

Nonparametric modeling of cosmological database on the χ^2 distribution

Maryam Vazirnia[✉] and Ahmad Mehrabi[✉]

Department of Physics, Bu-Ali Sina University, Hamedan 65178, 016016, Iran

 (Received 16 May 2021; revised 27 September 2021; accepted 23 November 2021; published 14 December 2021)

In the Λ CDM model, cosmological observations from the late and recent Universe reveal a puzzling $\sim 4.5\sigma$ tension in the current rate of Universe expansion. In addition to the various scenarios suggested to resolve the tension, a nonparametric modeling may provide useful insights. In this paper, we look at three well-known nonparametric methods: the smoothing method, the genetic algorithm, and the Gaussian process. Considering these three methods, we employ the recent Hubble parameter data to reconstruct the rate of Universe expansion and supernovae Pantheon sample to reconstruct the luminosity distance. In contrast to similar studies in the literature, the chi-squared distribution has been used to construct a reliable criterion to select a reconstruction. Finally, we compute the current rate of Universe expansion (H_0) for each method, provide some discussions regarding the performance of each approach, and compare the results.

DOI: [10.1103/PhysRevD.104.123530](https://doi.org/10.1103/PhysRevD.104.123530)

I. INTRODUCTION

The analysis of supernova type Ia (SNIa) data in Refs. [1,2] revealed for the first time that the expansion of the Universe is accelerating. Several independent measurements since then, including the cosmic microwave background (CMB) [3,4], baryon acoustic oscillation (BAO) [5–8], cosmic chronometers [9,10], and large-scale structures [11–13], confirmed such exotic behavior of the Universe at the recent time. Many attempts have been made to describe this phenomenon after these confirmations. There are mainly two avenues to find a solution: exotic matter with a negative equation of state (EOS) [14–18] or modification of gravity at large scales [19–23]. Among all the alternatives, a cosmological constant with EOS $w = -1$ and cold dark matter, the so-called Λ CDM, gives a good model to describe practically all datasets. However, the model suffers from severe theoretical problems which have not been resolved yet [24–26].

Considering a Friedman universe, the present rate of expansion (H_0) is one of the most essential characteristics in a cosmological scenario. Unfortunately, there is a strange discrepancy between measurement of H_0 considering late and early cosmological data. Assuming the Λ CDM model, the CMB data yield a relatively lower value of $H_0 \sim 67 \text{ Kms}^{-1} \text{ Mps}^{-1}$ [4], whereas local SNs anticipate $H_0 \sim 74 \text{ Kms}^{-1} \text{ Mps}^{-1}$ [27]. It is worth noting that the former requires a model to measure H_0 , whereas the latter relies solely on a distance ladder and the absolute magnitude of SNs. There have been numerous attempts to tackle this problem (see Ref. [28] and its references), but no adequate solution has yet been found.

Given a dataset, a model with some free parameters can be chosen to describe the data. The free parameters should

be constrained using a statistical method in this circumstance. In terms of statistics, the frequentist and Bayesian scenarios are the two options for constraining the free parameters. In the frequentist point of view, some estimators have been defined to estimate the values of parameters and their uncertainty. In the Bayesian scenario, on the other hand, the Bayes theorem has been used to determine the posterior probability distribution of the parameters. The distribution was then utilized to determine the optimal parameter values as well as their uncertainty [29].

On the other hand, using a nonparametric (NP) approach to describe data collection is a viable option. In machine learning scenarios, this type of modeling was created to have as much capacity to describe a dataset as possible. An NP is a unique way to investigate a data collection without having to assume any particular parametric shape. Furthermore, the NP generates a large number of reconstructions at once, some of which may have novel features. The Gaussian process (GP) is the most well-known NP modeling in cosmology, and it has been frequently employed in cosmic data analysis [30–38]. In this case, the data have been modeled by a sequence of Gaussian random values. The genetic algorithm (GA), on the other hand, given a set of base functions, can reconstruct multiple curves that are consistent with the data. This method has been used in Ref. [39] to study SNIa data and in Ref. [40] to investigate a null test on the cosmological constant. Moreover, this method has been used to study the dark energy in Ref. [41]. The smoothing method (SM) is the last procedure, which involves reconstructing a function using a smoothing kernel. In this case, a series of initial guesses at random sites has been improved with each iteration, resulting in a better fit to the data after each step.

This method was utilized in Refs. [42–44] to investigate various cosmological data.

We select these methods because they all rely on a sample of reconstructions, allowing us to compare them. However, with an NP approach, estimating errors is a challenging issue, which is a drawback in such cases. Reconstructions in the GP are obtained by sampling from a multivariate Gaussian distribution, with the standard deviation (mean) of a quantity at each point indicating the quantity’s uncertainty (central value). In contrast to the GP, there is no unique way to select a reconstruction and make a sample in the GA and SM. Considering the SM method, in Refs. [42–44], a reconstruction with $\chi^2 < \chi_{\text{ref}}^2$ has been selected, where χ_{ref}^2 is the χ^2 of a reference model, e.g., the Λ CDM. On the other hand, considering the GA, in Refs. [40,41,45,46], only the best reconstruction was obtained, and the uncertainty of that reconstruction was calculated in a different approach. In this paper, we introduce a more reliable sampling method to estimate the central value, with optimized chi-squared values that select the reconstructions.

The structure of this paper is as follows: We present the statistical tools needed for our analysis in Sec. II, as well as the chi-squared distribution and our new criterion for selecting a reconstruction in NP modeling. A brief discussion of parameter inference in a model-dependent technique has also been included. All three NP approaches are described in depth in Sec. III. Furthermore, given a dataset, we discuss how each method provides a consistent reconstruction. In Sec. IV, we present the results for different methods and discuss how different criterion affect the estimation of H_0 . Finally, in Sec. V, we provide our findings, explain, and compare them to similar efforts in the literature.

II. PARAMETER INFERENCE AND χ^2 DISTRIBUTION

Given a dataset as (x_i, y_i, σ_i) and knowing the distribution of noise, it is an easy task to construct a likelihood function. The likelihood contains all information relating to observed data and is a necessary requirement to infer the free parameters of a model. The Gaussian likelihood is given by

$$\mathcal{L}(\vec{\theta}) \propto \prod_{i=1}^N \exp\left(-\frac{1}{2} \frac{(f(x_i, \vec{\theta}) - y_i)^2}{\sigma_i^2}\right), \quad (1)$$

where $f(x_i, \vec{\theta})$ is the model prediction at x_i , $\vec{\theta}$ indicates all free parameters, and N is the number of data points. It is common to write the likelihood as

$$\mathcal{L}(\vec{\theta}) = \mathcal{L}_0 \exp\left(-\frac{1}{2} \chi^2\right), \quad (2)$$

where $\chi^2 = \sum_{i=1}^N \frac{(f(x_i, \vec{\theta}) - y_i)^2}{\sigma_i^2}$ and \mathcal{L}_0 is a normalization constant. Since the likelihood is Gaussian, the quantity

