

Complementary consistency test of the Copernican principle via Noether's theorem and machine learning forecasts

Rubén Arjona^{*} and Savvas Nesseris[†]

Instituto de Física Teórica UAM-CSIC, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain

 (Received 22 May 2021; accepted 2 November 2021; published 30 November 2021)

The Copernican principle (*CP*), i.e. the assumption that we are not privileged observers of the Universe, is a fundamental tenet of the standard cosmological model. A violation of this postulate implies the possibility that the apparent cosmic acceleration could be explained without the need of a cosmological constant, dark energy or covariant modifications of gravity. In this paper we present a new test of the *CP* relating the distance and the expansion rate, derived via Noether's theorem, which is complementary to other tests found in the literature. We also simulate fiducial data based on upcoming stage IV galaxy surveys and use them to reconstruct the Hubble rate $H(z)$ and the angular diameter distance $d_A(z)$ in order to forecast how well our null test can constrain deviations from the cosmological constant model. We find that our new test can easily rule out several scenarios based on the Lemaître-Tolman-Bondi void model at confidence of $\gtrsim 3\sigma$ at middle to high redshifts ($z > 0.5$).

DOI: [10.1103/PhysRevD.104.103532](https://doi.org/10.1103/PhysRevD.104.103532)

I. INTRODUCTION

The standard cosmological paradigm is based on two fundamental assumptions: first, that the dynamics of space-time are governed by Einstein's field equations and second, that the Universe is homogeneous and isotropic at scales larger than ~ 100 Mpc, a hypothesis normally referred to as the cosmological principle, which is considered to be a generalization of the Copernican principle (*CP*). The latter is one of the pillars of modern cosmology, stating that we do not occupy a special place in the Universe, or in other words, that any point in space must be equivalent to any other [1]. This leads to the framework of an homogeneous and isotropic background spacetime governed by the Friedmann-Lemaître-Robertson-Walker (FLRW) metric, which describes the geometry of the universe in terms of the scale factor $a(t)$, which obeys the Friedmann equation [2].

Clearly, any violations of the *CP* would disprove homogeneity and would provide an explanation for the observed accelerated expansion of the Universe without the need for a dark energy component. The latter could in fact have several possible explanations, such as modified gravity theories, global inhomogeneities such as a void model or novel dark fluid components currently unobserved in the laboratories [3–5].

Void models have the particularity that they do not employ any form of dark energy component as the accelerating expansion of the Universe is interpreted from the fact that we

live close to the center of a large underdense region, thus having the observer in a special place in the local Universe [6], or with backreaction which might mimic acceleration via the nonlinear effect of inhomogeneities [7,8].

As of now, the *CP* has been tested with different observations such as radio-astronomy [9], time drift of cosmological redshift [10], using type Ia supernovae [6,11], the integrated Sachs Wolfe effect [12], galaxy correlations and the baryon acoustic oscillations [13], the Hubble parameter [14], machine learning and cosmological distance probes [15], peculiar velocities [16], the cosmic microwave background (CMB) temperature and polarization spectrum [17], the spectral distortion of the CMB spectrum [18], galaxy surveys [19], the first order anisotropic kinetic Sunyaev Zel'dovich (kSZ) effect [20,21] and finally with a plethora of cosmological data that can be used to constrain spatial homogeneity [22]. Finally, for a representation of the current state of constraints on LTB models using various data see Ref. [23].

The simplest inhomogeneous models of the Universe are given by a spherically symmetric distribution of matter, which is mathematically described by a Lemaître-Tolman-Bondi (LTB) spacetime [24], which has been shown it can produce a Hubble diagram which in the past was consistent with observations, see Ref. [25] for a broad overview and Refs. [26,27] for some more recent constraints, but with more recent data it has been realized that simple void models cannot be used as an alternative to dark energy. Specifically, LTB models where decaying modes are not present produce a large kSZ signal [20,28,29], while models with large decaying modes, and correspondingly

^{*}ruben.arjona@uam.es
[†]savvas.nesseris@csic.es

a small kSZ signal, are not viable due to y distortions [30]. Specific cases where the void LTB models can be viable require fine-tuned initial conditions, thus leading to questions about the naturalness of these models.

Since in the near future we are going to have sufficiently good and rich cosmological data we will use the LTB models as a check for our null test to show how they can be ruled out with high confidence. Probing for deviations from the cosmological constant model (Λ CDM) is nontrivial in the absence of guiding principles or laboratory data [31]. Thus, over the years several consistency tests of the Λ CDM have appeared in the literature. In general these tests are constructed so that possible deviations from Λ CDM at any redshift are apparent and easy to quantify in the form of null tests. These are formulated such that they can be computed using reasonably directly observable quantities at any redshift, thus by computing the consistency test using data from multiple redshifts, one can examine the validity of the basic assumptions of the cosmological standard model. If these assumptions hold, the null test should be independent of redshift.

In this work we present a complementary test to the well-known curvature test of Ref. [1] that can be used to falsify the CP . This null test depends solely on distance and Hubble rate observations and is derived with the aid of Noether's theorem. The advantage of our new test is that it does not suffer from divergences and provides tighter constraints at high redshifts, as we will discuss in later sections.

