# –Title of my thesis–

Shreeprasad Bhat

A Thesis Submitted to

Indian Institute of Technology Hyderabad

In Partial Fulfillment of the Requirements for

The Degree of Master of Technology


Department of Artificial Intelligence

June 2022

# Declaration

I declare that this written submission represents my ideas in my own words, and where ideas or words of others have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the Institute and can also evoke penal action from the sources that have thus not been properly cited, or from whom proper permission has not been taken when needed.

_____

(Signature)

_____

(Shreeprasad Bhat)

_____

(Roll No.)

# Approval Sheet

This Thesis entitled –Title of my thesis– by Shreeprasad Bhat is approved for the degree of Master of Technology from IIT Hyderabad

---

(———-) Examiner
Dept. of Chem Eng
IITH

---

(———-) Examiner
Dept. Math
IITH

---

(Dr. Shantanu Desai) Adviser
Dept. of Physics
IITH

---

(———) Chairman
Dept. of Artificial Intelligence
IITH

# Abstract

This is not a document on how to use latex. It rather explains how to use iiththesis.cls file to write your thesis for PhD/M.Tech/MSc. This file is generated using the class iiththesis.cls. This document draws a broad picture of the structure and formatting of your thesis.

# Contents

# Chapter 1

# Model-independently calibrating the luminosity correlations of GRBs using deep learning



## 1.1 Introduction

The accelerating expansion of the universe is first found from the fact that the luminosity of type Ia supernovae (SNe Ia) is dimmer than expected [1]. This led to the discovery of Dark energy [2]. One of the few ways to measure properties of dark energy is to extend the Hubble Diagram(HD) to high redshift. The only way to extend HD to higher redshift is to Gamma Ray Burts (GRB). GRB have been found to be reasonably good standard candles in the usual sense that light curve and/or spectral properties are correlated to the luminosity, exactly as for Cepheids and supernovae, then simple measurements can be used to infer their luminosities and hence distances. The default expectation is the simplest model for the Dark Energy, where it does not change in time. This can be parametrized with the equation of state of the Dark Energy. The concordance case has w = -1 at all times, and this is the expectation of Einstein's cosmological constant, or if the Dark Energy arises from vacuum energy. Given the strong results from supernovae for redshifts of less than 1, the frontier has now been pushed to asking the question of whether the value of w changes with time (and redshift).

The best way to measure properties of the Dark Energy seems to be to measure the expansion history of our Universe and place significant constraints on models of the Universe. Hubble diagram can be used to measure it. The Hubble diagram (HD) is a plot of distance versus redshift, with the slope giving the expansion history of our Universe. been proposed to determine the distances and redshifts of two thousand supernovae per year out to redshift 1.7 with exquisite accuracy. The default expectation is the simplest model for the Dark Energy, where it does not change in time. This can be parameterized with the equation of state of the Dark Energy. The best way to measure whether dark energy changed with respect to redshift, is to measure it over wide range of redshifts, but supernovaes cannot be deteceted above 1.7 even with modern satelites. But GRBs offer means extend HD over redshift $> 6$. The reason is that GRBs are visible across much larger distances than

supernovae.

GRBs are now known to have several light curve and spectral properties from which the luminosity of the burst can be calculated (once calibrated), and these make GRBs into 'standard candles'. Several interesting correlations among Gamma Ray Burst (GRB) observables with available redshifts have been recently identified. Proper evaluation and calibration of these correlations may facilitate the use of GRBs as standard candles constraining the expansion history of the universe up to redshifts of $z > 6$.

## 1.2 Literature Survey

A remarkable progress in the observation of gamma-ray bursts (GRBs) has been the identification of several very good correlations among the GRB observables ($\tau_{lag} - L, V - L, E_{peak} - L, E_{peak} - E_{gamma}, \tau_{RT} - L, E_{peak} - E_{\gamma,iso}$)[3]. Since then GRBs are proposed to use standard candles. But, all the GRB correlations have been obtained by fitting a hybrid GRB sample without discriminating the redshift. Then, inevitably, the effect of the GRB evolution with the redshift, and the selection effects, have been ignored. [4] shows that not all luminosity correlations are applicable across all redshifts, particularly they show correlation parameters for $E_{iso} - E_{\gamma}$ varies significantly across redshifts. However, [5] finds no statistically significant evidence for redshift dependence of correlation parameters. They also find that one of the five correlation relations tested ($E_{peak} - E_{\gamma}$) has a significantly lower intrinsic dispersion compared to the other correlations. [6] calculates luminosity correlations for updated GRB data and found that finds no statistically significant evidence for redshift dependence of correlation parameters. They also find find that the intrinsic scatter of the $V - L$ correlation is too large and there seems no inherent correlation between the two parameters using the latest GRB data. However all the above assumed a flat universe model to test the luminosity dependence. [7] have proposed a model independet method to test luminosity correlations of Gamma Ray Bursts, and found that there is no evidence for redshift dependence for $E_{peak} - E_{gamma}$ relation.