$X_i = \frac{f(x_i, \vec{\theta}) - y_i}{\sigma_i}$ is a Gaussian random variable with zero mean and $\text{var} = 1$. Summing up the square of X_i gives a quantity $Q = \sum_{i=1}^k X_i^2$, which is distributed according to the chi-squared probability distribution function (PDF) for $x = \chi^2$:

$$P(x|k) = \frac{x^{\frac{k}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})}, \quad (3)$$

where k is the number of the degree of freedom (NDF). Furthermore, the PDF shows a peak at $\chi^2 \sim k$ for a moderately big k and the probability for $\chi^2 \gg k$ and $\chi^2 \ll k$ are negligible. In fact, if $\chi^2 \gg k$, the majority of data points are away from the model prediction, indicating an underfitting case. On the other hand, if $\chi^2 \ll k$, the majority of the data points are unexpectedly near to the model prediction, and we have an overfitting case. It is worth noting that neither of these scenarios is a good fit for the data. The chi-squared PDF could be used as a criterion to select a reconstruction based on its probability, according to the aforementioned reasoning. Given a percentage number, such as 95%, it is straightforward to compute the range $(\chi_{\text{min}}^2, \chi_{\text{max}}^2)$ of the interval that lie within this percentage:

$$Pr(\chi_{\text{min}}^2 < x < \chi_{\text{max}}^2) = \int_{\chi_{\text{min}}^2}^{\chi_{\text{max}}^2} dx P(x|k). \quad (4)$$

The probability of a reconstruction with a χ^2 within the interval $(\chi_{\text{min}}^2, \chi_{\text{max}}^2)$ is equal to the Pr . Reconstructions in a sample chosen based on the chi-squared are free from both underfitting and overfitting, indicating that the sample is reliable for estimating the value of a parameter. As we mentioned above, the authors of Refs. [44,47,48] utilized the condition $\chi^2 < \chi_{\text{ref}}^2$ to select a reconstruction for the SM technique. In such a procedure, there might be some overfitting reconstructions, making the sample not reliable for estimation of a quantity.

In contrast to an NP modeling, for a parametric modeling, observational data can be used to constrain its parameters. In a frequentist scenario, one might use the maximum likelihood estimator (MLE) to estimate the optimal values of the parameters. There are several ways for determining the MLE given a likelihood function. In addition, expanding the likelihood around its peak gives an estimation of the parameter uncertainties [29]. In contrast to this point of view, the Bayes theorem may be used to find the PDF of parameters, and, from the PDF, the best values and their uncertainties can be easily determined. The Bayes theorem is given by

$$P(\vec{\theta}|d) = \frac{\mathcal{L}(\vec{\theta})P(\vec{\theta})}{P(d)}, \quad (5)$$

where $P(\vec{\theta})$ is the prior information on parameters and $P(d)$ is the Bayesian evidence. With the exception of a few cases

that have an analytic solution, the majority of problems should be addressed numerically. In these cases, a numerical approach such as Markov chain Monte Carlo has been used to estimate the posterior $P(\bar{\theta}|d)$ by sampling from the numerator of the above equation.

III. NP MODELING OF A DATASET

In the machine learning domain, an NP approach is used to have as much as possible capacity to describe a dataset. The NP methods have been used in a number of studies (some of them are listed in the introduction) in cosmology to describe a dataset. Because little is known about dark energy, using NP modeling could provide new insights or show a new feature in the data. In this work, we consider three NP methods, all reliant upon a sample of reconstruction, and compare the results. Since the results depend on the sample, a reliable selection criterion is required. In this section, we go over the basics of three well-known NP methods: GP, GA, and SM. Then, in the following section, we compare the results after finding all viable reconstructions using two separate cosmological datasets. This study helps us understand the performance of each method as well as advantages and disadvantages of each approach.

A. Gaussian process

A Gaussian process is a sequence of Gaussian random variables (RVs) that can be presented by a multivariate Gaussian distribution. In this case, the diagonal entries represent uncertainty at each point, whereas off-diagonal terms represent correlation between points. Assuming that a dataset can be modeled by a Gaussian process, we have

$$f(x) \sim GP(\mu(x), K(x, \tilde{x})), \quad (6)$$

where $K(x, \tilde{x})$ is the kernel function and x and \tilde{x} are two different observational points. The $\mu(x)$ is the mean function which provides the mean of Gaussian RVs at each observational point. The kernel gives the covariance matrix of the multivariate Gaussian distribution and depends on some hyperparameters. The most well-known kernel is the squared exponential given by

$$K(x, \tilde{x}) = \sigma_f^2 \exp \frac{-(x - \tilde{x})^2}{2l^2},$$

where σ_f^2 and l are two hyperparameters. Given a GP at some observational points x , it is simple to find the GP at some arbitrary points x^* . The function's values at these points are given by sampling from a multivariate Gaussian $N(\mu^*, \Sigma^*)$, where the mean and covariance are given by [49]

$$\begin{aligned} \mu^* &= K(x, x^*)[K(x, x^*) + C_D]^{-1}Y, \\ \Sigma^* &= K(x^*, x^*) - K(x^*, x)[K(x, x^*) + C_D]^{-1}K(x, x^*), \end{aligned} \quad (7)$$

respectively, where C_D is the covariance matrix of the data and $Y = [y_i]$ is the column vector of observational data [as in the previous section, we assume a dataset as (x_i, y_i, σ_i)]. Notice that, in the above equations, a zero mean prior has been considered for the μ^* .

Given a dataset, one can easily compute the mean and covariance of the multivariate Gaussian distribution and then by sampling from that find many reconstructions. Notice that, since the derivative of a GP is another GP, it is also easy to find the derivative of reconstructed function (to see more details, refer to Refs. [34,49]). We use the GaussianProcessRegressor class of scikit learn to find the mean and covariance matrix of a GP [50].

B. Genetic algorithm

GA is an optimization method inspired by the process of natural selection. It relies on the biologically inspired operators such as crossover, mutation, and selection. The process starts from a population of individuals and produces a new generation through mutation, which is a random change in an individual, and crossover, which is a combination of many individuals. In this scenario, the probability of a next generation is given by a fitness function which is also our objective function for optimization. For more details on the algorithm, we refer the reader to Ref. [39].

One of the methods of GA is symbolic regression (SR). In this case, the method attempts to produce a mathematical expressions to describe a dataset. To do so, the SR generates an initial random population of individuals using a set of basic functions. Then individuals evolve through the GA process to find a new generation with a smaller fitness function. The process repeats until either a minimum fitness function or a certain number of generations is reached. We set the fitness function to the χ^2 , and, hence, the next generation provides a mathematical expression with smaller χ^2 . In our analysis, we use the public package gplearn, which is an extension of the scikit-learn machine learning library, to perform the SR. We have examined the hyperparameter space empirically to find the optimum values. In our code, we have set the tournament size = 30, probability of mutation = 0.03, probability of crossover = 0.9, and population size = 2000. In addition, we use only ‘‘add,’’ ‘‘subtract,’’ and ‘‘multiplication’’ for the base functions.

The GA method has been used to study the expansion of the Universe in Ref. [51], dark energy anisotropic stress in Ref. [52], the cosmic distance duality relation in Ref. [53], and null tests for the spatial curvature and homogeneity of the Universe in Ref. [54]. However, in these works, the authors found the best fit among all reconstructions and

estimated its uncertainty using some methods base on the path integral. In contrary, in our analysis, we run our code with several different random seeds to generate many reconstructions and then use our new selection criterion to make a reliable sample.

