier space, (e.g., Takada & Jain 2003, 2004; Semboloni et al. 2011; Fu et al. 2014), various estimators have been more recently used. Topological descriptors such as Minkowski functionals and Betti numbers have been applied to lensing convergence maps (Matsubara & Jain 2001; Sato et al. 2001; Taruya et al. 2002; Matsubara 2010; Kratochvil et al. 2011; Pratten & Munshi 2012; Petri et al. 2013; Shirasaki & Yoshida 2014; Ling et al. 2015; Vicinanza et al. 2019; Marques et al. 2019; Mawdsley et al. 2020; Parroni et al. 2020; Zürcher et al. 2021), and on three-dimensional Gaussian random fields to study the topology of the primordial density field (Park et al. 2013; Pranav et al. 2017, 2019), respectively. Moreover Hong et al. (2020) applied graph theory estimators to study the topological structure of clustering in N -body simulations corresponding to different cosmological models.

In the context of WL studies, the topological higher order estimators have usually been applied to lensing convergence maps. However, the convergence is not a direct observable, so to actually solve an inversion problem, it is necessary to start from the shear data. Even though several methods have already been conceived and a lot of progress has been made (e.g., Pires et al. 2009, 2020; Jullo et al. 2013; Jeffrey et al. 2018; Price et al. 2020, 2021), this reconstruction is still considered as a non-trivial problem that requires very accurate control of systematic effects coming from survey masking, borders, noise, and the fact that what we are actually observing is the galaxy ellipticity, which is a measure of the noisy reduced shear and not the shear itself. In order to circumvent this reconstruction

problem, higher-order statistics could be applied directly to ellipticity maps but the issue in this case would be the lack of theoretical predictions that the measurements could be compared to. Even in the cases where a theoretical study has indeed been carried out, as for the higher order moments and Minkowski functionals of convergence maps, we have to take into account the approximated calculations due to the challenge of modeling nonlinearities in the matter power spectrum and bispectrum. The mismatch between theoretical expected values and actual noisy observations can be dealt with through a calibration process performed on simulations (e.g., Vicinanza et al. 2018, 2019, Parroni et al. 2020) but this requires the creation of an appropriate parametrization, which is nevertheless a non-trivial approximation, adding nuisance parameters and possible degeneracy with the cosmological ones, hence weakening the constraints.

It is for these reasons that we chose to apply higher-order moments, Minkowski functionals, Betti numbers, and several statistics from graph theory to simulated noisy ellipticity maps, and to use different machine learning techniques to study the relation between those estimators and the cosmological parameters that were used to generate the simulations. In recent years, machine learning has proven to be a valuable tool in a variety of astrophysical studies and notably in some WL applications, such as the discrimination between different modified gravity cosmologies (e.g., Merten et al. 2019; Peel et al. 2019), the measurement of the (Ω_M, σ_8) parameters degeneracy (e.g., Gupta et al. 2018; Fluri et al. 2018, 2019), and mass maps reconstruction (e.g., Jeffrey et al. 2020). In our case, it allows us to bypass the theory issues discussed above and allows us to make direct use of noisy ellipticity maps, based upon which we calculated new and promising higher order estimators.

The paper is organized as follows. In Sect. 2, we describe how we obtained the simulated shear maps for the different set of cosmological parameters. In Sect. 3, we introduce the different higher-order estimators that we measured from the maps and we present the final datasets that we used for the training phase. In Sect. 4, we compare the results from different models. In Sect. 5, we use the best model obtained to study the effect of the dataset size, of the measurement accuracy, and of the smoothing scale on the score, as well as performing the training and the predictions using the different datasets. In Sect. 6, we study the relation between the individual estimators and each cosmological parameter. In Sect. 7, we discuss the limitations of this work and possible improvements. In Sect. 8, we draw our conclusions. In Appendix A, we briefly outline the different machine learning methods that we compared in this study.

2. Simulation of shear maps

In order to produce the simulated shear maps, we used The Full-sky Lognormal Astro-fields Simulation Kit (FLASK; Xavier et al. 2016), which is a fast and flexible public code that takes as its input the auto- and cross-power spectra to create random realizations of different astrophysical fields that follow a multivariate lognormal distribution, reproducing the expected cross-correlations between the input fields. The choice of a multivariate lognormal distribution is motivated by the better approximation that this distribution represents of the fields that we want to simulate, compared to a multivariate Gaussian distribution (e.g., Scaramella et al. 1993; Taruya et al. 2002; Hilbert et al. 2011; Clerkin et al. 2017). This is also a simpler approximation, which is capable of conveying the non-Gaussian information contained in the shear field that we are interested in measuring. Moreover, a non-negligible aspect of this kind of

simulation is the computational speed it offers, which allows FLASK to produce full-sky realization within minutes.

While we refer to [Xavier et al. \(2016\)](#) for the details on the inner workings of FLASK, we want to draw attention to some of the limitations of this approach here. In one of the two proposed solutions, FLASK computes the shear starting from the convergence, which is, in turn, calculated via an approximated line-of-sight integration of the simulated density field. This affects the choice of the redshift range of the simulations. In fact, this approximation consists of a weighted Riemann sum over the chosen redshift bins, which is able to reproduce the theoretical spectra within 3% for $z > 0.5$. Due to the small number of bins in the sum at low redshift, the precision of the approximation degrades quickly. For this reason, we decided to cut the source catalog at $z > 0.55$. Moreover, the line-of-sight integration solution produces a convergence field that follows a distribution of a sum of correlated lognormals, which is not exactly lognormal, even if it is very similar. Although it is possible to add a shift the convergence field generated by FLASK in order to match the third-order moment, this would artificially alter the convergence probability distribution function, so that moments higher than the third would be modified in an unpredictable way. Moreover, lacking a theoretical estimate for some of the statistics we go on to consider below, there is really no way to judge whether FLASK is able to reproduce them. However, running a number of full N -body and ray-tracing simulations as large as the one we need for our study is definitely not possible with the computing resources at our disposal. We therefore prefer to rely on FLASK for this preliminary study since we want to show the potentiality of the method we are proposing rather than apply it to real data.

We chose a *Euclid*-like source redshift distribution as was done for the [Euclid Collaboration \(2019a\)](#):

$$n(z) = \frac{3n_g}{2z_0} \left(\frac{z}{z_0}\right)^2 \exp\left[-\left(\frac{z}{z_0}\right)^{3/2}\right], \quad (1)$$

with n_g the number of galaxies per arcmin², and $z_0 = z_m/\sqrt{2}$ with z_m the median redshift. We set $n_g = 30$ gal arcmin⁻² and $z_m = 0.9$ as expected for *Euclid* given a limiting magnitude $\text{mag}_{\text{lim}} = 24.5$ in the imaging VIS filter.

We used CLASS ([Blas et al. 2011](#); [Dio et al. 2013](#)) to compute the input power spectra for 25 top-hat equispaced redshift bins over the range $0.0 \leq z \leq 2.5$ for a flat Λ CDM model, varying the cosmological parameters $\{H_0, \omega_b, \Omega_M, \Omega_\Lambda, w_0, n_s, \sigma_8\}$, in each simulation. Because we wanted to compare our measurements and errors on each of those parameters with state-of-the-art results, we chose to refer to the results presented in the constraints on [PC18](#). For each simulation, we randomly extracted each parameter from a Gaussian distribution with mean μ and width σ , corresponding to the measured values and errors of the constraints on [PC18](#) parameters, respectively, which are shown in [Table 1](#). With these settings, we created a first batch of ~ 1000 simulations, followed by a second batch of ~ 500 simulations for which we doubled the width of each Gaussian in order to cover a larger area in the parameter space.

Setting $N_{\text{side}} = 2048$ and giving as input to FLASK the source redshift distribution, the mask based on the private *Euclid* Flagship galaxy mock catalog version 1.6.18, and the angular auto- and cross-power spectra calculated as described, we obtained a catalog containing the coordinates, the redshift, and the noisy ellipticity components for each galaxy, for each simulation.

As described in [Xavier et al. \(2016\)](#), the complex ellipticity $\varepsilon = \varepsilon_1 + i\varepsilon_2$ is computed as

Table 1. Mean and width of the Gaussian distributions from which we extracted the cosmological parameter values for each simulation, corresponding to the constraints on [PC18](#) results and errors.

Parameter	μ	σ
H_0	67.66	0.42
ω_b	0.02242	0.00014
Ω_M	0.3111	0.0056
Ω_Λ	$1 - \Omega_M$	0.0056
w_0	-1.028	0.032
n_s	0.9665	0.0038
σ_8	0.8102	0.0060

$$\varepsilon = \begin{cases} \frac{\varepsilon_s + g}{1 + g^* \varepsilon_s}, & |g| \leq 1 \\ \frac{1 + g \varepsilon_s^*}{\varepsilon_s^* + g^*}, & |g| > 1, \end{cases} \quad (2)$$

where $g \equiv \gamma/(1 - \kappa)$, is the reduced shear, γ is the shear, and κ is the convergence, and $\varepsilon_s = \varepsilon_{s,1} + i\varepsilon_{s,2}$ is the source intrinsic ellipticity, whose components $(\varepsilon_{s,1}, \varepsilon_{s,2})$ are randomly drawn from a Gaussian distribution with zero mean and a standard deviation σ_{ε_s} that can be set by the user. We set $\sigma_{\varepsilon_s} = 0.3$ ¹. When a map is constructed, the ellipticity components are averaged inside each pixel in order to decrease the noise given by the intrinsic ellipticity. We clarify then that what we call shear maps are therefore maps of the noisy ellipticity, which is the only WL direct observable. More precisely, the value in each pixel is $|\varepsilon| = \varepsilon_1^2 + \varepsilon_2^2$, the mean of the norm of the complex ellipticity.

Each catalog was then split into redshift bins with equal width $\Delta z = 0.05$ and centered in z from 0.5 to 1.8 in steps of 0.3. According to the *Euclid* Red Book ([Laureijs et al. 2011](#)), photometric redshifts will be measured with an error $< 0.05(1 + z)$ allowing to separate sources in bins with a center determined with an accuracy better than $0.002(1 + z)$ as confirmed by more recent analyses ([Euclid Collaboration 2019a, 2020](#); [Joshi et al. 2019](#)). As such, our choice $\Delta z = 0.05$ is well within the realistic capabilities of *Euclid*. For each slice of redshift, we first obtained 100 maps of 5×5 deg² and 300×300 pixels, leaving a gap of $\sim 1^\circ$ between two consecutive maps so that we were able to consider them as independent realizations. We performed a gnomonic projection to project the maps onto the plane of the sky, under a flat sky approximation, which holds for the size of the maps that we used.

In [Fig. 1](#), we show one of the shear maps obtained using the catalog from the simulation corresponding to the constraints on [PC18](#) parameters and redshift bin $z = 0.6$. The map was smoothed using a Gaussian filter with scale $\theta_s = 2'$ and normalized subtracting the mean and dividing by the variance. In a second moment, in order to increase the training set and the signal-to-noise ratio (S/N), we decided to perform four more realizations for each combination of the cosmological parameters already calculated, therefore making 500 maps for each simulation, bringing the total to $\sim 750\,000$ maps per redshift bin.

¹ While in the literature, noise is usually added to κ or γ directly on the map pixels assuming a Gaussian distribution with zero mean and width $\sigma_n = \sigma_{\varepsilon_s} / \sqrt{n_g A_{\text{pixel}}}$, with $\sigma_{\varepsilon_s} = 0.3$, FLASK assigns the noise, in the form of ε_s , on each individual galaxy through Eq. (2).

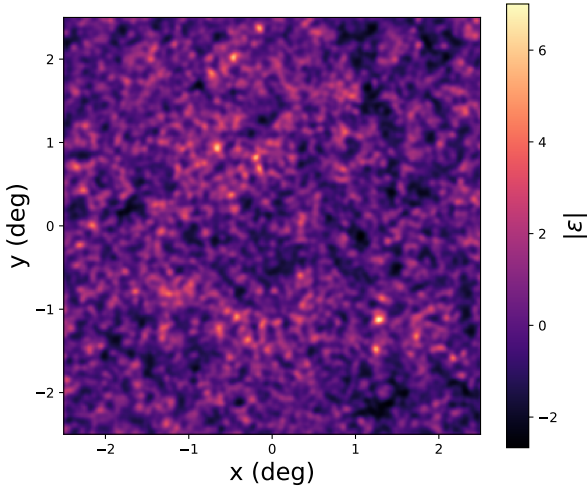


Fig. 1. One of the maps taken from the simulation corresponding to the constraints on **PC18** parameters, at $z = 0.6$, smoothed with a Gaussian filter with scale $\theta_s = 2'$ and normalized subtracting the mean and dividing by the variance.

3. Higher-order estimators

The simulated maps are the input for our investigation of the potentiality of high-order statistics to constrain cosmological parameters. Going beyond the second order opens up a wide range of possible choices, and it is not clear a priori which is the most promising one. For this reason, we considered many different alternatives, which we briefly describe in the following subsections.

3.1. Higher-order moments

Before measuring the higher order moments (HOM), we smoothed the shear maps using a Gaussian filter with scale $\theta_s = \{2', 4', 6'\}$ and subtracted the mean value to put all the maps to the same null mean value. Denoting with $\varepsilon(x, y)$ the resulting field, on each map, for all redshift bins, we estimated the third- and fourth-order centered moments, the skewness, and the kurtosis, respectively defined as:

$$\begin{aligned} k_3 &= \langle \varepsilon^3 \rangle, \\ k_4 &= \langle \varepsilon^4 \rangle, \\ S_3 &= k_3 / \langle \varepsilon^2 \rangle^{3/2}, \\ S_4 &= k_4 / \langle \varepsilon^2 \rangle^2, \end{aligned} \quad (3)$$

In Fig. 2, we show the HOM as a function of the redshift, calculated on the maps from the simulation corresponding to the constraints on **PC18** parameters, for the smoothing scale $\theta_s = 2'$. The values are averaged over 500 maps. We notice that all the moments have their minimum value at $z = 0.9$, the median redshift of the source distribution, and then they increase with the redshift. We obtained similar results but lower absolute values for the smoothing scales $\theta_s = \{4', 6'\}$. In particular, in the simulation corresponding to the constraints on **PC18** parameters, we measured a decrease in value between the measurements obtained with $\theta_s = 2'$ and $\theta_s = 6'$ of a factor ~ 40 . With the exception of k_4 , for which the minimum value shifts to higher redshift for increasing smoothing, the behavior of the HOM remains unaltered.