## 1.3 Observational Data

### 1.3.1 GRB

The GRB dataset we use is from [6]. In Table 1, we list the variables of 116 GRBs that we use in fitting luminosity correlations

### 1.3.2 Pantheon

Pantheon compilation [8] is the combined sample of SNe Ia discovered from different surveys to form the largest sample consisting of total of 1048 SNe Ia raning from $0.01 < z < 2.3$.

### 1.3.3 Union

The updated supernova Union2.1[9] compilation of 580 SNe is available at http://supernova.lbl.gov/Union

## 1.4 Methodology

### 1.4.1 Gaussian Processes

Gaussian Processes is a non-parametric regression technqiue, in the sense that we don't make any assumption about the function form. We define a prior probability distribution over functions y(x), such that set of values evaluated at $x_1, x_2...x_n$ follow gaussian distribution. This joint gaussian distributino is specified by mean and covariance. In most applications, we will not have any prior knowledge about the mean of y(x) and so by symmetry we take it to be zero. The covariance of y(x) evaluated at any two values of x, which is given by the kernel function $k(x, x')$.

### 1.4.2 Reccurent Neural Networks

to be written...

## 1.5 Reconstruction and calibration of distance modulus using Gaussian Processes



We first use Gaussian processes to reconstruct $\mu - z$ relation from pantheon data. Gaussian processes can construct function without involving any model assumption. The Gaussian processes only depend on the covariance function $k(x, x')$, which characterizes the correlation between the function value at $x$ to that at $x'$. There are many covariance functions available, but any covariance function should be positive definite and monotonously decreasing with the increment of distance between $x$ and $x'$. Here we use the following kernel

$$k(x, x') = ConstantKernel() + 1.0 * DotProduct(1) * *0.1 + 1.0 * WhiteKernel(1) \tag{1.1}$$

Our kernel (1.1) is a sum of linear, constant and whitekernels. Linear Kernel with exponent is used to capture relation in the data, constant kernel is used as scale magnitude and white kernel explains the noise in the input.

### 1.5.1 Training

We optimize the hyper-parameters of kernels by maximizing the marginal likelihood marginalized over function values f at the whole locations X. We use the publicly available python package sklearn[10] to reconstruct distance modulus as a function of redshift. The results are plotted in (1.2). The posterior samples drawn from kernel is shown in (1.1) .In the range where data points are sparse, the uncertainty of the reconstructed function is large. While training GP numerical issues are common to occur, hence we set $\alpha = 0.3$ and standardize the distance modulus before training. We also restart optimizer 100 times, parameters sampled log-uniform randomly from the space of allowed range.