C. SM method

In addition to the above approaches, SM is also a tool to study a dataset in a model-independent manner. It has only a hyperparameter, the smoothing width Δ , which depends on the number of data points as well as their qualities. In this scenario, SM begins with a sequence of arbitrary guess values and tries to generate a new smooth curve that is closer to the data points at each step.

Given the C_D , the covariance of data, to reconstruct a function at arbitrary x^* points, one can start from an initial guess and find the improved values $[\hat{f}_{n+1}(x^*)]$ at the next iteration by

$$\hat{f}_{n+1}(x^*) = \hat{f}_n(x^*) + \frac{\delta f_n^T \cdot C_D^{-1} \cdot W(x^*)}{1^T \cdot C_D^{-1} \cdot W(x^*)}, \quad (8)$$

where the kernel function $W(x^*)$ and δf_n^T are given, respectively, by

$$W_i(x^*) = \exp\left(\frac{-\ln^2\left(\frac{1+x^*}{1+x_i}\right)}{2\Delta^2}\right), \quad (9)$$

$$\delta f_n = f(x_i) - \hat{f}_n(x_i), \quad (10)$$

the 1^T is a unite column vector (its size is the same as the size of data), and $f(x_i) = y_i$ are the observed values. Notice that, similar to the GP, x_i (x^*) indicates observational (arbitrary) points. This method was used in Refs. [44,47,48] to reconstruct the expansion history using cosmological data. However, in these works, the authors chose reconstructions under condition $\chi^2 < \chi^2_{\Lambda\text{CDM}}$ to make a sample. Since the method relies on a sample of reconstructions, we also consider this scenario to investigate our datasets. In contrast to previous works, we use a different strategy to set the initial guess. The Taylor expansion of the Hubble parameter up to the fourth order in z gives us a Hubble parameter in terms of (H_0, q_0, j_0, l_0) (the cosmography parameters), which is independent of any model [55–58] (to see the expansion coefficients as a function of cosmography parameters, refer to Ref. [58]). Notice that, in order to expand the range of the convergence, the redshift has changed to the so-called y redshift $y = \frac{z}{1+z}$. We sample (H_0, q_0, j_0, l_0) from a wide uniform distribution for each initial guess and then utilize them to generate the initial Hubble parameter at x^* . Note that we have double-checked the results to ensure that they are robust and unaffected by the initial guess.

IV. DATASET AND RESULTS

In order to investigate the output of the above-mentioned NP methods, we consider two separate cosmological data. At the background level, the luminosity and angular diameter distances, as well as direct measurements of the Hubble parameter, have provided information about the Universe's expansion rate. From the luminosity and angular diameter distances, one can measure the source comoving distance and then derive the Hubble parameter. Assuming a flat geometry, the Hubble parameter is given by

$$D(z) = \int_0^z \frac{dz}{H(z)},$$

$$H(z) = \frac{1}{D'(z)}, \quad (11)$$

where $D(z)$ is the comoving distance at redshift z . Notice that, if $D'(z) = 0$ at a point, we cannot use the equation to find the Hubble parameter. For all reconstructions of the comoving distance, we check this condition and discard those that have $D'(z) = 0$ at one or more points.

The most up-to-date and precise measurement of luminosity distance at this moment is the SNIa Pantheon sample [59]. This sample contains 1048 spectroscopically confirmed SNIa up to redshift $z = 2.26$. To avoid the degeneracy between H_0 and the absolute magnitude of the SNIa, we set $M = -19.3$ through our study. Using this dataset, we generate a large number of reconstructions through each method, and then we used the chi-squared PDF with probability $pr = 68\%$ and $pr = 95\%$ to make a sample. These two intervals for the SNIa data are presented in Fig. 1. Finally using Eq. (11), these reconstructions converted to the Hubble parameter.

The second dataset is the measurement of the Hubble parameter data collected in Ref. [10]. The dataset includes

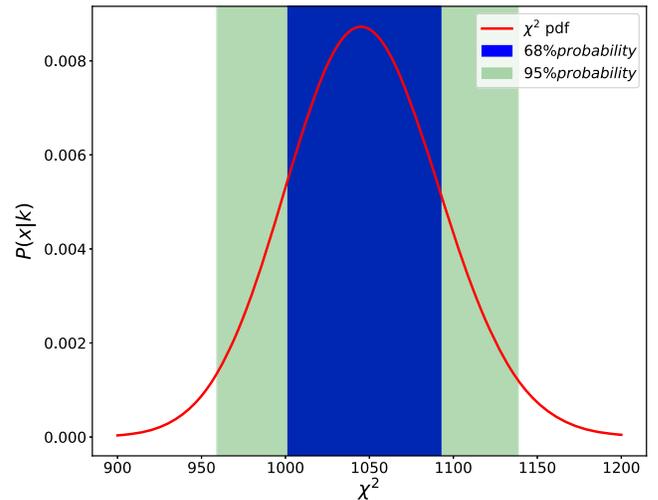


FIG. 1. The chi-squared PDF (solid red line) for SNIa data and the range of χ^2 for $pr = 68\%$ and $pr = 95\%$.