The two plots in the bottom, corresponding to S_3 and S_4 , are grayed out to indicate that those HOM were discarded because their scatter among different maps generated with the same cosmology was larger than the difference among maps with different cosmologies. In other words, we did not retain the features for which the ratio between the inter-simulation variance and the intra-simulation variance was smaller than one. In this case, one can not discriminate among cosmological models so that the corresponding probe is not expected to be of any help for our aims. In the following, we used the same criterion to discard several values of the other estimators.

3.2. Minkowski functionals

Given the smoothed two-dimensional shear field $\varepsilon(x, y)$ with zero mean and variance σ_0^2 , we define the excursion set Q_ν as the region where $\varepsilon/\sigma_0 > \nu$ holds for a given threshold ν . The three Minkowski functionals (MFs) are defined as:

$$\begin{aligned} V_0(\nu) &= \frac{1}{A} \int_{Q_\nu} da, \\ V_1(\nu) &= \frac{1}{4A} \int_{\partial Q_\nu} dl, \\ V_2(\nu) &= \frac{1}{2\pi A} \int_{\partial Q_\nu} dl \mathcal{K}, \end{aligned} \quad (4)$$

where A is the map area, ∂Q_ν the excursion set boundary, da and dl are the surface and line element along ∂Q_ν , and \mathcal{K} its curvature. Therefore, V_0 , V_1 , and V_2 are the area, the perimeter, and the genus characteristics of the excursion set Q_ν .

Defining $\epsilon \equiv \varepsilon/\sigma_0$, we can rewrite Eq. (4) as

$$\begin{aligned} V_0(\nu) &= \frac{1}{A} \int_A dx dy \Theta(\epsilon - \nu), \\ V_1(\nu) &= \frac{1}{4A} \int_A dx dy \delta_D(\epsilon - \nu) \sqrt{\epsilon_x^2 + \epsilon_y^2}, \\ V_2(\nu) &= \frac{1}{2\pi A} \int_A dx dy \delta_D(\epsilon - \nu) \frac{2\epsilon_x \epsilon_y \epsilon_{xy} - \epsilon_x^2 \epsilon_{yy} - \epsilon_y^2 \epsilon_{xx}}{\epsilon_x^2 + \epsilon_y^2}, \end{aligned} \quad (5)$$

where $\kappa_i = \partial\kappa/\partial x_i$, and $\kappa_{ij} = \partial^2\kappa/\partial x_i\partial x_j$ with $(i, j) = (x, y)$, are the first and second derivatives of the field.

In Fig. 3, we can see the three MFs obtained from the maps of the simulation corresponding to the constraints on **PC18** parameters averaged over 500 maps, for all redshift bins, with smoothing scale $\theta_s = 2'$, as a function of the threshold, with $\nu \in [-4, 7]$ and $\Delta\nu = 1$. The dashed black lines delimit the S/N range that was retained for the training sample, which corresponds to the range $\nu \in [-1, 1]$. The values outside of these lines, in the grayed out regions, were discarded following the variance ratio criterion previously defined. While the lines corresponding to V_0 at different redshift are almost indistinguishable, we can see that V_1 values increase with z , V_2 values for $\nu \geq 0$ have the same behavior, and V_2 values for $\nu < 0$ tend to decrease as a function of the redshift. For the smoothing scales $\theta_s = \{4', 6'\}$, we obtained qualitatively similar results but lower absolute values for V_1 and V_2 . Specifically, for the simulation corresponding to the constraints on **PC18** parameters, the measurements of V_0 with $\theta_s = 2'$ and with $\theta_s = 6'$ changed less than $\sim 5\%$, while for V_1 and V_2 we observed a decrease in value of a factor ~ 30 .

3.3. Betti numbers

The definition of Betti numbers (Betti 1870) requires the knowledge of some fundamental concepts of simplicial homology.

Higher order moments

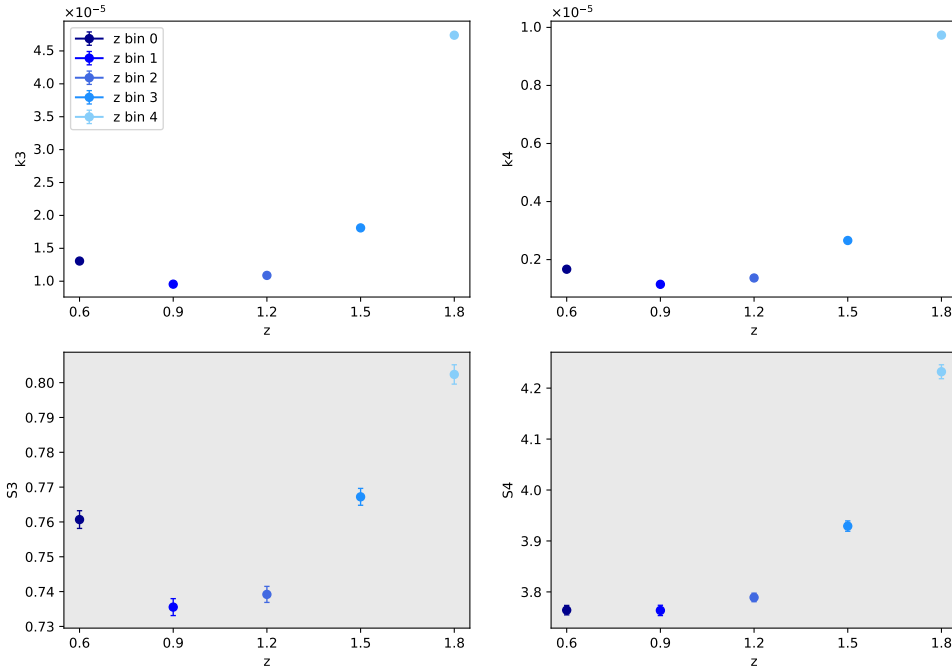


Fig. 2. HOM calculated on the maps obtained from the simulation corresponding to the constraints on **PC18** parameters, averaged over 500 maps, for all redshift bins, with smoothing scale $\theta_s = 2'$. The error bars correspond to the standard deviation divided by the square root of the number of maps. The grayed-out plots in the bottom show that the S_3 and S_4 HOM were discarded from the rest of the analysis.

Minkowski functionals

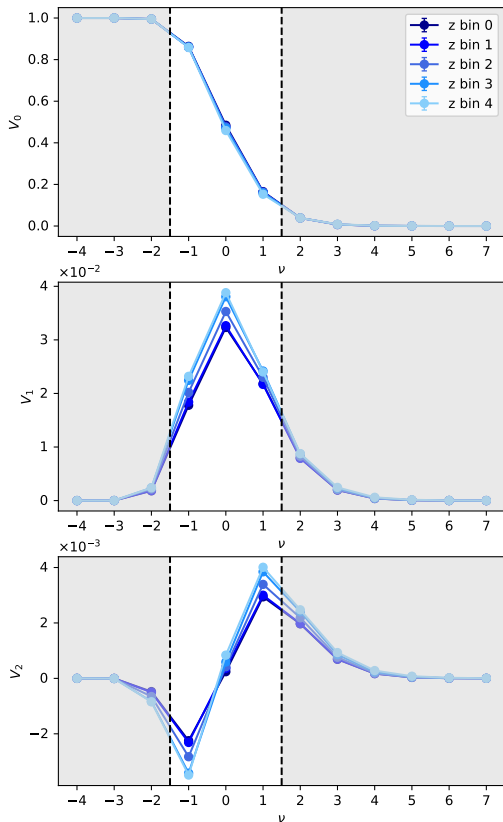


Fig. 3. MFs calculated on the maps obtained from the simulation corresponding to the constraints on **PC18** parameters, averaged over 500 maps, for all redshift bins, with smoothing scale $\theta_s = 2'$. The error bars correspond to the standard deviation divided by the square root of the number of maps. The dashed black lines contain the values selected for the training, the grayed out regions indicate the discarded values.

A proper treatment of this topic is beyond the scope of this paper, so we refer to more specific resources (e.g., [Munkres 1984](#); [Delfinado & Edelsbrunner 1993](#); [Edelsbrunner & Harer 2008](#)) for details and formal definitions.

Considering again $\epsilon(x, y)$ and the excursion set Q_ν , we define the Betti numbers in two dimensions, β_0 and β_1 , as the number of connected regions and the number of holes in the excursion set, respectively. In [Fig. 4](#), we show a working example obtained from an excursion set of a random Gaussian field on a 10×10 pixel map. We applied a Delaunay triangulation to the map and considered two points as connected if they touch each other horizontally, vertically, or diagonally, that is, if they are 8-connected. In [Fig. 4](#), the vertices are represented as numbered blue circles, the edges as pink lines, and the triangles as shaded pink regions enclosed by three edges. Every connected region is represented with a different shade of gray, and the holes are outlined in dark blue. Because we have seven different connected regions and one hole, the Betti numbers in this case will be $\beta_0 = 7$, and $\beta_1 = 1$.

In [Fig. 5](#), we show the two Betti numbers as a function of the threshold, with $\nu \in [-4, 7]$ and $\Delta\nu = 1$, estimated from the maps of the simulation with the constraints on **PC18** parameters, averaged over 500 maps, for the different redshift bins, and with smoothing scale $\theta_s = 2'$. As in [Fig. 3](#), the dashed black lines delimit the threshold range that was retained for the training, that is, the range $\nu \in [0, 2]$ for β_0 and the range $\nu \in [-2, 0]$ for β_1 , and the grayed out regions indicate the range discarded following the variance ratio criterion. We can see from the left panel that the number of connected regions for a given threshold increases with the redshift, while in the right panel, we note that the number of holes increases with z for $\nu < 0$ and it has the opposite behavior for $\nu \geq 0$. Increasing the smoothing to scales of $\theta_s = \{4', 6'\}$ does not change the behavior of the measured curves but their absolute value gets smaller. For the simulation corresponding to the constraints on **PC18** parameters, the decrease in value between the measurements obtained with $\theta_s = 2'$ and with $\theta_s = 6'$ is of a factor of ~ 10 .

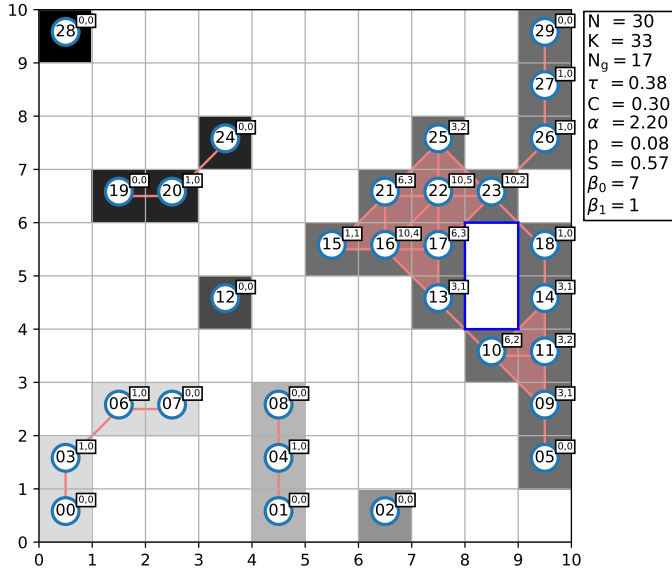


Fig. 4. Working example obtained from an excursion set of a random Gaussian field on a 10×10 pixel map, to illustrate the quantities defined in Sect. 3.3 for the Betti numbers, and in Sect. 3.4 for the graph statistics. Every numbered blue circle is a vertex, the pink lines are edges, the shaded pink regions are triangles. Different connected regions are represented with different shades of gray, and the holes are outlined in dark blue. The tag on each vertex shows the quantities $k_i(k_i - 1)/2$, and Δ_i , which are the number of connected triples, and the number of triangles, centered on the vertex. The parameters listed on the right are, in order, N the numbers of vertices, K the number of edges, N_g the number of vertices belonging to the giant component, τ the transitivity, C the LCC, α the average degree, p the edge density, S the fraction of vertices belonging to the giant component, and β_0, β_1 the Betti numbers.

3.4. Graph statistics

The simplicial complex structure defined in Sect. 3.3 can also be interpreted as a network or a graph so that some tools used in network science can be applied to it. Following Hong et al. (2020) we define the following basic graph quantities

$$\begin{aligned} \alpha &= \frac{2K}{N}, \\ p &= \frac{2K}{N(N-1)}, \\ S &= \frac{N_g}{N}, \end{aligned} \quad (6)$$

where N is the total number of vertices, N_g is the number of vertices belonging to the largest connected sub-graph in a network, called the “giant component”, and K is the total number of edges. Defining the “degree” as the number of neighbors for each vertex, we can call α the “average degree” while p is the fraction of connected edges over all pairwise combinations and it is, therefore, called the “edge density”, and S is the fraction of vertices belonging to the giant component. In Fig. 4, for example, $N = 30$ and the largest component is the structure on the right, composed by 17 vertices, so that $N_g = 17$ and $S = 0.57$. The total number of edges is $K = 33$, so that we can calculate also the average degree and the edge density, $\alpha = 2.2$ and $p = 0.08$.

Given three vertices, if they are connected by at least two edges, they are called a “connected triple”, whereas if they are connected by three edges, thereby forming a triangle, they are called a “closed triple”. With this definition, we also consider a

“closed triple” a “connected triple”. We can look at Fig. 4 to better understand this concept. For example, the vertices $\{19, 20, 24\}$ form one connected triple, centered on the vertex 20. The vertices $\{00, 03, 06, 07\}$ form two connected triples, one centered on the vertex 03, and one centered on the vertex 06. Taking as an example the triangle formed by the vertices $\{09, 10, 11\}$ and ignoring for the moment the other vertices connected to it (vertices 05, 13, and 14) we can count three triples which are closed and, therefore, also connected, $\{11, 09, 10\}$ centered on the vertex 09, $\{09, 10, 11\}$ centered on the vertex 10, and $\{10, 11, 09\}$ centered on the vertex 11. This means that a triangle always counts for three closed triples. We can then define the following quantity:

$$\tau = \frac{\text{No. of closed triples}}{\text{No. of connected triples}} = \frac{3 \times \text{No. triangles}}{\text{No. of connected triples}}, \quad (7)$$

which is referred to as “transitivity” or a “global clustering coefficient”. In Fig. 4, $\tau = 0.38$. We can also calculate the transitivity for each vertex i , a quantity referred to as the local clustering coefficient (LCC), as

$$C_i = \frac{2\Delta_i}{k_i(k_i - 1)} = \frac{\text{No. of closed triples centered on } i}{\text{No. of connected triples centered on } i}, \quad (8)$$

where k_i is the number of neighbors of the vertex i , so that $k_i(k_i - 1)/2$ is the total number of connected triples centered on the vertex, and Δ_i is the number of triangles centered on the vertex, that is, the number of closed triples centered on the vertex. In Fig. 4, every vertex has a tag with two numbers, the first corresponds to $k_i(k_i - 1)/2$, and the second is Δ_i . For example, the vertex 04 has two neighbors so that $k_{04} = 2$ and $k_{04}(k_{04} - 1)/2 = 1$, and it has no triangles centered on it so that $\Delta_{04} = 0$ and therefore $C_{04} = 0$. The vertex 16 has $k_{16} = 5$, $k_{16}(k_{16} - 1)/2 = 10$, $\Delta_{16} = 4$, so that $C_{16} = 0.4$. For vertices like 02, 12, and 28, which have no neighbors and no triangles centered on them, C_i is not defined. We define the average LCC as the averaged value of C_i over all N vertices and call it C . In Fig. 4, $C = 0.3$. We measured τ , C , α , p , and S on the graph obtained by applying a Delaunay triangulation to the $\epsilon(x, y)$ maps that we first downgraded to 100×100 pixels for computational speed reasons.

