# COSMOLOGICAL RECONSTRUCTIONS WITH ARTIFICIAL NEURAL NETWORKS

#### Isidro Gómez-Vargas

CICATA-Legaria, Instituto Politécnico Nacional 11500, Ciudad de México, México. ICF, Universidad Nacional Autónoma de México, 62210, Cuernavaca, Morelos, México. igomezv0701@alumno.ipn.mx

#### J. Alberto Vázquez\*

ICF, Universidad Nacional Autónoma de México, 62210, Cuernavaca, Morelos, México.
javazquez@icf.unam.mx

## Ricardo Medel Esquivel

CICATA-Legaria, Instituto Politécnico Nacional, 11500, Ciudad de México, México. ICF, Universidad Nacional Autónoma de México, 62210, Cuernavaca, Morelos, México. rmedele1500@alumno.ipn.mx

#### Ricardo García-Salcedo

CICATA-Legaria, Instituto Politécnico Nacional 11500, Ciudad de México, México. Also at the Guanajuato University on sabbatical stay. rigarcias@ipn.mx

April 2, 2021

#### **ABSTRACT**

The relevance of non-parametric reconstructions of cosmological functions lies in the possibility of analyzing the observational data independently of any theoretical model. Several techniques exist and, recently, Artificial Neural Networks have been incorporated to this type of analysis. By using Artificial Neural Networks we present a new strategy to perform non-parametric data reconstructions without any preliminary statistical or theoretical assumptions and even for small observational datasets. In particular, we reconstruct cosmological observables from cosmic chronometers,  $f\sigma_8$  measurements and the distance modulus of the Type Ia supernovae. In addition, we introduce a first approach to generate synthetic covariance matrices through a variational autoencoder, for which we employ the covariance matrix of the Type Ia supernovae compilation. To test the usefulness of our developed methods, with the neural network models we generated random data points mostly absent in the original datasets and performed a Bayesian analysis on some simple dark energy models. Some of our findings point out to slight deviations from the  $\Lambda$ CDM standard model, contrary to the expected values coming from the original datasets.

**Keywords** Dark energy · Artificial Neural Networks · Non-parametric Reconstructions · Machine Learning

# 1 Introduction

One of the biggest challenges for the cosmological community is the explanation of the current accelerated expansion of the Universe. A theoretical conception, commonly called Dark Energy (DE), is introduced to explain this mysterious

phenomenon and whose nature is still unravelled [1, 2, 3]. The cosmological constant  $\Lambda$ , being the simplest form of the dark energy, together with the Cold Dark Matter (CDM), which is a key component for structure formation in the universe, conform the standard cosmological model or  $\Lambda$ CDM. This model has had great achievements such as being in excellent agreement with most of the currently available data, for example measurements from the Cosmic Microwave Background radiation [4], Supernovae Ia (SNeIa) [5], Cosmic Chronometers (CC) [6] and Baryon Acoustic Oscillations (BAO) [7]. Nevertheless, the  $\Lambda$ CDM model has its own drawbacks: on theoretical grounds the cosmological constant faces several problems, i.e. fine tuning and cosmic coincidence [8, 9], and from an observational point of view, it also suffers to the so-called Hubble tension – a measurement disagreement of  $H_0$  among datasets [10]. The presence of these issues opens the possibility to extensions beyond the standard model by either considering a dynamical DE or modifications to the general theory of relativity [11].

The search of possible signatures for models beyond the  $\Lambda$ CDM has led to the creation of an impressive set of high accuracy surveys, already underway or being planned [12, 13, 14], to gather a considerable amount of information that constrains the properties of the universe. That is, a viable model that leads to the current accelerating universal expansion is demanded to comply with all the relevant observational data. Extensions to the cosmological constant which allow a redshift-dependent equation-of-state (EoS) w(z) include extra dimensions [15], modified gravity [16], scalar fields [17], scalar-tensor theories with non-minimal derivative coupling to the Einstein tensor [18] and graduated dark energy [19], just to mention a few. However, in the absence of a fundamental and well-defined theory of dark energy, a time-dependent behaviour can also be investigated by choosing an EoS mathematically appealing or a parameterised form in a simple way, examples of these forms in terms of redshift include a Taylor expansion [20], polynomial [21], logarithmic [22], oscillatory [23, 24] or in terms of cosmic time [25]. Nonetheless, the a priori assumption of a specific model may lead to misleading model-dependent results regardless of the dark energy properties, and hence, instead of committing to a particular model the non-parametric inference techniques allow to extract information directly from the observational data to detect features within cosmological functions, for instance w(z). That is, the main aim of a non-parametric approach is to infer (reconstruct) an unknown quantity based mainly on the data and making as few assumptions as possible [26, 8]. Several non-parametric techniques are used to reconstruct cosmological functions from the data directly, such as histogram density estimators [27], Principal Component Analysis (PCA) [28], smoothed step functions [29], gaussian processes [30, 31, 32, 33], Simulation Extrapolation method (SIMEX) [34] and Bayesian nodal free-form methods [35, 36]. After the reconstruction is performed, the function in place can be considered as a new model in order to look for possible deviations from the standard ACDM. In other words, the result of a non-parametric reconstruction may be used to analyse its similarity with different theoretical models and therefore to select its best description for the data. There are several examples of non-parametric reconstructions of cosmological functions, some of them focus on dark energy features [37, 27, 29, 38], cosmic expansion [34], deceleration parameter [33], growth rate of structure formation [32], luminosity distance [39, 40] and primordial power spectrum [41, 42], among many others.

The recent increase in computing power and the vast amount of coming data have allowed the incursion of machine learning methods as analysis tools in observational cosmology [43, 44, 45, 46, 47]. In this work we focus on the computational models called Artificial Neural Networks (ANNs). They have been used in a variety of applications, such as image analysis [48, 49], N-body simulations [50, 51] and statistical methods [52, 53, 54, 55]. Artificial Neural Networks have already been applied to non-parametric reconstructions of cosmological functions [56] and, unlike other methods such as gaussian processes, ANNs may not assume any statistical distribution of the data. The main aim of this paper is to present a new strategy for non-parametric reconstructions using directly the cosmological observations to feed neural networks and then to generate models based solely on the data without any statistical or cosmological assumptions. We show that our methodology can be applicable to any astronomical datasets and we propose a special treatment for those containing a full covariance matrix, *i.e.* including correlations between measurements.

The rest of the paper has the following structure. In Section 2, we briefly introduce the cosmological and statistical concepts used throughout this work: cosmological models, functions and observations in section 2.1; an overview of neural networks in section 2.2 and a short summary of Bayesian inference in section 2.3. Section 3 describes the methodology used during the neural network training to generate models based on cosmological data. Section 4 contains the results of the reconstructed functions, namely the Hubble distance H(z), a combination of the growth rate of cosmological perturbations times the matter power spectrum normalisation  $f\sigma_8(z)$  and the distance modulus  $\mu(z)$  along with its covariance matrix. We use the resulting models and trained neural networks to generate synthetic data, over which we perform a Bayesian inference analysis to test a couple of dark energy models. Finally, in Section 5 we expose our final comments. Furthermore, within the appendices a brief description of variational autocoders is included and how we train them to learn from the covariance matrix with systematic errors.

# 2 Cosmological and statistical background

In this section we introduce some of the cosmological models, functions and datasets used throughout this work. We also provide a brief overview of the relevant concepts of Bayesian inference and the essentials of Artificial Neural Networks. Along this paper we use the geometric unit system where  $\hbar = c = 8\pi G = 1$ .

#### 2.1 Cosmological models and datasets

#### Models

The Friedmann equation describing the late-time dynamical evolution for a flat-ΛCDM model can be written as

$$H(z)^{2} = H_{0}^{2} \left[ \Omega_{m,0} (1+z)^{3} + (1 - \Omega_{m,0}) \right], \tag{1}$$

where H is the Hubble parameter and  $\Omega_m$  is the matter density parameter; subscript 0 attached to any quantity denotes its present-day (z=0) value. In this case, the EoS is w(z)=-1.

A step further to the standard model is to consider the dark energy being dynamic, where the evolution of its EoS is usually parameterised. A commonly used form of w(z) is to take into account the next contribution of a Taylor expansion in terms of the scale factor  $w(a) = w_0 + (1-a)w_a$  or in terms of redshift  $w(z) = w_0 + \frac{z}{1+z}w_a$  (CPL model: [20, 57]). The parameters  $w_0$  and  $w_a$  are real numbers such that at the present epoch  $w|_{z=0} = w_0$  and  $dw/dz|_{z=0} = -w_a$ , and we recover  $\Lambda$ CDM when  $w_0 = -1$  and  $w_a = 0$ . Hence the Friedmann equation for the CPL parameterisation turns out to be:

$$H(z)^{2} = H_{0}^{2} \left[ \Omega_{m,0} (1+z)^{3} + (1 - \Omega_{m,0})(1+z)^{3(1+w_{0}+w_{a})} e^{-\frac{3w_{a}z}{1+z}} \right].$$
 (2)

As part of some simple models that allow deviations from  $\Lambda$ CDM we also use the polynomial-CDM model (PolyCDM) [58], that can be thought as a parameterisation of the Hubble function [59]. This model has the following Friedmann equation:

$$H(z)^{2} = H_{0}^{2} \left[ \Omega_{m,0} (1+z)^{3} + \Omega_{1,0} (1+z)^{2} + \Omega_{2,0} (1+z)^{1} + (1 - \Omega_{m,0} - \Omega_{1,0} - \Omega_{2,0}) \right], \tag{3}$$

where  $\Omega_{1,0}$  and  $\Omega_{2,0}$  are two additional parameters, which within the  $\Lambda$ CDM both of them remain absent. Nevertheless, in [60]  $\Omega_{2,0}$  is interpreted as a 'missing matter' component introduced to allow a symmetry that relates the big bang to the future conformal singularity. We recover  $\Lambda$ CDM when  $\Omega_{1,0}=0$  and  $\Omega_{2,0}=0$ .

## Datasets

Cosmic chronometers (CC) are galaxies that evolve slowly and allow direct measurements of the Hubble parameter H(z). These measurements have been collected along several years [61, 62, 6, 63, 64, 65, 66, 67], and now 31 data points are available within redshifts between 0.09 and 1.965, along with their statistical errors. Given a Friedmann equation from a cosmological model then a theoretical value for H(z) can be obtained and compared directly with these measurements.

The **growth rate measurement** is usually referred to the product of  $f\sigma_8(a)$  where  $f(a) \equiv d \ln \delta(a)/d \ln a$  is the growth rate of cosmological perturbations given by the density contrast  $\delta(a) \equiv \delta \rho/\rho$ , being  $\rho$  the energy density and  $\sigma_8$  the normalisation of the power spectrum on scales within spheres of  $8h^{-1}$ Mpc [68]. Therefore, the observable quantity  $f\sigma_8(a)$  [or equivalently  $f\sigma_8(z)$ ] is obtained by solving numerically

$$f\sigma_8(a) = a \frac{\delta'(a)}{\delta(1)} \sigma_{8,0}. \tag{4}$$

The  $f\sigma_8$  data are obtained through the peculiar velocities from Redshift Space Distortions (RSD) measurements [69] observed in redshift survey galaxies or by weak lensing [70], where the density perturbations of the galaxies are proportional to the perturbations of matter. An extended version of the Gold-2017 compilation is available at [71], with 22 independent measurements of  $f\sigma_8(z)$  from redshift space distortion measurements from a variety of surveys (see references therein).

**Supernovae** (SNeIa). Let us assume a spatially flat universe, for which the relationship between the luminosity distance  $d_L$  and the comoving distance D(z) is given by:

$$d_L(z) = \frac{1}{H_0}(1+z)D(z), \quad \text{with} \quad D(z) = H_0 \int \frac{dz}{H(z)},$$
 (5)

Thus, the observable quantity is computed by the distance modulus  $\mu(z) = 5 \log d_L(z) + 25$ .

The SNeIa dataset used in this work corresponds to the Joint Lightcurve Analysis (JLA), a compilation of 740 Type Ia supernovae. It is available in a binned version that consists in 31 data points with a covariance matrix  $C_{jla} \in \mathbb{R}^{31 \times 31}$  related to the systematic measurement errors [5].

#### 2.2 Artificial neural networks

The ANNs consist of several sets of neurons or nodes grouped in layers, and connections between them with an associated number called weight. In general, the learning mechanism of an ANN is as follows:

• The first layer of neurons reads the features of the dataset. In each connection between neurons is assigned a random number called weight (we use random numbers with normal distribution centred on 0 with standard deviation of 0.01). The input data make up a matrix  $X_1$  and provide the values for the first layer of nodes. The  $X_i$  refers to the values of nodes in the i-th layer. The weights make up another matrix  $W_i$  and they are the values for the connections between the i-th and the (i+1)-th layers. The product Z of these two matrices is the following:

$$Z_{i+1} = W_i^T X_i, (6)$$

where  $W_i \in \mathbb{R}^{m \times n}$ , with m and n the number of nodes in the i-th and (i+1)-th layers respectively.  $X_i$  corresponds to the i-th layer, therefore has m dimensions. It is worth to apply the transpose of  $W_i$  in order to allow the matrix product.

• A nonlinear activation (or transfer) function  $\phi$  modulates  $Z_i$  and assigns values to the next layer of neurons. This process, known as forward propagation, is repeated until the last layer is reached. The values of neurons in subsequent layers are given by:

$$X_{i+1} = \phi(Z_{i+1}). (7)$$

• The value of the neurons in the last layer must be evaluated by an error function (or loss function) which measures the difference between the value given by the ANN and the expected one. The loss function is minimised by an optimisation algorithm such as *gradient descent* combined with the *backpropagation* algorithm to calculate gradients [72, 73]. In this paper we use the mean squared error (MSE) as a loss function which is a usual selection in regression problems:

$$MSE = \frac{1}{n} \sum_{i}^{n} (Y_i - \hat{Y}_i)^2,$$
(8)

where  $Y_i$  is a vector with predictions of the ANN,  $\hat{Y}_i$  a vector with the expected values and n is the number of predictions (or the length of  $Y_i$  and  $\hat{Y}_i$ ). The goal of the training ANN is to minimise the loss function. The number of samples propagated through the network before updating the weights is known as *batch size* and each iteration of the entire data set constitutes an *epoch*.

During backward propagation the weights are updated, then forward propagation is performed again. This is
repeated until the loss function reaches the desired precision and then the neural network is trained and ready
to make predictions.

At the beginning of the ANN training, the original dataset is split in two parts: training and validation sets. A common choice is 80% and 20% respectively. The first set is used to train the ANN, while the validation set contains unseen values, therefore it is useful for testing the performance of the ANN and evaluating its ability to produce a good model to the input dataset. On the other hand, the *Universal Approximation Theorem* [74] states that an ANN with at least one hidden layer with a finite number of neurons can approach any continuous function if the activation function is continuous and nonlinear. Therefore an ANN is capable of learning the intrinsic functions inside cosmological datasets.

Two types of artificial neural networks are implemented in this work: FeedForward Neural Networks (FFNN) and AutoEncoders (AE). The FFNN, also called multilayer perceptrons or deep feedforward networks, are the quintessential deep learning models [75]. In this type of ANNs the connections between layers and the information flow are straightforward. They are composed of one input layer, at least one hidden layer and an output layer. The input is conformed by the independent variables (or features) of the dataset, while the output contains the dependent variables (or labels). On the other hand, the autoencoders [76] are trained to generate a copy of its input on its output. Autoencoders can be thought as two symmetrical coupled ANNs, where the first (encoder) makes a dimensional reduction for the input and obtains a coded representation (vector embedding or latent space) of the original data. The second part

(decoder) takes the coded representation of the data and recovers an instance with the same data type and dimension of the original input. The encoder is a function f that maps the input x with dimension l to an encoded vector l with dimension l, with l:

$$f: x \in \mathbb{R}^l \to h \in \mathbb{R}^m, \tag{9}$$

where  $h_i := f_i(x) = \phi(W_i^T X_i), \ i = 1, 2, ..., m$  with  $\phi$  being the activation function. The decoder is the following g function, that maps the encoded representation with dimension m into an output  $\hat{x}$  with the same dimension l as the original input x:

$$g: h \in \mathbb{R}^m \to \hat{x} \in \mathbb{R}^l. \tag{10}$$

If the activation function, used in the autoencoder, is the identity function, i.e.  $\phi(x) = x$ , then this type of neural network is analogous to the Principal Component Analysis (PCA) technique. In this work, we use a particular type called variational autoencoder (VAE) [77, 78] described in the Appendix A.