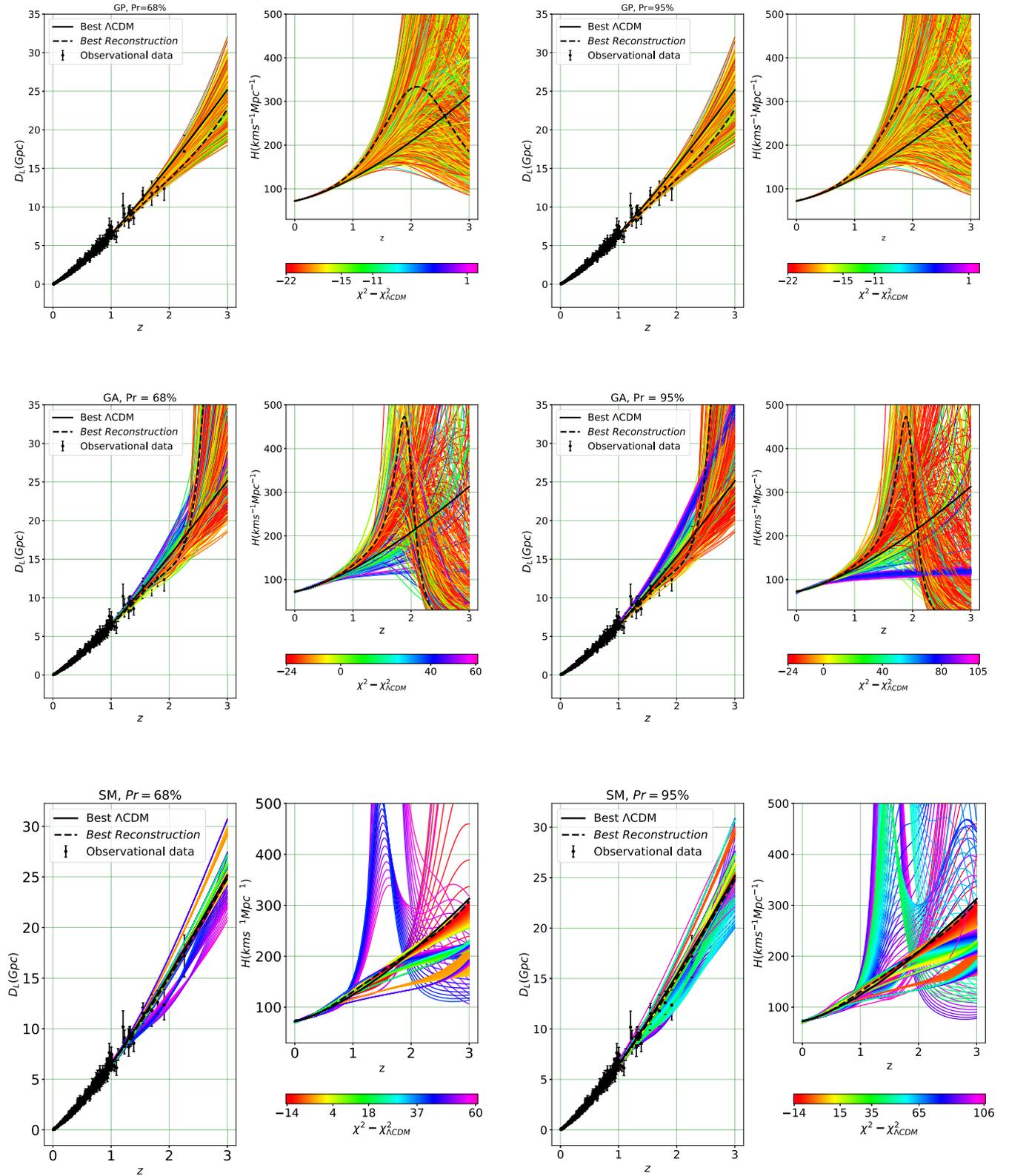


FIG. 2. Reconstructions of the luminosity distance and derived Hubble parameter considering the SNIa data. The upper, middle, and lower panels present the results for GP, GA, and SM methods, respectively. The best Λ CDM (the best reconstruction) is shown by a solid black (dashed black) curve in each panel. The color bar shows the difference between χ^2 and the best fit Λ CDM, $\chi^2_{\Lambda\text{CDM}}$. In all scenarios, the left panel (right panel) presents reconstruction for probability $pr = 68\%$ ($pr = 95\%$).

measurement of the cosmic chronometer as well as the radial BAO. In addition to the 38 data points in the collection, we add the measurement of H_0 from nearby SNIa [27]. Note that the BAO data points are correlated, but, for the sake of simplicity, we ignore such correlations in the current study.

A. Results for the SNIa data

As has been previously stated, the methods provide a reconstruction of the luminosity distance in the case of SNIa data, and the Hubble parameter is a derived parameter. In this subsection, we present the results of our analysis using the SNIa data. For all techniques, the reconstructions of the luminosity distance and their corresponding Hubble parameter are displayed in Fig. 2. Along with all reconstructions, the best fit Λ CDM and the best reconstructions (the lowest χ^2) have been presented in each panel. The solid black (dashed black) curve in each panel presents the best Λ CDM (best reconstruction), and the color bar shows the $\Delta\chi^2 = \chi^2 - \chi^2_{\Lambda\text{CDM}}$ quantity. The upper panels show the GP results for $pr = 68\%$ and $pr = 95\%$, which are almost the same. In fact, all of the reconstructions are within a small range of χ^2 , so adjusting the PDF probability has little effect on the outcome.

The best-reconstructed Hubble has a peak at $z \sim 2$ and a smaller value at $z = 3$ compared to the best Λ CDM (for the best reconstruction, $\Delta\chi^2 = -22$). Notice that almost all of these reconstructed curves have χ^2 values lower than the best Λ CDM and also the range of χ^2 values is narrower than in the other two scenarios.

Considering the GA method, the reconstructed luminosity distance and corresponding Hubble parameter have been presented in the middle panel in Fig. 2. The best reconstruction provides a Hubble parameter which has a peak around $z \sim 2$ similar to the GP but with an even smaller χ^2 value ($\Delta\chi^2 = -24$). Furthermore, the results for $pr = 68\%$ and $pr = 95\%$ are different, with more curves having a larger χ^2 value in the case of $pr = 95\%$. These findings suggest that GA provides more flexibility to investigate a dataset compared to GP.

Finally, the results of the SM method have been presented in the lower panel. In contrast to the other methods, the best reconstruction in the SM is very close to the best Λ CDM. In the SM method, the best reconstruction has $\Delta\chi^2 = -14$, which is relatively larger than other methods. Notice that, since in the case of SNIa the number of data points is larger than the number of Hubble data, a larger χ^2 could be selected based on the chi-squared PDF. For example, in the case of GA and SM the $\Delta\chi^2$ might be large up to 105, while for the Hubble data it is around 25.

B. Results for the Hubble data

The reconstructions of the Hubble parameter considering the Hubble data are shown in Fig. 3 for all the scenarios.

The upper panel shows the GP results for $pr = 68\%$ and $pr = 95\%$ cases. The range of χ^2 in $pr = 68\%$ is narrower than $pr = 95\%$ as is expected from the chi-squared PDF. Because of the constraint in the GP [see Eq. (7)], a considerable number of reconstructions have a χ^2 around or less than $\chi^2_{\Lambda\text{CDM}}$. Moreover, at high redshifts $z \sim 2.5-3$, reconstructions provide a lower $H(z)$ value compare to the Λ CDM, with the best reconstruction being roughly 25% smaller than the best Λ CDM at $z = 3$. Considering the GA method, the results have been shown in the middle panel in Fig. 3. Similar to GP, among all reconstructions, only those reconstructions with probability 68% and 95% have been selected and presented in the panels. While the results are similar to GP up to redshifts ($z < 1.5$), GA provides more scattered curves at $z = 3$ and the best reconstruction is closer to the best Λ CDM than GP. There are also two curves with $\chi^2 - \chi^2_{\Lambda\text{CDM}} \sim 10$ that have a minor peak around $z = 2$. These results suggest that GA may be more useful and adaptable than GP when it comes to discovering a new feature in a dataset.

Finally, the results of the SM method are displayed in the lower panel. The χ^2 range is similar to that of GA, although there are more curves with large χ^2 . Moreover, reconstructions with a high χ^2 had a lower $H(z)$ (purple curves) at $z \sim 1.5$ and a higher value at $z = 3$. The best reconstruction is nearly identical to the best Λ CDM up to redshift $z \sim 2$ but gives a slightly larger (smaller) value in the case $pr = 68\%$ ($pr = 95\%$) at redshift $z = 3$.

Our findings suggest that using the chi-squared PDF to select reconstructions provides a better insight in analyzing a dataset, and in this scenario some reconstructions may reveal a new feature. In contrary, selecting reconstructions with χ^2 smaller than a threshold value may not offer a comprehensive sample. It is important, since in the majority of cases our main objective is the estimation of a quantity (like H_0) from the sample.