In Fig. 6, we show the graph statistics estimated from the simulated maps with the constraints on PC18 parameters, averaged over 500 maps, as a function of the threshold, with $\nu \in [-4, 7]$ and $\Delta\nu = 1$, for each redshift bin, and with smoothing scale $\theta_s = 2'$. Again, as in Figs. 3 and 5, the dashed black lines delimit the values that were selected for the training phase, and the grayed out regions correspond to the discarded values, following the variance ratio criterion. As we can see, we used the values of τ , C , α , and p in the range $\nu \in [-1, 1]$, while S was entirely discarded. We notice that τ and C have a very similar behavior, as expected considering that C is the vertex-wise version of τ , and that they both increase with ν and with z , meaning that the clustering of the structures in the maps increases both globally and locally as a function of the redshift and of the threshold. On the other hand, α , p , and S tend to decrease with the redshift, so that while the structures tend to cluster more, they also get smaller. Regarding the behavior with the threshold, while S is too noisy and p is more or less constant in the range considered, α decreases. Therefore, as expected, structures get smaller for high S/N. Again, for smoothing scales $\theta_s = \{4', 6'\}$, the qualitative behavior of the different graph statistics remains unchanged but their absolute value slightly decreases. In fact, for the simulation corresponding to the constraints on PC18 parameters, we measured a difference of less than $\sim 15\%$, between the values obtained with $\theta_s = 2'$ and with $\theta_s = 6'$.

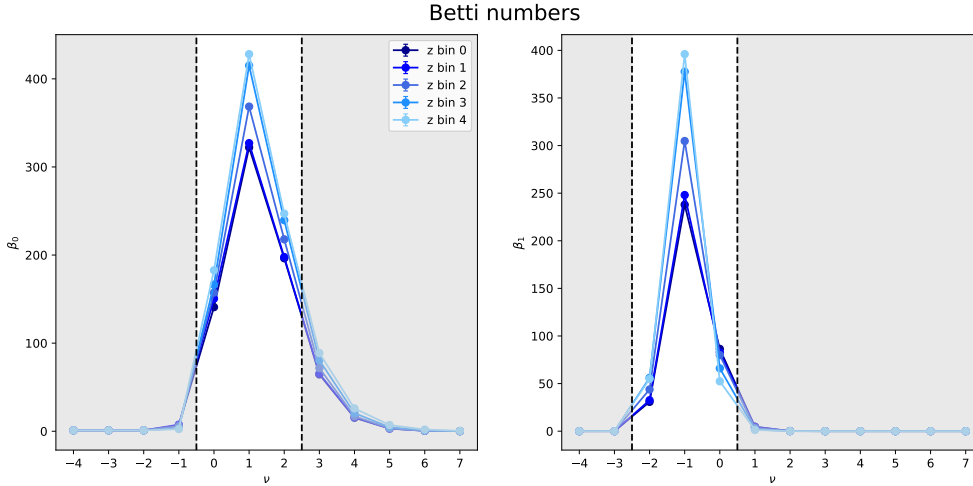


Fig. 5. Two Betti numbers calculated on the maps obtained from the simulation corresponding to the constraints on **PC18** parameters, averaged over 500 maps, for all redshift bins, with a smoothing scale $\theta_s = 2'$. The error bars correspond to the standard deviation divided by the square root of the number of maps. The dashed black lines contain the values selected for the training, the grayed out regions indicate the discarded values.

3.5. Training and test samples

In order to compile the final dataset, we collected the measurements of all the higher order estimators described so far. Including the k_3 and k_4 HOM, the three MFs in the threshold range of $\nu \in [-1, 1]$, the two Betti numbers in the range of $\nu \in [0, 2]$ for β_0 and in the range of $\nu \in [-2, 0]$ for β_1 , and the τ , C , α , and p graph statistics in the range of $\nu \in [-1, 1]$, we obtained 29 measurements for each redshift bin, making a total of 145 measurements for each simulation.

Because the measurements on a single map are dominated by the noise, we need to increase the S/N of the estimators by averaging them over multiple maps. In order to investigate whether we could obtain a better performance from a bigger but noisier dataset or from a smaller dataset with a higher S/N, we decided to average each measurement over 100, 300, and all 500 maps belonging to the same simulation, obtaining three versions of the dataset. In the first version (hereafter AVG100), we averaged each of the 145 measurements over 100 maps, corresponding to a total area of 2500 deg^2 . Because we have 500 maps for each simulation, we obtained five realizations of the set of 145 measurements for each combination of the cosmological parameters. Therefore, the dataset passed from 750 000 to 7500 independent realizations. In the second version (hereafter AVG300), we averaged each of the 145 measurements over 300 maps, corresponding to a total area of 7500 deg^2 , obtaining just one realization for each cosmological model. This means that there are 200 maps, among the 500 maps per simulation, that we did not use. This dataset passed from 750 000 to 1500 independent realizations. Finally, in the third version (hereafter AVG500), we averaged each of the 145 measurements over 500 maps, corresponding to a total area of $12\,500 \text{ deg}^2$, therefore, using all the maps available for each simulation and obtaining again just one realization of the set of measurements for each combination of the cosmological parameters. This dataset also passed from from 750 000 to 1500 independent realizations. A summary of the three datasets is given in Table 2.

Each of the three datasets is then divided into a training set, consisting of 80% of the respective original dataset, and a test set obtained with the remaining 20%. Hereafter we will refer to the estimator measurements in the datasets as “features”, and to the corresponding cosmological parameters as “labels”. We repeated this procedure for each smoothing scale, obtaining the three datasets, AVG100, AVG300, and AVG500 for $\theta_s = 2'$, and the datasets AVG100 and AVG300 for $\theta_s = \{4', 6'\}$. Considering that the application of Gaussian smoothing degrades part of

the information contained in the shear maps, we expect to obtain progressively worse results for increasing smoothing scale. For this reason and to save computational time, we chose not to apply the smoothing scales $\theta_s = \{4', 6'\}$ to the AVG500 dataset, deeming as exhaustive the comparison of the results coming from the AVG100 and AVG300 dataset with those from the smoothing.

A comment is in order here about the choice of the map size. Having a side length of 5 deg only gives us confidence that the flat sky approximation can be used, which is useful given that a full sky treatment of some of the above statistics is not available. Such small maps are, however, likely to be affected by cosmic variance which is a further motivation to average over a large number of them. In a realistic application, this can be done splitting the full survey area into $5 \times 5 \text{ deg}^2$ non-overlapping maps. This would demand an area of $(2500, 7500, 12\,500) \text{ deg}^2$ in order to create the AVG100, AVG300, AVG500 datasets. Among current ongoing Stage III surveys, DES is compliant with our requirements for AVG100 since Y3 and Y5 data releases will cover 5000 deg^2 . On the contrary, Stage IV surveys will be needed for AVG300 and AVG500 given the large area required. In particular, both *Euclid* ($15\,000 \text{ deg}^2$) and LSST ($18\,000 \text{ deg}^2$) will cover enough area for both cases. The methods we are presenting is therefore designed to fully exploit the potentiality of Stage IV surveys.

In Fig. 7, we show the correlation between the different features using the 500 maps from the simulation with the constraints on **PC18** parameters (300 maps for $\theta_s = \{4', 6'\}$). In the upper-left panel, we find the total covariance matrix for smoothing $\theta_s = 2'$, which includes all the selected features and all redshift bins, and, in the upper right panel, a zoom on the first redshift bin. We can observe that within the same simulation, the correlation between features at different redshift bins is quite small due to the adopted binning. On the other hand, for a given redshift bin some features appear to be more correlated than others. In particular, we notice higher correlations between the HOM and the MFs and between the MFs and the Betti numbers, while the graph statistics appear to have slightly lower correlations with the other set of estimators. Such a result was not entirely unexpected. At a lowest order, the MFs can be expressed as a perturbative series whose coefficients are related to the generalized moments whose analytical expression is quite similar to that for the HOM. This tells us that MFs are indeed related to the moments of the distribution so that a correlation can be anticipated. Similarly, the Betti numbers are known to be a generalization of the MFs which explains why they turn out to be correlated

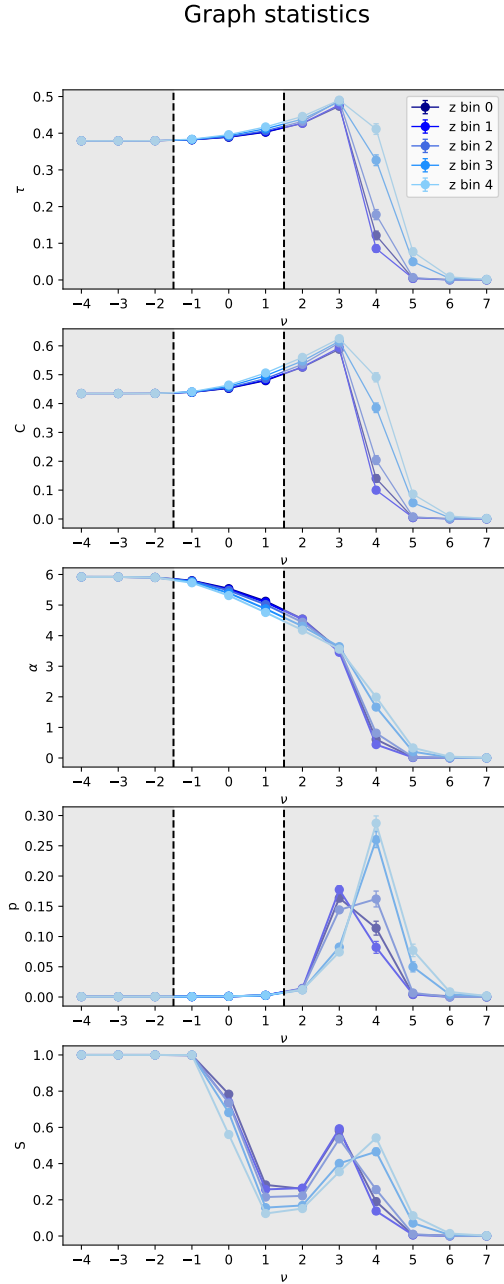


Fig. 6. Graph statistics calculated on the maps obtained from the simulation corresponding to the constraints on PC18 parameters, averaged over 500 maps, for all redshift bins, with smoothing scale $\theta_s = 2'$. The error bars correspond to the standard deviation divided by the square root of the number of maps. The dashed black lines contain the values selected for the training, the grayed out regions indicate the discarded values.

with them. On the contrary, the relation between graph statistics and the other estimators has not been investigated up to now, so the lack of correlation that we find is an interestingly novel property. In the bottom panels, we show again the zoomed covariance matrix of the first redshift bin but for different smoothing scales, $\theta_s = 4'$ on the left, and $\theta_s = 6'$ on the right. We notice that the correlation between the Betti numbers at different thresholds decreases, while the correlation of the graph statistics increases. On the other hand, the correlation between the graph statistics and the rest of the estimators further decreases, showing a decoupling into two sets of estimators.

4. Model selection

Using the training and the test sets, which we obtained as explained in the previous section, we compared different machine learning algorithms in order to establish the model that best describes the relation between the features that we measured on the shear maps and the cosmological parameters. Explaining the inner workings of the different algorithms and the particular implementations that we used is beyond the scope of this paper. We nevertheless briefly outline the methods used in Appendix A, referring the interested reader to the specific resources therein for further details. The algorithms that we tested are linear regression, ridge regression, kernel ridge regression, Bayesian ridge regression, lasso regression, support vector machine, K nearest neighbors, Gaussian processes, decision tree, random forests, and gradient boosting.

We used the Python scikit-learn (Pedregosa et al. 2011) library implementation of all the listed algorithms. A limitation of the Scikit-learn implementation in the majority of the methods used is the lack of a possibility to training a model to predict the entire set of labels at once. In fact, even if the labels were chosen randomly and independently, making them uncorrelated (with the exception of Ω_M and Ω_Λ , which are linked by the assumption of a flat universe, $\Omega_\Lambda = 1 - \Omega_M$), they act simultaneously on the maps and some features could be sensitive to particular parameter combinations. This problem could have been solved by varying one cosmological parameter at a time for each simulation. The downside is that the computational time needed to generate the same amount of simulations would be multiplied by the number of cosmological parameters that we are interested in and, more importantly, this approach would produce a training set that would not be representative of observations. In fact, making vary one parameter at a time requires to fix the remaining parameters to some value that, with observations, we do not know a priori. Therefore, we decided to perform the training separately for each cosmological parameter and considered the effect of the variation of the remaining parameters as additional noise on the features. This means that we trained seven different machine learning models for each algorithm and used them to predict the respective cosmological parameters.

We performed a three-fold cross-validation to choose the values of the hyperparameters that determine the best model for each method. We evaluated the performance of each model using the R^2 score, defined as:

$$R^2 = (1 - \text{RSS}/\text{TSS})$$

$$\text{RSS} = \sum (y_{\text{true}} - y_{\text{pred}})^2$$

$$\text{TSS} = \sum (y_{\text{true}} - \langle y_{\text{true}} \rangle)^2,$$
(9)

where RSS is the residual sum of squares, TSS is the total sum of squares, y_{true} are the true labels, and y_{pred} are the predicted labels. With this definition, the best score is 1 and a constant model that always predicts the expected value of y , disregarding the input features, would give a score of 0. The score can also be negative, because a model can be arbitrarily worse than the constant model.