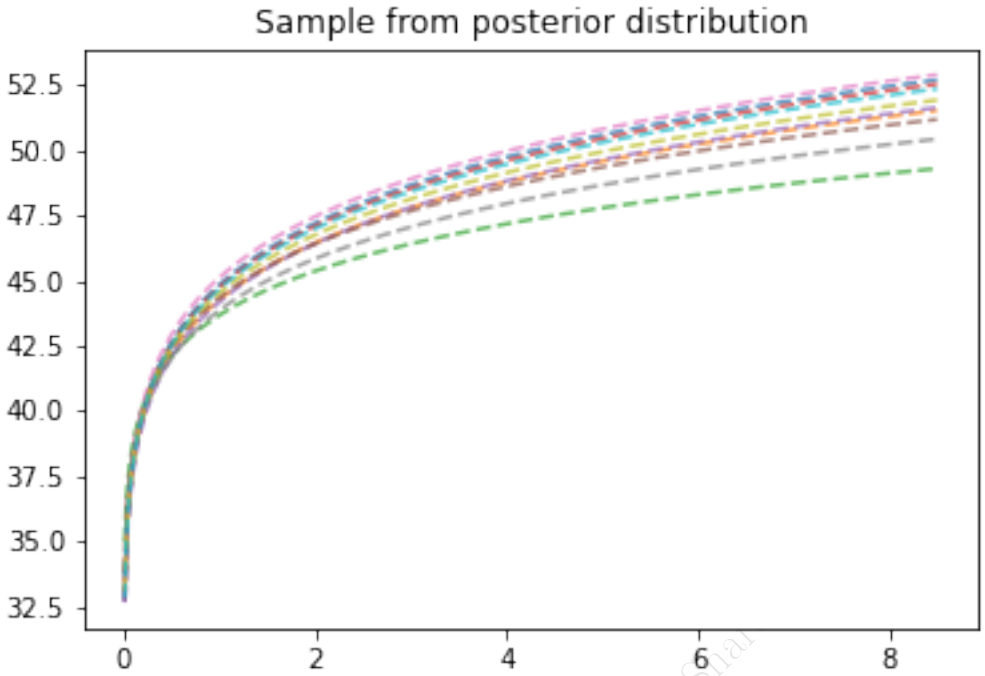


Figure 1.1: Posterior samples drawn from GP



The error bars with predictions are shown below

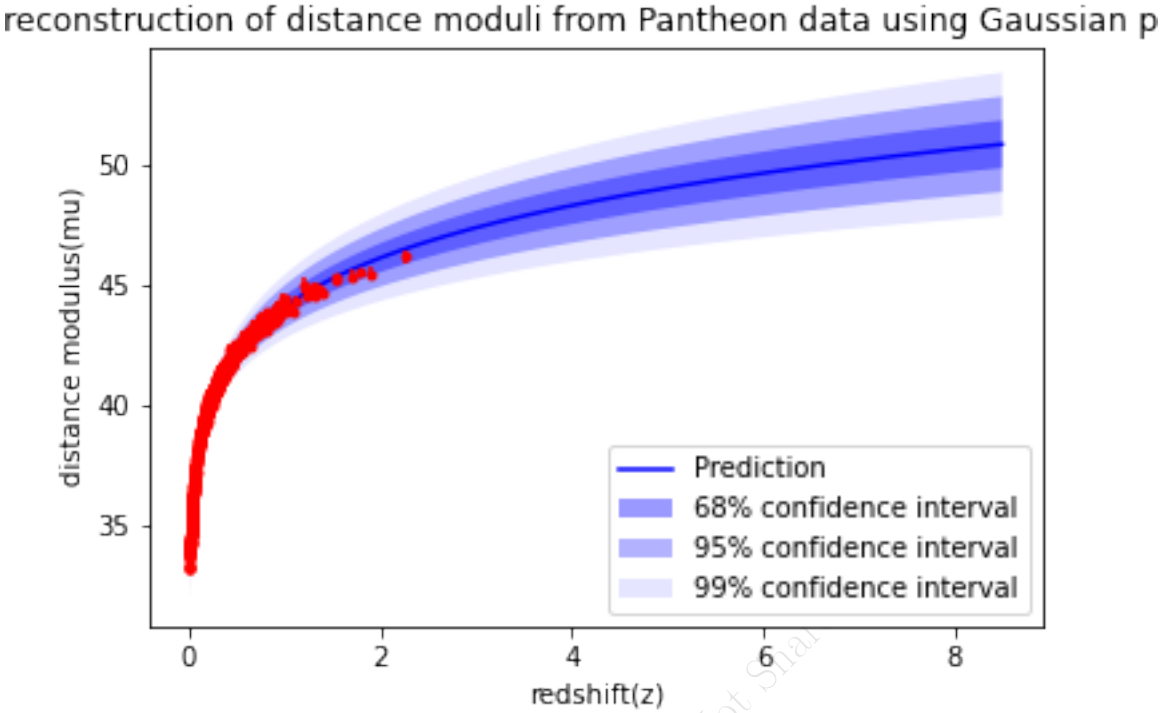


Figure 1.2: The reconstruction of distance moduli from Pantheon data set using GP. The red dots with $1\sigma$ error bars are the Pantheon data points. The light-blue dots are the central values of reconstruction. The shaded regions are the $1\sigma$, $2\sigma$ and $3\sigma$ uncertainties.