As previously mentioned, the ANNs have the ability to approximate any function, in other words, they can generate computational models that generalise the input data. However, an ANN model has several intrinsic parameters (known as hyperparameters) such as the number of layers, number of nodes, batch size, optimiser algorithm and number of epochs, among others. It is worth to select carefully a good combination of them to guarantee that the ANN model has the capability of generalisation, an incorrect choice of them can produce undesirable models, either underfitted or overfitted with respect to the data.

Although the goal of the neural network training is to minimise the loss function, in our case the MSE, the following relationship should be taken into account:

$$MSE = bias^2 + variance, \tag{11}$$

where the bias measures how far away the neural network predictions are from the actual value, while the variance refers to how much the prediction varies at nearby points. As the ANN model gets more complex, the bias can decrease while the variance can increase, this is called the bias-variance dilemma [79]. A model with high variance will be overfitted, while a model with high bias will be insufficient to learn the complexity of the data (underfitting). In both cases, the model generated by the neural network have inaccurate predictions. One way to avoid this problem is by monitoring the behaviour of the loss function throughout the training epochs, both in the training set and in the validation set. A common practice to prevent an incorrect fitting in the ANN model is to increase the size of the training set, otherwise we need to calibrate carefully the hyperparameters of our ANN models to achieve acceptable results. There are several approaches to tune the hyperparameters [80, 81, 64, 82]. Because our ANNs have relatively simple architectures (between two and five hidden layers and just a few thousands of neurons) we use a common empirical strategy based on a grid of hyperparameters [82]. The hyperparameter grid consists of a selection of possible values for all the ANN parameters to be adjusted. The ANN is trained for all combinations included on the grid to find the one that provides the lowest value for the loss function in the validation set. In addition, it is necessary to verify that the loss function of both, the validation set and the training set, have a convergent behaviour to ensure that the ANN model is well trained (neither overfitting nor underfitting). Whenever underfitting occurs it is mainly because the behaviour of the loss function in the test and training sets shows that the model has a high error in its predictions and, therefore, also a high bias. On the other hand, overfitting occurs when the loss function evaluated at the validation shows an increasing trend, or because there is a considerable gap with the loss function of the training set, thus, it has a high variance. We use the difference between the predictions of the last two epochs ( $\Delta MSE_{val}$ ) to get an idea about the variance of the ANN model, the smaller this error, the smaller the variance.

Another important concept is the dropout (DO), a regularisation technique [83] that allows smaller values to be achieved in the loss function and prevents overfitting. It consists in randomly turning off neurons during training, so the neurons that operate at each epoch are different. The associated hyperparameter is a scalar value that indicates the probability of turning off a neuron in each epoch. Due to its random nature, the dropout can be used as a Monte Carlo simulation [84]. When an ANN is trained, the dropout can be implemented in such a way that each prediction is different because the active neurons are different at each epoch. Therefore, it is possible to make several predictions, and thus obtain the average and standard deviations. Using this formalism, dubbed Monte Carlo dropout (MC-DO) [84] we can obtain a statistical uncertainty of a trained ANN model. We apply the dropout method to the FFNNs implemented in this work and compare the results with those solely with FFNNs.

### 2.3 Bayesian inference

Given a set of observational data and a mathematical expression for a cosmological model, a conditional probability density function can be constructed in terms of the model parameters and the observables. There are many ways to infer the combination of parameters that best fit to the data. In cosmology, Bayesian inference algorithms have been

100

Output

 $\mu(z)$ 

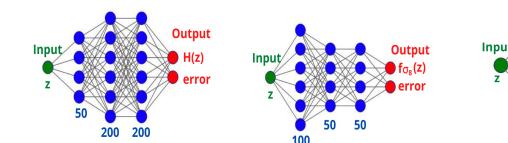


Figure 1: Neural network architectures chosen for cosmic chronometers (CC),  $f\sigma_8$  measurements and distance modulus in JLA respectively; the batch size found for each case was: 16, 1 and 1. In the last architecture, there is only one node in the output layer because the errors are computed with a variational autoencoder (described in the Appendix A) given the original covariance matrix of the systematic errors. Blue numbers indicate the nodes in each layer.

used prominently [85, 86, 87]; however, methods such as the Laplace approximation [88], genetic algorithms [89, 44], simulated annealing [90] or particle swarm optimisation [91] have also been explored.

Here, we use the Bayesian inference to test the quality of our synthetic data generated by the trained ANNs in comparison with the real data. Bayesian statistics is a paradigm in which probabilities are computed given the prior knowledge of the data [92, 93]. It is based on Bayes' Theorem:

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)},\tag{12}$$

where D represents the observational dataset and  $\theta$  the set of free parameters in the theoretical model.  $P(\theta)$  is the prior probability density function and represents the previous knowledge of the parameters  $\theta$ .  $P(D|\theta)$  is the likelihood function and indicates the conditional probability density function of the data given the model. Finally, P(D) is a normalisation constant, that is, the likelihood marginalisation and is called the Bayesian evidence and it is very useful in model comparison, for example it has been used in several papers to compare dark energy models through the Bayes factor and the Jeffrey's scale [17, 36].

# 3 Methodology

Throughout this work we use FFNNs for datasets with diagonal covariance matrices and autoencoders when correlations between the measurements are present, for instance within the JLA dataset.

In general, for the three types of cosmological observations (CC,  $f\sigma_8$  and SNeIa) we have followed the next steps to find out a suitable neural network model for the corresponding data:

- We train several neural network configurations to gain insights about the complexity that their architecture require to model the data. According to the results of the loss function, we choose a number of layers.
- Several values are suggested for each hyperparameter of the neural network, based on the intuition achieved in the first step, a grid is formed that must be traversed to find the combination that provides the minimum value of the loss function. Among the hyperparameters it is the batch size, the number of nodes per layer and, in some cases, the dropout.
- The best ANN architectures found for each case are shown in Figure 1. The first two correspond to the CC and  $f\sigma_8$  datasets respectively, for which 320 combinations were tested up to three hidden layers: number of nodes in  $\{50, 100, 150, 200\}$  and the batch size in  $\{1, 4, 8, 16, 32\}$ . We found that for the compressed JLA dataset a one-layer neural network works best, so we refined the third architecture among 20 combinations, varying the number of nodes in  $\{30, 50, 100, 150, 200\}$  and the batch size in  $\{1, 2, 4, 8\}$ .
- We train the neural network with the combination of hyperparameters chosen in the previous step with a correct number of epochs. We verify the behaviour of the loss function in the training and validation sets to check that our model is neither underfitted nor overfitted. The effect of the epochs in the learning process using the first two ANN architectures of Figure 1 is shown in Figure 2.
- Once the neural network is trained, we can generate synthetic data points with absent redshift in the original datasets.

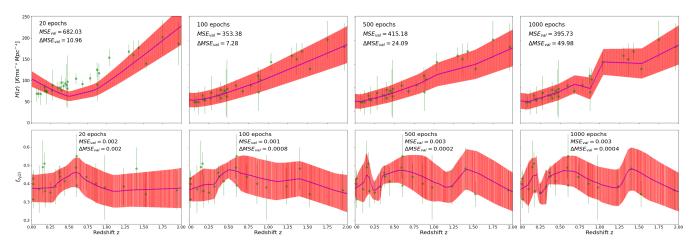


Figure 2: Effect of the number of epochs in the training with the CC dataset (top) and with the  $f\sigma_8(z)$  measurements (bottom). The first case (20 epochs) shows underfitting, while considering 1000 epochs shows overfitting. In the  $f\sigma_8$  dataset, the cases for 500 and 1000 epochs present overfitting. In both cases, we choose 100 epochs due to the lowest value of MSE and  $\Delta$ MSE in the validation set. Green points display real data-points with error bars, and in purple synthetic data along with red error bars.

- By making several predictions with the neural network, the reconstruction of the data can be appreciated and compared with the original data. If the statistical behaviour of the synthetic data is not consistent, the neural networks must be retrained.
- We store the output, for a certain number of predicted data points as well as the new covariance matrix in order to be able to do the Bayesian inference of the cosmological models with these artificial data.
- We compare the parameter estimation of the synthetic data with the original set to verify they are statistically consistent and to analyse their differences. For this purpose, we use the SimpleMC<sup>1</sup> package [94], initially released at [58], along with a modified version of the *dynesty* nested sampling library [95], which allows to do the parameter estimation and Bayesian evidence calculation.

We have developed a package called CRANN<sup>2</sup> that contains the ANNs models already trained to produce synthetic cosmological data given a set of arbitrary redshifts. All the ANNs used in this work and their hyperparameter tuning were based on Tensorflow<sup>3</sup> and Keras<sup>4</sup> Python libraries.

In the case of cosmic chronometers and  $f\sigma_8$  measurements we use FFNNs. These types of data have a diagonal covariance matrix and hence it can be arranged into a single column of the same length as the number of measurements. Therefore, these two datasets have three features (columns): the redshift z, the function f(z), and the related error (taken from the diagonal of the covariance matrix). Once the FFNN is trained, for a given of set of points (redshifts) we can output the cosmological function together with its simulated statistical error. It should be noted that the neural networks learn to generate the cosmological function and the error, where the latter is the result of modelling the original statistical errors from the observational data set.

On the other hand, through the analysis of the JLA SNeIa compilation, we also use a FFNN to learn the behaviour of distance modulus  $\mu(z)$  in a similar fashion we did for the CC and  $f\sigma_8$ . However, in order to handle the full covariance matrix we use a VAE as described in the Appendix A; using this type of neural network allows us to map the distribution of the distance modulus to the distribution of the coded representation of the autoencoder to generate new covariance matrices. One restriction of this method to bear in mind is that the new matrix has to have the same dimension as the original one. However, we can generate a matrix given any combination of new redshifts, provided that this set has the same length as the original measurements.

In addition to the above procedure, we slightly modify the FFNNs to implement an epistemic calculation of their uncertainties using the dropout [84]. In this way, we add dropout between the layers of the FFNNs and run the Monte Carlo dropout several times to obtain average values and uncertainties for each prediction (as described in Section

www.github.com/ja-vazquez/SimpleMC

<sup>&</sup>lt;sup>2</sup>Cosmological Reconstructions with Artificial Neural Networks (CRANN). The link will be available upon acceptance.

<sup>&</sup>lt;sup>3</sup>www.tensorflow.org

<sup>&</sup>lt;sup>4</sup>www.keras.io

2.2). We combine our FFNN designs with the implementation of MC-DO layers from astronn<sup>5</sup>[96], and compare the results of this method with the previous ANNs implementations. Because dropout is a regularisation technique, the number of epochs is irrelevant for a large enough set. The error predictions and the uncertainties are independent, therefore the total standard deviation is:

$$\sigma = \sqrt{er_p + \sum_i u_i^2},\tag{13}$$

where  $u_i$  is the epistemic uncertainty involved with the FFNN used and  $er_p$  is the error prediction.

As mentioned above, once the non-parametric reconstruction is obtained we can generate synthetic data and use them to perform a Bayesian inference procedure. For these purposes, we use the following flat priors: for the matter density parameter today  $\Omega_m \in [0.05,\ 0.5]$ , for the physical baryon density parameter  $\Omega_b h^2 \in [0.02,\ 0.025]$ , for the reduced Hubble constant  $h \in [0.4,\ 0.9]$ , and for the amplitude of the (linear) power spectrum  $\sigma_8 \in [0.6,\ 1.0]$ . When assuming the CPL parameterisation, we use  $w_0 \in [-2.0,\ 0.0]$  and  $w_a \in [-2.0,\ 2.0]$ ; and for the PolyCDM model, we use  $\Omega_1 \in [-1.0,\ 1]$  and  $\Omega_2 \in [-1.0,\ 1]$ . The h parameter refers to the dimensionless reduced Hubble parameter today  $H/100 {\rm kms}^{-1} {\rm Mpc}^{-1}$ .

## 4 Results

In order to perform the reconstructions of the Hubble parameter H(z), the growth rate measurement  $f\sigma_8(z)$  and the distance modulus  $\mu(z)$  we apply two different methods implemented with the feedforward neural networks shown in Figure 1: i) using the trained neural network (FFNN) and ii) along with the FFNN, by considering uncertainties with the Monte Carlo dropout (FFNN+MC-DO). To test the quality of our ANNs predictions we perform the Bayesian inference procedure for the CPL and PolyCDM models with the original data and with the data generated by the neural networks. Then, and to improve the constraints on the free parameters, in some cases we also include a compressed version of Planck-15 information (treated as a BAO experiment located at redshift z=1090 [58]).

### **Reconstruction of** H(z)

Once the chosen FFNN is trained with the CC dataset (green points in Figure 3), we input new redshift values (z) and let the network to predict the corresponding values for H(z) and their errors. In the left-panel of Figure 3, we generate 1000 synthetic data (magenta points) with their respective error bars (red lines) with the FFNN. These errors are the result of the ANN modelling the errors contained in the original dataset.

Besides the intrinsic error associated with the datasets, we consider an uncertainty related with the FFNN by adding a Monte Carlo dropout between each layer of the chosen FNNN architecture, under the method described in [84]. Among several tests to dropout values between [0,0.5], we found that a good value for the dropout is 0.3 and we trained the FFNN with MC-DO along 1000 epochs. After training the new FFNN with dropout between each layer, we can make the predictions to 1000 unseen redshifts and, with 100 executions of MC-DO for each prediction. We obtain the right-panel of Figure 3, that contains the total standard deviation considering the uncertainties of the neural network. In this case, the result has variations caused by the probabilistic nature of the MC-DO and, now, the error bars contain information about the statistical uncertainty of the FFNN trained model; indeed the error is larger than in the FFNN alone case.

An interesting feature shown in both panels of Figure 3 is that, despite the original dataset does not contain a measurement for H(z=0), the FFNN prediction is  $H_{pred}(z=0)=75.09\pm15.49~{\rm km~s^{-1}~Mpc^{-1}}$  and the prediction of the FFNN with Monte Carlo dropout is  $H_{pred}(z=0)=77.07\pm15.91~{\rm km~s^{-1}~Mpc^{-1}}$ . Both results have a better matching to the one measured with Cepheid variables [97]:  $H_0=73.24~{\rm km~s^{-1}~Mpc^{-1}}$ . In addition, it can be appreciated that as the redshift increases, the model generated by the FFNNs approaches  $\Lambda$ CDM with  $H_0=67.40~{\rm km~s^{-1}~Mpc^{-1}}$  measured by Planck mission [4]. It is worth to remember that the ANN models do not have any prior cosmological or statistical assumption so they have been built solely from the data.

# **Reconstruction of** $f\sigma_8(z)$

Similarly to the Hubble parameter case, we apply the same methodology but now with measurements of the  $f\sigma_8$  function [71]. In order to generate the reconstruction plot shown in the left panel of Figure 4, once the FFNN is trained, we generate 1000 synthetic data points (red points). Besides, we put the evaluation of  $f\sigma_8(z)$  from CPL model in three different scenarios of  $w_0$  and  $w_a$ .