All the penalized models (i.e., ridge, kernel ridge, Bayesian ridge, lasso, and support vector machine) obtained the best score with a small value of the penalty hyperparameter ($\alpha \leq 10^{-4}$) tending therefore to a simple linear regression model. All the models employing a kernel (i.e., kernel ridge, support vector machine, and Gaussian processes) gave the best performance when using a linear kernel, compared to the other kernels that we

Table 2. Summary of the three datasets.

Dataset	No. of maps per average	No. of realizations per cosmology	Total no. of realizations	Equivalent area
AVG100	100	5	7500	2500
AVG300	300	1	1500	7500
AVG500	500	1	1500	12 500

Notes. Starting from 750 000 independent realizations (500 maps for each of the 1500 cosmologies): in AVG100 each feature is averaged over 100, corresponding to a total area of 2500 deg², making five realizations for each cosmology and a total of 7500 realizations; in AVG300 each feature is averaged over 300, corresponding to a total area of 7500 deg², making one realization for each cosmology and a total of 1500 realizations; in AVG500 each feature is averaged over 500, corresponding to a total area of 12 500 deg², making one realization for each cosmology and a total of 1500 realizations.

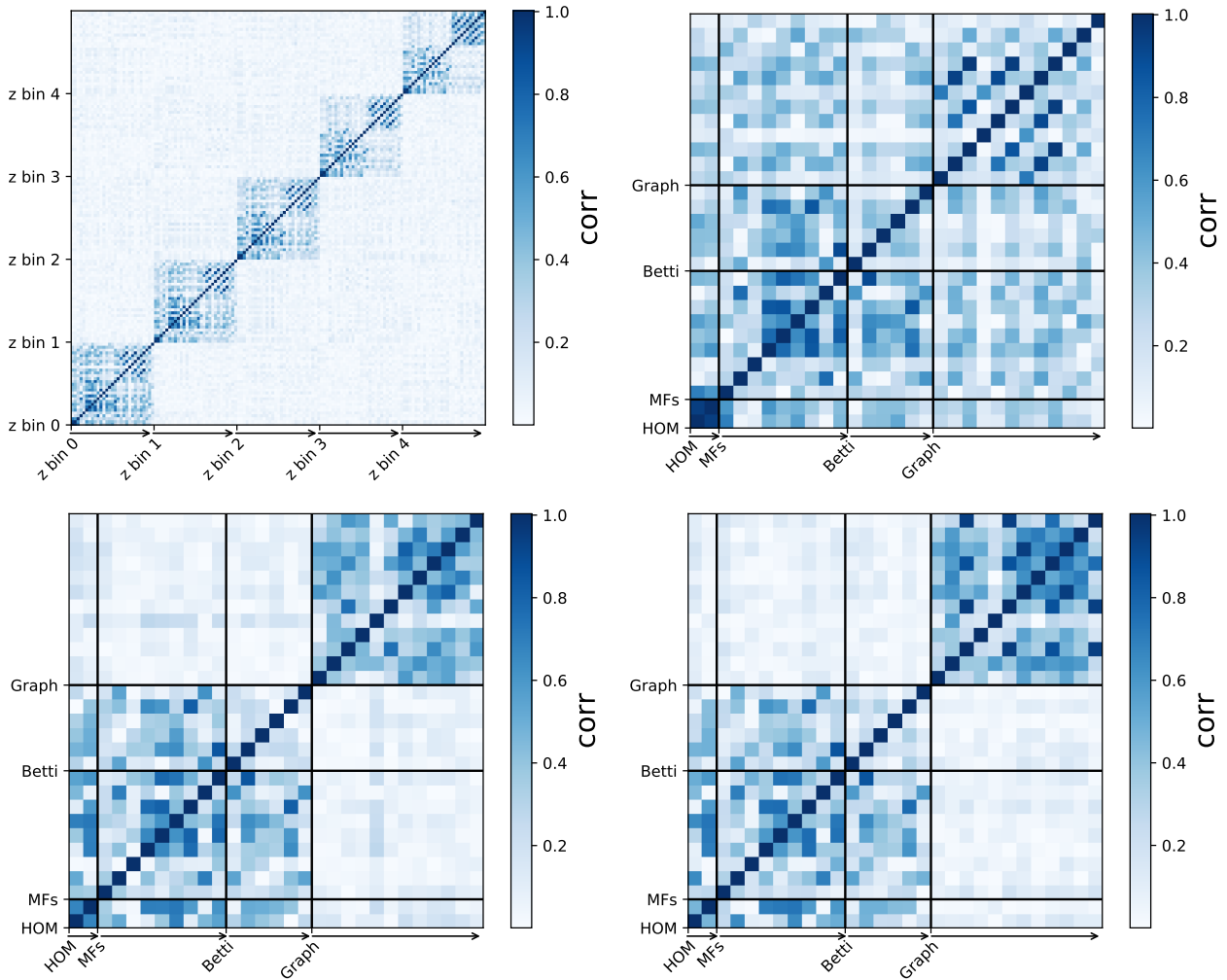


Fig. 7. *Top left:* covariance matrix obtained using all 29 features for each redshift bin, measured on the 500 maps from the simulation with the constraints on PC18 parameters, with smoothing $\theta_s = 2'$. *Top right:* same as in the top left panel, but for the first redshift bin only. *Bottom left:* same as in the top right panel, but measured on 300 maps for smoothing $\theta_s = 4'$. *Bottom right:* same as in the top right panel, but measured on 300 maps for smoothing $\theta_s = 6'$. We notice that within the same simulation, the correlation between features at different redshift bins is quite small due to the adopted binning, while for a given redshift bin some features appear to be more correlated than others. The correlation between the graph statistics and the rest of the estimators further decreases with increasing smoothing scale.

tested (i.e., polynomial of degree 2 and 3, RBF, sigmoid, Matern, and rational quadratic kernels from the Scikit-learn library), again making the models tend to a linear regression.

In Table 3, we show the score obtained on the test set for the best model for each method and for each cosmological parameter, with the dash indicating a negative score. The results were obtained using the AVG100 dataset with $\theta_s = 2'$. As expected,

we can see that all the models tending to the linear model obtained the same score, with the exception of support vector machine, which uses a different minimization function compared to linear, ridge, kernel ridge, Bayesian ridge, lasso, and Gaussian processes. Overall, the best score is obtained with the linear regression, followed by Gaussian processes and random forests, while K nearest neighbors, decision tree, and support

Table 3. R^2 scores of the best model on the test set for: linear regression, ridge regression, kernel ridge regression, Bayesian ridge regression, lasso regression, support vector machine (SVM), K nearest neighbors (KNN), Gaussian processes (GP), decision tree (DT), random forests (RF), and gradient boosting (GB), for each cosmological parameter.

	Linear	Ridge	k. ridge	b. ridge	Lasso	SVM	KNN	GP	DT	RF	GB
H_0	0.16	0.17	0.17	0.17	0.18	0.16	–	0.16	0.05	0.15	0.16
ω_b	–	–	–	–	–	–	–	–	–	–	–
Ω_M	0.61	0.61	0.61	0.61	0.61	–	0.31	0.61	0.19	0.49	0.55
Ω_Λ	0.61	0.61	0.61	0.61	0.61	–	0.31	0.60	0.19	0.49	0.55
w_0	0.65	0.65	0.65	0.65	0.65	0.57	0.49	0.65	0.42	0.59	0.60
n_s	0.16	0.16	0.16	0.16	0.16	–	–	0.16	0.02	0.10	0.11
σ_8	0.56	0.56	0.56	0.56	0.56	–	0.30	0.56	0.25	0.48	0.49

Notes. The dash represents a negative value of the score. The dataset used is the AVG100 with $\theta_s = 2'$. All models tending to the linear regression obtained the same score, with the exception of support vector machine. Overall, the best score is obtained with the linear regression.

vector machine perform progressively poorly. The fact that linear regression performs better than the other models can be explained considering the small interval of variation of each cosmological parameter. We are looking at a very zoomed-in region of the hyperspace spanned by the features so that locally the relation with the cosmological parameters tends to linearity. The parameter that is best predicted is w_0 with a promising score of 0.65, with Ω_M and Ω_Λ coming next with a still fairly good score of 0.61. We obtained a slightly lower score of 0.56 for σ_8 , while H_0 , and n_s obtained a much lower scores, and ω_b cannot be predicted at all, having only negative scores for all models. The scores that we obtained confirm the ability of WL to probe certain parameters more than others as was expected from previous cosmic shear studies (e.g., Takada & Jain 2004; Munshi et al. 2008) in which Ω_M and σ_8 were tightly constrained and, using external CMB measurements of H_0 , n_s , and ω_b , it was possible to also improve the constraints on the dark energy equation of state parameters. Increasing the smoothing scale to $\theta_s = \{4', 6'\}$, we obtain the same qualitative results in terms of the performance of one model with respect to another but overall progressively lower scores. We investigate the effect of the smoothing in more detail in the next section.

5. Predictions of the cosmological parameters

Considering the results outlined in the previous section, we decided to retain for the rest of this work only the best-performing model, which turned out to be linear regression. Using this model, we want to study the impact that the size of the training set (i.e., the number of independent realizations), the S/N of the features (i.e., the number of maps used to average the features), and the smoothing have on the results. We also want to determine which is the best predicted parameter and verify if the results are consistent using the different training sets.

We performed the training and the prediction for each cosmological parameter on all three datasets, AVG100, AVG300, and AVG500 with smoothing $\theta_s = 2'$ and on the AVG100 and the AVG300 datasets with the additional smoothing scales $\theta_s = \{4', 6'\}$, in order to compare the results.

In Tables 4–6, we show the R^2 score that we obtained in each case. The dash represents a negative value of the score. In the following, we only discuss the R^2 score because we found that given our dataset, it is the most meaningful statistic with regard to evaluating the model performance. In Appendix B, we also consider two additional statistics, the mean squared error, and the mean absolute percentage error.

Table 4. R^2 score for the three datasets, AVG100, AVG300, and AVG500 with smoothing scale $\theta_s = 2'$, for each cosmological parameter.

	AVG100, $\theta_s = 2'$	AVG300, $\theta_s = 2'$	AVG500, $\theta_s = 2'$
H_0	0.16	0.09	0.03
ω_b	–	–	–
Ω_M	0.61	0.64	0.70
Ω_Λ	0.61	0.64	0.70
w_0	0.65	0.75	0.81
n_s	0.16	0.23	0.32
σ_8	0.56	0.73	0.80

Notes. The dash represents a negative value of the score. As expected, increasing the number of maps per simulation improves the performance of the algorithm.

Table 5. Same as in Table 4, but for the AVG100 and AVG300 datasets with smoothing scale $\theta_s = 4'$.

	AVG100, $\theta_s = 4'$	AVG300, $\theta_s = 4'$
H_0	0.16	–
ω_b	–	–
Ω_M	0.55	0.60
Ω_Λ	0.55	0.60
w_0	0.62	0.68
n_s	0.06	–
σ_8	0.52	0.64

Table 6. Same as in Table 4, but for the AVG100 and AVG300 datasets with smoothing scale $\theta_s = 6'$.

	AVG100, $\theta_s = 6'$	AVG300, $\theta_s = 6'$
H_0	0.18	0.06
ω_b	–	–
Ω_M	0.55	0.61
Ω_Λ	0.55	0.61
w_0	0.60	0.60
n_s	0.04	–
σ_8	0.47	0.52

Notes. Comparing this results with those in Tables 4 and 5, we observe that overall the score decreases with increasing smoothing scale.

Starting from Table 4, we compare the results obtained with the AVG100, AVG300, and AVG500 datasets with smoothing $\theta_s = 2'$. We can see that while H_0 and ω_b cannot be predicted with any dataset version, the results for the other parameters improve progressively. The best measured parameter is still w_0 for which we obtained the improved score of 0.75 using the AVG300 dataset, and the even better score of 0.81 with the AVG500 dataset, corresponding to a score improvement of $\sim 25\%$. The two parameters which obtained the next best score with the AVG100 dataset, Ω_M and Ω_Λ , both got a smaller but still exhibited a consistent improvement of $\sim 15\%$ in score reaching a value of 0.64 and 0.70, with the AVG300 and the AVG500 datasets respectively. The parameters that obtained the largest score improvement are σ_8 which went from a value of 0.56 for the AVG100 dataset, to 0.73 for the AVG300 dataset, reaching the very good score of 0.80 for the AVG500 dataset with a total score improvement of $\sim 43\%$, making it the second best predicted parameter, and n_s which improved its score of $\sim 100\%$, going from 0.16 to 0.23 for the AVG300 dataset, and 0.32 for the AVG500 dataset.

Moving on to Table 5, we can see the results obtained from the AVG100 and the AVG300 datasets with smoothing $\theta_s = 4'$ and observe an overall decrease in score. Comparing the AVG100 results with $\theta_s = 2'$ and $\theta_s = 4'$, we notice that the most affected parameters are Ω_M and Ω_Λ , which got their score degraded of $\sim 10\%$. On the other hand, the same comparison for the AVG300 datasets shows that we obtained the greatest decrease in score for w_0 and σ_8 , corresponding to $\sim 10\%$ and $\sim 14\%$, respectively. We also remark that n_s cannot be predicted with $\theta_s = 4'$ with neither the AVG100 nor the AVG300 dataset.

In Table 6, we show the results obtained from the AVG100 and the AVG300 datasets with smoothing $\theta_s = 6'$. We notice that while the scores of Ω_M and Ω_Λ are stable compared to the results presented in Table 5, the score of w_0 and σ_8 further decreased of $\sim 10\text{--}20\%$. Again, n_s cannot be measured with this smoothing scale.

Considering the behavior of the score for the different cosmological parameters that we observed with increasing smoothing, we can conclude that Ω_M and Ω_Λ are less affected by the degradation of the spatial resolution and of the information that we can measure on the shear maps compared to w_0 , σ_8 , and n_s . In fact, as we show in Sect. 6, the features that most contribute to the measurement of Ω_M and Ω_Λ are the graph statistics, which turned out to be much less sensitive to the smoothing scale compared to the other estimators. We also remark that the decoupling between the graph statistics estimators and the HOM, MFs, and Betti numbers, which we observed in the correlation matrix presented in Fig. 7 for increasing smoothing, did not help the algorithm to extract additional information, having probably been counterbalanced by the increased correlation inside the set of graph statistics estimators.