Log Marginal Likelihood = -20.3

The coefficient of determination $R^2$ = 0.9951

### 1.5.2 Testing redshift dependence of luminosity correlations

The luminosity relations are connections between measurable parameters of the light curves and/or spectra with the GRB luminosity. Specifically, I will be using the power law relationships between explained below. This section will discuss the calibration of all six relations. The calibration will essentially be a fit on a log-log plot of the luminosity indicator versus the luminosity. For this calibration process, the burst's luminosity distance must be known to convert Pbolo to L (or Sbolo to Egamma) and this is known only for bursts with measured redshifts. However, an important point is that the conversion from the observed redshift to a luminosity distance is done by machine learning model. The observed luminosity indicators will have different values from those that would be observed in the rest frame of the GRB. That is, the light curves and spectra seen by Earth-orbiting satellites suffer time-dilation and redshift. The physical connection between the indicators and the luminosity is in the GRB rest frame, so we must take our observed indicators and correct them to the rest frame of the GRB. For the two times (Tlag and TRT ), the observed quantities must be divided by 1+z to correct for time dilation. The observed V value varies as the inverse of the time stretching, so our measured value must be multiplied by 1 + z to correct to the GRB rest frame. The observed Epeak value must be multiplied by 1 + z to correct for the redshift of the spectrum.

The number of peaks in the light curve is defined in such a way as to have no z dependance. The dilation and redshift effects on thetajet and Egamma,iso have already been corrected in equations 1 and 2. A possibly substantial problem for the Tlag, V , and TRT relations is that we are in practice limited to the available energy bands (c.f. Table 5) whereas these correspond to different energy bands in the GRB reference frame. Ideally, we would want to measure these indicators in observed energy bands that correspond to some consistent band in the GRB frame

1. Lag versus Luminosity ($T_{lag} - L$)
2. Variability versus Luminosity ($V - L$)
3. $E_{peak}$ versus Luminosity ($E_{peak} - L$)
4. $E_{peak}$ versus $E_\gamma$ ($E_{peak} - E_\gamma$)
5. $T_{RT}$ versus Luminosity ($T_{RT} - L$)
6. $E_{peak}$ versus $E_{iso}$ ($E_{peak} - E_{iso}$)



$$\log \frac{L}{\mathrm{erg\,s^{-1}}} = a_1 + b_1 \log \frac{\tau_{\mathrm{lag},i}}{0.1s}, \tag{1.2}$$

$$\log \frac{L}{\mathrm{erg\ s^{-1}}} = a_2 + b_2 \log \frac{V_i}{0.02}, \tag{1.3}$$

$$\log \frac{L}{\mathrm{erg\ s^{-1}}} = a_3 + b_3 \log \frac{E_{p,i}}{300\mathrm{keV}} \tag{1.4}$$

$$\log \frac{E_\gamma}{\mathrm{erg}} = a_4 + b_4 \log \frac{E_{p,i}}{300\mathrm{keV}}, \tag{1.5}$$

$$\log \frac{L}{\mathrm{erg\ s}} = a_5 + b_5 \log \frac{\tau_{\mathrm{RT},i}}{0.1\ \mathrm{s}}, \tag{1.6}$$

$$\log \frac{E_{\mathrm{iso}}}{\mathrm{erg}} = a_6 + b_6 \log \frac{E_{p,i}}{300\mathrm{keV}} \tag{1.7}$$

Assuming that GRBs radiate isotropically, the isotropic equivalent luminosity can be derived from the bolometric peak flux $P_{\mathrm{bolo}}$ by (Schaefer 2007)

$$L = 4\pi d_L^2 P_{\mathrm{bolo}}\ ,$$

where $d_L$ is the luminosity distance of GRB, which can be obtained from the reconstructed distance moduli of Pantheon presented in section B with the relation

$$\mu = 5 \log \frac{d_L}{\mathrm{Mpc}} + 25.$$

Hence, the uncertainty of $L$ propagates from the uncertainties of $P_{\mathrm{bolo}}$ and $d_L$. The isotropic equivalent energy $E_{\mathrm{iso}}$ can be obtained from the bolometric fluence $S_{\mathrm{bolo}}$ by