<sup>&</sup>lt;sup>5</sup>https://github.com/henrysky/astroNN

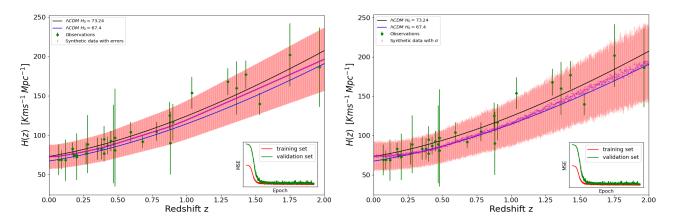


Figure 3: H(z) reconstructions produced with 1000 synthetic data points generated with FFNNs. Left: Purple line represents the FFNN predictions for H(z) along with their error bars in red colour. Right: Similarly to FFNN but adding MC-DO, we executed 100 times the Monte Carlo dropout to compute the uncertainties of the predictions, therefore the purple points are the average predictions of the MC-DO executions and the red error bars are the uncertainties of the FFNN plus the error predictions (see Equation 13). In both cases, we compare the non-parametric reconstruction with the original cosmic chronometers (green bars) and H(z) from  $\Lambda$ CDM, as shown in the labels. The small panels displayed the receptive behaviour of the loss function (MSE) in the training (red) and validation (green) sets.

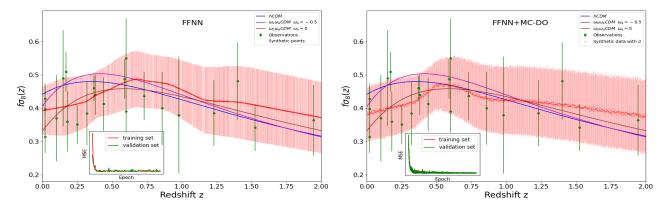


Figure 4: Reconstruction of  $f\sigma_8(z)$  with 1000 synthetic data points (red dots) and their respective errors (red bars) learned by ANNs. Left: FFNN alone, red line is conformed by the predictions of  $f\sigma_8$ . Right: FFNN using Monte Carlo dropout, the averages of 100 executions of MC-DO are indicated with the red line and their standard deviations are added to the error predictions. In both cases the small panels displayed the receptive behaviour of the loss function (MSE) in the training (red curve) and validation (green curve) sets.

We added Monte Carlo dropout to the FFNN, shown in the second ANN architecture of Figure 1, to be able to calculate the uncertainties of the ANN. We train this new FFNN along 2000 epochs. In this case, among several tests to dropout values between [0,0.5], we choice a dropout of 0.1 because it had the best performance. Then we obtain, with 1000 synthetic  $f\sigma_8$  data points, the reconstruction of the right-panel of Figure 4, where the purple line is the average obtained by MC-DO predictions and the error bars contain an error conformed by the standard deviations (uncertainties) of MC-DO for each prediction plus the error predictions. We can notice, that in both cases, the models plotted are within the reconstruction and hence this dataset by itself may provide loose constraints on the CPL parameters. However, the values  $w_0 = -0.8$  and  $w_a = -0.4$  (brown line) seem to have a better agreement with the reconstruction, as we shall see below.

### Distance modulus $\mu(z)$ reconstructions

Regarding to the distance modulus  $\mu(z)$ , we train a FFNN (last ANN in Figure 1) for a given redshift. We assume a gaussian distribution for the predictions of the distance modulus and using a trained VAE we produce a new covariance matrix (see Appendix A). Once the FFNN is trained, we can generate synthetic data points for unseen redshifts and reconstruct the  $\mu(z)$  function as it can be appreciate in the left panel of Figure 5. In particular, we use 31 log-uniformly

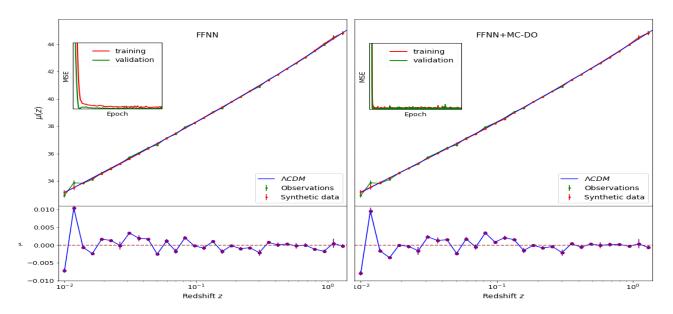


Figure 5: *Left*: 31 new data points (red dots) generated with FFNN. *Right*: 31 new data points (red dots) generated with FFNN+MC-DO. In both cases it is shown their receptive behaviour of the loss function (MSE) in the training (red curve) and validation (green curve) sets along the chosen number of epochs for each case (300 and 1800). The 31 green dots are the original points from the binned version of JLA.

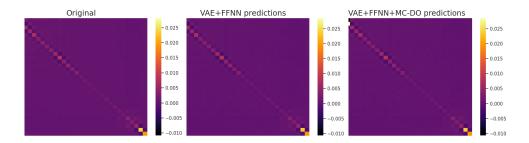


Figure 6: *Left*: Original covariance matrix with systematic errors from JLA compilation (binned version). Covariance matrices predicted by the VAE (*Middle*) and VAE with MC-DO (*Right*).

distributed redshifts over the interval  $z \in [0.01, 1.3]$ . To apply MC-DO, we did several tests to dropout values in the interval [0, 0.5] and we found that a dropout with 0.01 value has a good performance. We executed 100 times MC-DO to obtain the right panel of Figure 5.

An important point to bear in mind is that when using a full covariance matrix we need to restrict to 31 synthetic data points in order to generate the covariance matrix with the VAE for mapping to the new points in the latent space. The synthetic covariance matrix (Figure 6) was generated by the VAE, with the distance modulus predictions from the FFNNs (refer to the appendix for a detail description).

## Parameter estimation with synthetic data

Neural networks allow us to produce data models with several parameters (neural network weights) which are difficult to interpret, however with the use of synthetic data generated by these models we can analyse them with Bayesian inference and compare their results with those obtained from the original data. Thus we performed the Bayesian inference analysis to estimate the best fit parameters of the CPL and PolyCDM models. The aim of this procedure is to look for possible deviations of the  $\Lambda$ CDM model with the neural networks approach.

In addition to the three original datasets (cosmic chronometers,  $f\sigma_8$  measurements and binned JLA compilation), we have created two datasets for each type of observation from the trained FFNNs with and without MC-DO. As a proof of the concept, the new datasets for CC and  $f\sigma_8$  consisted of 50 random uniformly distributed points in redshift, while for SNeIa they were 31 log-uniformly distributed in redshift (same size as the original dataset). For the SNeIa case we also generated its respective covariance matrix with the decoder part of the trained VAE.

We have used the data from CC,  $f\sigma_8$  measurements and JLA separately, and also some combinations of them. The most representative results are in Figure 7 along with Table 1, which contains mean values and standard deviations, and they have been sorted according to the datasets used as a source (original, FFNN, and FFNN+MC-DO), and to the models involved ( $\Lambda$ CDM, CPL, and PolyCDM). It is indicated when the Planck point has been added to the data sets. Results are displayed for the reduced Hubble parameter h,  $\sigma_8$ ,  $w_0$  and  $w_a$  measurements for the CPL model, and  $\Omega_1$  with  $\Omega_2$  when the model is PolyCDM. In addition, the last column of the Table 1 contains the  $-2\log\mathcal{L}_{max}$  of the Bayesian inference process for each case. Before analysing each scenario separately, it is worth mentioning that there are some generalities in the results. In general, it can be noted that when using a single source separately (no Planck information added) the constraints are consistent among each other, that is, they all have a similar best-fit (maximum likelihood) and are in agreement with the  $\Lambda$ CDM model.

In the case of parameter estimation using exclusively the CC dataset, the first two panels of Figure 7 (and first block in Table 1) show that the best-fit values are mutually contained within their  $1\sigma$  standard deviations and are in agreement with the  $\Lambda$ CDM values. However, we can notice that when Planck information is added the reduced Hubble parameter value slightly increases for the ANNs. For instance, for the CPL model with original data the constraints are  $h=0.673\pm0.046$  whereas for the synthetic data it increases to  $0.713\pm0.059$  for the FFNN and  $0.726\pm0.063$  for the FFNN+MC-DO; in fact these values obtained by the synthetic data are closer to the Hubble parameter value of the Cepheid variables than to the Planck mission value. This issue, as a supplement to Figure 3, shows that the neural network models generated by cosmic chronometers are sensing the Hubble tension, although considering the size of the standard deviation values, all the results of the parameter estimation are still statistically consistent with each other. Something similar happens when added Planck information to JLA SNeIa and assuming the CPL model. With the original data the constraints are  $h=0.695\pm0.021$  whereas for the synthetic data they increase to  $0.704\pm0.025$  for the FFNN and  $0.712\pm0.026$  for the FFNN+MC-DO. However both datasets are still statistically consistent, within  $1\sigma$ , with the  $\Lambda$ CDM parameters. If taken into account the JLA+CC combination with the Planck information, the increment of the reduced Hubble parameter is still present but also a small deviation of  $\Lambda$ CDM (about  $1\sigma$ ) for FFNN+MC-DO, with constraints of  $w_0=-0.957\pm0.141$  and  $w_a=-0.563\pm0.669$ .

On the other hand, considering only measurements of  $f\sigma_8$ +Planck for the synthetic data, the  $w_0$  and  $w_a$  constraints suggest a slight deviation from  $\Lambda$ CDM. With the FFNN data the values are  $w_0 = -0.657 \pm 0.172$  and  $w_a = -0.493 \pm 0.265$ , and for FFNN+MC-DO we have  $w_0 = -0.673 \pm 0.183$  and  $w_a = -0.364 \pm 0.221$ . In fact, it can be seen in the Figure 7 that the cosmological constant is right on the limits of the  $2\sigma$  contours.

By using all datasets combined CC+ $f\sigma_8$ +JLA we have performed a Bayesian inference to the three models  $\Lambda$ CDM, CPL and PolyCDM. With the original datasets we found consistency throughout the models with the  $\Lambda$ CDM parameters and slight shift when using the synthetic data, for instance higher values for the reduced Hubble parameter, lower for the  $\sigma_8$  parameter and for  $w_0$ :  $-0.916 \pm 0.065$  (FFNN) and  $-0.925 \pm 0.068$  (FFNN+MC-DO). Deviations of the standard values are enhanced once we use synthetic data along with Planck information. This can be seen on the constraints of the PolyCDM model for the FFNN source:  $\Omega_1 = 0.272 \pm 0.194$ ,  $\Omega_2 = -0.092 \pm 0.058$ . Also, based on the improvement in the fit alone ( $\sqrt{2\Delta \ln \mathcal{L}_{max}}$ ), and using the same source, we found a preference to the data for the CPL model of  $1.5\sigma$  and  $1.7\sigma$  for the PolyCDM. That is, the Artificial Neural Network by itself is finding deviations from the standard cosmological model.

The above discussion suggests that, if the models generated by the neural networks are correct, hypothetical new observations within the range of the existing ones would tend to move away from  $\Lambda$ CDM. Therefore, parameter estimation in the CPL and PolyCDM models in conjunction to the models generated by the neural networks suggest that  $\Lambda$ CDM does not have the best match to the data. In all cases, the addition of the Planck point increases the tension and the need for a model beyond  $\Lambda$ CDM.

To reinforce the idea that models generated by the neural networks depart from  $\Lambda$ CDM, from the posterior distribution samples for CPL, we obtained the posterior distribution of its corresponding EoS, as shown in Figure 8 (using fgivenx Python library [98]). From these plots it can be seen that w=-1 lies in the most probable region within  $1\sigma$  with the original data; however, in the case of the synthetic data the cosmological constant moves away from the most probable region, still within  $1\sigma$  without considering the Planck point, and outside  $1\sigma$  when it is taken into account.

Finally, as part of the Bayesian analysis, we perform a model comparison with Bayesian evidences Z through the Bayes' factor B and the Jeffrey's scale [17]. Table 2 shows the log-Bayes' factor of  $\Lambda$ CDM compared to CPL and PolyCDM

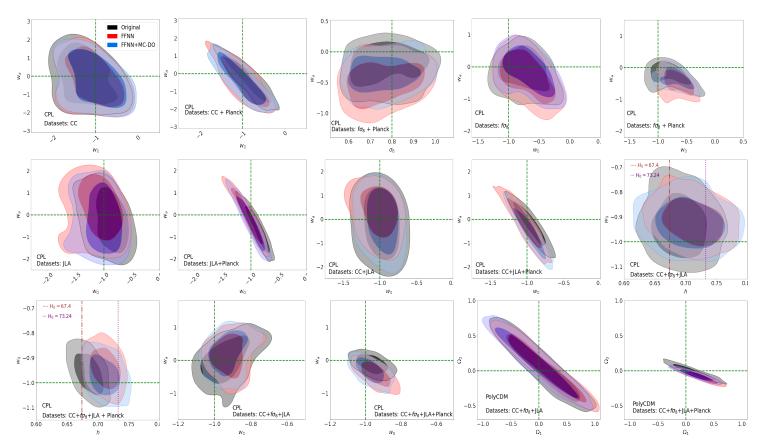


Figure 7: 2D marginalised posterior distributions from different combinations of datasets: original data, synthetic datasets from FFNN and FFNN+MC-DO. The green dashed lines ( $w_0=-1, w_a=0$ ) and ( $\Omega_1=0, \Omega_2=0$ ) correspond to the  $\Lambda$ CDM model. The constraints are plotted with  $1\sigma$  and  $2\sigma$  confidence contours.

models using the different sources of data. It can be seen that with the synthetic data the penalisation of having extra parameters decreases from strong advantage to an inconclusive advantage due to the improvement of the fit in both models. However, it is worth noting that  $\Lambda$ CDM stays with a slight advantage.

Source	CPL	PolyCDM
Original	3.651	2.837
FFNN	1.823	0.687
FFNN+MC-DO	2.464	1.159

Table 2: Log-Bayes' factor  $\ln(B) = \ln(Z_{\Lambda CDM}) - \ln(Z)$  of  $\Lambda \text{CDM}$  with respect the other models using the same data source for each case. The combined dataset used in this table is JLA+CC+ $f_{\sigma 8}$ +Planck

## 5 Conclusions

Throughout this work, we have shown that well-calibrated artificial neural networks have the capacity to produce non-parametric reconstructions from which synthetic cosmological data, statistically consistent with the originals, can be generated even when the datasets are small.

We have explored the generation of synthetic covariance matrices through VAE, and the results have allowed us to carry out Bayesian inference without drawbacks. However, for larger datasets, we believe that it will be convenient to use convolutional layers in the autoencoder and a slightly different approach to dealing with the computing demand.

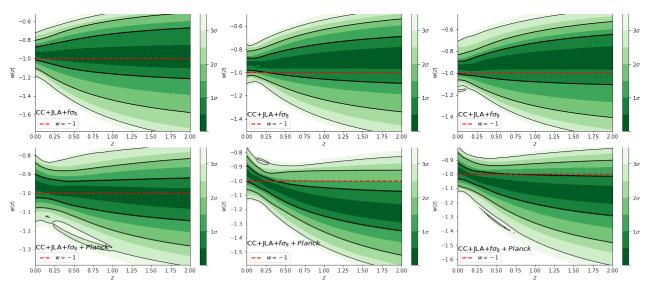


Figure 8: Posterior probability distribution functions of the Dark Energy EoS considering CPL parameterisation, by using original data, FFNN and FFNN+MC-DO respectively. Planck-15 point information is additionally included in the lower panels.  $1-3\sigma$  confidence intervals are plotted as black lines.

Using the Monte Carlo dropout method allows to have more information about the predictions of neural networks through their epistemic estimation of uncertainty. The results obtained have also contributed to both the methodological and the cosmological analysis and to validate the outputs from the ANNs without this method as they are very similar.

The models generated by the neural networks were produced exclusively from the data, therefore they offer the possibility of reconstructing cosmological functions without assumptions about the data distribution and without assuming any cosmological model as a starting point. From the non-parametric reconstructions produced with the neural networks, we were able to observe how the Hubble parameter changes as the cosmic chronometers are at higher redshifts, as suggested by the current Hubble tension. We could also note that SNeIa are observations very much in accordance with  $\Lambda$ CDM and, in contrast, that  $\Lambda$ CDM is not the best model to describe  $f\sigma_8$  measurements. Overall, using Bayesian inference on the CPL and PolyCDM models with the synthetic neural network data, we have observed that the  $\Lambda$ CDM model does not perfectly match these data and loses some of the advantage given by the original observations.