Figure 8 shows the true labels versus the predicted labels, for each cosmological parameter, using the AVG100, the AVG300, and the AVG500 datasets, with smoothing $\theta_s = 2'$. The dots represent individual predictions while the shaded areas correspond to the 1σ region, obtained dividing the test sample into 10 bins of the true label values and calculating the mean and standard deviation of the predictions inside each bin. The more the colored region for each given parameter aligns along the dashed black diagonal, the better will be the prediction obtained with such model. The black dots are the values of the constraints on PC18 parameters with the respective error, reported also on the y -axis as reference. We can see that the alignment along the diagonal improves passing from the blue, to the red, and to the

orange regions, corresponding to the AVG100, the AVG300, and the AVG500 datasets, and that we obtain the best results for the Ω_M , Ω_Λ , w_0 , and σ_8 parameters, confirming what is shown in Tables 4–6.

In the top-left panel of Fig. 9, we show the learning curve for the w_0 parameter, which corresponds to the score as a function of the training set size (as percentage values of the complete dataset size, reported in Table 2), for the training and the test set of the AVG500 dataset, with smoothing $\theta_s = 2'$. We performed three-fold cross-validation and plotted the mean values, with the error bars corresponding to the standard deviation of the different realizations. We see that as the training set size increases, the train score decreases and the test score increases. For example, using $\sim 20\%$ of the entire dataset for training and the remaining $\sim 80\%$ for testing results in a training score of >0.9 and in a test score of ~ 0.5 . When the training score is much higher than the test score, we are in a situation of overfitting, meaning that the model performs almost perfectly on the data that it has seen before but very poorly on new data. Around a training set size of 50% the two curves start to flatten and come out of the overfitting region. We can verify that for our default choice of a training set size of ≥ 80 , our model does not overfit the data because the difference between the training and the test score is small.

In the top-right panel of Fig. 9, we show the learning curve for the w_0 parameter for the test set of the three datasets, with smoothing $\theta_s = 2'$. We notice that the blue curve corresponding to the AVG100 dataset flattens out at a training set size of $\sim 50\%$ and a test score of ~ 0.6 . The red and the orange curves, representing the AVG300 and the AVG500 datasets, respectively, exhibit a different behavior. We can see that for a training set size of $<3\%$, we obtain low or negative scores, while for a training set size $>30\%$ the test set curves show an increasing score that reaches values of ~ 0.7 for the AVG300 dataset and ~ 0.8 for the AVG500 dataset, yet they do not reach a plateau value. This is due to the smaller size of the AVG300 and the AVG500 datasets, which is one fifth of the AVG100 dataset. We can conclude that the number of maps used to average the features in order to increase the S/N of the measurements has a sizable impact on the results and it is proportional to the R^2 score obtained. On the other hand, while increasing the dataset size for the AVG100 version would not change the results; for the AVG300 and the AVG500, we can envisage a margin of improvement of a few percent.

In the bottom-left and bottom-right panels of Fig. 9, we show again the learning curve for w_0 and Ω_M , respectively, but this time we compare the results obtained on the test set using the AVG300 dataset with smoothing $\theta_s = \{2', 4', 6'\}$. Looking at the plot for w_0 , we notice the same increasing trend for the three curves as for the AVG300 and AVG500 curves in the top panel for smoothing $\theta_s = 2'$. Confirming the results shown in Tables 5 and 6, the red and orange curves reach a lower score compared to the blue curve. Even if the margin of improvement of a few percent with increasing dataset size is apparent for all three cases, we remark that the plateau value will be lower for increasing smoothing. From the same analysis for the Ω_M plot, we observe that the curves corresponding to smoothing $\theta_s = \{4', 6'\}$ are more or less superposed and that they reach a score value very close to the blue curve. This consolidates the conclusions that we drew when comparing the results of Tables 5 and 6 for the different cosmological parameters regarding the higher sensitivity of w_0 , σ_8 , and n_s to some particular bits of information contained in the shear maps that is degraded by the smoothing, as compared to Ω_M and Ω_Λ .

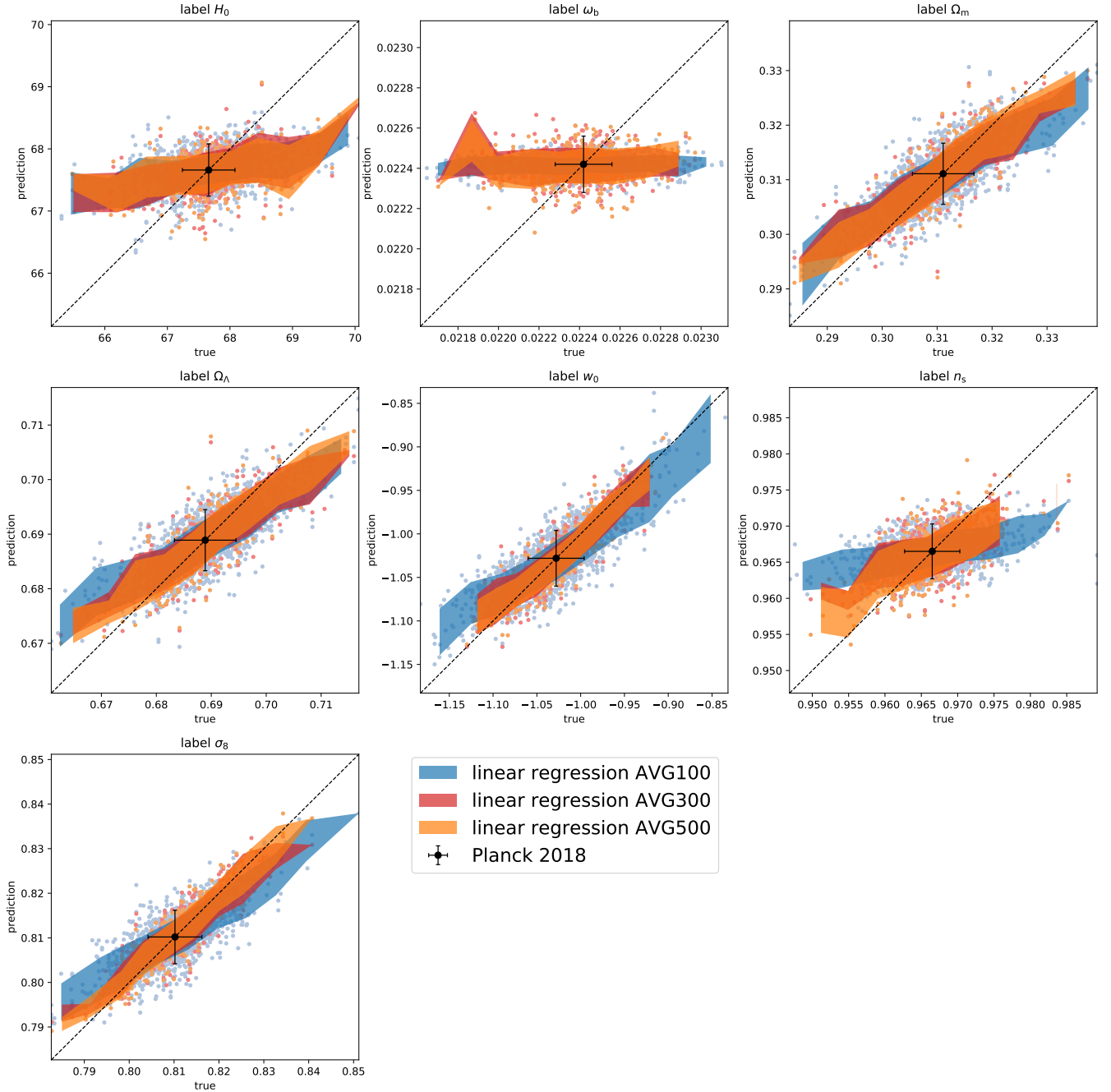


Fig. 8. Performance on the test set of the linear regression model for each cosmological parameter, and for the three datasets, AVG100, AVG300, and AVG500, with $\theta_s = 2'$. The dots represent individual predictions. The shaded regions delimit the 1σ interval obtained dividing the test sample into 10 bins of the true label values and calculating the mean and standard deviation of the predictions. We obtain the best predictions with the AVG500 dataset and for the Ω_M , Ω_Λ , w_0 , and σ_8 parameters.

In summary, we conclude that overall the model performance increases as a function of the S/N of the features (or the number of maps used for the averages, i.e., the total area) and decreases as a function of the smoothing scale.

6. Feature importance

Once we obtained the predictions of the cosmological parameters, we wanted to investigate their relation with the higher order estimators used to obtain them. In order to estimate which features contribute more to the measurement of each cosmological parameter, we repeated the training using random subsets of the

features in the AVG500 dataset, with smoothing $\theta_s = 2'$. We then assigned a value to each feature summing the score divided by number of features in the subset, for each random realization, increasing the number of iterations until convergence, for a total of 800 000 realizations. We used random subsets with sizes between 3 and 15 features. We then normalized the feature contributions so that they assume a value between 0 and 1.

In Fig. 10, we show a color map that indicates the importance of each feature, that is, the contribution of each feature to the final score, for each cosmological parameter. We remark that as the feature contributions are normalized separately for each color map, they offer a measure of the importance of each

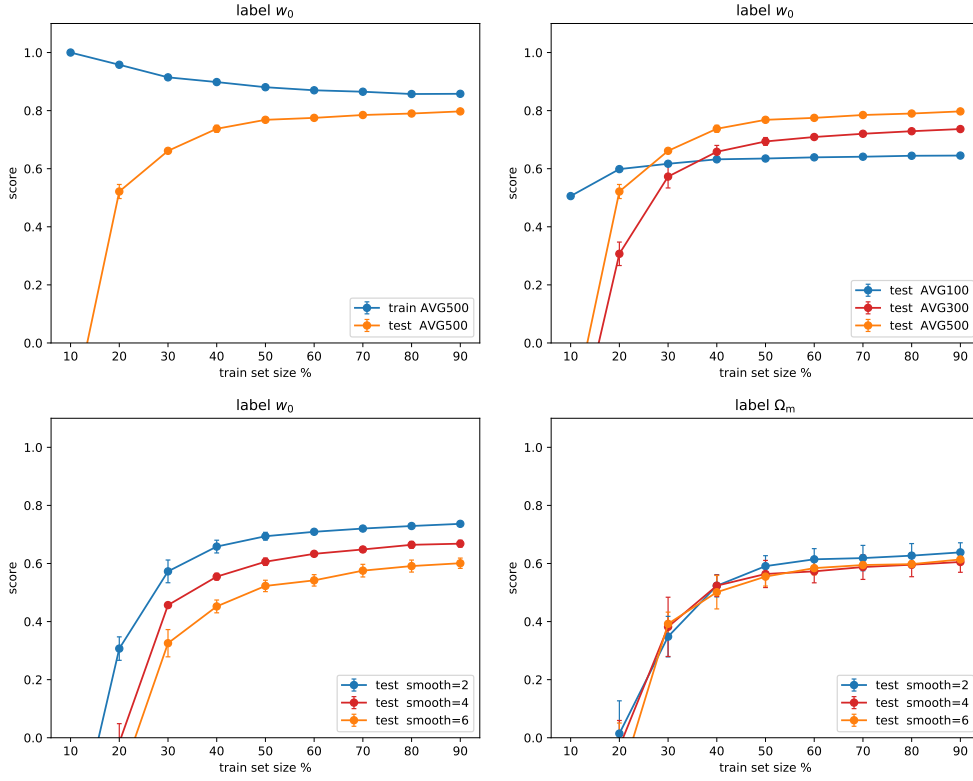


Fig. 9. *Top left:* learning curve for the train and the test set of the AVG500 dataset, with smoothing $\theta_s = 2'$, for the w_0 parameter. *Top right:* learning curve for the test set of the AVG100, AVG300, and AVG500 datasets, with smoothing $\theta_s = 2'$, for the w_0 parameter. *Bottom left:* learning curve for the test set of the AVG300 dataset, with smoothing $\theta_s = \{2', 4', 6'\}$, for the w_0 parameter. *Bottom right:* same as in the middle panel but for the Ω_M parameter. In all cases, we performed a three-fold cross-validation.

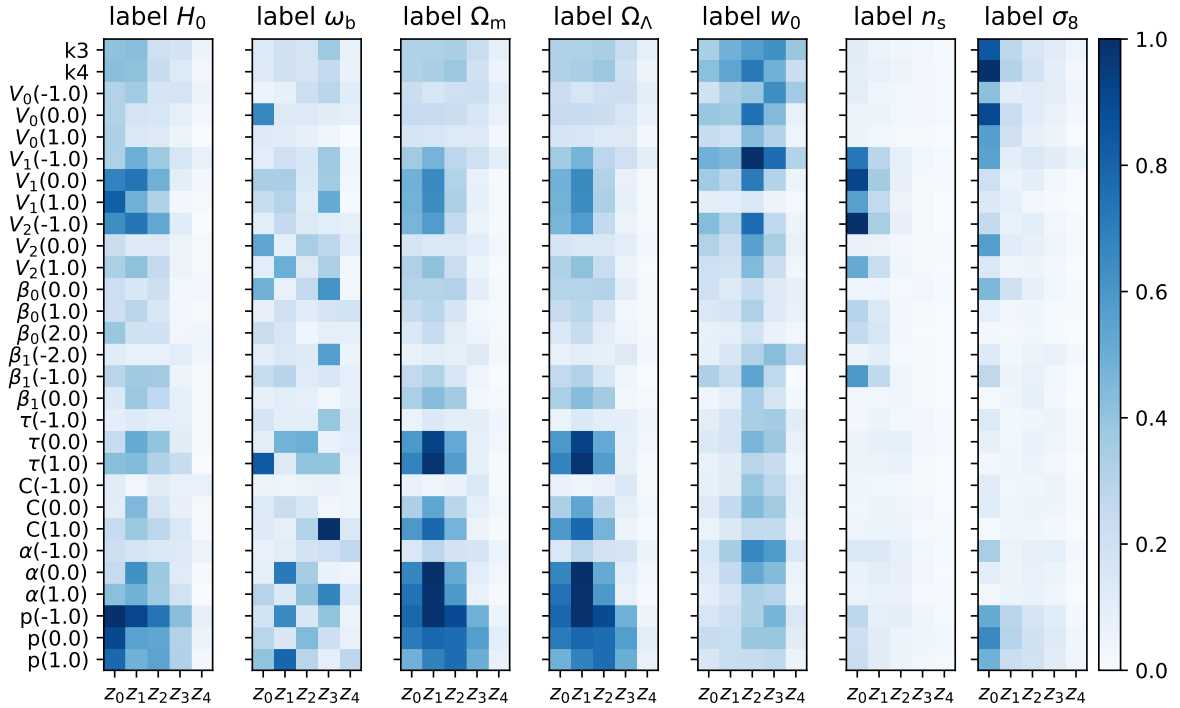


Fig. 10. Normalized feature importance for each cosmological parameter. Graph statistics have a fundamental importance for the measurement of Ω_M and Ω_Λ , as do the MFs for n_s , and the HOM for σ_8 , while w_0 slightly favors HOM and MFs. Overall the information measured from the different estimators increases with decreasing redshift. From left to right, the R^2 scores are $\{0.03, -, 0.70, 0.70, 0.81, 0.32, 0.80\}$, as reported in Table 4.

feature compared to the others, in the context of the prediction of a given cosmological parameter. When comparing the color maps between them, we have to keep in mind the R^2 scores for the AVG500 dataset, which are reported in Table 4 for each cosmological parameter. Color maps for parameters with low R^2

score should be taken with caution, which is the case for H_0 , ω_b , and n_s . We nevertheless report them for completeness.