Our paper is organized as follows: In Sec. II we present the theoretical formalism, our null test named $\mathcal{O}(z)$ and a description of the LTB models used to check the consistency test. In Sec. III we describe our simulated data based on an optimistic Stage IV galaxy survey and the machine learning (ML) algorithm used to reconstruct the data, namely the genetic algorithms [32]. Finally, in Sec. IV we present our results and in Sec. V we summarize our conclusions.

II. ANALYSIS

Under the assumption of a Friedmann-Lemaître-Robertson-Walker (FLRW) metric, the luminosity distance can be written as

$$d_L(z) = \frac{c(1+z)}{H_0\sqrt{-\Omega_k}} \sin\left(\sqrt{-\Omega_k} \int_0^z dz' \frac{H_0}{H(z')}\right), \quad (1)$$

where Ω_k is the curvature parameter today and $H(z)$ is the expansion rate. The luminosity distance $d_L(z)$ is related to the angular diameter distance $d_A(z)$ through the Etherington relation, i.e. $d_L(z) = (1+z)^2 d_A(z)$. Using the comoving angular diameter distance $D(z)$, defined as $D(z) = \frac{H_0}{c}(1+z)d_A(z)$, we can regroup Eq. (1) to solve for the curvature parameter Ω_k in terms of $H(z)$ and $D(z)$ as [1]

$$\Omega_k = \frac{[E(z)D'(z)]^2 - 1}{[D(z)]^2}, \quad (2)$$

where $E(z) \equiv H(z)/H_0$ is the dimensionless Hubble parameter and the prime is a derivative with respect to the redshift $' = d/dz$. The above relation allows us to estimate the spatial curvature parameter from distance and Hubble rate observations, without having to assume any particular dark energy model or other model parameters. It also allows us to test the curvature at any single redshift as it has been reconstructed in several works, see for instance Refs. [1,15,31,33–39]. Since the curvature parameter Ω_k does not depend on redshift, we can differentiate this to obtain a relation that must always equal zero. This can be expressed as [1]

$$C(z) = 1 + E(z)^2(D(z)D''(z) - D'(z)^2) + E(z)E'(z)D(z)D'(z), \quad (3)$$

where $C(z)$ has to be zero at all redshifts in any model described by a FLRW metric, as was originally shown in Ref. [1]. In Sec. IV we will present constraints on this test with an upcoming Stage IV survey along with a complementary CP test inspired from Noether's theorem. The advantage of using Noether's theorem to make a complementary test is that by taking into account the symmetries of the system of equations that describe the expansion of the Universe, we can reduce the order of the differential equations that appear in the final test. This allows us to keep the errors of the reconstructions smaller, as higher order derivatives of noise data tend to make the reconstructions less robust at high redshifts, as we demonstrate in what follows.

A. Lagrangian formalism and null test

We now present a complementary test of $C(z)$ to probe the CP . Using Eq. (2) we can solve for $D'(z)^2$, which will be given by

$$D'(z)^2 = \frac{1 + D(z)^2\Omega_k}{E(z)^2}, \quad (4)$$

and inserting this relation into Eq. (3) we have

$$D''(z) + \frac{E'(z)}{E(z)}D'(z) - \frac{\Omega_k}{E(z)^2}D(z) = 0. \quad (5)$$

To find a null test that involves the distance measure $D(z)$, we will make use of the Lagrangian formalism. The first step is to find a Lagrangian for Eq. (5) and, with the help of Noether's theorem, to find an associated conserved quantity. For a description of the Noether symmetry approach and applications for null tests see [40,41].

In a nutshell, if we assume that the Lagrangian can be written as $\mathcal{L} = \mathcal{L}(z, D(z), D'(z))$. Then, the Euler-Lagrange equations are

$$\frac{\partial \mathcal{L}}{\partial D} - \frac{d}{dz} \frac{\partial \mathcal{L}}{\partial D'} = 0. \quad (6)$$

So, let us assume a Lagrangian of the form

$$\mathcal{L} = T - V, \quad (7)$$

$$T = \frac{1}{2} f_1(z, E(z)) D'(z)^2, \quad (8)$$

$$V = \frac{1}{2} f_2(z, E(z)) D(z)^2, \quad (9)$$

where the f_1 and f_2 are arbitrary functions that need to be determined so that the resulting equation after implementing the Euler-Lagrange Eq. (6) is exactly Eq. (5). Therefore, substituting the former Lagrangian in the Euler-Lagrange Eq. (6) and comparing the result with Eq. (5) we are able to get the two functions f_1 and f_2 and consequently to build the Lagrangian \mathcal{L} of the system:

$$\mathcal{L} = \frac{1}{2} E(z) D'(z)^2 + \frac{1}{2} \frac{\Omega_k}{E(z)} D(z)^2. \quad (10)$$