It is worth mentioning that the cosmological results obtained in this work are limited to current cosmological observations and have been sufficient to show some interesting cosmological features from the data alone. We have shown that our method can be a good complement to the traditional Bayesian analysis and, moreover, could be applied to other types of cosmological observations and models, as well as to perform some forecasts. In this way, we can see that the use of neural networks, from their models created for the data and the generation of synthetic data, can complement the analysis of cosmological models and improve the interpretations of their behaviours. We plan to apply similar techniques to other types of cosmological data, including a complete set of covariance matrices.

# Acknowledgements

This work was partially supported by CONACyT-Mexico scholarship, BEIFI-IPN. RG-S acknowledges the support provided by SIP20200666-IPN and SIP20210500-IPN grants and FORDECYT-PRONACES-CONACYT CF-MG-2558591. JAV acknowledges the support provided by FOSEC SEP-CONACYT Investigación Básica A1-S-21925, FORDECYT-PRONACES-CONACYT 304001 and UNAM-DGAPA-PAPIIT IA104221.

### A Reconstruction of covariance matrix with VAE

In this appendix we briefly describe the idea of a variational autoencoder (VAE) [77, 78], however for an extended review see [99, 100]. In addition, we explain a first approach method used to generate synthetic covariance matrices from the original covariance matrix of the JLA SNeIa binned version.