$$E_{\mathrm{iso}}\ = 4\pi d_L^2 S_{\mathrm{bolo}}\ (1+z)^{-1},$$

| Correlation | sample | N | a | $a_err$ | b | $b_err$ | $\sigma$ | $\sigma_{int}$ |
|---|---|---|---|---|---|---|---|---|
| $T_{lag} - L$ | low-z | 37 | 52.09 | 0.11 | -0.78 | 0.16 | 0.51 | 0.09 |
| | high-z | 32 | 52.59 | 0.07 | -0.65 | 0.12 | 0.22 | 0.09 |
| | All-z | 69 | 52.32 | 0.07 | -0.76 | 0.11 | 0.47 | 0.06 |
| $V - L$ | low-z | 47 | 52.1 | 0.25 | 0.65 | 0.37 | 0.93 | 0.14 |
| | high-z | 57 | 52.8 | 0.15 | 0.34 | 0.14 | 0.62 | 0.07 |
| | All-z | 104 | 52.38 | 0.14 | 0.6 | 0.15 | 0.76 | 0.07 |
| $E_{peak} - L$ | low-z | 50 | 51.87 | 0.09 | 1.47 | 0.19 | 0.59 | 0.07 |
| | high-z | 66 | 52.48 | 0.06 | 1.15 | 0.15 | 0.3 | 0.06 |
| | All-z | 116 | 52.17 | 0.06 | 1.44 | 0.14 | 0.55 | 0.05 |
| $E_{peak} - E_{\gamma}$ | low-z | 12 | 50.63 | 0.08 | 1.56 | 0.19 | 0.23 | 0.09 |
| | high-z | 12 | 50.74 | 0.14 | 1.17 | 0.43 | 0.39 | 0.14 |
| | All-z | 24 | 50.67 | 0.07 | 1.47 | 0.17 | 0.26 | 0.07 |
| $T_{RT} - L$ | low-z | 39 | 52.69 | 0.13 | -1.34 | 0.19 | 0.48 | 0.07 |
| | high-z | 40 | 52.86 | 0.08 | -0.81 | 0.17 | 0.34 | 0.07 |
| | All-z | 79 | 52.77 | 0.08 | -1.23 | 0.13 | 0.45 | 0.05 |
| $E_{peak} - E_{iso}$ | low-z | 40 | 52.56 | 0.1 | 1.6 | 0.2 | 0.6 | 0.08 |
| | high-z | 61 | 53.0 | 0.06 | 1.27 | 0.14 | 0.38 | 0.04 |
| | All-z | 101 | 52.8 | 0.06 | 1.53 | 0.13 | 0.52 | 0.04 |

Table 1.1: A test caption



the uncertainty of $E_{iso}$ propagates from the uncertainties of $S_{bolo}$ and $d_L$. If on the other hand, GRBs radiate in two symmetric beams, then we can define the collimation-corrected energy $E_{\gamma}$ as

$$E_{\gamma} \equiv E_{\rm iso}\, F_{\rm beam}\ ,$$

where $F_{\rm beam} \equiv 1 - \cos\theta_{\rm jet}$ is the beaming factor, $\theta_{\rm jet}$ is the jet opening angle. The uncertainty of $E_{\gamma}$ propagates from the uncertainties of $E_{\rm iso\ and}\ F_{\rm beam}$ .

In order to test if the correlations discussed in the above section vary with redshift, we divide the GRB samples into two subsamples corresponding to the following redshift bins: the low-z sample ($z \leq 1.4$) which consists of 50 GRBs, and the high-z sample ($z > 1.4$) which consists of 66 GRBs. We investigate the redshift dependence of luminosity correlations for this two subsamples, as well as for the full GRBs sample. To fit the six luminosity correlations, we apply the D'Agostini's liklihood[11]

$$\mathcal{L}\left(\sigma_{\rm int}, a, b\right) \propto \prod_i \frac{1}{\sqrt{\sigma_{\rm int}^2 + \sigma_{yi}^2 + b^2\sigma_{xi}^2}} \times \exp\left[-\frac{\left(y_i - a - bx_i\right)^2}{2\left(\sigma_{\rm int}^2 + \sigma_{yi}^2 + b^2\sigma_{xi}^2\right)}\right]$$

For each correlation and each redshift bin, By maximizing this joint likelihood function, we can derive the best-fitting parameters $a$, $b$ and the intrinsic scatter $\sigma_{int}$, where the intrinsic scatter $\sigma_{int}$ denotes any other unknown errors except for the measurement errors. The results of the fits and the number of GRBs used in each fit are summarized in (1.1).