It is easy to see that substituting Eq. (10) into Eq. (6) results exactly to Eq. (5). Now that we have a Lagrangian we can use Noether's theorem to find a conserved quantity that will be later translated to the null test. So, if we have an infinitesimal transformation \mathbf{X} with a generator

$$\mathbf{X} = \alpha(D) \frac{\partial}{\partial D} + \frac{d\alpha(D)}{dz} \frac{\partial}{\partial D'}, \quad (11)$$

$$\frac{d\alpha(D)}{dz} \equiv \frac{\partial \alpha}{\partial D} D'(z) = \alpha'(z), \quad (12)$$

such that for the Lie derivative of the Lagrangian we have $L_{\mathbf{X}} \mathcal{L} = 0$, then

$$\Sigma = \alpha(z) \frac{\partial \mathcal{L}}{\partial D'} \quad (13)$$

is a constant of "motion" for the Lagrangian of Eq. (10). From Eq. (13) we get that

$$\Sigma = \alpha(D(z)) E(z) D'(z), \quad (14)$$

while from the Lie derivative we also obtain

$$\alpha(z) = \alpha_0 e^{-\int_0^z \frac{\Omega_k D(x)}{E(x)^2 D'(x)} dx}, \quad (15)$$

where α_0 is an integration constant. Then, the constant Σ becomes

$$\Sigma = \frac{E(z) D'(z)}{D'(0)} e^{-\int_0^z \frac{\Omega_k D(x)}{E(x)^2 D'(x)} dx}, \quad (16)$$

where we absorbed α_0 into Σ and normalized the above equation so that the null test must be equal to unity for all values of z .

Finally, to write the above null test only as a function of $E(z)$ and $D(z)$ we substitute Ω_k from Eq. (2) into Eq. (16), then the null test is given by

$$\mathcal{O}(z) = \frac{E(z) D'(z)}{D'(0)} e^{-\int_0^z \frac{E(x)^2 D'(x)^2 - 1}{E(x)^2 D(x) D'(x)} dx}. \quad (17)$$

B. LTB model

An alternative explanation, besides the cosmological constant Λ , for the current phase of accelerated expansion of the Universe is the idea of inhomogeneous universe models, where this expansion can be seen as an effective acceleration induced by our special position as observers residing inside a huge underdense region of space. These models violate the CP and a simple toy model which has been studied extensively in the literature is the spherically symmetric Lemaître-Tolman-Bondi model [24,42,43] (LTB) which describes a local void. It actually represents a family of models coming from a spherically symmetric solution of Einstein equations exerted by pressureless matter and no cosmological constant, as one still needs to provide a matter density profile [10]. The metric for our model of interest is given by

$$ds^2 = -dt^2 + X^2(r, t) dr^2 + A^2(r, t) d\Omega^2, \quad (18)$$

where $d\Omega^2 = d\theta^2 + \sin^2 \theta d\phi^2$ and the function $A(r, t)$ is analogous to the scale factor of the FLRW metric, albeit it also has a dependence on both time and the radial coordinate r . One can find a relation between $X(r, t)$ and $A(r, t)$ through the $0-r$ component of the Einstein equations, i.e $X(r, t) = A'(r, t) / \sqrt{1 - k(r)}$, where a prime denotes a derivative with respect to the coordinate r and $k(r)$ represents an arbitrary function, being similar to the role of the spatial curvature parameter.

This model can be totally described by the matter density $\Omega_m(r)$ and the Hubble expansion rate $H(r)$. We will check our consistency test with a particular LTB model known as the GBH parametrization [25]. In this case the matter and Hubble parameter profiles are given by

$$\Omega_m(r) = \Omega_{\text{out}} + (\Omega_{\text{in}} - \Omega_{\text{out}}) \frac{1 - \tanh[(r - r_0)/2\Delta r]}{1 + \tanh[r_0/2\Delta r]}, \quad (19)$$

$$H_0(r) = H_0 \left(\frac{1}{\Omega_k(r)} - \frac{\Omega_m(r)}{\sqrt{\Omega_k(r)^3}} \sinh^{-1} \sqrt{\frac{\Omega_k(r)}{\Omega_m(r)}} \right), \quad (20)$$

TABLE I. The parameters for the LTB models, where in all cases $\Omega_{m,\text{out}} = 1$ and $H_0 = 77$ km/s/Mpc. Note that the actual value of the Hubble rate today as measured by a comoving observer at $z = 0$, depends on the specific profile used. Here we assume the constrained GBH LTB profile of Ref. [25], given by Eqs. (19) and (20).

$\Omega_{m,\text{in}}$	r [Gpc]	Δr [Gpc]	LTB models
0.298	1.0	0.30	LTB1
0.197	1.5	0.45	LTB2
0.156	1.8	0.54	LTB3
0.200	2.0	0.60	LTB4

where $\Omega_k(r) = 1 - \Omega_m(r)$, Ω_{out} is the value of the matter density at infinity, Ω_{in} is the value of the matter density at the center of the void, r_0 is the size of the void and Δr represents a scale that characterises the transition to uniformity. In Table I we show the four GBH parameters used in our analysis, which correspond to characteristic voids of sizes of a few Gpc, as suggested in Ref. [25].

Note however, that in Ref. [23] the LTB model, albeit with a different profile than that of Eq. (19) and including a cosmological constant, was confronted against a plethora of cosmological data, including the Planck measurements of the CMB spectrum, BAO data, the Pantheon compilation of type Ia supernovae, local H_0 measurements, the $H(z)$ cosmic chronometers and finally, the Compton y distortion and kinetic Sunyaev-Zeldovich effect. In summary, it was found that the aforementioned data can tightly constrain the LTB model, almost at the cosmic variance level. In particular, on scales of ~ 100 Mpc structures can have a small non-Copernican contrast of just $\delta \sim 0.01$.

III. RECONSTRUCTIONS

We now describe both the mock data used and the machine learning process used to reconstruct the null test, namely the genetic algorithms.