Source	Model		Datasets:	CC			$-2 \ln \mathcal{L}_{max}$
0::1	A CDM		h	$w_0$	$w_a$		14.500
Original	$\Lambda  ext{CDM}$ $ ext{CPL}$	-	$0.678 \pm 0.039$ $0.703 \pm 0.064$	$$ $-1.223 \pm 0.447$	$$ $-0.061 \pm 1.075$		14.502 $14.290$
	CPL	+ Planck	$0.673 \pm 0.004$ $0.673 \pm 0.046$	$-0.867 \pm 0.326$	$-0.325 \pm 0.824$		14.638
FFNN	ΛCDM		$0.698 \pm 0.057$				0.176
TTININ	CPL	-	$0.703 \pm 0.037$ $0.703 \pm 0.071$	$-1.072 \pm 0.431$	$-0.179 \pm 1.025$		0.170
	CPL	+ Planck	$0.713 \pm 0.059$	$-0.962 \pm 0.337$	$-0.485 \pm 0.890$		0.120
FFNN+MC-DO	$\Lambda$ CDM	_	$0.699 \pm 0.063$				0.346
TTWWW DO	CPL	-	$0.689 \pm 0.078$	$-1.014 \pm 0.450$	$-0.227 \pm 1.003$		0.284
	CPL	+ Planck	$0.726 \pm 0.063$	$-1.029 \pm 0.355$	$-0.377 \pm 0.897$		0.808
			Datasets:	$f\sigma_8$			$-2 \ln \mathcal{L}_{max}$
Original	ACDM		$\frac{h}{0.648 \pm 0.147}$	$w_0$	$w_a$	σ <sub>8</sub>	11.932
Original	ΛCDM CPL	-	$0.648 \pm 0.147$ $0.638 \pm 0.135$	$-0.742 \pm 0.264$	$-0.144 \pm 0.468$	$0.787 \pm 0.115$ $0.777 \pm 0.111$	11.932
	CPL	+ Planck	$0.648 \pm 0.062$	$-0.801 \pm 0.229$	$-0.225 \pm 0.254$	$0.771 \pm 0.111$ $0.771 \pm 0.109$	11.944
FFNN	ΛCDM	-	$0.650 \pm 0.144$	_		$0.694 \pm 0.172$	0.292
	CPL	-	$0.648 \pm 0.142$	$-0.701 \pm 0.271$	$-0.290 \pm 0.540$	$0.034 \pm 0.112$ $0.777 \pm 0.111$	0.284
	CPL	+ Planck	$0.628 \pm 0.046$	$-0.657 \pm 0.172$	$-0.493 \pm 0.265$	$0.756 \pm 0.109$	0.374
FFNN+MC-DO	$\Lambda$ CDM	_	$0.651 \pm 0.147$			$0.652 \pm 0.170$	0.984
	CPL	-	$0.632 \pm 0.140$	$-0.674 \pm 0.270$	$-0.156 \pm 0.489$	$0.775 \pm 0.110$	0.960
	CPL	+ Planck	$0.622 \pm 0.046$	$-0.673 \pm 0.183$	$-0.364 \pm 0.221$	$0.756 \pm 0.103$	1.038
			Datasets:	JLA			$-2 \ln \mathcal{L}_{max}$
Original	ΛCDM	_	$\frac{h}{0.638 \pm 0.146}$		$w_a$		33.214
O'I'giillii	CPL	-	$0.652 \pm 0.141$	$-0.901 \pm 0.238$	$-0.216 \pm 0.899$		32.354
	CPL	+ Planck	$0.695 \pm 0.021$	$-0.880 \pm 0.140$	$-0.606 \pm 0.696$		30.528
FFNN	$\Lambda$ CDM	-	$0.645 \pm 0.144$				14.670
	CPL	-	$0.640 \pm 0.137$	$-1.092 \pm 0.277$	$0.287 \pm 0.957$		13.888
	CPL	+ Planck	$0.704 \pm 0.025$	$-1.061 \pm 0.178$	$-0.018 \pm 0.811$		14.808
FFNN+MC-DO	$\Lambda$ CDM	-	$0.643 \pm 0.142$				16.446
	CPL	-	$0.641 \pm 0.135$	$-1.037 \pm 0.248$	$-0.245 \pm 0.996$		16.274
	CPL	+ Planck	$0.712 \pm 0.026$	$-0.994 \pm 0.165$	$-0.395 \pm 0.802$		16.504
			Datasets:	CC+JLA			$-2 \ln \mathcal{L}_{max}$
			h.		$w_{\alpha}$		max
Original	ΛCDM	-	$h = 0.690 \pm 0.030$	$w_0$	$egin{array}{c} w_a \ \end{array}$		47.822
Original	CPL	-	$0.690 \pm 0.030$ $0.687 \pm 0.030$	$w_0$ $-0.980 \pm 0.173$	$$ $-0.156 \pm 0.939$		47.822 47.830
Original		- - + Planck	$0.690 \pm 0.030$	$w_0$			47.822
Original FFNN	CPL CPL ЛСDМ	-	$0.690 \pm 0.030$ $0.687 \pm 0.030$ $0.687 \pm 0.018$ $0.705 \pm 0.037$	$w_0$ 0.980 ± 0.173 -0.946 ± 0.137	$$ $-0.156 \pm 0.939$ $-0.232 \pm 0.592$ $$		47.822 47.830 47.918 14.846
-	CPL CPL ΛCDM CPL	- + Planck - -	$\begin{array}{c} 0.690 \pm 0.030 \\ 0.687 \pm 0.030 \\ 0.687 \pm 0.018 \\ 0.705 \pm 0.037 \\ 0.695 \pm 0.037 \end{array}$	$w_0$ $-0.980 \pm 0.173$ $-0.946 \pm 0.137$ $-1.010 \pm 0.165$	$$ $-0.156 \pm 0.939$ $-0.232 \pm 0.592$ $$ $0.315 \pm 0.715$		47.822 47.830 47.918 14.846 14.096
-	CPL CPL ACDM CPL CPL	-	$0.690 \pm 0.030$ $0.687 \pm 0.030$ $0.687 \pm 0.018$ $0.705 \pm 0.037$	$\begin{array}{c} w_0 \\ \\ -0.980 \pm 0.173 \\ -0.946 \pm 0.137 \\ \\ -1.010 \pm 0.165 \\ -1.031 \pm 0.150 \end{array}$	$$ $-0.156 \pm 0.939$ $-0.232 \pm 0.592$ $$		47.822 47.830 47.918 14.846
-	CPL CPL ACDM CPL CPL ACDM	- + Planck - -	$\begin{array}{c} 0.690 \pm 0.030 \\ 0.687 \pm 0.030 \\ 0.687 \pm 0.018 \\ 0.705 \pm 0.037 \\ 0.695 \pm 0.037 \\ 0.708 \pm 0.021 \\ 0.703 \pm 0.038 \end{array}$	$\begin{array}{c} w_0 \\ \\ -0.980 \pm 0.173 \\ -0.946 \pm 0.137 \\ \\ -1.010 \pm 0.165 \\ -1.031 \pm 0.150 \\ \end{array}$	$ \begin{array}{c}\\ -0.156 \pm 0.939\\ -0.232 \pm 0.592\\\\ 0.315 \pm 0.715\\ -0.167 \pm 0.688\\\\ \end{array} $		47.822 47.830 47.918 14.846 14.096 15.478 16.808
FFNN	CPL CPL ACDM CPL CPL ACDM CPL	- + Planck - - + Planck -	$\begin{array}{c} 0.690 \pm 0.030 \\ 0.687 \pm 0.030 \\ 0.687 \pm 0.018 \\ 0.705 \pm 0.037 \\ 0.695 \pm 0.037 \\ 0.708 \pm 0.021 \\ 0.703 \pm 0.038 \\ 0.698 \pm 0.039 \\ \end{array}$	$\begin{array}{c} w_0 \\ \\ -0.980 \pm 0.173 \\ -0.946 \pm 0.137 \\ \\ -1.010 \pm 0.165 \\ -1.031 \pm 0.150 \\ \\ -0.968 \pm 0.155 \end{array}$	$\begin{array}{c}\\ -0.156 \pm 0.939\\ -0.232 \pm 0.592\\\\ 0.315 \pm 0.715\\ -0.167 \pm 0.688\\\\ -0.046 \pm 0.859 \end{array}$		47.822 47.830 47.918 14.846 14.096 15.478 16.808 16.688
FFNN	CPL CPL ACDM CPL CPL ACDM	- + Planck - -	$\begin{array}{c} 0.690 \pm 0.030 \\ 0.687 \pm 0.030 \\ 0.687 \pm 0.018 \\ 0.705 \pm 0.037 \\ 0.695 \pm 0.037 \\ 0.708 \pm 0.021 \\ 0.703 \pm 0.038 \\ 0.698 \pm 0.039 \\ 0.717 \pm 0.019 \\ \end{array}$	$\begin{array}{c} w_0 \\ \\ -0.980 \pm 0.173 \\ -0.946 \pm 0.137 \\ \\ -1.010 \pm 0.165 \\ -1.031 \pm 0.150 \\ \\ -0.968 \pm 0.155 \\ -0.957 \pm 0.141 \\ \end{array}$	$ \begin{array}{c}\\ -0.156 \pm 0.939\\ -0.232 \pm 0.592\\\\ 0.315 \pm 0.715\\ -0.167 \pm 0.688\\\\ \end{array} $		47.822 47.830 47.918 14.846 14.096 15.478 16.808 16.688 17.252
FFNN	CPL CPL ACDM CPL CPL ACDM CPL	- + Planck - - + Planck -	$\begin{array}{c} 0.690 \pm 0.030 \\ 0.687 \pm 0.030 \\ 0.687 \pm 0.018 \\ 0.705 \pm 0.037 \\ 0.695 \pm 0.037 \\ 0.708 \pm 0.021 \\ 0.703 \pm 0.038 \\ 0.698 \pm 0.039 \\ \end{array}$	$\begin{array}{c} w_0 \\ \\ -0.980 \pm 0.173 \\ -0.946 \pm 0.137 \\ \\ -1.010 \pm 0.165 \\ -1.031 \pm 0.150 \\ \\ -0.968 \pm 0.155 \\ -0.957 \pm 0.141 \\ \hline \\ \textbf{CC+JLA+} \ fo_8 \end{array}$	$\begin{array}{c}\\ -0.156 \pm 0.939\\ -0.232 \pm 0.592\\\\ 0.315 \pm 0.715\\ -0.167 \pm 0.688\\\\ -0.046 \pm 0.859\\ -0.563 \pm 0.669 \end{array}$	$\sigma_8$	47.822 47.830 47.918 14.846 14.096 15.478 16.808 16.688
FFNN	CPL CPL ACDM CPL CPL ACDM CPL	- + Planck - - + Planck -	$0.690 \pm 0.030$ $0.687 \pm 0.030$ $0.687 \pm 0.018$ $0.705 \pm 0.037$ $0.695 \pm 0.037$ $0.708 \pm 0.021$ $0.703 \pm 0.038$ $0.698 \pm 0.039$ $0.717 \pm 0.019$ Datasets:	$\begin{array}{c} w_0 \\ \\ -0.980 \pm 0.173 \\ -0.946 \pm 0.137 \\ \\ -1.010 \pm 0.165 \\ -1.031 \pm 0.150 \\ \\ -0.968 \pm 0.155 \\ -0.957 \pm 0.141 \\ \end{array}$	$\begin{array}{c}\\ -0.156 \pm 0.939\\ -0.232 \pm 0.592\\\\ 0.315 \pm 0.715\\ -0.167 \pm 0.688\\\\ -0.046 \pm 0.859 \end{array}$	$\frac{\sigma_8}{0.795 \pm 0.115}$	47.822 47.830 47.918 14.846 14.096 15.478 16.808 16.688 17.252
FFNN FFNN+MC-DO	CPL CPL ACDM CPL CPL ACDM CPL ACDM CPL ACDM ACDM ACDM	- + Planck - - + Planck -	$\begin{array}{c} 0.690 \pm 0.030 \\ 0.687 \pm 0.030 \\ 0.687 \pm 0.018 \\ 0.705 \pm 0.037 \\ 0.695 \pm 0.037 \\ 0.708 \pm 0.021 \\ 0.703 \pm 0.038 \\ 0.698 \pm 0.039 \\ 0.717 \pm 0.019 \\ \hline \text{Datasets:} \\ h \\ 0.695 \pm 0.032 \\ 0.690 \pm 0.013 \\ \hline \end{array}$	$\begin{array}{c} w_0 \\ \\ -0.980 \pm 0.173 \\ -0.946 \pm 0.137 \\ \\ -1.010 \pm 0.165 \\ -1.031 \pm 0.150 \\ \\ -0.968 \pm 0.155 \\ -0.957 \pm 0.141 \\ \hline {\bf CC+JLA+}  f\sigma_8 \\ w_0 \\ \\ \\ \\ \\ \end{array}$		$0.795 \pm 0.115$ $0.790 \pm 0.109$	$\begin{array}{c} 47.822 \\ 47.830 \\ 47.918 \\ 14.846 \\ 14.096 \\ 15.478 \\ 16.808 \\ 16.688 \\ 17.252 \\ \hline -2 \ln \mathcal{L}_{max} \\ 60.244 \\ 60.33 \end{array}$
FFNN FFNN+MC-DO	CPL CPL ACDM CPL CPL ACDM CPL ACDM CPL CPL ACDM CPL CPL	- + Planck - - + Planck - + Planck	$\begin{array}{c} 0.690 \pm 0.030 \\ 0.687 \pm 0.030 \\ 0.687 \pm 0.030 \\ 0.687 \pm 0.018 \\ 0.705 \pm 0.037 \\ 0.695 \pm 0.037 \\ 0.708 \pm 0.021 \\ 0.703 \pm 0.038 \\ 0.698 \pm 0.039 \\ 0.717 \pm 0.019 \\ \hline \\ Datasets: \\ h \\ 0.695 \pm 0.032 \\ 0.690 \pm 0.013 \\ 0.692 \pm 0.029 \\ \hline \end{array}$	$\begin{array}{c} w_0 \\ \\ -0.980 \pm 0.173 \\ -0.946 \pm 0.137 \\ \\ -1.010 \pm 0.165 \\ -1.031 \pm 0.150 \\ \\ -0.968 \pm 0.155 \\ -0.957 \pm 0.141 \\ \hline \textbf{CC+JLA+} \ f\sigma_8 \\ w_0 \\ \\ -0.933 \pm 0.086 \\ \end{array}$		$0.795 \pm 0.115$ $0.790 \pm 0.109$ $0.763 \pm 0.111$	$\begin{array}{c} 47.822 \\ 47.830 \\ 47.918 \\ 14.846 \\ 14.096 \\ 15.478 \\ 16.808 \\ 16.688 \\ 17.252 \\ \hline -2 \ln \mathcal{L}_{max} \\ 60.244 \\ 60.33 \\ 59.840 \\ \end{array}$
FFNN+MC-DO  Original	CPL CPL ACDM CPL ACDM CPL ACDM CPL CPL ACDM CPL CPL CPL	- + Planck - - + Planck - - + Planck	$\begin{array}{c} 0.690 \pm 0.030 \\ 0.687 \pm 0.030 \\ 0.687 \pm 0.018 \\ 0.705 \pm 0.037 \\ 0.695 \pm 0.037 \\ 0.708 \pm 0.021 \\ 0.703 \pm 0.038 \\ 0.698 \pm 0.039 \\ 0.717 \pm 0.019 \\ \hline \\ Datasets: \\ h \\ 0.695 \pm 0.032 \\ 0.690 \pm 0.013 \\ 0.692 \pm 0.029 \\ 0.685 \pm 0.015 \\ \hline \end{array}$	$\begin{array}{c} w_0 \\ \\ -0.980 \pm 0.173 \\ -0.946 \pm 0.137 \\ \\ -1.010 \pm 0.165 \\ -1.031 \pm 0.150 \\ \\ -0.968 \pm 0.155 \\ -0.957 \pm 0.141 \\ \hline {\bf CC+JLA+}  f\sigma_8 \\ w_0 \\ \\ \\ \\ \\ \end{array}$		$0.795 \pm 0.115$ $0.790 \pm 0.109$ $0.763 \pm 0.111$ $0.763 \pm 0.112$	$\begin{array}{c} 47.822 \\ 47.830 \\ 47.918 \\ 14.846 \\ 14.096 \\ 15.478 \\ 16.808 \\ 16.688 \\ 17.252 \\ \hline -2 \ln \mathcal{L}_{max} \\ 60.244 \\ 60.33 \\ 59.840 \\ 59.832 \\ \end{array}$
FFNN FFNN+MC-DO	CPL CPL ACDM CPL ACDM CPL CPL ACDM CPL CPL CPL ACDM ACDM ACDM ACDM CPL CPL ACDM	- + Planck - - + Planck - + Planck - + Planck - + Planck	$\begin{array}{c} 0.690 \pm 0.030 \\ 0.687 \pm 0.030 \\ 0.687 \pm 0.018 \\ 0.687 \pm 0.018 \\ 0.705 \pm 0.037 \\ 0.695 \pm 0.037 \\ 0.708 \pm 0.021 \\ 0.703 \pm 0.038 \\ 0.698 \pm 0.039 \\ 0.717 \pm 0.019 \\ \hline \\ Datasets: \\ h \\ 0.695 \pm 0.032 \\ 0.690 \pm 0.013 \\ 0.692 \pm 0.029 \\ 0.685 \pm 0.015 \\ 0.721 \pm 0.034 \\ \hline \end{array}$	$\begin{array}{c} w_0 \\ \\ -0.980 \pm 0.173 \\ -0.946 \pm 0.137 \\ \\ -1.010 \pm 0.165 \\ -1.031 \pm 0.150 \\ \\ -0.968 \pm 0.155 \\ -0.957 \pm 0.141 \\ \hline \textbf{CC+JLA+} \ f\sigma_8 \\ w_0 \\ \\ -0.933 \pm 0.086 \\ -0.961 \pm 0.057 \\ \\ \end{array}$		$0.795 \pm 0.115$ $0.790 \pm 0.109$ $0.763 \pm 0.111$ $0.763 \pm 0.112$ $0.786 \pm 0.117$	$\begin{array}{c} 47.822 \\ 47.830 \\ 47.918 \\ 14.846 \\ 14.096 \\ 15.478 \\ 16.808 \\ 16.688 \\ 17.252 \\ \hline -2 \ln \mathcal{L}_{max} \\ 60.244 \\ 60.33 \\ 59.840 \\ 59.832 \\ 16.278 \\ \end{array}$
FFNN+MC-DO  Original	CPL CPL ACDM CPL CPL ACDM CPL CPL ACDM CPL CPL ACDM ACDM ACDM ACDM ACDM ACDM ACDM	- + Planck - - + Planck - + Planck	$\begin{array}{c} 0.690 \pm 0.030 \\ 0.687 \pm 0.030 \\ 0.687 \pm 0.018 \\ 0.687 \pm 0.018 \\ 0.705 \pm 0.037 \\ 0.695 \pm 0.037 \\ 0.708 \pm 0.021 \\ 0.703 \pm 0.038 \\ 0.698 \pm 0.039 \\ 0.717 \pm 0.019 \\ \hline \\ Datasets: \\ h \\ 0.695 \pm 0.032 \\ 0.690 \pm 0.013 \\ 0.692 \pm 0.029 \\ 0.685 \pm 0.015 \\ 0.721 \pm 0.034 \\ 0.704 \pm 0.013 \\ \hline \end{array}$	$\begin{array}{c} w_0 \\ -0.980 \pm 0.173 \\ -0.946 \pm 0.137 \\ -0.946 \pm 0.137 \\\\ -1.010 \pm 0.165 \\ -1.031 \pm 0.150 \\\\ -0.968 \pm 0.155 \\ -0.957 \pm 0.141 \\ \hline {\bf CC+JLA} + f\sigma_8 \\ w_0 \\\\ -0.933 \pm 0.086 \\ -0.961 \pm 0.057 \\\\\\\\\\\\\\\\\\\\ $	$ \begin{array}{c}\\ -0.156 \pm 0.939\\ -0.232 \pm 0.592\\\\ 0.315 \pm 0.715\\ -0.167 \pm 0.688\\\\ -0.046 \pm 0.859\\ -0.563 \pm 0.669\\ \hline\\ w_a\\\\ -0.009 \pm 0.476\\ -0.122 \pm 0.194\\\\\\\\\\\\\\\\\\\\ -$	$\begin{array}{c} 0.795 \pm 0.115 \\ 0.790 \pm 0.109 \\ 0.763 \pm 0.111 \\ 0.763 \pm 0.112 \\ 0.786 \pm 0.117 \\ 0.7500 \pm 0.105 \end{array}$	$\begin{array}{c} 47.822 \\ 47.830 \\ 47.918 \\ 14.846 \\ 14.096 \\ 15.478 \\ 16.808 \\ 16.688 \\ 17.252 \\ \hline -2 \ln \mathcal{L}_{max} \\ \hline 60.244 \\ 60.33 \\ 59.840 \\ 59.832 \\ 16.278 \\ 19.191 \\ \end{array}$
FFNN+MC-DO  Original	CPL CPL ACDM CPL CPL ACDM CPL CPL ACDM CPL CPL ACDM ACDM ACDM ACDM CPL ACDM ACDM ACDM ACDM ACDM ACDM ACDM ACDM	- + Planck	$\begin{array}{c} 0.690 \pm 0.030 \\ 0.687 \pm 0.030 \\ 0.687 \pm 0.018 \\ 0.705 \pm 0.037 \\ 0.695 \pm 0.037 \\ 0.708 \pm 0.021 \\ 0.703 \pm 0.038 \\ 0.698 \pm 0.039 \\ 0.717 \pm 0.019 \\ \hline \\ Datasets: \\ h \\ 0.695 \pm 0.032 \\ 0.690 \pm 0.013 \\ 0.692 \pm 0.029 \\ 0.685 \pm 0.015 \\ 0.721 \pm 0.034 \\ 0.704 \pm 0.013 \\ 0.712 \pm 0.032 \\ 0.712 \pm 0.032 \\ 0.712 \pm 0.034 \\ 0.712 \pm 0.032 \\ 0.018 \\ 0.0$	$\begin{array}{c} w_0 \\ \\ -0.980 \pm 0.173 \\ -0.946 \pm 0.137 \\ \\ -1.010 \pm 0.165 \\ -1.031 \pm 0.150 \\ \\ -0.968 \pm 0.155 \\ -0.957 \pm 0.141 \\ \hline {\bf CC+JLA} + f\sigma_8 \\ w_0 \\ \\ \\ -0.933 \pm 0.086 \\ -0.961 \pm 0.057 \\ \\ \\ -0.916 \pm 0.065 \\ \end{array}$		$\begin{array}{c} 0.795 \pm 0.115 \\ 0.790 \pm 0.109 \\ 0.763 \pm 0.111 \\ 0.763 \pm 0.112 \\ 0.786 \pm 0.117 \\ 0.7500 \pm 0.105 \\ 0.786 \pm 0.114 \end{array}$	$\begin{array}{c} 47.822 \\ 47.830 \\ 47.918 \\ 14.846 \\ 14.096 \\ 15.478 \\ 16.808 \\ 16.688 \\ 17.252 \\ \hline -2 \ln \mathcal{L}_{max} \\ 60.244 \\ 60.33 \\ 59.840 \\ 59.832 \\ 16.278 \\ 19.191 \\ 15.076 \end{array}$
FFNN+MC-DO  Original  FFNN	CPL CPL ACDM CPL CPL ACDM CPL CPL ACDM ACDM ACDM ACDM CPL CPL ACDM ACDM CPL CPL ACDM ACDM ACDM CPL CPL	- + Planck - - + Planck - + Planck - + Planck - + Planck	$\begin{array}{c} 0.690 \pm 0.030 \\ 0.687 \pm 0.030 \\ 0.687 \pm 0.018 \\ 0.705 \pm 0.037 \\ 0.695 \pm 0.037 \\ 0.708 \pm 0.021 \\ 0.703 \pm 0.038 \\ 0.698 \pm 0.039 \\ 0.717 \pm 0.019 \\ \hline \\ Datasets: \\ h \\ 0.695 \pm 0.032 \\ 0.690 \pm 0.013 \\ 0.692 \pm 0.029 \\ 0.685 \pm 0.015 \\ 0.721 \pm 0.034 \\ 0.704 \pm 0.013 \\ 0.712 \pm 0.032 \\ 0.712 \pm 0.032 \\ 0.712 \pm 0.032 \\ 0.712 \pm 0.015 \\ \end{array}$	$\begin{array}{c} w_0 \\ \\ -0.980 \pm 0.173 \\ -0.946 \pm 0.137 \\ \\ -1.010 \pm 0.165 \\ -1.031 \pm 0.150 \\ \\ -0.968 \pm 0.155 \\ -0.957 \pm 0.141 \\ \hline {\bf CC+JLA+}  f\sigma_8 \\ w_0 \\ \\ \\ -0.933 \pm 0.086 \\ -0.961 \pm 0.057 \\ \\ -0.916 \pm 0.065 \\ -0.941 \pm 0.057 \\ \end{array}$		$\begin{array}{c} 0.795 \pm 0.115 \\ 0.790 \pm 0.109 \\ 0.763 \pm 0.111 \\ 0.763 \pm 0.111 \\ 0.763 \pm 0.112 \\ 0.786 \pm 0.117 \\ 0.7500 \pm 0.105 \\ 0.786 \pm 0.114 \\ 0.733 \pm 0.100 \\ \end{array}$	$\begin{array}{c} 47.822 \\ 47.830 \\ 47.918 \\ 14.846 \\ 14.096 \\ 15.478 \\ 16.808 \\ 16.688 \\ 17.252 \\ \hline 00000000000000000000000000000000000$