It turns out that the color maps for H_0 , Ω_M , and Ω_Λ show the same qualitative features. For all of them, the HOM, the zeroth-order MF V_0 , and the Betti numbers give a negligible

contribution to the final score, which is dominated by the other two MFs and the (τ, C, α, p) graph statistics. Low-to-medium redshift bins are preferred with the bin at $z = 0.9$ for graph statistics playing a major role in determining Ω_M and Ω_Λ .

It is, on the contrary, hard to interpret the color map for ω_b with its almost random distribution of colors. This is, however, not surprising given the negative score in every version of the feature dataset used. Again, this could have been anticipated since ω_b is mainly responsible for modulating the BAO wiggles in the matter power spectrum which are smoothed out by the lensing kernel. As such, WL is not sensible to this parameter no matter which estimators (second or higher order) that we rely on.

The color map for w_0 points toward the HOM and MFS as the main contributors with no particular preference for a redshift bin, while graph statistics and Betti numbers follow with the medium-to-high redshift range contributing the most. The need to use all the redshift bins (although with different estimators) is likely related to the need to follow the growth of structures whose evolution is determined by the w_0 value. Which estimator is best suited to this depends on the level of non-Gaussianity. At low redshift, the nonlinear collapse of structures enhances the non-Gaussianity of the field which can be quantified by the ease to measure HOM and MFs. On the contrary, at larger z , one is approaching the linear regime hence the need for more advanced tools to spot the residual non-Gaussianity.

Moving to n_s , the color map points at V_1 , V_2 , and β_1 at low redshift as leading contributor with β_0 and p giving the residual contribution. It is important to remember, however, that n_s is poorly determined overall so that the color map is less informative. Finally, the color map for σ_8 shows some similarity with the one for w_0 , the main contribution coming from the HOM, followed by the MFs and the p parameter. However, the importance of the parameters now increases as z decreases, with the first redshift bin bringing the majority of the information.

Excluding ω_b for its erratic behavior from the rest of this discussion, we can observe that overall the information measured from the different estimators increases with decreasing redshift (dramatically so for n_s and σ_8), with the exception of w_0 . While for Ω_M , and Ω_Λ the peak is reached at $z = 0.9$, the most contributing redshift bin for n_s and σ_8 is clearly $z = 0.6$. On the other hand, for w_0 the importance appears to be more uniform in redshift, with a slight preference for $z = 1.2 - 1.5$.

While MFs seems to be in some measure sensitive to all the cosmological parameters, generally, the Betti numbers appear to contribute the least, with a small exception at some thresholds for n_s and σ_8 . Graph statistics have a fundamental importance for the measurement of Ω_M and Ω_Λ , as do the MFs for n_s , and the HOM for σ_8 . Once again, w_0 shows a more uniform behavior also in terms of estimators contribution, slightly favoring HOM and MFs. This explains to some extent the decrease in score for the different cosmological parameters with increasing smoothing scale. As we discuss in Sect. 5, the score of Ω_M and Ω_Λ is more stable as a function of the smoothing, compared to the score of w_0 , σ_8 , and n_s , and this behavior is reflected by the most important features for each parameter. In fact, while the HOM, MFs, and Betti numbers change with the smoothing of a factor ~ 40 , ~ 30 , and ~ 10 , respectively, the graph statistics only vary of less than $\sim 15\%$.

The feature importance can also be interpreted considering the physical meaning of the different estimators. The measurement of Ω_M , and Ω_Λ parameters is mainly due to τ , C , α , and p , making these parameters sensitive to overall degree of connectivity or clustering measured on the shear maps. The n_s parameter is linked to the information that is contained in the derivatives

of the shear field, through V_1 and V_2 , which are in turn connected to the matter power spectrum and bispectrum, while in addition to this, σ_8 is also related to a greater extent to the variance of the shear field through V_0 and to the three- and fourth-point correlation functions, through k_3 and k_4 . Finally, all of the above contributes to the measurement of w_0 .

7. Improving the methodology

The main aim of this paper is to present a new methodology applicable to WL higher order statistics starting from the shear field without the need for convergence reconstruction or for a theoretical formulation of the relation between the estimators used and the underlying cosmology. The interesting results discussed above are a good reason to reconsider the limitations of this first step in order to understand how to make the method still more appealing.

One point worth improving is related to the realism of the training set of simulations. We have indeed approximated the shear field as lognormal and used FLASK to quickly generate a large set of maps varying the cosmological parameters. Both these aspects can be ameliorated. First, we note that the requirement that the lognormal approximation is a good representation of the shear field has forced us to consider only bins with $z > 0.5$, thus cutting out the low redshift regime, which is the one dominated by the dark energy we want to investigate. Since we only considered models with constant equation of state, it has not been of paramount importance to investigate where the transition from accelerated to decelerated expansion takes place. Adding w_a to the list of parameters would probably call for the inclusion of lower redshift bins. Also the fact that FLASK does not make any assumption on the shear higher order moments, making them unreliable for realistic simulations, implies that some of the estimators that we considered could lead to different results on actual observations. Moving beyond FLASK is therefore necessary in order to create a training set which is as similar as possible to the underlying true universe. For this same reason, it is of fundamental importance to adopt the correct source redshift distribution and account for the errors due to photo- z . We note that both these aspects are survey-dependent so that a reliable training exercise can be obtained only with a good knowledge of the survey specifics. Moreover, intrinsic alignment, which in the weak regime linearly adds to the lensing shear, should also be taken into account. Even if intrinsic alignment is a local effect that should not alter the global topology of the maps and it is subdominant at high redshift, it increases the correlation among close redshift bins and therefore might also increase the correlation among features at different z , decreasing their constraining power. It also represents an additional source of noise that should be modeled and included in the simulations (e.g., Bruderer et al. 2016; Hildebrandt et al. 2017; Wei et al. 2018; Ghosh et al. 2020). Another missing ingredient are baryonic effects, which affects the cosmic shear signal at medium angular scales, where the total matter power spectrum is subjected to a suppression of power up to $\sim 30\%$, and at very small scales, where the power is enhanced because of efficient baryon cooling and star formation in the halo centers. These effects have been modeled both numerically and analytically but different implementations lead to different results. While the general trend is reproduced in most simulations, there is still no agreement on the quantitative level so that more work need to be done to reach a self-consistent treatment of these processes

in the cosmological context (e.g., Harnois-Déraps et al. 2015; Chisari et al. 2018; Euclid Collaboration 2019b; Kacprzak et al. 2019; Schneider et al. 2019, 2020a,b).

Another aspect worth improving is the range which the cosmological parameters are varied over to generate the simulated maps. Our initial goal was to compare the constraints that can be obtained from WL high-order statistics with those from the joint use of CMB and other probes reported in the constraints on PC18. This motivated us to choose the constraints on PC18 values as reference and the corresponding 1σ errors as width of the Gaussian distribution that we used to randomly extract the simulations parameters. Machine learning methods can not really assign an error to the estimate of a parameter so that we decided to quantify the uncertainty by analyzing the statistics of the test results. As shown in Fig. 8, the constraints we get in this way are comparable to those in the constraints on PC18 suggesting that using high-order statistics is as efficient as the CMB joint with other probes. However, in order to strengthen this promising result, it would be interesting to probe a much wider range in the parameter hyperspace to see whether the preference for multilinear regression we have found is genuine or an artifact of having used a so small range that all deviations from the fiducial can be parameterized through linear relations. It is entirely possible, indeed, that in this case, more sophisticated machine learning methods would stand out as the preferred ones and would eventually improve the constraints.

A final point to discuss concerns the precision of the estimators measurements. As we have seen, higher scores ask for higher precision which can be obtained by averaging over a large number of maps. Ideally, one could generate as many maps as needed from the same initial simulation, but this is no more the case if one wants to rely on real maps. Indeed, for a *Euclid* like survey, cutting the total 15 000 sq deg survey area leads to ~ 300 maps which must be taken as an upper limit preventing to increase precision through averaging over an ad libitum number of maps. Luckily, Fig. 9 shows that, even keeping fixed the number of maps to ~ 300 , we can still improve our predictions while increasing the dataset size. For this project, we were able to employ only $\sim 200\,000$ hours of CPU time. The use of few million hours of computation time on more powerful hardware (such as a national supercomputer) could allow us to build a dataset 10–100 times larger. As an alternative, it would be feasible to rely on a decent number of realistic N -body simulations to cut more maps investing the same amount of computational resources. Such a strategy, however, would ask for a preliminary narrowing of the parameter space so that additional probes should be used to avoid wasting time in exploring models that have already failed in fitting other data.

Therefore, before this method can be applied to observations there are two main technical aspects that need to be addressed. First of all, the simulations must be realistic, including all the effects that we have been discussed above, and more specifically, they must be as close as possible to the actual data that we want to use in terms of survey characteristics such as source redshift distribution, noise, photo- z errors, and mask. The survey mask in particular, would determine the available area and the maximum number of maps that can be generated. Second, such simulations would require a significant amount of computation time so that the definition and the sampling of the cosmological parameter space must be optimized, for instance, with Latin hypercube sampling (McKay et al. 1979; Tang 1993; Euclid Collaboration 2019b).

8. Conclusions

The search for new statistical methods capable of shedding further light on the nature and nurture of dark energy is becoming more and more important as the promise of unprecedented high quality data from Stage IV lensing surveys (such as *Euclid*) moves towards realistic applications. Motivated by this consideration, we have here investigated a machine learning approach to higher-order statistics of the shear field. Going beyond second-order allows us to better probe the non-Gaussianity imprinted on the shear field by the nonlinear collapse of structures, thereby allowing us to alleviate degeneracy among cosmological parameters.

Our proposed method is innovative in three aspects. First, we directly work on the shear field as reconstructed from noisy galaxy ellipticity field which is the only quantity that is actually measured from images. This makes it possible to circumvent the non-trivial problem of map making, that is, the need to reconstruct the convergence field from noisy shear data. As a second novelty, we added some graph statistics measurements to the list of the estimators which (to the best of our knowledge) have never been used before in the context of WL studies and, in particular, on shear maps. The third novel aspect of the proposed methodology is the decision to use machine learning techniques to infer the almost complete set cosmological parameters solely from the shear higher-order estimators. We were motivated by the consideration that a theoretical formalism to compute the adopted quantities is available only for some of them and, in any case, based on a number of approximations and hypotheses that limits their application and may run the risk of introducing uncontrolled bias. The use of machine learning, on the contrary, is free of any assumption, flexible enough to include as many estimators as we want, and as reliable as the training set. While the application of machine learning for the prediction of cosmological parameters in the context of WL is not a new concept per se, our work diverges from the rest of the literature, as previous WL machine learning studies have been limited to the use of convergence maps, to the variation of only two parameters (Ω_M, σ_8) in a wide range sampled in big steps, and to the training of neural networks. To our knowledge, ours is the first work in which several machine learning methods were applied to noisy ellipticity maps, while more than two cosmological parameters were made vary and the contribution of each estimator to the measurements of each parameter was investigated.

Using CLASS and FLASK, we produced 500 simulated noisy ellipticity maps, at five redshift bins, for 1500 sets of cosmological parameters, making each parameter vary randomly within 1σ , and 2σ for a smaller sample, from the values measured by the constraints on PC18. On each map, we measured the HOM, MFs, Betti numbers, and graph statistics higher order estimators at different thresholds for a total of 29 features per redshift bin. We created several datasets to investigate how the size, the accuracy, and the smoothing of the training sample affects the results obtained on the test set. In the AVG100 dataset, we averaged each feature over 100 maps (corresponding to a total area of 2500 deg^2 , obtaining five independent realizations for cosmology and a total of 7500 independent realizations) and we applied Gaussian smoothing with $\theta_s = \{2', 4', 6'\}$. Both the AVG300 and AVG500 datasets contain one independent realizations for cosmology and a total of 1500 independent realizations, but in the first dataset we averaged each feature over 300 maps (corresponding to a total area of 7500 deg^2), with smoothing scales $\theta_s = \{2', 4', 6'\}$, while in the second the average was performed

using 500 maps (corresponding to a total area of $12\,500\text{ deg}^2$), for smoothing $\theta_s = 2'$ only.

We performed the model selection comparing different machine learning algorithms and found out that the best performing model is also the simplest one, that is, the linear regression. As we expected, the score decreases increasing the smoothing scale, and more severely so for the w_0 , σ_8 , and n_s cosmological parameters, which appear to be more affected by the loss of information due to the smoothing, compared to Ω_M and Ω_Λ . We observed that the precision of the feature measurements (i.e., the S/N) has to be favored over the number of independent realizations per cosmology in the training dataset because the score generally increases with the number of maps used to perform the averages, namely, with the total survey area. In fact, we obtained a better performance with the AVG500 dataset, which contains only one realization per cosmology but a higher feature S/N, compared to the AVG100 dataset, which contains five realizations per cosmology but lower feature S/N.