A. Mock data

Our mock baryon acoustic oscillations (BAO) data for the angular diameter distance $d_A(z)$ and the Hubble rate $H(z)$ are based on a future upgrade of dark energy spectroscopic instrument (DESI) [44]. DESI is a survey with the goal of probing the expansion rate and large-scale structure (LSS) of the universe and can complement other future BAO surveys by extending the probed redshift range [45].

The DESI survey, whose operations started at the end of 2019, is expected to obtain optical spectra for tens of millions of galaxies and quasars up to redshift $z \sim 4$, which will allow for BAO and redshift-space distortion cosmological analyses. Our forecast data will cover the redshift range $z \in [0.05, 3.55]$, but their precision will also depend on the target population. The blue galaxies (BGs) will cover

the redshift range $z \in [0.05, 0.45]$ in five equispaced redshift bins, the luminous red galaxies (LRGs) and emission line galaxies (ELGs) will focus on $z \in [0.65, 1.85]$ with 13 equispaced redshift bins, while the Ly- α forest quasar survey will cover $z \in [1.96, 3.55]$ with 11 equispaced redshift bins [44]. The difference in the redshift distributions of the various targets (BGs, ELGs, quasi-stellar objects, etc.) are due to the fact that each of these targets needs different selection methods to accumulate sufficiently large samples of spectroscopic targets from photometric data, see for example Sec. III in Ref. [44].

To create the mocks we assume the $H(z)$ and $d_A(z)$ are uniformly distributed in the range $z \in [0.1, 3.55]$, divided into 20 equally spaced binds of step $dz = 0.2$. The $H(z_i)$ and $d_A(z_i)$ function was estimated as its theoretical value from the different cosmological models plus a Gaussian error (which can be either negative or positive) and assigning an error of 0.5% of its value to $H(z)$ and for $d_A(z)$ an error of 0.28% for $z < 1.1$ and 0.39% for $z > 1.1$, which is in agreement with a similar setup to [46]. We further assume these measurements to be uncorrelated.

Finally, we have assumed a fiducial cosmology of a flat Λ CDM model with matter density $\Omega_{m,0} = 0.3$ and Hubble constant of $H_0 = 70$ kms $^{-1}$ Mpc $^{-1}$.

B. Genetic algorithms

In this section we explain the implementation of the genetic algorithms (GAs) in our analysis. They have successfully been applied in cosmology for several reconstructions on a wide range of data, for further details see Refs. [15,47–56]. Other applications of the GA cover particle physics [57–59] and astronomy and astrophysics [60–62]. Also, other symbolic regression methods applied in physics and cosmology can be found at [63–70].

The GA can be seen as a particular type of ML methods mainly constructed to perform unsupervised regression of data, i.e. it carries out nonparametric reconstructions finding an analytic function of one or more variable that describes the data extremely well.

The GA operates by simulating the notion of biological evolution through the principle of natural selection, as conveyed by the genetic operations of mutation and crossover. Essentially, a set of test functions evolves as time goes by through the effect of the stochastic operators of crossover, i.e the joining of two or more candidate functions to form another one, and mutation, i.e a random alteration of a candidate function. This procedure is then repeated thousands of times so as to ensure convergence, while different random seeds can be used to further explore the functional space.

Due to the construction of the GA as a stochastic approach, the probability that a population of functions will give rise to offspring is normally assumed to be proportional to its fitness to the data, where in our analysis is given by a χ^2 statistic and conveys how good every

TABLE II. The grammars used in the GA analysis. Other complex forms are automatically produced by the mutation and crossover operations as described in the text.

Grammar type	Functions
Polynomials	$c, x, 1 + x$
Fractions	$\frac{x}{1+x}$
Trigonometric	$\sin(x), \cos(x), \tan(x)$
Exponentials	$e^x, x^x, (1+x)^{1+x}$
Logarithms	$\log(x), \log(1+x)$

individual agrees with the data. For the simulated data in our analysis we are assuming that the likelihoods are sufficiently Gaussian that we use the χ^2 in our GA approach. In the GA, the probability to have offspring and the fitness of each individual is proportional to the likelihood, causing an evolutionary pressure that favors the best-fitting functions in every population, hence driving the fit towards the minimum in a few generations.

In our analysis we reconstruct the Hubble rate $H(z)$ and the angular diameter distance $d_A(z)$ from the mock data created, and the course of action to its reconstruction is as follows. First, our predefined grammar was formed on the following functions: exp, log, polynomials etc. and a set of operations $+, -, \times, \div$; see Table II for the complete list.

As a prior for our $H(z)$ reconstruction we imposed that $H(z=0) = H_0$ and for the $d_A(z)$ reconstruction we assumed that $d_A(z \ll 1) = c/H_0 z + O(z^2)$, where both priors have been motivated by physical reasons, namely the fact that the value of the Hubble parameter today is the one that we actually infer from observations and the Hubble law respectively. However, we make no assumptions on the curvature of the Universe or any modified gravity or dark energy model. Furthermore, in order to avoid overfitting or any spurious reconstructions we required that all functions reconstructed by the GA are continuous and differentiable, without any singularities in the redshift probed by the data.