FFNN+MC-DO  Original	CPL CPL ACDM CPL CPL ACDM CPL CPL ACDM ACDM ACDM CPL CPL ACDM ACDM ACDM ACDM ACDM ACDM ACDM ACDM	- + Planck	$\begin{array}{c} 0.690 \pm 0.030 \\ 0.687 \pm 0.030 \\ 0.687 \pm 0.018 \\ 0.705 \pm 0.037 \\ 0.695 \pm 0.037 \\ 0.708 \pm 0.021 \\ 0.703 \pm 0.038 \\ 0.698 \pm 0.039 \\ 0.717 \pm 0.019 \\ \hline \\ Datasets: \\ h \\ 0.695 \pm 0.032 \\ 0.690 \pm 0.013 \\ 0.692 \pm 0.029 \\ 0.685 \pm 0.015 \\ 0.721 \pm 0.034 \\ 0.704 \pm 0.013 \\ 0.712 \pm 0.032 \\ 0.712 \pm 0.034 \\ 0.712 \pm 0.032 \\ 0.712 \pm 0.035 \\ 0.712 \pm 0.035 \\ 0.713 \pm 0.035 \\ \end{array}$	$\begin{array}{c} w_0 \\ \\ -0.980 \pm 0.173 \\ -0.946 \pm 0.137 \\ \\ -1.010 \pm 0.165 \\ -1.031 \pm 0.150 \\ \\ -0.968 \pm 0.155 \\ -0.957 \pm 0.141 \\ \hline {\bf CC+JLA} + f\sigma_8 \\ w_0 \\ \\ \\ -0.933 \pm 0.086 \\ -0.961 \pm 0.057 \\ \\ \\ -0.916 \pm 0.065 \\ \end{array}$		$\begin{array}{c} 0.795 \pm 0.115 \\ 0.790 \pm 0.109 \\ 0.763 \pm 0.111 \\ 0.763 \pm 0.112 \\ 0.786 \pm 0.117 \\ 0.7500 \pm 0.105 \\ 0.786 \pm 0.114 \\ 0.733 \pm 0.100 \\ 0.775 \pm 0.116 \end{array}$	$\begin{array}{c} 47.822 \\ 47.830 \\ 47.918 \\ 14.846 \\ 14.096 \\ 15.478 \\ 16.808 \\ 16.688 \\ 17.252 \\ \hline -2 \ln \mathcal{L}_{max} \\ 60.244 \\ 60.33 \\ 59.840 \\ 59.832 \\ 16.278 \\ 19.191 \\ 15.076 \\ 17.044 \\ 18.096 \\ \end{array}$
FFNN+MC-DO  Original  FFNN	CPL CPL ACDM CPL CPL ACDM CPL CPL ACDM ACDM ACDM ACDM CPL CPL ACDM ACDM CPL CPL ACDM ACDM ACDM CPL CPL	- + Planck	$\begin{array}{c} 0.690 \pm 0.030 \\ 0.687 \pm 0.030 \\ 0.687 \pm 0.018 \\ 0.705 \pm 0.037 \\ 0.695 \pm 0.037 \\ 0.708 \pm 0.021 \\ 0.703 \pm 0.038 \\ 0.698 \pm 0.039 \\ 0.717 \pm 0.019 \\ \hline \\ Datasets: \\ h \\ 0.695 \pm 0.032 \\ 0.690 \pm 0.013 \\ 0.692 \pm 0.029 \\ 0.685 \pm 0.015 \\ 0.721 \pm 0.034 \\ 0.704 \pm 0.013 \\ 0.712 \pm 0.032 \\ 0.712 \pm 0.032 \\ 0.712 \pm 0.032 \\ 0.712 \pm 0.015 \\ \end{array}$	$\begin{array}{c} w_0 \\ \\ -0.980 \pm 0.173 \\ -0.946 \pm 0.137 \\ \\ -1.010 \pm 0.165 \\ -1.031 \pm 0.150 \\ \\ -0.968 \pm 0.155 \\ -0.957 \pm 0.141 \\ \hline {\bf CC+JLA+}  f\sigma_8 \\ w_0 \\ \\ -0.933 \pm 0.086 \\ -0.961 \pm 0.057 \\ \\ -0.916 \pm 0.065 \\ -0.941 \pm 0.057 \\ \\ \\ \\ \\ \\ \\ \\ $		$\begin{array}{c} 0.795 \pm 0.115 \\ 0.790 \pm 0.109 \\ 0.763 \pm 0.111 \\ 0.763 \pm 0.111 \\ 0.763 \pm 0.112 \\ 0.786 \pm 0.117 \\ 0.7500 \pm 0.105 \\ 0.786 \pm 0.114 \\ 0.733 \pm 0.100 \\ \end{array}$	$\begin{array}{c} 47.822 \\ 47.830 \\ 47.918 \\ 14.846 \\ 14.096 \\ 15.478 \\ 16.808 \\ 16.688 \\ 17.252 \\ \hline -2 \ln \mathcal{L}_{max} \\ \hline 60.244 \\ 60.33 \\ 59.840 \\ 59.832 \\ 16.278 \\ 19.191 \\ 15.076 \\ 17.044 \\ \end{array}$
FFNN+MC-DO  Original  FFNN	CPL CPL ACDM CPL ACDM CPL ACDM CPL CPL ACDM ACDM ACDM CPL CPL ACDM ACDM ACDM ACDM ACDM ACDM ACDM ACDM	- + Planck	$\begin{array}{c} 0.690 \pm 0.030 \\ 0.687 \pm 0.030 \\ 0.687 \pm 0.030 \\ 0.687 \pm 0.018 \\ 0.705 \pm 0.037 \\ 0.695 \pm 0.037 \\ 0.708 \pm 0.021 \\ 0.703 \pm 0.038 \\ 0.698 \pm 0.039 \\ 0.717 \pm 0.019 \\ \hline \\ Datasets: \\ h \\ 0.695 \pm 0.032 \\ 0.690 \pm 0.013 \\ 0.692 \pm 0.029 \\ 0.685 \pm 0.015 \\ 0.721 \pm 0.034 \\ 0.704 \pm 0.013 \\ 0.712 \pm 0.032 \\ 0.712 \pm 0.032 \\ 0.712 \pm 0.032 \\ 0.712 \pm 0.035 \\ 0.712 \pm 0.035 \\ 0.712 \pm 0.015 \\ 0.713 \pm 0.035 \\ 0.702 \pm 0.012 \\ \hline \end{array}$	$\begin{array}{c} w_0 \\ -0.980 \pm 0.173 \\ -0.946 \pm 0.137 \\ -0.946 \pm 0.137 \\ -1.010 \pm 0.165 \\ -1.031 \pm 0.150 \\ -0.968 \pm 0.155 \\ -0.957 \pm 0.141 \\ \hline {\bf CC+JLA+}  f\sigma_8 \\ w_0 \\ -0.961 \pm 0.057 \\ -0.916 \pm 0.065 \\ -0.941 \pm 0.057 \\ -0.9025 \pm 0.068 \\ -0.970 \pm 0.055 \\ \hline \end{array}$		$\begin{array}{c} 0.795 \pm 0.115 \\ 0.790 \pm 0.109 \\ 0.763 \pm 0.111 \\ 0.763 \pm 0.112 \\ 0.786 \pm 0.117 \\ 0.7500 \pm 0.105 \\ 0.786 \pm 0.114 \\ 0.733 \pm 0.100 \\ 0.775 \pm 0.116 \\ 0.753 \pm 0.100 \\ \end{array}$	$\begin{array}{c} 47.822 \\ 47.830 \\ 47.918 \\ 14.846 \\ 14.096 \\ 15.478 \\ 16.808 \\ 16.688 \\ 17.252 \\ \hline 00000000000000000000000000000000000$
FFNN+MC-DO  Original  FFNN  FFNN+MC-DO	CPL CPL ACDM CPL CPL ACDM CPL CPL ACDM ACDM ACDM ACDM CPL CPL	- + Planck	$\begin{array}{c} 0.690 \pm 0.030 \\ 0.687 \pm 0.030 \\ 0.687 \pm 0.018 \\ 0.705 \pm 0.037 \\ 0.695 \pm 0.037 \\ 0.708 \pm 0.021 \\ 0.703 \pm 0.038 \\ 0.698 \pm 0.039 \\ 0.717 \pm 0.019 \\ \hline \\ Datasets: \\ h \\ 0.695 \pm 0.032 \\ 0.690 \pm 0.013 \\ 0.692 \pm 0.029 \\ 0.685 \pm 0.015 \\ 0.721 \pm 0.034 \\ 0.704 \pm 0.013 \\ 0.712 \pm 0.032 \\ 0.712 \pm 0.032 \\ 0.712 \pm 0.034 \\ 0.702 \pm 0.015 \\ 0.713 \pm 0.035 \\ 0.702 \pm 0.015 \\ 0.713 \pm 0.035 \\ 0.702 \pm 0.012 \\ 0.706 \pm 0.037 \\ 0.711 \pm 0.017 \\ \hline \end{array}$	$\begin{array}{c} w_0 \\ \\ -0.980 \pm 0.173 \\ -0.946 \pm 0.137 \\ \\ -1.010 \pm 0.165 \\ -1.031 \pm 0.150 \\ \\ -0.968 \pm 0.155 \\ -0.957 \pm 0.141 \\ \hline {\bf CC+JLA+}  f\sigma_8 \\ w_0 \\ \\ \\ -0.933 \pm 0.086 \\ -0.961 \pm 0.057 \\ \\ \\ -0.916 \pm 0.065 \\ -0.941 \pm 0.057 \\ \\ \\ -0.925 \pm 0.068 \\ -0.970 \pm 0.055 \\ \hline \Omega_1 \\ \end{array}$	$ \begin{array}{c}\\ -0.156 \pm 0.939\\ -0.232 \pm 0.592\\\\ 0.315 \pm 0.715\\ -0.167 \pm 0.688\\\\ -0.046 \pm 0.859\\ -0.563 \pm 0.669\\ \\ \hline\\ w_a\\\\ 0.009 \pm 0.476\\ -0.122 \pm 0.194\\\\ 0.150 \pm 0.432\\ -0.417 \pm 0.246\\\\ 0.222 \pm 0.443\\ -0.318 \pm 0.247\\ \hline\\ \Omega_2\\ \end{array} $	$\begin{array}{c} 0.795 \pm 0.115 \\ 0.790 \pm 0.109 \\ 0.763 \pm 0.111 \\ 0.763 \pm 0.111 \\ 0.763 \pm 0.112 \\ 0.786 \pm 0.117 \\ 0.7500 \pm 0.105 \\ 0.786 \pm 0.114 \\ 0.733 \pm 0.100 \\ 0.775 \pm 0.116 \\ 0.753 \pm 0.100 \\ 0.763 \pm 0.105 \\ 0.723 \pm 0.097 \\ \end{array}$	$\begin{array}{c} 47.822 \\ 47.830 \\ 47.918 \\ 14.846 \\ 14.096 \\ 15.478 \\ 16.808 \\ 16.688 \\ 17.252 \\ \hline 00000000000000000000000000000000000$
FFNN+MC-DO  Original  FFNN	CPL CPL ACDM CPL ACDM CPL CPL ACDM CPL CPL ACDM ACDM ACDM CPL CPL ACDM ACDM ACDM ACDM CPL CPL ACDM ACDM CPL CPL	- + Planck	$\begin{array}{c} 0.690 \pm 0.030 \\ 0.687 \pm 0.030 \\ 0.687 \pm 0.030 \\ 0.687 \pm 0.038 \\ 0.695 \pm 0.037 \\ 0.708 \pm 0.037 \\ 0.708 \pm 0.032 \\ 0.703 \pm 0.038 \\ 0.698 \pm 0.039 \\ 0.717 \pm 0.019 \\ \hline \\ Datasets: \\ h \\ 0.695 \pm 0.032 \\ 0.690 \pm 0.013 \\ 0.692 \pm 0.029 \\ 0.685 \pm 0.015 \\ 0.721 \pm 0.034 \\ 0.704 \pm 0.013 \\ 0.712 \pm 0.032 \\ 0.712 \pm 0.032 \\ 0.712 \pm 0.035 \\ 0.702 \pm 0.015 \\ 0.713 \pm 0.035 \\ 0.702 \pm 0.012 \\ 0.706 \pm 0.037 \\ \hline \end{array}$	$\begin{array}{c} w_0 \\ -0.980 \pm 0.173 \\ -0.946 \pm 0.137 \\ -0.946 \pm 0.137 \\ -1.010 \pm 0.165 \\ -1.031 \pm 0.150 \\ -0.968 \pm 0.155 \\ -0.957 \pm 0.141 \\ \hline {\bf CC+JLA+}  f\sigma_8 \\ w_0 \\ -0.961 \pm 0.057 \\ -0.916 \pm 0.065 \\ -0.941 \pm 0.057 \\ -0.9025 \pm 0.068 \\ -0.970 \pm 0.055 \\ \hline \end{array}$		$\begin{array}{c} 0.795 \pm 0.115 \\ 0.790 \pm 0.109 \\ 0.763 \pm 0.111 \\ 0.763 \pm 0.112 \\ 0.786 \pm 0.117 \\ 0.7500 \pm 0.105 \\ 0.786 \pm 0.114 \\ 0.733 \pm 0.100 \\ 0.775 \pm 0.116 \\ 0.753 \pm 0.100 \\ 0.763 \pm 0.105 \\ \end{array}$	$\begin{array}{c} 47.822 \\ 47.830 \\ 47.918 \\ 14.846 \\ 14.096 \\ 15.478 \\ 16.808 \\ 16.688 \\ 17.252 \\ \hline -2 \ln \mathcal{L}_{max} \\ \hline 60.244 \\ 60.33 \\ 59.840 \\ 59.832 \\ 16.278 \\ 19.191 \\ 15.076 \\ 17.044 \\ 18.096 \\ 20.842 \\ 17.734 \\ 19.034 \\ \end{array}$
FFNN+MC-DO  Original  FFNN+MC-DO  Original	CPL CPL ACDM CPL CPL ACDM CPL CPL ACDM ACDM ACDM CPL CPL ACDM ACDM ACDM ACDM ACDM ACDM CPL CPL ACDM ACDM ACDM ACDM ACDM ACDM ACDM ACDM	- + Planck	$\begin{array}{c} 0.690 \pm 0.030 \\ 0.687 \pm 0.030 \\ 0.687 \pm 0.018 \\ 0.705 \pm 0.037 \\ 0.695 \pm 0.037 \\ 0.708 \pm 0.021 \\ 0.703 \pm 0.038 \\ 0.698 \pm 0.039 \\ 0.717 \pm 0.019 \\ \hline \\ Datasets: \\ h \\ 0.695 \pm 0.032 \\ 0.690 \pm 0.013 \\ 0.690 \pm 0.013 \\ 0.692 \pm 0.029 \\ 0.685 \pm 0.015 \\ 0.721 \pm 0.034 \\ 0.704 \pm 0.013 \\ 0.712 \pm 0.032 \\ 0.712 \pm 0.032 \\ 0.712 \pm 0.035 \\ 0.712 \pm 0.015 \\ 0.713 \pm 0.035 \\ 0.702 \pm 0.012 \\ 0.706 \pm 0.017 \\ \hline \\ 0.693 \pm 0.029 \\ 0.696 \pm 0.017 \\ \hline \end{array}$	$\begin{array}{c} w_0 \\ \\ -0.980 \pm 0.173 \\ -0.946 \pm 0.137 \\ \\ -1.010 \pm 0.165 \\ -1.031 \pm 0.150 \\ \\ -0.968 \pm 0.155 \\ -0.957 \pm 0.141 \\ \hline {\bf CC+JLA+}  f\sigma_8 \\ w_0 \\ \\ \\ -0.933 \pm 0.086 \\ -0.961 \pm 0.057 \\ \\ \\ -0.916 \pm 0.065 \\ -0.941 \pm 0.057 \\ \\ \\ -0.925 \pm 0.068 \\ -0.970 \pm 0.055 \\ \hline \Omega_1 \\ 0.089 \pm 0.416 \\ 0.147 \pm 0.238 \\ \end{array}$	$ \begin{array}{c}\\ -0.156 \pm 0.939\\ -0.232 \pm 0.592\\\\ 0.315 \pm 0.715\\ -0.167 \pm 0.688\\\\ -0.046 \pm 0.859\\ -0.563 \pm 0.669\\ \\ \hline\\ w_a\\\\\\ 0.009 \pm 0.476\\ -0.122 \pm 0.194\\\\\\ 0.150 \pm 0.432\\ -0.417 \pm 0.246\\\\\\ 0.222 \pm 0.443\\ -0.318 \pm 0.247\\ \hline\\ \Omega_2\\ -0.034 \pm 0.302\\ -0.020 \pm 0.071\\ \end{array} $	$\begin{array}{c} 0.795 \pm 0.115 \\ 0.790 \pm 0.109 \\ 0.763 \pm 0.111 \\ 0.763 \pm 0.112 \\ 0.786 \pm 0.117 \\ 0.7500 \pm 0.105 \\ 0.786 \pm 0.114 \\ 0.733 \pm 0.100 \\ 0.775 \pm 0.116 \\ 0.753 \pm 0.100 \\ 0.763 \pm 0.105 \\ 0.723 \pm 0.097 \\ \hline \\ 0.788 \pm 0.108 \\ 0.779 \pm 0.107 \\ \hline \end{array}$	$\begin{array}{c} 47.822 \\ 47.830 \\ 47.918 \\ 14.846 \\ 14.096 \\ 15.478 \\ 16.808 \\ 16.688 \\ 17.252 \\ \hline 00000000000000000000000000000000000$
FFNN+MC-DO  Original  FFNN  FFNN+MC-DO	CPL CPL ACDM CPL CPL ACDM CPL CPL ACDM ACDM ACDM CPL CPL ACDM ACDM ACDM ACDM CPL CPL	- + Planck	$\begin{array}{c} 0.690 \pm 0.030 \\ 0.687 \pm 0.030 \\ 0.687 \pm 0.018 \\ 0.705 \pm 0.037 \\ 0.695 \pm 0.037 \\ 0.708 \pm 0.021 \\ 0.703 \pm 0.038 \\ 0.698 \pm 0.039 \\ 0.717 \pm 0.019 \\ \hline \\ Datasets: \\ h \\ 0.695 \pm 0.032 \\ 0.690 \pm 0.013 \\ 0.690 \pm 0.013 \\ 0.692 \pm 0.029 \\ 0.685 \pm 0.015 \\ 0.721 \pm 0.034 \\ 0.704 \pm 0.013 \\ 0.712 \pm 0.034 \\ 0.712 \pm 0.035 \\ 0.712 \pm 0.035 \\ 0.712 \pm 0.015 \\ 0.713 \pm 0.035 \\ 0.702 \pm 0.012 \\ 0.706 \pm 0.037 \\ 0.711 \pm 0.017 \\ \hline \end{array}$	$\begin{array}{c} w_0 \\ \\ -0.980 \pm 0.173 \\ -0.946 \pm 0.137 \\ \\ -1.010 \pm 0.165 \\ -1.031 \pm 0.150 \\ \\ -0.968 \pm 0.155 \\ -0.957 \pm 0.141 \\ \hline {\bf CC+JLA+}  f\sigma_8 \\ w_0 \\ \\ \\ -0.933 \pm 0.086 \\ -0.961 \pm 0.057 \\ \\ \\ -0.916 \pm 0.065 \\ -0.941 \pm 0.057 \\ \\ \\ -0.925 \pm 0.068 \\ -0.970 \pm 0.055 \\ \hline \Omega_1 \\ \\ 0.089 \pm 0.416 \\ \hline \end{array}$		$\begin{array}{c} 0.795 \pm 0.115 \\ 0.790 \pm 0.109 \\ 0.763 \pm 0.111 \\ 0.763 \pm 0.112 \\ 0.786 \pm 0.117 \\ 0.7500 \pm 0.105 \\ 0.786 \pm 0.114 \\ 0.733 \pm 0.100 \\ 0.775 \pm 0.116 \\ 0.753 \pm 0.100 \\ 0.763 \pm 0.105 \\ 0.723 \pm 0.097 \\ \hline \end{array}$	$\begin{array}{c} 47.822 \\ 47.830 \\ 47.918 \\ 14.846 \\ 14.096 \\ 15.478 \\ 16.808 \\ 16.688 \\ 17.252 \\ \hline 00000000000000000000000000000000000$
FFNN+MC-DO  Original  FFNN+MC-DO  Original	CPL CPL ACDM CPL ACDM CPL ACDM CPL CPL ACDM ACDM ACDM ACDM CPL CPL	- + Planck	$\begin{array}{c} 0.690 \pm 0.030 \\ 0.687 \pm 0.030 \\ 0.687 \pm 0.030 \\ 0.687 \pm 0.018 \\ 0.705 \pm 0.037 \\ 0.695 \pm 0.037 \\ 0.708 \pm 0.021 \\ 0.703 \pm 0.038 \\ 0.698 \pm 0.039 \\ 0.717 \pm 0.019 \\ \hline \\ Datasets: \\ h \\ 0.695 \pm 0.032 \\ 0.690 \pm 0.013 \\ 0.692 \pm 0.029 \\ 0.685 \pm 0.015 \\ 0.721 \pm 0.034 \\ 0.704 \pm 0.013 \\ 0.712 \pm 0.034 \\ 0.712 \pm 0.035 \\ 0.712 \pm 0.035 \\ 0.712 \pm 0.015 \\ 0.713 \pm 0.035 \\ 0.702 \pm 0.012 \\ 0.706 \pm 0.017 \\ 0.693 \pm 0.029 \\ 0.696 \pm 0.017 \\ 0.696 \pm 0.017 \\ 0.712 \pm 0.035 \\ \hline \end{array}$	$\begin{array}{c} w_0 \\ \\ -0.980 \pm 0.173 \\ -0.946 \pm 0.137 \\ \\ -1.010 \pm 0.165 \\ -1.031 \pm 0.150 \\ \\ -0.968 \pm 0.155 \\ -0.957 \pm 0.141 \\ \hline {\bf CC+JLA+}  f\sigma_8 \\ w_0 \\ \\ \\ -0.933 \pm 0.086 \\ -0.961 \pm 0.057 \\ \\ \\ -0.916 \pm 0.057 \\ \\ \\ -0.916 \pm 0.057 \\ \\ \\ -0.925 \pm 0.068 \\ -0.941 \pm 0.057 \\ \\ \\ -0.970 \pm 0.055 \\ \Omega_1 \\ 0.089 \pm 0.416 \\ 0.147 \pm 0.238 \\ 0.110 \pm 0.444 \\ \end{array}$	$ \begin{array}{c}\\ -0.156 \pm 0.939\\ -0.232 \pm 0.592\\\\ 0.315 \pm 0.715\\ -0.167 \pm 0.688\\\\ -0.046 \pm 0.859\\ -0.563 \pm 0.669\\ \hline\\ \hline\\ w_a\\\\\\ 0.009 \pm 0.476\\ -0.122 \pm 0.194\\\\\\ 0.150 \pm 0.432\\ -0.417 \pm 0.246\\\\\\ 0.222 \pm 0.443\\ -0.318 \pm 0.247\\ \Omega_2\\ 0.034 \pm 0.302\\ -0.020 \pm 0.071\\ 0.048 \pm 0.302\\ \end{array} $	$\begin{array}{c} 0.795 \pm 0.115 \\ 0.790 \pm 0.109 \\ 0.763 \pm 0.111 \\ 0.763 \pm 0.112 \\ 0.786 \pm 0.112 \\ 0.786 \pm 0.114 \\ 0.7500 \pm 0.105 \\ 0.786 \pm 0.114 \\ 0.733 \pm 0.100 \\ 0.775 \pm 0.116 \\ 0.753 \pm 0.100 \\ 0.763 \pm 0.105 \\ 0.723 \pm 0.097 \\ \hline \\ 0.788 \pm 0.108 \\ 0.779 \pm 0.107 \\ 0.776 \pm 0.109 \\ \hline \end{array}$	$\begin{array}{c} 47.822 \\ 47.830 \\ 47.918 \\ 14.846 \\ 14.096 \\ 15.478 \\ 16.808 \\ 16.688 \\ 17.252 \\ \hline 00000000000000000000000000000000000$