C. Estimation of H_0

For both Hubble and SNIa data, the distribution of χ^2 for all reconstructions, as well as the χ^2 PDF, is presented in Fig. 4. In these diagrams, the probability has been set to $pr = 95\%$, the red solid line indicates the χ^2 PDF, and the vertical black line shows the location of the best Λ CDM. For both datasets, SM produces a nearly uniform distribution, but GA and GP provide more reconstructions with smaller χ^2 . Notice that the results of SM (GA) are dependent on the number of iterations (generations), and raising this number results in more reconstructions with less χ^2 (albeit χ^2 does not reduce significantly after some steps). The distribution of χ^2 in GP, on the other hand, covers a narrow region and is unaffected by sample size. Moreover, whereas the Hubble data reconstruction covers the whole range of $\chi^2_{\min} < \chi^2 < \chi^2_{\max}$ for $pr = 95\%$, the SNIa reconstructions are unable to reach the minimum

value χ^2_{\min} , and all GP reconstructions have a χ^2 smaller than the best Λ CDM.

As we mentioned above, the reconstructions may be used to estimate the value of H_0 as well as its uncertainty in a model-independent manner. In fact, estimating the central value is an easy task, but estimating the uncertainty is more difficult. While the central value can be easily estimated from the mean or median, there are different approaches for estimation of the uncertainty. Considering the SM method,

in Refs. [42–44,47,48], the authors used the maximum and minimum values at each redshift to find an interval of the uncertainty. In the GA scenario, on the other hand, the path integral approach has been used to estimate the uncertainty [41,60]. Note that in this case the method gives only the uncertainty of the best reconstruction, not the overall uncertainty. In a GP approach, however, estimate of these quantities is simple, and the mean (standard deviation) at each redshift gives the central value (uncertainty). In order

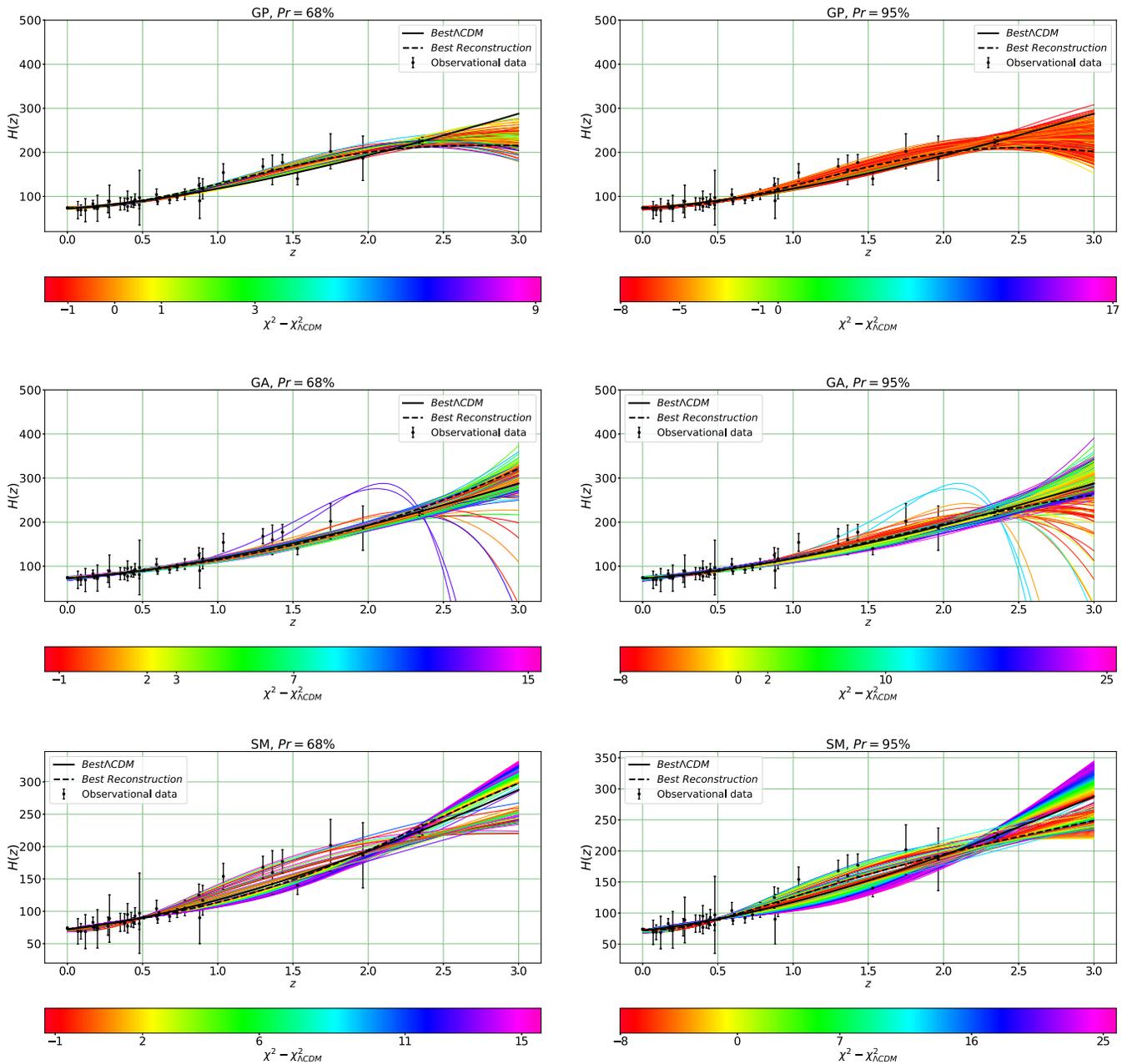


FIG. 3. Reconstructions of the Hubble parameter considering the Hubble data. The upper, middle, and lower panels present the results for GP, GA, and SM methods, respectively. The best Λ CDM (the best reconstruction) is shown by a solid black (dashed black) curve in each panel. The color bar shows the difference between χ^2 and the best fit Λ CDM, $\chi^2_{\Lambda\text{CDM}}$. In all scenarios, the left panel (right panel) presents reconstruction for probability $pr = 68\%$ ($pr = 95\%$).

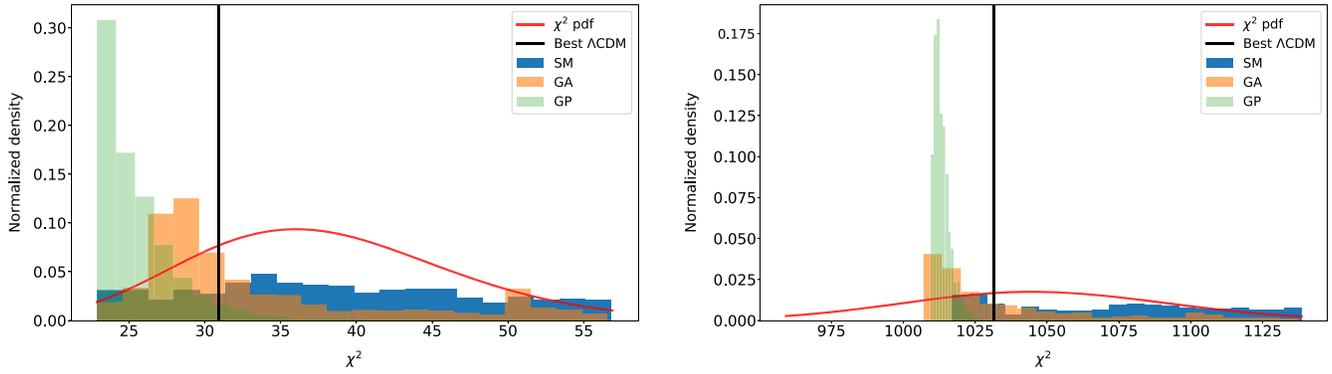


FIG. 4. Distribution of the reconstructions χ^2 in each method for $pr = 95\%$. The solid red line indicates the χ^2 PDF corresponding to the number of degrees of freedom for each dataset. The left panel (right panel) shows the results for the Hubble data (the SNIa data).

to estimate the value of H_0 as well as its uncertainty, we use a similar process as GP.