Once the initial population has been constructed, the fitness of each member is computed by a χ^2 statistic, using the $H(z)$ and $d_A(z)$ data points directly as input. Then, using a tournament selection process, see Ref. [47] for more details, the best-fitting functions in each generation are selected and the two stochastic operations of crossover and mutation are used. To guarantee convergence, the GA operation is repeated thousands of times and with various random seeds, to explore properly the functional space. The final output of the code is a set of functions of $H(z)$ and $d_A(z)$ that describes the Hubble rate and the angular diameter distance respectively.

The error estimates of the reconstructed function are obtained via the path integral approach, originally implemented in Refs. [49,50]. This approach consists of having an analytical estimate of the error of the reconstructed quantity by computing a path integral over all possible functions around the best fit GA that may contribute to the

likelihood, and it has been shown that this can be performed whether the data points are correlated or uncorrelated. This error reconstruction method has been exhaustively examined and tested against a bootstrap Monte Carlo by Ref. [49]. More explicitly, given a reconstructed function $f(x)$ from the GA, the path integral approach of Ref. [49] gives us the 1σ error $\delta f(x)$. This can be also compared to error propagation if one assumes that the error in a quantity is taken as $\sigma_f = f'(p)\delta p$, where p would represent a parameter. We have extensively compared our approach finding that this assumption is appropriate for the data set used here.

Finally, the reason we need the GA is that traditional inference approaches, such as Markov chain Monte Carlo simulations require a specific model for the expansion history, e.g. Λ CDM, to fit the data and the choice of this model will clearly bias the results or even miss critical features of the data. The main advantage of the GA is that, given some data, we allow the algorithm to determine the best-fit model, thus we can remain theory agnostic, as we do not have to assume a dark energy model.

IV. RESULTS

In this section we present our GA fits to the simulated data and the corresponding consistency test obtained from our reconstructions on $H(z)$ and $d_A(z)$. We want to stress that the aim of this work is not simply to rule out LTB void models as contenders of dark energy models, which has already been done so, but to present a complementary consistency test of the CP in a nontrivial but still interesting setting.

In Fig. 1 we show our new null test $\mathcal{O}(z)$ given by Eq. (17) and the $C(z)$ test given by Eq. (3), both of which can be used to find deviations from the CP through our reconstructions on the Hubble rate $H(z)$ and the angular diameter distance $d_A(z)$. Our reconstructed functions start at $z = 0.1$ which is the redshift of our first mock data point. In both cases the black solid line and the gray region correspond to the GA best fit and its 1σ error respectively. The black dashed line represents the flat Λ CDM model and we see that our reconstructions recover well the null hypothesis of the FLRW metric used to make the mocks. Recall that the Λ CDM curves are theoretical and so are precisely at $\mathcal{O}(z) = 1$ and $C(z) = 0$. The colored lines represent the four different theoretical LTB models, concretely defined in Table I.

In the left panel of Fig. 1 we show our $\mathcal{O}(z)$ test and as can be seen, it is a good discriminator of all LTB models at high and intermediate redshifts, i.e. $z \sim 0.6$ and beyond, as the errors remain consistently low at all redshifts. On the other hand, in the right panel of Fig. 1 we present the $C(z)$ test which can discriminate the LTB model from Λ CDM at intermediate redshifts $0.5 < z < 1.5$ but at low redshifts ($z < 0.3$) and high redshifts above $z \sim 1.5$ the values of

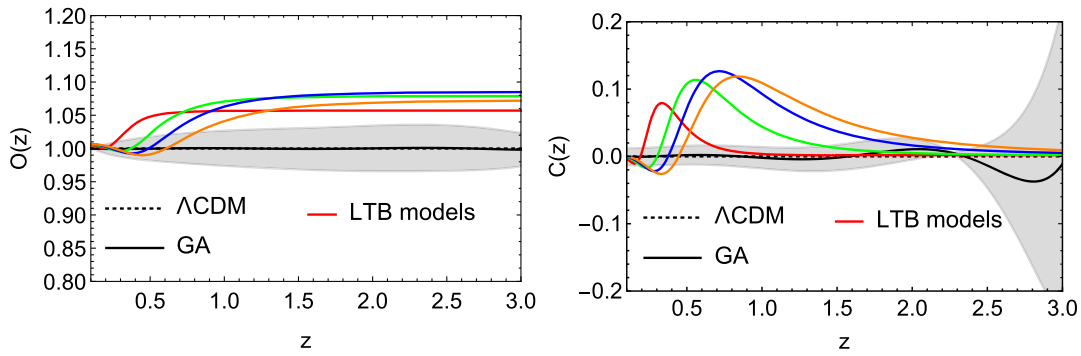


FIG. 1. Reconstructions of the tests of the Copernican principle: the $\mathcal{O}(z)$ test of Eq. (17) (left panel) and the $C(z)$ test of Eq. (5) (right panel). The black solid line corresponds to the GA best fit, the gray shaded region corresponds to the 68.3% confidence regions and the colored lines to the various LTB profiles described in Table I.

$C(z)$ of the fiducial LTB models asymptote to zero, thus being dominated by the errors.

Therefore, by comparing both panels we may infer that our test manages to detect deviations from the CP particularly well and consistently at middle to high redshifts, when the traditional $C(z)$ test does not perform equally well. Hence, our null test presented serves as a complementary consistency check of the CP and is especially useful at high redshifts.