Table 1: Parameter estimation using Bayesian inference with datasets from different sources: original, FFNN alone and FFNN using Monte Carlo dropout.

Variational autoencoders use variational inference to sample the compressed representation (or latent space) and, therefore, allow to know the probability density function associated, precisely, to the compressed representation. Unlike

classical autoencoders, such as those described earlier in this work, two layers of the same dimension as the latent space are designed before the compressed representation, whose function is to generate values to sample the mean  $\mu$  and variance  $\sigma$  which are the parameters of the statistical distribution that produces an input data (matrix or image) of the VAE to generate a point z of the latent space. A diagram of classical and variational autoencoders is shown in the Figure 9.

As a way to construct a latent space distribution similar to the proposed Gaussian distribution, the Kullback-Leiber divergence (KL) is used. Thus, the selection of the relevant loss function to train the VAE is as follows:

$$loss_{VAE} = MSE + \mathbf{KL}(q(z|x)||p(z)), \tag{14}$$

where q(z|x) is the probability density function to generate a z point of the latent space given an input x. On the other hand, we can assume that p(z) = N(0, I) with p a probability density function of the z points in latent space and N a normal distribution centred at 0 with covariance matrix equal to the identity matrix. Because VAEs are widely used in image processing, it is more common to choose *binary cross entropy* [101] instead of MSE, however our interest is in the numerical information of the covariance matrices and not just in a classification problem that takes place in image generation. Then to have a dataset to train our VAE, we generated thousands of matrices by adding Gaussian noise of

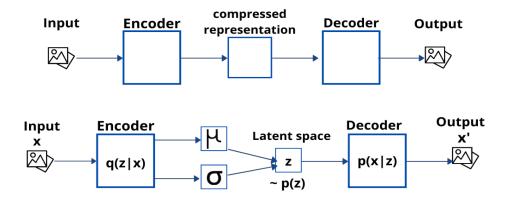


Figure 9: Top: A classical autoencoder has an encoder, a decoder and a discrete compressed representation. *Bottom*: A variational autoencoder has a continuous latent space that is sampled with the mean and variance layers. The notation is the same as in Equation 14 with z the latent space variable, x is the input variable and the encoder and decoder are associated with conditional probability density functions.

the same order of magnitude for each entry of the original covariance matrix. With this data set and the architecture shown in Figure 10 we train the autoencoder over 1000 epochs.

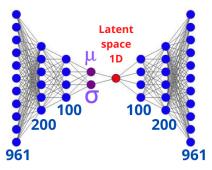


Figure 10: VAE architecture designed to generate synthetic covariance matrices from a point in the latent space. This VAE was trained on a dataset created from the systematic error covariance matrix of the JLA binned version.  $\mu$  and  $\sigma$  represent two layers connected to the last layer of the encoder and to the latent space; in this case both layers have a single neuron (the same dimension as the latent space). We use a batch size of 32 and the hyperbolic tangent as the activation function.

Since we are interested on mapping the distribution of the distance modulus to the latent space, we design the VAE with a 1-dimensional latent space, so its mean  $\mu$  and variance  $\sigma$  are also 1-dimensional. Thus, when training the VAE, we

can use the decoder part to generate new covariance matrices that traverse the latent space. Once the VAE is trained, we can explore the mean, variance and latent space layers, as can be seen in the left and middle panel of Figure 11. To generate covariance matrices from the predictions of the modular distances coming from the FFNNs, using their means and standard deviations we have assigned them a Gaussian distribution (right panel of Figure 11). We have related the original measurements to the most likely region of the latent space, and the deviations from the original measurements can be linearly mapped to the latent space to generate a new covariance matrix as shown in Figure 6.

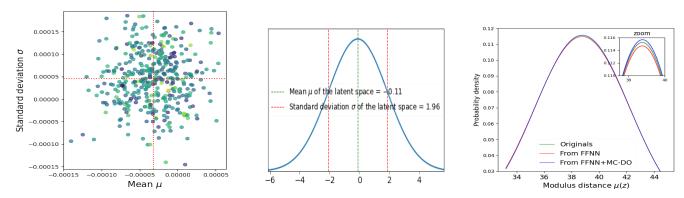


Figure 11: After the VAE is trained, each element of the training set generates a value in the mean layer and in the variance layer. *Left*: Samples of the mean and variance layers. *Middle*: Sampled distribution of the latent space. *Right*: Comparison between the distributions assigned for the modular distances from different sources, these distributions are mapped into the latent space to generate a new covariance matrix with the VAE decoder.

# References

- [1] Edmund J Copeland, Mohammad Sami, and Shinji Tsujikawa. Dynamics of dark energy. *International Journal of Modern Physics D*, 15(11):1753–1935, 2006. [arXiv: hep-th/0603057].
- [2] Luca Amendola and Shinji Tsujikawa. *Dark energy: theory and observations*. Cambridge University Press, 2010.
- [3] Pilar Ruiz-Lapuente. Dark energy: observational and theoretical approaches. Cambridge University Press, 2010.
- [4] Nabila Aghanim, Yashar Akrami, M Ashdown, J Aumont, C Baccigalupi, M Ballardini, AJ Banday, RB Barreiro, N Bartolo, S Basak, et al. Planck 2018 results-vi. cosmological parameters. *Astronomy & Astrophysics*, 641:A6, 2020. [arXiv:1807.06209].
- [5] M et al Betoule, R Kessler, J Guy, J Mosher, D Hardin, R Biswas, P Astier, P El-Hage, M Konig, S Kuhlmann, et al. Improved cosmological constraints from a joint analysis of the sdss-ii and snls supernova samples. *Astronomy & Astrophysics*, 568:A22, 2014. [arXiv:1401.4064].
- [6] Daniel Stern, Raul Jimenez, Licia Verde, Marc Kamionkowski, and S Adam Stanford. Cosmic chronometers: constraining the equation of state of dark energy. i: h(z) measurements. *Journal of Cosmology and Astroparticle Physics*, 2010(02):008, 2010. [arXiv:0907.3149].
- [7] Shadab Alam, Metin Ata, Stephen Bailey, Florian Beutler, Dmitry Bizyaev, Jonathan A Blazek, Adam S Bolton, Joel R Brownstein, Angela Burden, Chia-Hsun Chuang, et al. The clustering of galaxies in the completed sdss-iii baryon oscillation spectroscopic survey: cosmological analysis of the dr12 galaxy sample. *Monthly Notices of the Royal Astronomical Society*, 470(3):2617–2652, 2017. [arXiv:1607.03155].
- [8] Varun Sahni. The cosmological constant problem and quintessence. Classical and Quantum Gravity, 19(13):3435, 2002. [arXiv: astro-ph/0202076].
- [9] P James E Peebles and Bharat Ratra. The cosmological constant and dark energy. *Reviews of modern physics*, 75(2):559, 2003. [arXiv: astro-ph/0207347].
- [10] Stephen M Feeney, Daniel J Mortlock, and Niccolo Dalmasso. Clarifying the hubble constant tension with a bayesian hierarchical model of the local distance ladder. *Monthly Notices of the Royal Astronomical Society*, 476(3):3861–3882, 2018. [arXiv:1707.00007].

- [11] Austin Joyce, Lucas Lombriser, and Fabian Schmidt. Dark energy versus modified gravity. *Annual Review of Nuclear and Particle Science*, 66:95–122, 2016. [arXiv:1601.06133].
- [12] Amir Aghamousa, Jessica Aguilar, Steve Ahlen, Shadab Alam, Lori E Allen, Carlos Allende Prieto, James Annis, Stephen Bailey, Christophe Balland, Otger Ballester, et al. The desi experiment part i: science, targeting, and survey design. 2016. [arXiv:1611.00036].
- [13] J Anthony Tyson. Large synoptic survey telescope: overview. In *Survey and Other Telescope Technologies and Discoveries*, volume 4836, pages 10–20. International Society for Optics and Photonics, 2002. [arXiv: astro-ph/0302102].
- [14] Luca Amendola, Stephen Appleby, Anastasios Avgoustidis, David Bacon, Tessa Baker, Marco Baldi, Nicola Bartolo, Alain Blanchard, Camille Bonvin, Stefano Borgani, et al. Cosmology and fundamental physics with the euclid satellite. *Living reviews in relativity*, 21(1):2, 2018. [arXiv:1606.00180].
- [15] Philippe Brax and Carsten van de Bruck. Cosmology and brane worlds: a review. *Classical and Quantum Gravity*, 20(9):R201, 2003. [arXiv: hep-th/0303095].
- [16] Timothy Clifton, Pedro G Ferreira, Antonio Padilla, and Constantinos Skordis. Modified gravity and cosmology. *Physics reports*, 513(1-3):1–189, 2012. [arXiv:1106.2476].
- [17] J Alberto Vázquez, David Tamayo, Anjan A Sen, and Israel Quiros. Bayesian model selection on scalar *ϵ*-field dark energy. *Phys. Rev. D*, 103(4):043506, 2021. [arXiv:2009.01904].
- [18] Israel Quiros, Tame Gonzalez, Ulises Nucamendi, Ricardo Garcia-Salcedo, Francisco Antonio Horta-Rangel, and Joel Saavedra. On the phantom barrier crossing and the bounds on the speed of sound in non-minimal derivative coupling theories. *Classical and Quantum Gravity*, 35(7):075005, 2018. [arXiv:1707.03885].
- [19] Özgür Akarsu, John D. Barrow, Luis A. Escamilla, and J. Alberto Vazquez. Graduated dark energy: Observational hints of a spontaneous sign switch in the cosmological constant. *Phys. Rev. D*, 101(6):063528, 2020. [arXiv:1912.08751].
- [20] Michel Chevallier and David Polarski. Accelerating universes with scaling dark matter. *International Journal of Modern Physics D*, 10(02):213–223, 2001. [arXiv: gr-qc/0009008].
- [21] Irene Sendra and Ruth Lazkoz. Supernova and baryon acoustic oscillation constraints on (new) polynomial dark energy parametrizations: current results and forecasts. *Monthly Notices of the Royal Astronomical Society*, 422(1):776–793, 2012. [arXiv:1105.4943].
- [22] Sergei D Odintsov, VK Oikonomou, AV Timoshkin, Emmanuel N Saridakis, and R Myrzakulov. Cosmological fluids with logarithmic equation of state. *Annals of Physics*, 398:238–253, 2018. [arXiv:1810.01276].
- [23] David Tamayo and J. Alberto Vazquez. Fourier-series expansion of the dark-energy equation of state. *Mon. Not. Roy. Astron. Soc.*, 487(1):729–736, 2019. [arXiv:1901.08679].
- [24] Jie Liu, Hong Li, Jun-Qing Xia, and Xinmin Zhang. Testing oscillating primordial spectrum and oscillating dark energy with astronomical observations. *Journal of Cosmology and Astroparticle Physics*, 2009(07):017, 2009. [arXiv:0901.2033].
- [25] Özgur Akarsu, Tekin Dereli, and J. Alberto Vazquez. A divergence-free parametrization for dynamical dark energy. *JCAP*, 06:049, 2015. [arXiv:1501.07598].
- [26] Larry Wasserman. All of nonparametric statistics. Springer Science & Business Media, 2006.
- [27] Varun Sahni and Alexei Starobinsky. Reconstructing dark energy. *International Journal of Modern Physics D*, 15(12):2105–2132, 2006. [arXiv: astro-ph/0610026].
- [28] Ranbir Sharma, Ankan Mukherjee, and HK Jassal. Reconstruction of late-time cosmology using principal component analysis. 2020. [arXiv:2004.01393].
- [29] Francesca Gerardi, Matteo Martinelli, and Alessandra Silvestri. Reconstruction of the dark energy equation of state from latest data: the impact of theoretical priors. *Journal of Cosmology and Astroparticle Physics*, 2019(07):042, 2019. [arXiv:1902.09423].
- [30] Ryan E. Keeley, Arman Shafieloo, Gong-Bo Zhao, Jose Alberto Vazquez, and Hanwool Koo. Reconstructing the universe: Testing the mutual consistency of the pantheon and SDSS/eBOSS BAO data sets with gaussian processes. *The Astronomical Journal*, 161(3):151, Feb 2021. [arXiv:2010.03234].
- [31] Christopher KI Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.