We perform a Markov Chain Monte Carlo analysis to calculate the posterior probability density function (PDF) of parameter space. We assume a flat prior on all the free parameters and limit $\sigma_{\rm int} > 0$. Note that not all GRBs can be used to analyze each luminosity correlation, because not all the necessary quantities are measurable for some GRBs. For example, GRBs without measurement of the spectrum lag can not used in the $\tau_{\rm lag} - L$ analysis. Hence, we present the best-fitting parameters, together with the number of available GRBs in each fitting in Table 1 In Figure 5 we

plot all the six luminosity correlations in logarithmic coordinates. Low- $z$ and high- $z$ GRBs are represented by blue and red dots with the error bars denoting $1\sigma$ uncertainties. The blue line, red line and black line stand for the best-fitting results for low- $z$ GRBs, high- $z$ GRBs and all- $z$ GRBs, respectively. The $1\sigma$ and $2\sigma$ contours and the PDFs for parameter space are plotted in Figure 6



As shown in Table 1 low- $z$ GRBs have a smaller intercept, but a sharper slope than high- $z$ GRBs for all the six luminosity correlations. All- $z$ GRBs have the parameter values between that of low- $z$ and high- $z$ subsamples. For the intrinsic scatter, low- $z$ GRBs have larger value than high- $z$ GRBs, and the $E_p - E_\gamma$ relation has the smallest intrinsic scatter hence we can only obtain its upper limit. The $V - L$ relation has the largest intrinsic scatter, thus it can not be fitted well with a simple line, which is legible in Figure [5 In Figure 6 the contours in the $(a, b)$ plane indicate that the $E_p - E_\gamma$ relation of low- $z$ GRBs is consistent with that of high- $z$ GRBs at $1\sigma$ confidence level. For the rest luminosity correlations, however, the intercepts and slopes for low- $z$ GRBs differ from that of high- $z$ GRBs at more than $2\sigma$ confidence level.