The distributions of H_0 for both datasets are illustrated in Fig. 5. The solid vertical line shows the location of the best Λ CDM, and distributions of H_0 from different methods are presented by different colors. For the Hubble data, the SM method provides a peak near to the Λ CDM, whereas results from GA and GP are scattered over a large area. In contrast to the other two approaches, the peak of the distribution in GP gives a substantially greater value $H_0 \sim 74$. This is a direct consequence of the SHOES [27] data point, which shifts H_0 toward a larger value. Since GP has been widely utilized in the literature, it is important to remember that one data point (especially one with a small error) can significantly alter the results. Unlike GP, GA is unaffected by this data point and provides a relatively smaller H_0 . On the other hand, for the SNIa data, we see a narrow distribution around the value of Λ CDM in GP and a relatively wider distribution for both SM and GA.

In contrast to Hubble data, the results of GP in the case of SNIa show a tight peak around the best Λ CDM, indicating that the central value is close to the Λ CDM and its uncertainty is smaller than other methods. In this case, the SM method yields a relatively wider distribution, and we have H_0 in a range of (68.5–74.5). Since the results might depend on the probability in our selection criterion, we perform our analysis with both $pr = 68\%$ and $pr = 95\%$. The results are presented schematically in Fig. 6 and quantitatively in Table I. The main points regarding the H_0 estimation are as follows.

- (i) The results of all methods are consistent with Λ CDM at 1σ level.
- (ii) The results for $pr = 68\%$ and $pr = 95\%$ are consistent with each other, but uncertainties are around 10%–20% larger for $pr = 95\%$ in GA and SM.
- (iii) The uncertainties in GP for the SNIa data are the same for both $pr = 68\%$ and $pr = 95\%$. This is mainly due to the fact that reconstructions in GP are concentrated in a small area of the chi-squared PDF,

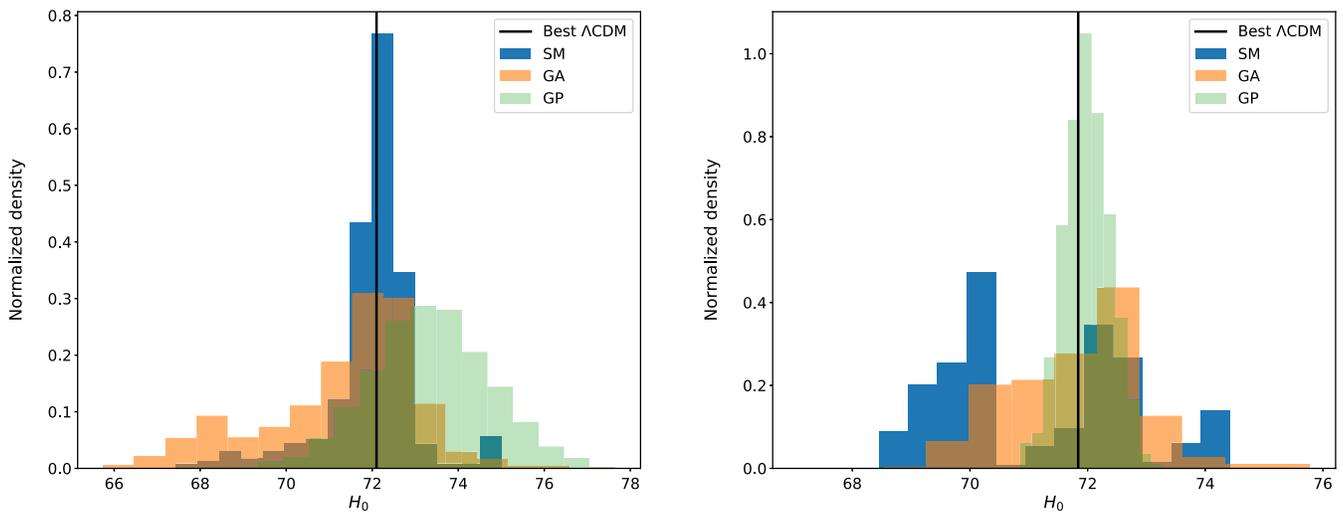


FIG. 5. The distribution of estimated H_0 considering $pr = 95\%$ in different methods. The left panel (right panel) shows the results for the Hubble parameter (the SNIa data).

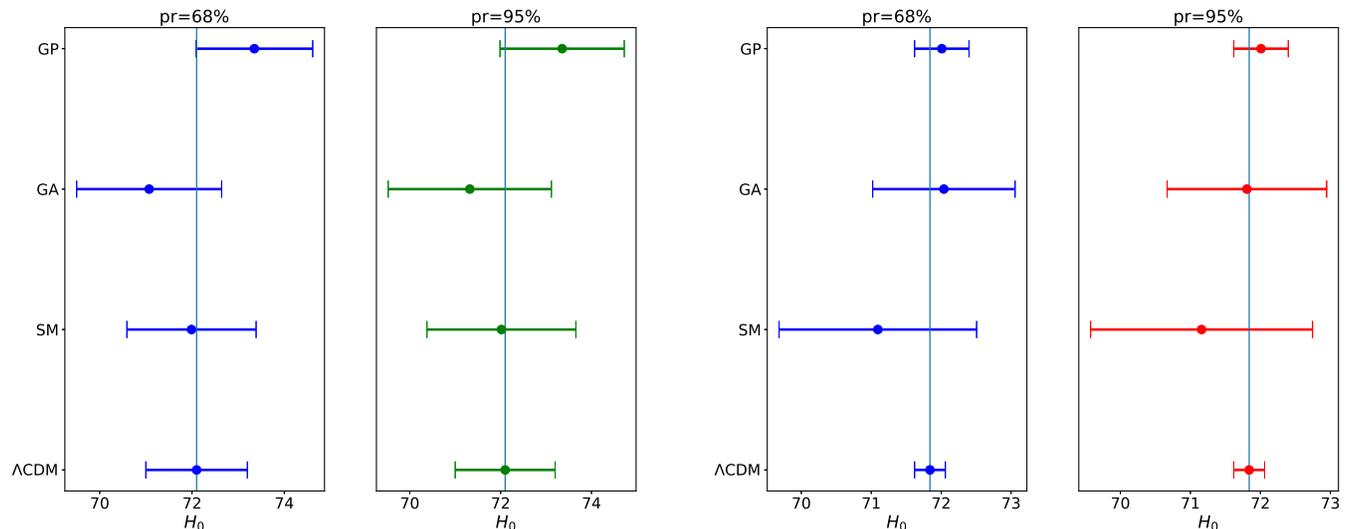


FIG. 6. Estimation of H_0 and its 1σ uncertainty in GP, GA, and SM for $pr = 68\%$ and $pr = 95\%$. The left panel (right panel) shows the results for the Hubble data (the SNIa data).

which does not change with probability. On the other hand, $pr = 68\%$ offers a 6% lower uncertainty for Hubble data than $pr = 95\%$.