V. CONCLUSIONS

In summary, we have presented a new consistency test of the Copernican principle, which is complementary to the curvature test of Ref. [1]. In particular, we used the Noether's theorem approach in order to obtain a conserved quantity that can be written in terms of the Hubble rate $H(z)$ and the comoving distance $D(z)$.

In order to forecast how well our new test, given by Eq. (17), can constrain deviations from the Copernican principle at large scales, we created mock datasets based on specifications of the DESI survey and using the Λ CDM model for the fiducial cosmology, for a variety of different profiles. This approach allows us to quantify any deviations using forecast mock data and plausible scenarios.

Then, to reconstruct the $\mathcal{O}(z)$ statistic given by Eq. (17) from the mock data, we preferred to use the machine learning approach, namely the GA, as this will allow us to obtain nonparametric and theory agnostic reconstructions of the data, in the form of $H(z)$ and $D(z)$, that we can in

turn use to reconstruct $\mathcal{O}(z)$. With the same functions we also reconstructed the $C(z)$ function of Ref. [1] given by Eq. (5).

Following this approach, we find that the GA with the $\mathcal{O}(z)$ statistic can correctly predict the underlying fiducial cosmology at all redshifts covered by the data, as seen in the left panel of Fig. 1 and can easily rule out several LTB scenarios at confidence of $\gtrsim 3\sigma$ at middle to high redshifts ($z > 0.5$). On the other hand, the $C(z)$ statistic, while it successfully rules out the same LTB profiles at small redshifts at a confidence of $\sim 8\sigma$ at intermediate redshifts ($0.5 < z < 1.0$), it does not fare equally well at higher redshifts ($z > 1.5$) as the errors become larger and the value of $C(z)$ asymptotes to zero, thus diminishing its predictive power.

To conclude, we find that the $\mathcal{O}(z)$ test provides complementary to other tests, information on possible deviations of homogeneity at different redshift regimes and can help test one of the fundamental assumptions of the standard cosmological model at high redshifts, something which is the goal of several current and upcoming surveys in the coming years.

ACKNOWLEDGMENTS

The authors acknowledge support from the Research Project No. PGC2018-094773-B-C32 and the Centro de Excelencia Severo Ochoa Program No. SEV-2016-0597. S. N. also acknowledges support from the Ramón y Cajal program through Grant No. RYC-2014-15843.

- [1] C. Clarkson, B. Bassett, and T. H.-C. Lu, A General Test of the Copernican Principle, *Phys. Rev. Lett.* **101**, 011301 (2008).
 [2] S. Weinberg, The cosmological constant problem, *Rev. Mod. Phys.* **61**, 1 (1989).

- [3] K. Tomita, A local void and the accelerating universe, *Mon. Not. R. Astron. Soc.* **326**, 287 (2001).
 [4] R. K. Barrett and C. A. Clarkson, Undermining the cosmological principle: Almost isotropic observations in