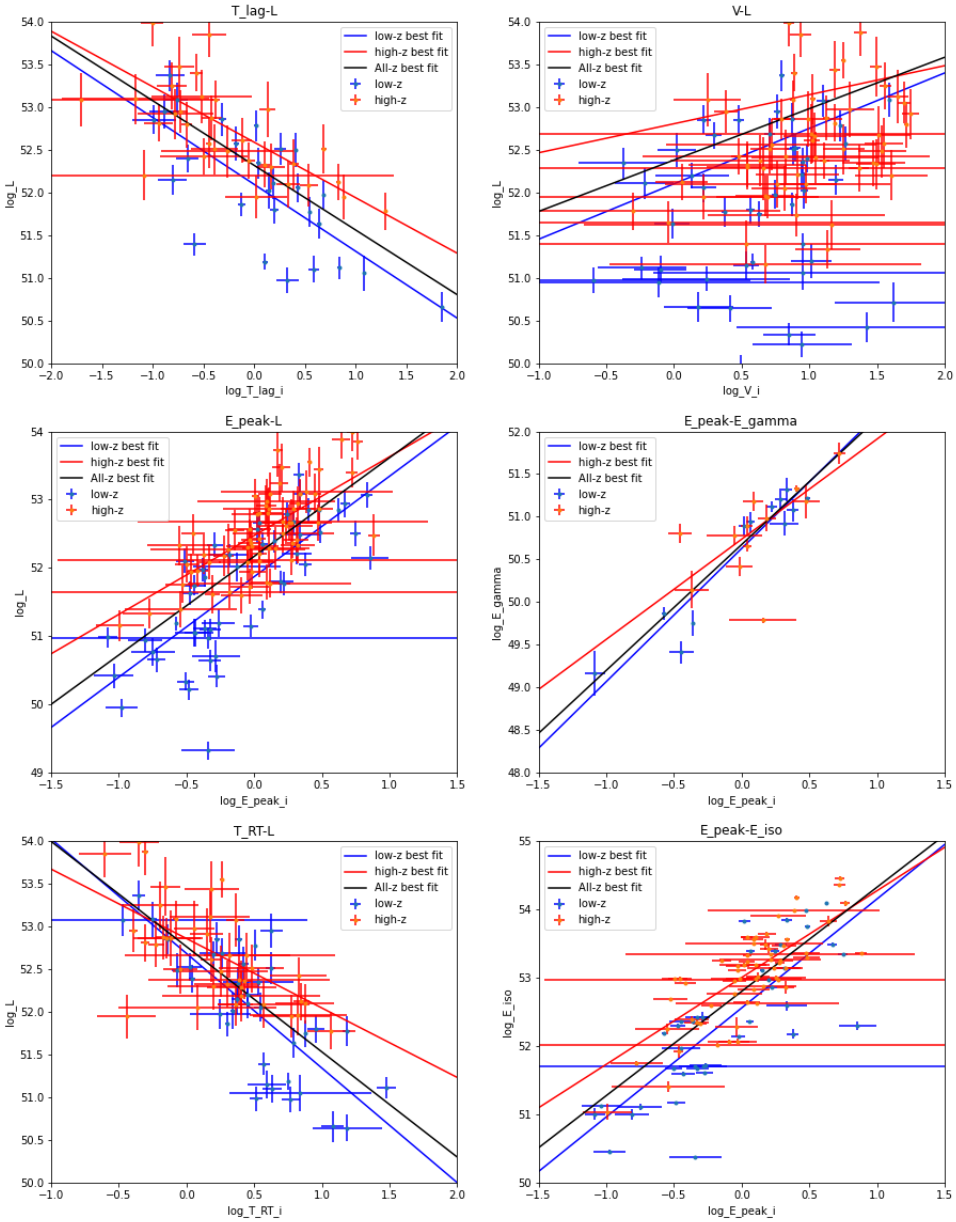


Figure 1.3: Luminsosity correlations best fit

### 1.5.3 Calibrating distance modulus from $E_{peak} - E_{gamma}$ relation

Having luminosity correlations calibrated, we can conversely using these correlations to calibrate the distance of GRBs, and further use GRBs to constrain cosmological models. Since our calibration of luminosity correlations is independent of cosmological model, the circularity problem is avoided. As we have seen, the $E_p - E_\gamma$ relation is not significantly evolving with redshift, so we use this relation to calibrate the distance of GRBs. Due to that the TABLE 1

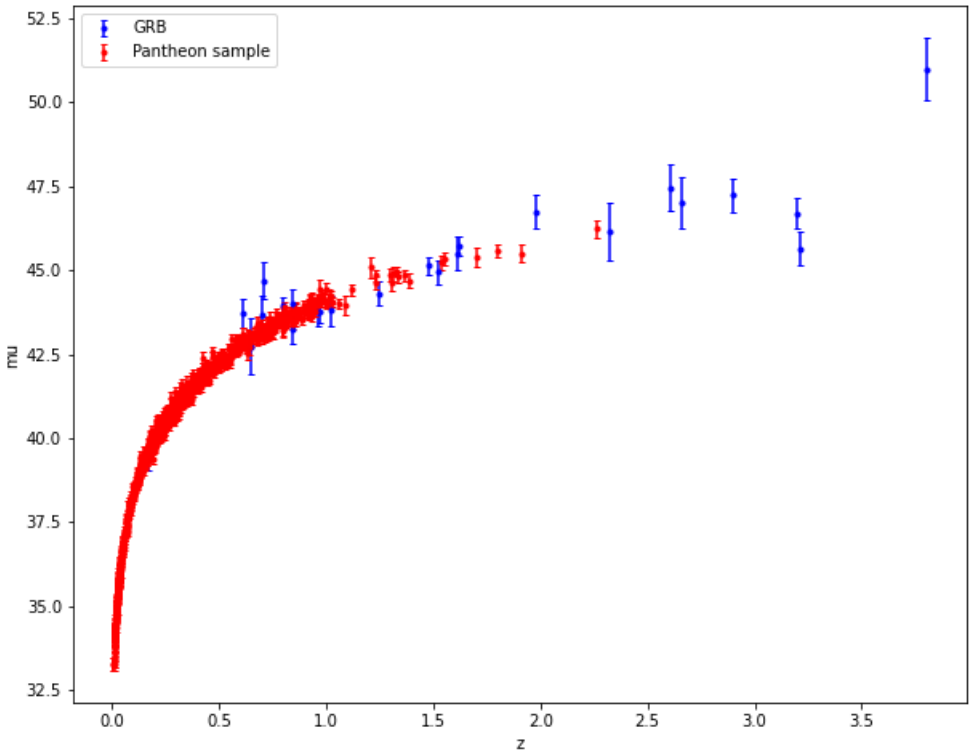


Figure 1.4: GRB Hubble Diagram

### 1.5.4 Constraints on the dark energy

Luminosity distance can be written as

$$d_L = c(1+z)\int_0^z \frac{1}{H(z)}dz \tag{1.8}$$

For flat ΛCDM, $H(z)$ can be written as

$$H(z) = H_0\sqrt{\Omega_M(1+z)^2 + 1 - \Omega_M} \tag{1.9}$$

We use emcee[12] to fit the dark energy equation. With the Pantheon dataset, the matter density of the flat ΛCDM model is constrained to be $\Omega_M = 0.278 \pm 0.007$. With 24 long GRBs alone, the matter density is constrained to be $\Omega_M = 0.307 \pm 0.065$. It indicates that the Hubble diagram in high redshift is consistent with the ΛCDM model