- (iv) The SHOES data point shifts the H_0 toward a larger value in GP, but the other two methods are not sensitive to this data point.
- (v) GP has the least uncertainty among all the methods, and uncertainties in GP are only slightly larger than Λ CDM.

V. CONCLUSION

In this work, we consider three well-known NP methods, all reliant upon a sample of reconstructions, namely, GP, GA, and SM methods, and introduce a novel approach to select a consistent reconstruction. We compare the results of employing the SNIa (the Pantheon sample) and a recent collection of Hubble parameters to the NP techniques and check the consistency of our selection criterion. Unlike previous studies, we use the probability of each reconstruction based on the chi-squared PDF to select a

consistent reconstruction. Given the NDF of a dataset, it is straightforward to find the probability of each reconstruction (according to its χ^2 value), and setting a probability threshold, such as $pr = 68\%$ or $pr = 95\%$, one can separate reconstructions with these probabilities according to the chi-squared PDF. However, a different strategy has been utilized in Refs. [42–44,47,48] to select a reconstruction. In these works, the authors have considered a reference χ^2_{ref} as a threshold and selected all reconstructions with a χ^2 smaller than the threshold.

Considering the Hubble data, we reconstruct the Hubble parameter directly and compare the results for the cases of $pr = 68\%$ and $pr = 95\%$ in all three methods. On the other hand, the Hubble parameter is computed from the reconstructed luminosity distance for the SNIa data. According to our results, GA is more flexible in finding a new feature compare to the other methods. Moreover, for both datasets, GP provides a smaller range of χ^2 with the majority of them having a χ^2 smaller than the concordance Λ CDM. In addition, our analysis indicates that GP is more efficient than the other two methods in terms of computational time, whereas SM is the slowest.

The reconstructions could also be used to estimate the value of H_0 , an essential quantity in cosmology. It is worth noting that estimating H_0 requires a reliable sample of reconstructions. For all reconstructions in the sample, we obtain H_0 and investigate its distribution for each dataset considering all methods. Notice that the estimating uncertainty in an NP technique is a difficult task.

Specifically, in Refs. [42–44,47,48], the authors have used the maximum and minimum values at each redshift to determine an uncertainty interval. The authors of Refs. [41,51,53,60], on the other hand, have employed a path-integral-based method to estimate the uncertainty in the GA. In this paper, we adopt the same procedure as GP

TABLE I. Estimation of H_0 (km/s/Mpc) and its 1σ uncertainty in GP, GA, and SM for $pr = 68\%$ and $pr = 95\%$. The left column (right column) shows the results for the Hubble data (the SNIa data).

Method/data	Hubble data	SNIa
SM ($Pr = 68\%$)	71.99 ± 1.43	71.09 ± 1.41
SM ($Pr = 95\%$)	72.02 ± 1.64	71.16 ± 1.59
GA ($Pr = 68\%$)	71.07 ± 1.57	72.04 ± 1.02
GA ($Pr = 95\%$)	71.32 ± 1.80	71.80 ± 1.14
GP ($Pr = 68\%$)	73.33 ± 1.27	72.02 ± 0.38
GP ($Pr = 95\%$)	73.35 ± 1.35	72.02 ± 0.38
Λ CDM	72.1 ± 1.1	71.85 ± 0.22

and estimate the central value using the sample mean at each redshift and the uncertainty using the standard deviation. Our results indicate that GP has the lowest uncertainties, which are slightly higher than those found in the concordance Λ CDM. On the other hand, while all the estimated central values are consistent with each other, in SM and GA, the estimation of uncertainty for $pr = 95\%$ is roughly 10%–20% more than the results for $pr = 68\%$. Furthermore, changing the chi-squared probability has no significant effect on the estimated uncertainty in GP for both datasets. This is primarily owing to the fact that GP gives a limited range of χ^2 compared to the other two methods. Based on our findings, it may be preferable to use GA or SM in conjunction with GP when studying a dataset.

Estimation of $H(z)$ in a model-independent manner is not new. In particular, the authors of Ref. [61] have combined the Pantheon and BAO measurements and utilized a parametric form of $H(z)$ to find the value of H_0 . They have assumed some priors on the sound horizon

at the drag epoch and obtained two H_0 estimates that were significantly closer to the Planck Λ CDM estimate than the SHOES estimation. Our estimate is around $1.5 - 3\sigma$ larger than theirs, owing to the fact that our results are independent of the sound horizon. On the other hand, our results are very similar to those presented in Ref. [62], which used the Hubble data alone as well as a combination of Hubble data and the SNIa to estimate H_0 from a GP. In addition, the authors of Ref. [51] employed GA to estimate the H_0 using a similar dataset. They found a considerably larger uncertainty (~ 12 km/s/Mpc) using the path integral approach than what we obtained (~ 2 km/s/Mpc).

Finally, our results indicate that the SHOES data point has a significant impact on estimation of H_0 in GP, whereas the other two methods are less affected by this data point. In fact, when GP is used instead of the other two approaches, the data point shifts the peak of the H_0 distribution toward a bigger value. This finding also suggests that considering GA or SM along with GP may be beneficial.