- inhomogeneous cosmologies, *Classical Quantum Gravity* **17**, 5047 (2000).
- [5] M.-N. Celerier, Do we really see a cosmological constant in the supernovae data?, *Astron. Astrophys.* **353**, 63 (2000).
- [6] T. Clifton, P. G. Ferreira, and K. Land, Living in a Void: Testing the Copernican Principle with Distant Supernovae, *Phys. Rev. Lett.* **101**, 131302 (2008).
- [7] T. Buchert, On average properties of inhomogeneous fluids in general relativity. I. Dust cosmologies, *Gen. Relativ. Gravit.* **32**, 105 (2000).
- [8] D. L. Wiltshire, Exact Solution to the Averaging Problem in Cosmology, *Phys. Rev. Lett.* **99**, 251101 (2007).
- [9] H. L. Bester, J. Larena, and N. T. Bishop, Testing the Copernican principle with future radio-astronomy observations, [arXiv:1705.00994](https://arxiv.org/abs/1705.00994).
- [10] J.-P. Uzan, C. Clarkson, and G. F. R. Ellis, Time Drift of Cosmological Redshifts as a Test of the Copernican Principle, *Phys. Rev. Lett.* **100**, 191303 (2008).
- [11] K. Bolejko and J. S. B. Wyithe, Testing the Copernican principle via cosmological observations, *J. Cosmol. Astropart. Phys.* **02** (2009) 020.
- [12] K. Tomita and K. T. Inoue, Probing violation of the Copernican principle via the integrated Sachs-Wolfe effect, *Phys. Rev. D* **79**, 103505 (2009).
- [13] S. February, C. Clarkson, and R. Maartens, Galaxy correlations and the BAO in a void universe: Structure formation as a test of the Copernican principle, *J. Cosmol. Astropart. Phys.* **03** (2013) 023.
- [14] Z.-S. Zhang, T.-J. Zhang, H. Wang, and C. Ma, Testing the Copernican principle with the Hubble parameter, *Phys. Rev. D* **91**, 063506 (2015).
- [15] D. Sapone, E. Majerotto, and S. Nesseris, Curvature versus distances: Testing the FLRW cosmology, *Phys. Rev. D* **90**, 023012 (2014).
- [16] W. A. Hellwing, A. Nusser, M. Feix, and M. Bilicki, Not a Copernican observer: Biased peculiar velocity statistics in the local Universe, *Mon. Not. R. Astron. Soc.* **467**, 2787 (2017).
- [17] J. P. Zibin, A. Moss, and D. Scott, Can We Avoid Dark Energy?, *Phys. Rev. Lett.* **101**, 251303 (2008).
- [18] R. R. Caldwell and A. Stebbins, A Test of the Copernican Principle, *Phys. Rev. Lett.* **100**, 191302 (2008).
- [19] F. S. Labini and Y. V. Baryshev, Testing the Copernican and Cosmological principles in the local universe with galaxy surveys, *J. Cosmol. Astropart. Phys.* **06** (2010) 021.
- [20] P. Zhang and A. Stebbins, Confirmation of the Copernican Principle at Gpc Radial Scale and above from the Kinetic Sunyaev Zel'dovich Effect Power Spectrum, *Phys. Rev. Lett.* **107**, 041301 (2011).
- [21] C.-M. Yoo, T. Kai, K.-I. Nakao, and M. Sasaki, Testing the Copernican principle with the kSZ effect, in the 19th Workshop on General Relativity and Gravitation in Japan (2009).
- [22] W. Valkenburg, V. Marra, and C. Clarkson, Testing the Copernican principle by constraining spatial homogeneity, *Mon. Not. R. Astron. Soc.* **438**, L6 (2014).
- [23] D. Camarena, V. Marra, Z. Sakr, and C. Clarkson, The Copernican principle in light of the latest cosmological data, [arXiv:2107.02296](https://arxiv.org/abs/2107.02296).
- [24] H. Bondi, Spherically symmetrical models in general relativity, *Mon. Not. R. Astron. Soc.* **107**, 410 (1947).
- [25] J. Garcia-Bellido and T. Haugboelle, Confronting Lemaitre-Tolman-Bondi models with observational cosmology, *J. Cosmol. Astropart. Phys.* **04** (2008) 003.
- [26] V. V. Luković, B. S. Haridasu, and N. Vittorio, Exploring the evidence for a large local void with supernovae Ia data, *Mon. Not. R. Astron. Soc.* **491**, 2075 (2020).
- [27] L. Kazantzidis and L. Perivolaropoulos, Hints of a local matter underdensity or modified gravity in the low z Pantheon data, *Phys. Rev. D* **102**, 023520 (2020).
- [28] J. Garcia-Bellido and T. Haugboelle, Looking the void in the eyes—The kSZ effect in LTB models, *J. Cosmol. Astropart. Phys.* **09** (2008) 016.
- [29] J. P. Zibin and A. Moss, Linear kinetic Sunyaev-Zel'dovich effect and void models for acceleration, *Classical Quantum Gravity* **28**, 164005 (2011).
- [30] J. P. Zibin, Can decaying modes save void models for acceleration?, *Phys. Rev. D* **84**, 123508 (2011).
- [31] A. Shafieloo and C. Clarkson, Model independent tests of the standard cosmological model, *Phys. Rev. D* **81**, 083537 (2010).
- [32] <https://github.com/RubenArjona>.
- [33] E. Mortsell and J. Jonsson, A model independent measure of the large scale curvature of the Universe, [arXiv:1102.4485](https://arxiv.org/abs/1102.4485).
- [34] S. Räsänen, K. Bolejko, and A. Finoguenov, New Test of the Friedmann-Lemaître-Robertson-Walker Metric Using the Distance Sum Rule, *Phys. Rev. Lett.* **115**, 101301 (2015).
- [35] B. L'Huillier and A. Shafieloo, Model-independent test of the FLRW metric, the flatness of the Universe, and nonlocal measurement of $H_0 r_d$, *J. Cosmol. Astropart. Phys.* **01** (2017) 015.
- [36] M. Denissenya, E. V. Linder, and A. Shafieloo, Cosmic curvature tested directly from observations, *J. Cosmol. Astropart. Phys.* **03** (2018) 041.
- [37] C.-G. Park and B. Ratra, Using the tilted flat- Λ CDM and the untilted nonflat Λ CDM inflation models to measure cosmological parameters from a compilation of observational data, *Astrophys. J.* **882**, 158 (2019).
- [38] S. Cao, J. Ryan, and B. Ratra, Using Pantheon and DES supernova, baryon acoustic oscillation, and Hubble parameter data to constrain the Hubble constant, dark energy dynamics, and spatial curvature, [arXiv:2101.08817](https://arxiv.org/abs/2101.08817).
- [39] N. Khadka and B. Ratra, Determining the range of validity of quasar x-ray and UV flux measurements for constraining cosmological model parameters, [arXiv:2012.09291](https://arxiv.org/abs/2012.09291).
- [40] S. Nesseris and D. Sapone, Novel null-test for the Λ cold dark matter model with growth-rate data, *Int. J. Mod. Phys. D* **24**, 1550045 (2015).
- [41] S. Nesseris, D. Sapone, and J. García-Bellido, Reconstruction of the null-test for the matter density perturbations, *Phys. Rev. D* **91**, 023004 (2015).
- [42] A. G. Lemaître, The expanding universe, *Gen. Relativ. Gravit.* **29**, 641 (1997).
- [43] R. C. Tolman, Effect of inhomogeneity on cosmological models, *Proc. Natl. Acad. Sci. U.S.A.* **20**, 169 (1934).
- [44] A. Aghamousa *et al.* (DESI Collaboration), The DESI experiment part I: Science, targeting, and survey design, [arXiv:1611.00036](https://arxiv.org/abs/1611.00036).