- [32] Benjamin L'Huillier, Arman Shafieloo, David Polarski, and Alexei A Starobinsky. Defying the laws of gravity i: model-independent reconstruction of the universe expansion from growth data. *Monthly Notices of the Royal Astronomical Society*, 494(1):819–826, 2020. [arXiv:1906.05991].
- [33] Purba Mukherjee and Narayan Banerjee. Revisiting a non-parametric reconstruction of the deceleration parameter from observational data. 2020. [arXiv:2007.15941].
- [34] Ariadna Montiel, Ruth Lazkoz, Irene Sendra, Celia Escamilla-Rivera, and Vincenzo Salzano. Nonparametric reconstruction of the cosmic expansion with local regression smoothing and simulation extrapolation. *Physical Review D*, 89(4):043007, 2014. [arXiv:1401.4188].
- [35] Sonke Hee, JA Vázquez, WJ Handley, MP Hobson, and AN Lasenby. Constraining the dark energy equation of state using bayes theorem and the kullback–leibler divergence. *Monthly Notices of the Royal Astronomical Society*, 466(1):369–377, 2017. [arXiv:1607.00270].
- [36] J. Alberto Vazquez, M. Bridges, M.P. Hobson, and A.N. Lasenby. Reconstruction of the Dark Energy equation of state. JCAP, 09:020, 2012. [arXiv:1205.0847].
- [37] Gong-Bo Zhao, Marco Raveri, Levon Pogosian, Yuting Wang, Robert G Crittenden, Will J Handley, Will J Percival, Florian Beutler, Jonathan Brinkmann, Chia-Hsun Chuang, et al. Dynamical dark energy in light of the latest observations. *Nature Astronomy*, 1(9):627–632, 2017. [arXiv:1701.08165].
- [38] Tracy Holsclaw, Ujjaini Alam, Bruno Sanso, Herbert Lee, Katrin Heitmann, Salman Habib, and David Higdon. Nonparametric dark energy reconstruction from supernova data. *Physical Review Letters*, 105(24):241302, 2010. [arXiv:1011.3079].
- [39] Jun-Jie Wei and Xue-Feng Wu. An improved method to measure the cosmic curvature. *The Astrophysical Journal*, 838(2):160, 2017. [arXiv:1611.00904].
- [40] Hai-Nan Lin, Xin Li, and Li Tang. Non-parametric reconstruction of dark energy and cosmic expansion from the pantheon compilation of type ia supernovae. *Chinese Physics C*, 43(7):075101, 2019. [arXiv:1905.11593].
- [41] Will J. Handley, Anthony N. Lasenby, Hiranya V. Peiris, and Michael P. Hobson. Bayesian inflationary reconstructions from Planck 2018 data. *Phys. Rev. D*, 100(10):103511, 2019. [arXiv:1908.00906].
- [42] J.Alberto Vazquez, M. Bridges, M.P. Hobson, and A.N. Lasenby. Model selection applied to reconstruction of the Primordial Power Spectrum. *JCAP*, 06:006, 2012. [arXiv:1203.1252].
- [43] Austin Peel, Florian Lalande, Jean-Luc Starck, Valeria Pettorino, Julian Merten, Carlo Giocoli, Massimo Meneghetti, and Marco Baldi. Distinguishing standard and modified gravity cosmologies with machine learning. *Physical Review D*, 100(2):023508, 2019. [arXiv:1810.11030].
- [44] Rubén Arjona and Savvas Nesseris. What can machine learning tell us about the background expansion of the universe? *Physical Review D*, 101(12):123525, 2020. [arXiv:1910.01529].
- [45] Guo-Jian Wang, Xiao-Jiao Ma, and Jun-Qing Xia. Machine learning the cosmic curvature in a model-independent way. 2020. [arXiv:2004.13913].
- [46] Henry W Lin, Max Tegmark, and David Rolnick. Why does deep and cheap learning work so well? *Journal of Statistical Physics*, 168(6):1223–1247, 2017. [arXiv:1608.08225].
- [47] Isidro Gómez-Vargas, Ricardo Medel Esquivel, Ricardo García-Salcedo, and J Alberto Vázquez. Neural network within a bayesian inference framework. *Journal of Physics: Conference Series*, 1723(1):012022, 2021.
- [48] Sander Dieleman, Kyle W Willett, and Joni Dambre. Rotation-invariant convolutional neural networks for galaxy morphology prediction. *Monthly notices of the royal astronomical society*, 450(2):1441–1459, 2015. [arXiv:1503.07077].
- [49] Michelle Ntampaka, J ZuHone, D Eisenstein, D Nagai, A Vikhlinin, L Hernquist, F Marinacci, D Nelson, R Pakmor, A Pillepich, et al. A deep learning approach to galaxy cluster x-ray masses. *The Astrophysical Journal*, 876(1):82, 2019. [arXiv:1810.07703].
- [50] Andres C Rodríguez, Tomasz Kacprzak, Aurelien Lucchi, Adam Amara, Raphael Sgier, Janis Fluri, Thomas Hofmann, and Alexandre Réfrégier. Fast cosmic web simulations with generative adversarial networks. *Computational Astrophysics and Cosmology*, 5(1):4, 2018. [arXiv:1801.09070].
- [51] Siyu He, Yin Li, Yu Feng, Shirley Ho, Siamak Ravanbakhsh, Wei Chen, and Barnabás Póczos. Learning to predict the cosmological structure formation. *Proceedings of the National Academy of Sciences*, 116(28):13825–13832, 2019. [arXiv:1811.06533].
- [52] T Auld, Michael Bridges, MP Hobson, and SF Gull. Fast cosmological parameter estimation using neural networks. *Monthly Notices of the Royal Astronomical Society: Letters*, 376(1):L11–L15, 2007. [arXiv: astro-ph/0608174].

- [53] Justin Alsing, Tom Charnock, Stephen Feeney, and Benjamin Wandelt. Fast likelihood-free cosmology with neural density estimators and active learning. *Monthly Notices of the Royal Astronomical Society*, 488(3):4440– 4458, 2019. [arXiv:1903.00007].
- [54] Celia Escamilla-Rivera, Maryi A Carvajal Quintero, and Salvatore Capozziello. A deep learning approach to cosmological dark energy models. *Journal of Cosmology and Astroparticle Physics*, 2020(03):008, 2020. [arXiv:1910.02788].
- [55] Shi-Yu Li, Yun-Long Li, and Tong-Jie Zhang. Model comparison of dark energy models using deep network. *Research in Astronomy and Astrophysics*, 19(9):137, 2019. [arXiv:1907.00568].
- [56] Guo-Jian Wang, Xiao-Jiao Ma, Si-Yao Li, and Jun-Qing Xia. Reconstructing functions and estimating parameters with artificial neural networks: A test with a hubble parameter and sne ia. *The Astrophysical Journal Supplement Series*, 246(1):13, 2020. [arXiv:1910.03636].
- [57] Eric V Linder. Exploring the expansion history of the universe. *Physical Review Letters*, 90(9):091301, 2003. [arXiv: astro-ph/0208512].
- [58] Éric Aubourg, Stephen Bailey, Julian E Bautista, Florian Beutler, Vaishali Bhardwaj, Dmitry Bizyaev, Michael Blanton, Michael Blomqvist, Adam S Bolton, Jo Bovy, et al. Cosmological implications of baryon acoustic oscillation measurements. *Physical Review D*, 92(12):123516, 2015. [arXiv:1411.1074].
- [59] Zhongxu Zhai, Michael Blanton, Anže Slosar, and Jeremy Tinker. An evaluation of cosmological models from the expansion and growth of structure measurements. *The Astrophysical Journal*, 850(2):183, 2017. [arXiv:1705.10031].
- [60] J Alberto Vázquez, S Hee, MP Hobson, AN Lasenby, M Ibison, and M Bridges. Observational constraints on conformal time symmetry, missing matter and double dark energy. *Journal of Cosmology and Astroparticle Physics*, 2018(07):062, 2018. [arXiv:1208.2542].
- [61] Raul Jimenez, Licia Verde, Tommaso Treu, and Daniel Stern. Constraints on the equation of state of dark energy and the hubble constant from stellar ages and the cosmic microwave background. *The Astrophysical Journal*, 593(2):622, 2003. [arXiv: astro-ph/0302560].
- [62] Joan Simon, Licia Verde, and Raul Jimenez. Constraints on the redshift dependence of the dark energy potential. *Physical Review D*, 71(12):123001, 2005. [arXiv: astro-ph/0412269].
- [63] Michele Moresco, Licia Verde, Lucia Pozzetti, Raul Jimenez, and Andrea Cimatti. New constraints on cosmological parameters and neutrino properties using the expansion rate of the universe to  $z\sim 1.75$ . *Journal of Cosmology and Astroparticle Physics*, 2012(07):053, 2012. [arXiv:1201.6658].
- [64] Cong Zhang, Han Zhang, Shuo Yuan, Siqi Liu, Tong-Jie Zhang, and Yan-Chun Sun. Four new observational h(z) data from luminous red galaxies in the sloan digital sky survey data release seven. *Research in Astronomy and Astrophysics*, 14(10):1221, 2014. [arXiv:1207.4541].
- [65] Michele Moresco. Raising the bar: new constraints on the hubble parameter with cosmic chronometers at  $z \sim 2$ . *Monthly Notices of the Royal Astronomical Society: Letters*, 450(1):L16–L20, 2015. [arXiv:1503.01116].
- [66] Michele Moresco, Lucia Pozzetti, Andrea Cimatti, Raul Jimenez, Claudia Maraston, Licia Verde, Daniel Thomas, Annalisa Citro, Rita Tojeiro, and David Wilkinson. A 6% measurement of the hubble parameter at  $z \sim 0.45$ : direct evidence of the epoch of cosmic re-acceleration. *Journal of Cosmology and Astroparticle Physics*, 2016(05):014, 2016. [arXiv:1601.01701].
- [67] AL Ratsimbazafy, SI Loubser, SM Crawford, CM Cress, BA Bassett, RC Nichol, and P Väisänen. Agedating luminous red galaxies observed with the southern african large telescope. *Monthly Notices of the Royal Astronomical Society*, 467(3):3239–3254, 2017. [arXiv:1702.00418].
- [68] Khaled Said, Matthew Colless, Christina Magoulas, John R Lucey, and Michael J Hudson. Joint analysis of 6dfgs and sdss peculiar velocities for the growth rate of cosmic structure and tests of gravity. *Monthly Notices of* the Royal Astronomical Society, 497(1):1275–1293, 2020. [arXiv:2007.04993].
- [69] Nick Kaiser. Clustering in real space and in redshift space. *Monthly Notices of the Royal Astronomical Society*, 227(1):1–21, 1987.
- [70] Luca Amendola, Martin Kunz, and Domenico Sapone. Measuring the dark side (with weak lensing). *Journal of Cosmology and Astroparticle Physics*, 2008(04):013, 2008. [arXiv:0704.2421].
- [71] Bryan Sagredo, Savvas Nesseris, and Domenico Sapone. Internal robustness of growth rate data. *Physical Review D*, 98(8):083543, 2018. [arXiv:1806.10822].
- [72] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.

- [73] Yann A LeCun, Léon Bottou, Genevieve B Orr, and Klaus-Robert Müller. Efficient backprop. In *Neural networks: Tricks of the trade*, pages 9–48. Springer, 2012.
- [74] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural networks*, 3(5):551–560, 1990.
- [75] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.
- [76] Pierre Baldi and Kurt Hornik. Neural networks and principal component analysis: Learning from examples without local minima. *Neural networks*, 2(1):53–58, 1989.
- [77] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. 2013. [arXiv:1312.6114].
- [78] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *International conference on machine learning*, pages 1278–1286. PMLR, 2014.
- [79] Stuart Geman, Elie Bienenstock, and René Doursat. Neural networks and the bias/variance dilemma. *Neural computation*, 4(1):1–58, 1992.
- [80] Rémi Bardenet, Mátyás Brendel, Balázs Kégl, and Michele Sebag. Collaborative hyperparameter tuning. In *International conference on machine learning*, pages 199–207, 2013.
- [81] Frank Hutter, Holger H Hoos, Kevin Leyton-Brown, and Thomas Stützle. Paramils: an automatic algorithm configuration framework. *Journal of Artificial Intelligence Research*, 36:267–306, 2009. [arXiv:1401.3492].
- [82] Hugo Larochelle, Dumitru Erhan, Aaron Courville, James Bergstra, and Yoshua Bengio. An empirical evaluation of deep architectures on problems with many factors of variation. In *Proceedings of the 24th international conference on Machine learning*, pages 473–480, 2007.
- [83] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929– 1958, 2014.
- [84] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Insights and applications. In *Deep Learning Workshop, ICML*, volume 1, page 2, 2015.
- [85] Roberto Trotta. Bayes in the sky: Bayesian inference and model selection in cosmology. *Contemporary Physics*, 49(2):71–104, 2008. [arXiv:0803.4089].
- [86] F Feroz, MP Hobson, and M Bridges. Multinest: an efficient and robust bayesian inference tool for cosmology and particle physics. *Monthly Notices of the Royal Astronomical Society*, 398(4):1601–1614, 2009. [arXiv:0809.3437].
- [87] Florent Leclercq. Bayesian optimization for likelihood-free cosmological inference. *Physical Review D*, 98(6):063511, 2018. [arXiv:1805.07152].
- [88] AN Taylor and TD Kitching. Analytic methods for cosmological likelihoods. *Monthly Notices of the Royal Astronomical Society*, 408(2):865–875, 2010. [arXiv:1003.1136].
- [89] Savvas Nesseris and Juan Garcia-Bellido. A new perspective on dark energy modeling via genetic algorithms. *Journal of Cosmology and Astroparticle Physics*, 2012(11):033, 2012. [arXiv:1205.0364].
- [90] Steen Hannestad. Stochastic optimization methods for extracting cosmological parameters from cosmic microwave background radiation power spectra. *Physical Review D*, 61(2):023002, 1999.
- [91] Jayanti Prasad and Tarun Souradeep. Cosmological parameter estimation using particle swarm optimization. *Physical Review D*, 85(12):123008, 2012. [arXiv:1108.5600].
- [92] Luis E Padilla, Luis O Tellez, Luis A Escamilla, and J Alberto Vazquez. Cosmological parameter inference with bayesian statistics. 2019. [arXiv:1903.11127].
- [93] Ricardo Medel Esquivel, Isidro Gómez-Vargas, J Alberto Vázquez, and Ricardo García Salcedo. An introduction to markov chain monte carlo. *Boletín de Estadística e Investigación Operativa*, 1(37):47–74, 2021.
- [94] JA Vazquez, I Gomez-Vargas, and A Slosar. Updated version of a simple mcmc code for cosmological parameter estimation where only expansion history matters. https://github.com/ja-vazquez/SimpleMC, 2020.
- [95] Joshua S Speagle. dynesty: a dynamic nested sampling package for estimating bayesian posteriors and evidences. *Monthly Notices of the Royal Astronomical Society*, 493(3):3132–3158, 2020. [arXiv:1904.02180].
- [96] Henry W Leung and Jo Bovy. Deep learning of multi-element abundances from high-resolution spectroscopic data. *Monthly Notices of the Royal Astronomical Society*, 483(3):3255–3277, 2019. [arXiv:1808.04428].

- [97] Adam G Riess, Lucas M Macri, Samantha L Hoffmann, Dan Scolnic, Stefano Casertano, Alexei V Filippenko, Brad E Tucker, Mark J Reid, David O Jones, Jeffrey M Silverman, et al. A 2.4% determination of the local value of the hubble constant. *The Astrophysical Journal*, 826(1):56, 2016. [arXiv:1604.01424].
- [98] Will Handley. fgivenx: Functional posterior plotter. The Journal of Open Source Software, 3(28), Aug 2018.
- [99] Carl Doersch. Tutorial on variational autoencoders. 2016. [arXiv:1606.05908].
- [100] Diederik P Kingma and Max Welling. An introduction to variational autoencoders. 2019. [arXiv:1906.02691].
- [101] Daniel Ramos, Javier Franco-Pedroso, Alicia Lozano-Diez, and Joaquin Gonzalez-Rodriguez. Deconstructing cross-entropy for probabilistic binary classifiers. *Entropy*, 20(3):208, 2018.