We found the best scores for the AVG500 dataset with smoothing $\theta_s = 2'$. We were able to accurately predict w_0 and σ_8 with a score $R^2 \sim 0.8$, followed by Ω_M and Ω_Λ with $R^2 \sim 0.7$, and n_s with $R^2 \sim 0.3$. The remaining parameters, H_0 and ω_b , could not be measured with our approach. On one hand, this confirms the greatest sensitivity of WL to certain cosmological parameters, as expected from previous cosmic shear studies. On the other hand, considering the lack of external constraints on H_0 , ω_b , and n_s , it could be surprising to an extent that one of the best measured parameters is indeed w_0 , even if this result could probably be due in part to the fact that we kept fixed w_a , the parameter that controls the evolution of the dark energy equation of state. The other interesting aspect of this work consists in the investigation of the importance of each feature in the measurement of the different cosmological parameters. The new estimators that we introduced, the graph statistics, resulted to be very promising, contributing effectively to the prediction of all parameters (remarkably so for Ω_M and Ω_Λ), along with MFs that confirmed their utility even when applied to shear maps. The HOM are important for the measurement of w_0 and especially of σ_8 , while the Betti numbers contribute less compared to the other estimators. In terms of redshift, the majority of the information comes for low-medium z for Ω_M and Ω_Λ , low z for n_s and σ_8 , and from all bins but with a peak at medium-high z for w_0 .

We also discussed the limitations of this work, which consist mainly of the approximations made on the simulations side, the reduced redshift and cosmological parameters range used, and the limited dataset size that we were able to produce. This work was performed using only a couple hundred thousands of computation hours but with additional resources it would be possible to increase the size of the dataset and to explore a larger portion of the cosmological parameters hyperspace from which potentially more complex relations between the features and the labels could emerge. On the other hand, such resources could also be invested in the production of more realistic simulations.

Finally, we want to stress the potential of this approach in terms of its flexibility. The fact that we do not need to develop a theoretical treatment of the statistics that we want to use in order to express their relation to cosmology and their expected value with regard to the universe we are considering opens the way to exploring different interesting estimators and even to create new ones. This method allows us to easily introduce new features and to study their relevance in the measurement of a specific cosmological parameter, along with the particular redshift range that we need to probe to access the majority of the information. Moreover, this regression model could be turned

into a classification model that we could employ to distinguish between alternative cosmologies, such as $f(R)$ modified gravity models or really any model that differs from the standard Λ CDM model. We conclude that in considering the results that we obtained with this first and somehow rough attempt at the application of this method, we believe that it is worthwhile to take the investigation of this promising approach further in the context of higher-order lensing statistics analyses in future works.

Acknowledgements. CP and VFC are funded by Italian Space Agency (ASI) through contract *Euclid* - IC (I/031/10/0) and acknowledge financial contribution from the agreement ASI/INAF/I/023/12/0. We acknowledge the support from the grant MIUR PRIN 2015 Cosmology and Fundamental Physics: illuminating the Dark Universe with *Euclid*.

References

- Aihara, H., Arimoto, N., Armstrong, R., et al. 2017, *PASJ*, 70, s4
 Alam, S., Ata, M., Bailey, S., et al. 2017, *MNRAS*, 470, 2617
 Bartelmann, M., & Maturi, M. 2017, *Scholarpedia*, 12, 32440
 Bentley, J. L. 1975, *Commun. ACM*, 18, 509
 Betti, E. 1870, *Annali di Matematica Pura ed Applicata* (1867–1897), 4, 140
 Bishop, C. M. 2006, *Pattern Recognition and Machine Learning* (Springer)
 Blas, D., Lesgourgues, J., & Tram, T. 2011, *J. Cosmol. Astropart. Phys.*, 2011, 034
 Breiman, L. 2001, *Mach. Learn.*, 45, 5
 Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. 1984, *Classification and Regression Trees*, 432 (Belmont, CA: Wadsworth International Group), 151
 Bruderer, C., Chang, C., Refregier, A., et al. 2016, *ApJ*, 817, 25
 Chisari, N. E., Richardson, M. L. A., Devriendt, J., et al. 2018, *MNRAS*, 480, 3962
 Clerkin, L., Kirk, D., Manera, M., et al. 2017, *MNRAS*, 466, 1444
 de Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., & Valentijn, E. A. 2012, *Exp. Astron.*, 35, 25
 Delfinado, C. J. A., & Edelsbrunner, H. 1993, *Proceedings of the Ninth Annual Symposium on Computational Geometry*, 232
 DES Collaboration (Abbott, T., et al.) 2005, *The Dark Energy Survey*, Tech. rep.
 DES Collaboration (Abbott, T. M. C., et al.) 2018, *Phys. Rev. D*, 98, 043526
 Dio, E. D., Montanari, F., Lesgourgues, J., & Durrer, R. 2013, *J. Cosmol. Astropart. Phys.*, 2013, 044
 Dunkley, J., Komatsu, E., Nolta, M. R., et al. 2009, *ApJS*, 180, 306
 Edelsbrunner, H., & Harer, J. 2008, *Discrete Comput. Geom. - DCG*, 453
 Euclid Collaboration (Blanchard, A., et al.) 2019a, *Euclid Preparation: VII. Forecast validation for Euclid Cosmological Probes*
 Euclid Collaboration (Knabenhans, M., et al.) 2019b, *Euclid preparation: II. The EUCLIDEMULATOR - A Tool to Compute the Cosmology Dependence of the Nonlinear Matter Power Spectrum*
 Euclid Collaboration (Guglielmo, V., et al.) 2020, *Euclid preparation: VIII. The Complete Calibration of the Colour-Redshift Relation survey: VLT/KMOS observations and data release*
 Fawcett, T. 2001, *Proceedings of the 2001 IEEE International Conference on Data Mining, ICDM '01* (USA: IEEE Computer Society), 131
 Fluri, J., Kacprzak, T., Refregier, A., et al. 2018, *Phys. Rev. D*, 98, 123518
 Fluri, J., Kacprzak, T., Lucchi, A., et al. 2019, *Phys. Rev. D*, 100, 063514
 Friedman, J. 2002, *Comput. Stat. Data Anal.*, 38, 367
 Fu, L., Kilbinger, M., Erben, T., et al. 2014, *MNRAS*, 441, 2725
 Ghosh, B., Durrer, R., & Schaefer, B. M. 2020, *ArXiv e-prints* [arXiv:2005.04604]
 Green, J., Schechter, P., Baltay, C., et al. 2012, *ArXiv e-prints* [arXiv:1208.4012]
 Gupta, A., Matilla, J. M. Z., Hsu, D., & Haiman, Z. 2018, *Phys. Rev. D*, 97, 103515
 Hamana, T., Shirasaki, M., Miyazaki, S., et al. 2020, *PASJ*, 72, 16
 Harnois-Déraps, J., van Waerbeke, L., Viola, M., & Heymans, C. 2015, *MNRAS*, 450, 1212
 Hastie, T., Tibshirani, R., & Friedman, J. 2009, *The Elements of Statistical Learning: Data Mining, Inference and Prediction* (Springer Science & Business Media)
 Hikage, C., Oguri, M., Hamana, T., et al. 2019, *PASJ*, 71, 43
 Hilbert, S., Hartlap, J., & Schneider, P. 2011, *A&A*, 536, A85
 Hildebrandt, H., Viola, M., Heymans, C., et al. 2017, *MNRAS*, 465, 1454
 Hong, S., Jeong, D., Hwang, H. S., et al. 2020, *MNRAS*, 493, 5972

- Jeffrey, N., Abdalla, F. B., Lahav, O., et al. 2018, *MNRAS*, **479**, 2871
- Jeffrey, N., Lanusse, F., Lahav, O., & Starck, J.-L. 2020, *MNRAS*, **492**, 5023
- Joshi, B. A., Cohen, S., Windhorst, R. A., et al. 2019, *ApJ*, **883**, 157
- Jullo, E., Pires, S., Jauzac, M., & Kneib, J.-P. 2013, *MNRAS*, **437**, 3969
- Kacprzak, G. G., Vander Vliet, J. R., Nielsen, N. M., et al. 2019, *ApJ*, **870**, 137
- Kilbinger, M. 2015, *Rep. Prog. Phys.*, **78**, 086901
- Kim, S.-J., Koh, K., Lustig, M., Boyd, S., & Gorinevsky, D. 2008, *IEEE J. Select. Top. Signal Process.*, **1**, 606
- Köhlinger, F., Viola, M., Joachimi, B., et al. 2017, *MNRAS*, **471**, 4412
- Kratochvil, J. M., Lim, E. A., Wang, S., et al. 2011, in *American Astronomical Society Meeting Abstracts #217*, *Bull. Am. Astron. Soc.*, **43**, 225.02
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, *ArXiv e-prints* [arXiv:1110.3193]
- Ling, C., Wang, Q., Li, R., et al. 2015, *Phys. Rev. D*, **92**, 064024
- LSST Science Collaboration (Abell, P. A., et al.), 2009, *ArXiv e-prints* [arXiv:0912.0201]
- Marques, G. A., Liu, J., Matilla, J. M. Z., et al. 2019, *J. Cosmol. Astropart. Phys.*, **2019**, 019
- Matsubara, T. 2010, *Phys. Rev. D*, **81**, 083505
- Matsubara, T., & Jain, B. 2001, *ApJ*, **552**, L89
- Mawdsley, B., Bacon, D., Chang, C., et al. 2020, *MNRAS*, **565**,
- McKay, M. D., Beckman, R. J., & Conover, W. J. 1979, *Technometrics*, **21**, 239
- Merten, J., Giocoli, C., Baldi, M., et al. 2019, *MNRAS*, **487**, 104
- Munkres, J. R. 1984, *Elements of Algebraic Topology*, Addison (Wesley Publishing Company, Inc)
- Munshi, D., Valageas, P., van Waerbeke, L., & Heavens, A. 2008, *Phys. Rep.*, **462**, 67
- Murphy, K. P. 2012, *Machine Learning: a Probabilistic Perspective* (MIT press)
- Omohundro, S. M. 1989, *Five Balltree Construction Algorithms*, *Tech. rep.* International Computer Science Institute 1947 Center Street, Suite 600, Berkeley, California 94704
- Park, C., Pranav, P., Chingangbam, P., et al. 2013, *J. Korean Astron. Soc.*, **46**, 125
- Parroni, C., Cardone, V. F., Maoli, R., & Scaramella, R. 2020, *A&A*, **633**, A71
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *J. Mach. Learn. Res.*, **12**, 2825
- Peel, A., Lalande, F., Starck, J.-L., et al. 2019, *Phys. Rev. D*, **100**, 023508
- Petri, A., Haiman, Z., Hui, L., May, M., & Kratochvil, J. M. 2013, *Phys. Rev. D*, **88**, 123002
- Pires, S., Starck, J. L., Amara, A., et al. 2009, *MNRAS*, **1265**
- Pires, S., Vandenbussche, V., Kansal, V., et al. 2020, *A&A*, **638**, A141
- Planck Collaboration XVI. 2014, *A&A*, **571**, A16
- Planck Collaboration XIII. 2016, *A&A*, **594**, A13
- Planck Collaboration VI. 2020, *A&A*, **641**, A6
- Pranav, P., Edelsbrunner, H., van de Weygaert, R., et al. 2017, *MNRAS*, **465**, 4281
- Pranav, P., van de Weygaert, R., Vegter, G., et al. 2019, *MNRAS*, **485**, 4167
- Pratten, G., & Munshi, D. 2012, *MNRAS*, **423**, 3209
- Price, M. A., Cai, X., McEwen, J. D., et al. 2020, *MNRAS*, **492**, 394
- Price, M. A., McEwen, J. D., Pratley, L., & Kitching, T. D. 2021, *MNRAS*, **500**, 5436
- Rasmussen, C. E., & Williams, C. K. I. 2006, *Gaussian Processes for Machine Learning* (The MIT Press)
- Riess, A. G., Casertano, S., Yuan, W., et al. 2018, *ApJ*, **861**, 126
- Rifkin, R. M., & Lippert, R. A. 2007, *Notes on regularized least squares*, *Tech. rep.*, *Technical Report MIT-CSAIL-TR-2007-025* (Computer Science and Artificial Intelligence Laboratory, MIT)
- Sato, J., Takada, M., Jing, Y. P., & Futamase, T. 2001, *ApJ*, **551**, L5
- Scaramella, R., Cen, R., & Ostriker, J. P. 1993, *ApJ*, **416**, 399
- Schneider, A., Teyssier, R., Stadel, J., et al. 2019, *JCAP*, **2019**, 020
- Schneider, A., Refregier, A., Grandis, S., et al. 2020a, *JCAP*, **2020**, 020
- Schneider, A., Stoira, N., Refregier, A., et al. 2020b, *JCAP*, **2020**, 019
- Semboloni, E., Schrabback, T., van Waerbeke, L., et al. 2011, *MNRAS*, **410**, 143
- Shirasaki, M., & Yoshida, N. 2014, *ApJ*, **786**, 43
- Smola, A. J., & Schölkopf, B. 2004, *Stat. Comput.*, **14**, 199
- Takada, M., & Jain, B. 2003, *MNRAS*, **344**, 857
- Takada, M., & Jain, B. 2004, *MNRAS*, **348**, 897
- Tang, B. 1993, *J. Am. Stat. Assoc.*, **88**, 1392
- Taruya, A., Takada, M., Hamana, T., Kayo, I., & Futamase, T. 2002, *ApJ*, **571**, 638
- Tibshirani, R., Hastie, T., & Friedman, J. 2010, *J. Stat. Software*, **33**
- Troxel, M. A., MacCrann, N., Zuntz, J., et al. 2018, *Phys. Rev. D*, **98**, 043528
- Vicinanza, M., Cardone, V. F., Maoli, R., Scaramella, R., & Er, X. 2018, *Phys. Rev. D*, **97**, 023519
- Vicinanza, M., Cardone, V. F., Maoli, R., et al. 2019, *Phys. Rev. D*, **99**, 043534
- Wei, C., Li, G., Kang, X., et al. 2018, *ApJ*, **853**, 25
- Xavier, H. S., Abdalla, F. B., & Joachimi, B. 2016, *MNRAS*, **459**, 3693
- Zürcher, D., Fluri, J., Sgier, R., Kacprzak, T., & Refregier, A. 2021, *JCAP*, **01**, 028

Appendix A: Models description

Here, we briefly overview the methods that we use and compare in Sect. 4. We recall that we used the scikit-learn (Pedregosa et al. 2011) library implementation of all the listed algorithms. We refer to comprehensive machine learning books for the theoretical background (e.g., Bishop 2006; Hastie et al. 2009; Murphy 2012) and to the Python Scikit-learn library page on regression problems² for the technical description of each method. When possible, we will refer to more specific resources for the details on the particular algorithm implementation that we used, contained in the library.