- [45] M. Martinelli *et al.* (EUCLID Collaboration), Euclid: Forecast constraints on the cosmic distance duality relation with complementary external probes, [arXiv:2007.16153](#).
- [46] M. Vargas-Magaña, D. D. Brooks, M. M. Levi, and G. G. Tarle (DESI Collaboration), Unraveling the Universe with DESI, in *53rd Rencontres de Moriond on Cosmology* (2018).
- [47] C. Bogdanos and S. Nesseris, Genetic algorithms and supernovae type Ia analysis, *J. Cosmol. Astropart. Phys.* **05** (2009) 006.
- [48] S. Nesseris and A. Shafieloo, A model independent null test on the cosmological constant, *Mon. Not. R. Astron. Soc.* **408**, 1879 (2010).
- [49] S. Nesseris and J. Garcia-Bellido, A new perspective on dark energy modeling via genetic algorithms, *J. Cosmol. Astropart. Phys.* **11** (2012) 033.
- [50] S. Nesseris and J. García-Bellido, Comparative analysis of model-independent methods for exploring the nature of dark energy, *Phys. Rev. D* **88**, 063521 (2013).
- [51] R. Arjona, Machine learning meets the redshift evolution of the CMB temperature, *J. Cosmol. Astropart. Phys.* **08** (2020) 009.
- [52] R. Arjona and S. Nesseris, Hints of dark energy anisotropic stress using machine learning, [arXiv:2001.11420](#).
- [53] R. Arjona and S. Nesseris, What can machine learning tell us about the background expansion of the Universe?, *Phys. Rev. D* **101**, 123525 (2020).
- [54] R. Arjona and S. Nesseris, Novel null tests for the spatial curvature and homogeneity of the Universe and their machine learning reconstructions, [arXiv:2103.06789](#).
- [55] R. Arjona and S. Nesseris, Machine learning and cosmographic reconstructions of quintessence and the Swampland conjectures, [arXiv:2012.12202](#).
- [56] R. Arjona, H.-N. Lin, S. Nesseris, and L. Tang, Machine learning forecasts of the cosmic distance duality relation with strongly lensed gravitational wave events, [arXiv:2011.02718](#).
- [57] S. Abel, D. G. Cerdeño, and S. Robles, The power of genetic algorithms: What remains of the pMSSM?, [arXiv:1805.03615](#).
- [58] B. C. Allanach, D. Grellscheid, and F. Quevedo, Genetic algorithms and experimental discrimination of SUSY models, *J. High Energy Phys.* **07** (2004) 069.
- [59] Y. Akrami, P. Scott, J. Edsjo, J. Conrad, and L. Bergstrom, A profile likelihood analysis of the constrained MSSM with genetic algorithms, *J. High Energy Phys.* **04** (2010) 057.
- [60] M. Wahde and K. Donner, Determination of the orbital parameters of the m 51 system using a genetic algorithm, *Astron. Astrophys.* **379**, 115 (2001).
- [61] V. Rajpaul, Genetic algorithms in astronomy and astrophysics, in *Proceedings, 56th Annual Conference of the South African Institute of Physics (SAIP 2011): Gauteng, South Africa, 2011* (2012), pp. 519–524.
- [62] M. Ho, M. M. Rau, M. Ntampaka, A. Farahi, H. Trac, and B. Poczos, A robust and efficient deep learning method for dynamical mass measurements of galaxy clusters, [arXiv:1902.05950](#).
- [63] S.-M. Udrescu and M. Tegmark, AI Feynman: A physics-inspired method for symbolic regression, [arXiv:1905.11481](#).
- [64] Y. Setyawati, M. Pürer, and F. Ohme, Regression methods in waveform modeling: A comparative study, *Classical Quantum Gravity* **37**, 075012 (2020).
- [65] H. Vaddireddy, A. Rasheed, A. E. Staples, and O. San, Feature engineering and symbolic regression methods for detecting hidden physics from sparse sensors, [arXiv:1911.05254](#).
- [66] K. Liao, A. Shafieloo, R. E. Keeley, and E. V. Linder, A model-independent determination of the Hubble constant from lensed quasars and supernovae using Gaussian process regression, *Astrophys. J. Lett.* **886**, L23 (2019).
- [67] E. Belgacem, S. Foffa, M. Maggiore, and T. Yang, Gaussian processes reconstruction of modified gravitational wave propagation, *Phys. Rev. D* **101**, 063505 (2020).
- [68] Y. Li, Y. Hao, W. Cheng, and R. Rainer, Analysis of beam position monitor requirements with Bayesian Gaussian regression, [arXiv:1904.05683](#).
- [69] M. Bernardini, L. Mayer, D. Reed, and R. Feldmann, Predicting dark matter halo formation in N-body simulations with deep regression networks, [arXiv:1912.04299](#).
- [70] A. Gómez-Valent and L. Amendola, H_0 from cosmic chronometers and type Ia supernovae, with Gaussian processes and the weighted polynomial regression method, in the *15th Marcel Grossmann Meeting on Recent Developments in Theoretical and Experimental General Relativity, Astrophysics, and Relativistic Field Theories* (2019).