-
- [1] A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner *et al.*, *Astron. J.* **116**, 1009 (1998).
- [2] S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom *et al.*, *Astrophys. J.* **517**, 565 (1999).
- [3] Planck Collaboration XIII, *Astron. Astrophys.* **594**, A13 (2016).
- [4] N. Aghanim *et al.* (Planck Collaboration), *Astron. Astrophys.* **641**, A6 (2020).
- [5] B. A. Reid, L. Samushia, M. White, W. J. Percival, M. Manera *et al.*, *Mon. Not. R. Astron. Soc.* **426**, 2719 (2012).
- [6] T. M. C. Abbott *et al.* (DES Collaboration), *Mon. Not. R. Astron. Soc.* **483**, 4866 (2019).
- [7] S. Alam *et al.* (BOSS Collaboration), *Mon. Not. R. Astron. Soc.* **470**, 2617 (2017).
- [8] H. Gil-Marn *et al.*, *Mon. Not. R. Astron. Soc.* **477**, 1604 (2018).
- [9] D. Stern, R. Jimenez, L. Verde, M. Kamionkowski, and S. A. Stanford, *J. Cosmol. Astropart. Phys.* **02** (2010) 008.
- [10] O. Farooq, F. R. Madiyar, S. Crandall, and B. Ratra, *Astrophys. J.* **835**, 26 (2017).
- [11] S. Alam, M. Ata, S. Bailey, F. Beutler, D. Bizyaev, J. A. Blazek, A. S. Bolton, J. R. Brownstein, A. Burden, C.-H. Chuang *et al.*, *Mon. Not. R. Astron. Soc.* **470**, 2617 (2017).
- [12] T. Abbott, F. Abdalla, A. Alarcon, J. Alekski, S. Allam, S. Allen, A. Amara, J. Annis, J. Asorey, S. Avila *et al.*, *Phys. Rev. D* **98**, 043526 (2018).
- [13] S. Alam *et al.* (eBOSS Collaboration), *Phys. Rev. D* **103**, 083533 (2021).
- [14] E. J. Copeland, M. Sami, and S. Tsujikawa, *Int. J. Mod. Phys. D* **15**, 1753 (2006).
- [15] T. Chiba, S. Dutta, and R. J. Scherrer, *Phys. Rev. D* **80**, 043517 (2009).
- [16] L. Amendola and S. Tsujikawa, *Dark Energy: Theory and Observations* (Cambridge University Press, Cambridge, England, 2010).
- [17] A. Mehrabi, *Phys. Rev. D* **97**, 083522 (2018).
- [18] A. Mehrabi and S. Basilakos, *Eur. Phys. J. C* **78**, 889 (2018).
- [19] G. Magnano and L. M. Sokolowski, *Phys. Rev. D* **50**, 5039 (1994).
- [20] A. Dobado and A. L. Maroto, *Phys. Rev. D* **52**, 1895 (1995).
- [21] S. Capozziello, S. Carloni, and A. Troisi, *Recent Res. Dev. Astron. Astrophys.* **1**, 625 (2003).
- [22] S. M. Carroll, V. Duvvuri, M. Trodden, and M. S. Turner, *Phys. Rev. D* **70**, 043528 (2004).
- [23] L. Sebastiani, S. Vagnozzi, and R. Myrzakulov, *Adv. High Energy Phys.* **2017**, 1 (2017).
- [24] T. Padmanabhan, *Phys. Rep.* **380**, 235 (2003).
- [25] L. Perivolaropoulos, [arXiv:0811.4684](https://arxiv.org/abs/0811.4684).
- [26] A. Padilla, [arXiv:1502.05296](https://arxiv.org/abs/1502.05296).
- [27] A. G. Riess, S. Casertano, W. Yuan, L. M. Macri, and D. Scolnic, *Astrophys. J.* **876**, 85 (2019).
- [28] E. D. Valentino, O. Mena, S. Pan, L. Visinelli, W. Yang, A. Melchiorri, D. F. Mota, A. G. Riess, and J. Silk, *Classical Quantum Gravity* **38**, 153001 (2021).
- [29] R. Trotta, [arXiv:1701.01467](https://arxiv.org/abs/1701.01467).
- [30] A. Shafieloo, A. G. Kim, and E. V. Linder, *Phys. Rev. D* **85**, 123530 (2012).
- [31] K. Liao, A. Shafieloo, R. E. Keeley, and E. V. Linder, *Astrophys. J.* **886**, L23 (2019).
- [32] A. Gmez-Valent and L. Amendola, *J. Cosmol. Astropart. Phys.* **04** (2018) 051.
- [33] A. M. Pinho, S. Casas, and L. Amendola, *J. Cosmol. Astropart. Phys.* **11** (2018) 027.

- [34] A. Mehrabi and S. Basilakos, *Eur. Phys. J. C* **80**, 632 (2020).
- [35] S. Dhawan, J. Alsing, and S. Vagnozzi, *Mon. Not. R. Astron. Soc.* **506**, L1 (2021).
- [36] C. Escamilla-Rivera, J. Levi Said, and J. Mifsud, *J. Cosmol. Astropart. Phys.* **10** (2021) 016.
- [37] R. C. Bernardo and J. Levi Said, *J. Cosmol. Astropart. Phys.* **08** (2021) 027.
- [38] R. Briffa, S. Capozziello, J. Levi Said, J. Mifsud, and E. N. Saridakis, *Classical Quantum Gravity* **38**, 055007 (2020).
- [39] C. Bogdanos and S. Nesseris, *J. Cosmol. Astropart. Phys.* **05** (2009) 006.
- [40] S. Nesseris and A. Shafieloo, *Mon. Not. R. Astron. Soc.* **408**, 1879 (2010).
- [41] S. Nesseris and J. Garca-Bellido, *J. Cosmol. Astropart. Phys.* **11** (2012) 033.
- [42] A. Shafieloo, *Mon. Not. R. Astron. Soc.* **380**, 1573 (2007).
- [43] A. Shafieloo and C. Clarkson, *Phys. Rev. D* **81**, 083537 (2010).
- [44] A. Shafieloo, B. LHuillier, and A. A. Starobinsky, *Phys. Rev. D* **98**, 083526 (2018).
- [45] R. Arjona, H.-N. Lin, S. Nesseris, and L. Tang, *Phys. Rev. D* **103**, 103513 (2021).
- [46] R. Arjona and S. Nesseris, *Phys. Rev. D* **103**, 063537 (2021).
- [47] A. Shafieloo, *J. Cosmol. Astropart. Phys.* **08** (2012) 002.
- [48] B. L'Huillier, A. Shafieloo, and H. Kim, *Mon. Not. R. Astron. Soc.* **476**, 3263 (2018).
- [49] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)* (The MIT Press, Cambridge, MA, 2005).
- [50] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, *J. Mach. Learn. Res.* **12**, 2825 (2011).
- [51] R. Arjona and S. Nesseris, *Phys. Rev. D* **101**, 123525 (2020).
- [52] R. Arjona and S. Nesseris, *J. Cosmol. Astropart. Phys.* **11** (2020) 042.
- [53] R. Arjona, H.-N. Lin, S. Nesseris, and L. Tang, *Phys. Rev. D* **103**, 103513 (2021).
- [54] R. Arjona and S. Nesseris, *Phys. Rev. D* **103**, 103539 (2021).
- [55] S. M. Feeney, H. V. Peiris, A. R. Williamson, S. M. Nissanke, D. J. Mortlock, J. Alsing, and D. Scolnic, *Phys. Rev. Lett.* **122**, 061105 (2019).
- [56] G. Risaliti and E. Lusso, *Nat. Astron.* **3**, 272 (2019).
- [57] E. Lusso, G. Risaliti, E. Nardini, G. Bargiacchi, M. Benetti, S. Bisogni, S. Capozziello, F. Civano, L. Eggleston, M. Elvis *et al.*, *Astron. Astrophys.* **642**, A150 (2020).
- [58] M. Rezaei, S. Pour-Ojaghi, and M. Malekjani, *Astrophys. J.* **900**, 70 (2020).
- [59] D. M. Scolnic *et al.*, *Astrophys. J.* **859**, 101 (2018).
- [60] S. Nesseris and J. Garca-Bellido, *Phys. Rev. D* **88**, 063521 (2013).
- [61] P. Lemos, E. Lee, G. Efstathiou, and S. Gratton, *Mon. Not. R. Astron. Soc.* **483**, 4803 (2019).
- [62] A. Gómez-Valent and L. Amendola, *J. Cosmol. Astropart. Phys.* **04** (2018) 051.