– Linear regression: Linear model with as many coefficients as the number of features. It aims to minimize the residual sum of squares between the true labels and the labels predicted by the linear approximation. Calling \mathbf{X} the feature matrix, \mathbf{w} the coefficient vector, and \mathbf{y} the labels, we want to solve a problem of the form:

$$\min_w \|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2.$$

– Ridge regression: Penalized linear model with as many coefficients as the number of features. It adds a L2-norm penalty term, which controls the size of the coefficients, to the residual sum of squares that has to be minimized. The penalty term is controlled by a hyperparameter α . This corresponds to the following solution of the problem (see Rifkin & Lippert 2007, for theoretical and implementation details):

$$\min_w \|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2 + \alpha\|\mathbf{w}\|_2^2.$$

– Kernel ridge regression: Ridge regression with the application of a kernel, i.e., a function $k(\mathbf{x}, \mathbf{x}') : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ that measures similarity between any two points \mathbf{x}, \mathbf{x}' of the feature space \mathcal{X} . It allows for the learning of a linear function in the space induced by the kernel that corresponds to a nonlinear function in the original space. The model is determined by the kernel choice and by the regularization hyperparameter α .

– Bayesian ridge regression: Probabilistic regression model with as many coefficients as the number of features. It imposes a prior over the coefficients \mathbf{w} in the form of a spherical Gaussian

$$p(\mathbf{w}|\lambda) = \mathcal{N}(\mathbf{w}|0, \lambda^{-1}I_p),$$

and the output is assumed to be a Gaussian distribution around $\mathbf{X}\mathbf{w}$,

$$p(\mathbf{y}|\mathbf{X}, \mathbf{w}, \alpha) = \mathcal{N}(\mathbf{y}|\mathbf{X}\mathbf{w}, \alpha),$$

where α and λ are two regularization hyperparameters, which control the precision of the estimate and are computed from the data with the assumption of uninformative priors. The aim is to maximize the log marginal likelihood of the model.

– Lasso regression: Penalized linear model with sparse coefficients. It adds a L1-norm penalty term, which reduces the number of features used in the regression, to the residual sum of squares that has to be minimized. The penalty term is controlled by a hyperparameter α . This corresponds to the solution of the problem (see Tibshirani et al. 2010 and Kim et al. 2008 for theoretical and implementation details):

$$\min_w \|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2 + \alpha\|\mathbf{w}\|_1.$$

² https://scikit-learn.org/stable/supervised_learning.html#supervised-learning

– Support vector machine: Similar to kernel ridge regression, but instead of the squared error function it uses the ϵ -insensitive error function which learns a sparse model, ignoring errors that are smaller than ϵ . This corresponds to the solution of the problem:

$$\min_w V_\epsilon(\mathbf{X}\mathbf{w} - \mathbf{y}) + \alpha\|\mathbf{w}\|_2^2,$$

with V_ϵ the ϵ -insensitive error function,

$$V_\epsilon(r) = \begin{cases} 0, & |r| \leq \epsilon \\ |r| - \epsilon. & \text{otherwise} \end{cases}$$

The model is therefore given by the choice of the kernel, the ϵ hyperparameter, and the regularization hyperparameter. See Smola & Schölkopf (2004) for a detailed description of support vector machine regression theory and algorithms.

– K nearest neighbors: The label corresponding to the set of input features is given by the mean of the label values of the k nearest neighbors points in the feature space. The number of neighbors to consider is a hyperparameter of the model and different metrics can be used to calculate the distance between points in the feature space. We used the default standard Euclidean distance. See Bentley (1975) for the K-D Tree algorithm and Omohundro (1989) for the Ball Tree algorithm.

– Gaussian processes: Given the features \mathbf{x}_i and the corresponding label y_i , we want to find the function:

$$y_i = f(\mathbf{x}_i) + \epsilon,$$

where ϵ is the noise, assumed to be Gaussian $\mathcal{N}(0, \sigma_n)$. We need then to infer a distribution over functions given the data $p(f|\mathbf{X}, \mathbf{y})$ and from that make predictions on test points (\mathbf{x}_*, y_*) calculating the mean of the conditional distribution $p(f_*|\mathbf{x}_*, \mathbf{X}, \mathbf{y})$. We can use Gaussian Processes to solve this problem. A Gaussian Process is a collection of random variables, any finite subset of which have a joint Gaussian distribution, and it is completely specified by its mean function and its covariance function. In the noisy case, given a kernel, we can write the covariance of the prior distribution over the target functions $\mathbf{f} = f(\mathbf{x})$ as

$$\text{cov}(f(\mathbf{x}_i), f(\mathbf{x}_j)) = k(\mathbf{x}_i, \mathbf{x}_j) + \sigma_n^2 \delta_{ij},$$

where k is a kernel function. In order to model the predictive distribution, we can apply the Bayes theorem and use the prior to condition the training data to model the joint distribution $p(\mathbf{f}, \mathbf{f}_*)$ of the training functions \mathbf{f} and the functions in test points \mathbf{f}_* , which, by definition of Gaussian processes, will be a joint Gaussian

$$\begin{pmatrix} \mathbf{f} \\ \mathbf{f}_* \end{pmatrix} = \mathcal{N} \sim \left(\mathbf{0}, \begin{pmatrix} K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 I & K(\mathbf{X}, \mathbf{X}_*) \\ K(\mathbf{X}_*, \mathbf{X}) & K(\mathbf{X}_*, \mathbf{X}_*) \end{pmatrix} \right),$$

where K is the covariance matrix and $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$. From $p(\mathbf{f}, \mathbf{f}_*)$ we can then calculate the posterior predictive distribution $p(f_*|\mathbf{x}_*, \mathbf{X}, \mathbf{y})$. The model is defined by the kernel choice. The kernel hyperparameters are fitted from the data. See Rasmussen & Williams (2006) for a theoretical overview and Algorithm 2.1 therein for the implementation details.

– Decision tree: Non-parametric method that creates a model applying a series of binary decision rules to the features, through a series of nodes. At each node, the sample is divided in two subsamples using the best split, i.e., the split that corresponds to the binary decision that minimizes the mean squared error

between the true labels and the predicted labels, among all possible decisions (one for each feature). Each subsample is in turn split in two and the procedure is iterated until the terminal nodes, called leaf nodes, are reached and a prediction is given. The structure created in this way is called the tree. The maximum depth of the tree is a hyperparameter of the model. See [Breiman et al. \(1984\)](#) for a description of classification and regression tree (CART) algorithms.

– Random forests: An ensemble of decision trees. Each tree is built from a bootstrap sample of the training data. The best split at each node of the tree is determined using a random subset of the input features. The prediction is given by averaging the results of all the trees in the ensemble. Among others, the number of trees, the maximum depth of each tree, and the number of features used to choose the split are hyperparameters of the model. See [Breiman \(2001\)](#) for a detailed description of random forests theory and implementation.

– Gradient boosting: Similar to random forests, but at each iteration the current tree is trained using the residual error of the previous tree, in order to refine the prediction of the labels. The contribution of each tree is controlled by a learning rate hyperparameter. See [Fawcett \(2001\)](#) and [Friedman \(2002\)](#) for algorithm details.

Appendix B: MSE and MAPE results

In Sect. 5, we present and discuss the results that we obtained performing the training and the prediction for each cosmological parameter on all three datasets, AVG100, AVG300, and AVG500, with smoothing $\theta_s = 2'$ and on the AVG100 and the AVG300 datasets with the additional smoothing scales $\theta_s = \{4', 6'\}$, using the R^2 score as the only metric to evaluate the different model performances.

Here, we expand on those results, focusing on the analysis of the R^2 score values in conjunction with the mean squared error (MSE) and with the mean absolute percentage error (MAPE), which are defined as

$$\begin{aligned} \text{MSE} &= \frac{1}{N} \sum (y_{\text{pred}} - y_{\text{true}})^2 \\ \text{MAPE} &= \frac{1}{N} \sum \left| \frac{(y_{\text{pred}} - y_{\text{true}})}{y_{\text{true}}} \right| \times 100, \end{aligned} \quad (\text{B.1})$$

where y_{true} are the true labels, y_{pred} are the predicted labels, and N is the total number of data points (i.e., independent realizations in the dataset).

In Tables B.1–B.3, we add to the scores already given in Sect. 5 (in Tables 4–6) the MSE, and the MAPE that we obtained in each case. The dash in the R^2 column represents a negative value of the score, we do not report the values of the MSE and MAPE in those cases. We notice that parameters that have a low-medium score ($R^2 < 0.5$), as H_0 , n_s , and σ_8 in some instances in Tables B.1–B.3 present a low value of the MSE and/or of the MAPE. This is somewhat counter-intuitive because we would expect high MSE and MAPE for low scores.

We can explain this result by rewriting Eq. (9) using the following relation:

$$\begin{aligned} \text{TSS} &= \text{RSS} + \text{ESS} \\ \text{ESS} &= \sum (y_{\text{pred}} - \langle y_{\text{true}} \rangle)^2, \end{aligned} \quad (\text{B.2})$$

where RSS is the residual sum of squares, TSS is the total sum of squares, as defined in Eq. (9), and ESS is the explained sum of squares or the sum of squares due to regression. Equation (9) then becomes

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}} = \frac{\text{ESS}}{\text{TSS}}, \quad (\text{B.3})$$

which can be interpreted as the total variance explained by the model over the total variance or the proportion of the variance in the true label values that is predictable from the features. In other words, the ESS and the TSS measure how much variation there is in the predicted label values and in the true label values, respectively.

This means that a low R^2 score corresponds to $\text{ESS} \ll \text{TSS}$ so that the model will produce predicted values that are too close to the mean value of the true label distribution compared to the actual variance of the true label distribution. An example is a constant or almost constant model that always predicts the mean value of the true label distribution. However, if at the same time the true label distribution has a low TSS value, that is, low variance (as is the case when considering the small range of variation of the cosmological parameters in the simulations), this model could lead to low MSE and MAPE because, on average, the predicted labels will be close to the true labels, being both concentrated around the mean. This means that the MSE and MAPE alone have no real meaning in evaluating the performance of a model and have to be considered in conjunction with the R^2 score.

We can look at Fig. 8 to better understand this concept. As explained in Sect. 5, the true labels are shown versus the predicted labels in Fig. 8 for each cosmological parameter, using the AVG100, the AVG300, and the AVG500 datasets, with smoothing $\theta_s = 2'$. The dots represent individual predictions while the shaded areas correspond to the 1σ region, obtained dividing the test sample into ten bins of the true label values and calculating the mean and standard deviation of the predictions inside each bin. The more the colored region for each given parameter aligns along the dashed black diagonal, the better will be the prediction obtained with such model. The black dots are the values of the constraints on PC18 parameters with the respective error, reported also on the y -axis as reference. We can see that H_0 and ω_b in all three versions, and n_s in the AVG100 version and slightly in the AVG300 version, fall in the situation described above. The predicted model, in fact, is represented by a more or less horizontal region and the individual predictions correspond to a cloud of dots concentrated around the mean value. This results in a bad model that is not able to predict the true label values over the entire range considered, especially for extreme values, which explains the low R^2 score. Because the cloud of individual predictions is centered around the mean with a small scatter, the MSE and MAPE have also a low value.

The remaining parameters show an almost perfect alignment along the diagonal, especially for the AVG500 version, confirming the good R^2 score results presented in Table B.1.

Table B.1. R^2 score, MSE, and MAPE for the three datasets, AVG100, AVG300, and AVG500 with smoothing scale $\theta_s = 2'$, for each cosmological parameter.

	AVG100, $\theta_s = 2'$			AVG300, $\theta_s = 2'$			AVG500, $\theta_s = 2'$		
	R^2	MSE	MAPE	R^2	MSE	MAPE	R^2	MSE	MAPE
H_0	0.16	0.30	0.6%	0.09	0.32	0.6%	0.03	0.34	0.7%
ω_b	–	–	–	–	–	–	–	–	–
Ω_M	0.61	2.30×10^{-5}	1.2%	0.64	2.29×10^{-5}	1.2%	0.70	1.89×10^{-5}	1.1%
Ω_Λ	0.61	2.30×10^{-5}	0.5%	0.64	2.29×10^{-5}	0.5%	0.70	1.89×10^{-5}	0.5%
w_0	0.65	7.34×10^{-4}	2.0%	0.75	4.57×10^{-4}	1.6%	0.81	3.51×10^{-4}	1.5%
n_s	0.16	2.22×10^{-5}	0.4%	0.23	1.78×10^{-5}	0.3%	0.32	1.58×10^{-5}	0.3%
σ_8	0.56	2.74×10^{-5}	0.5%	0.73	1.68×10^{-5}	0.4%	0.80	1.25×10^{-5}	0.4%

Notes. The dash in the R^2 column represents a negative value of the score, we do not report the values of the MSE and MAPE in those cases. As expected, increasing the number of maps per simulation improves the performance of the algorithm.

Table B.2. Same as in Table 4, but for the AVG100 and AVG300 datasets with smoothing scale $\theta_s = 4'$.

	AVG100, $\theta_s = 4'$			AVG300, $\theta_s = 4'$		
	R^2	MSE	MAPE	R^2	MSE	MAPE
H_0	0.16	0.35	0.7%	–	–	–
ω_b	–	–	–	–	–	–
Ω_M	0.55	2.57×10^{-5}	1.2%	0.60	2.51×10^{-5}	1.2%
Ω_Λ	0.55	2.57×10^{-5}	0.6%	0.60	2.51×10^{-5}	0.5%
w_0	0.62	8.53×10^{-4}	2.3%	0.68	5.82×10^{-4}	1.9%
n_s	0.06	2.60×10^{-5}	0.4%	–	–	–
σ_8	0.52	3.17×10^{-5}	0.5%	0.64	2.27×10^{-5}	0.5%

Table B.3. Same as in Table 4 but for the AVG100 and AVG300 datasets with smoothing scale $\theta_s = 6'$.

	AVG100, $\theta_s = 6'$			AVG300, $\theta_s = 6'$		
	R^2	MSE	MAPE	R^2	MSE	MAPE
H_0	0.18	0.34	0.7%	0.06	0.33	0.7%
ω_b	–	–	–	–	–	–
Ω_M	0.55	2.57×10^{-5}	1.3%	0.61	2.46×10^{-5}	1.2%
Ω_Λ	0.55	2.57×10^{-5}	0.6%	0.61	2.46×10^{-5}	0.5%
w_0	0.60	9.17×10^{-4}	2.3%	0.60	7.37×10^{-4}	2.1%
n_s	0.04	2.68×10^{-5}	0.4%	–	–	–
σ_8	0.47	3.50×10^{-5}	0.6%	0.52	3.00×10^{-5}	0.5%

Notes. Comparing this results with those in Tables 4 and 5, we observe that overall the score decreases with increasing smoothing scale.