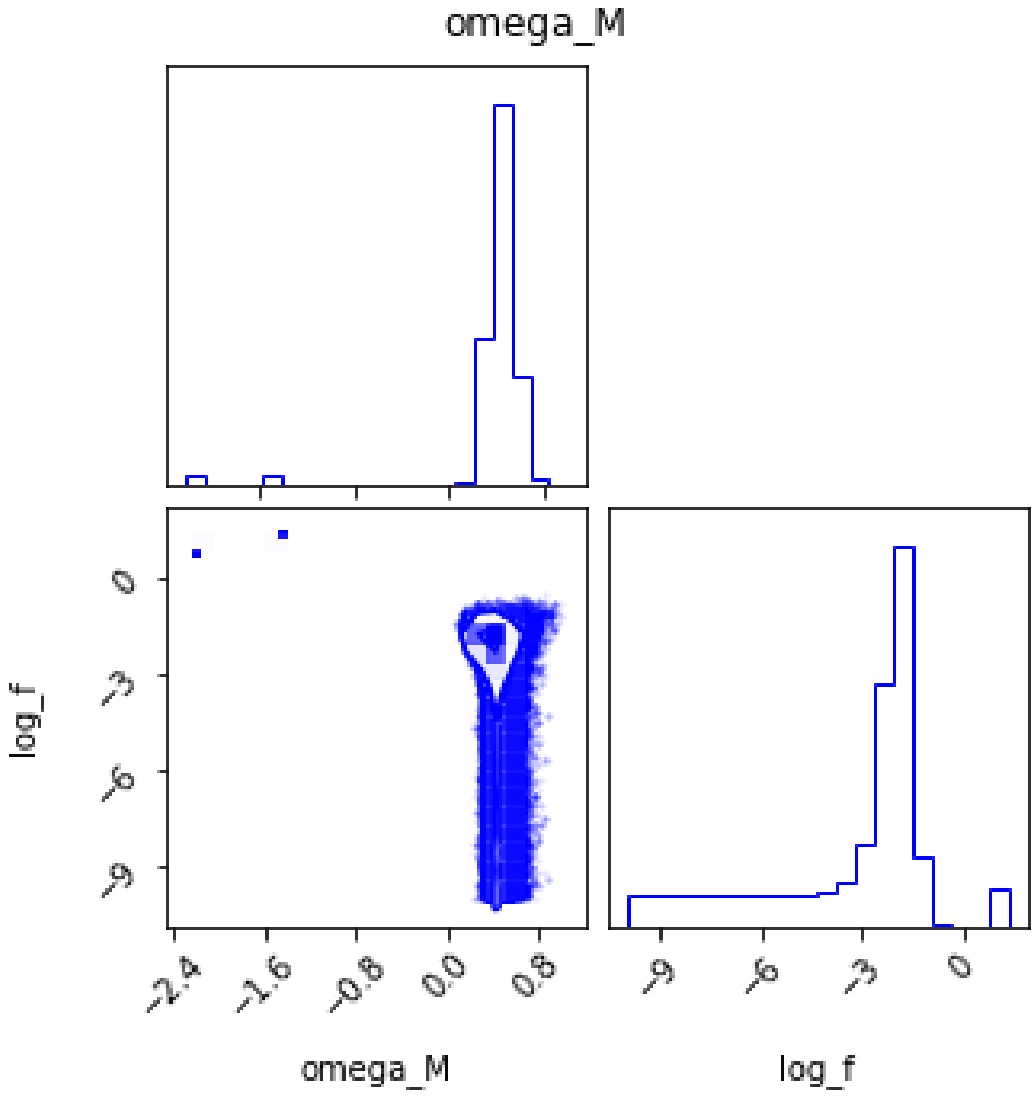


Figure 1.5: GRB Hubble Diagram

## 1.6 Reconstruction and calibration of distance modulus using Deep Learning

We construct the RNN+BNN network and train it with the package TensorFlow2[13]. For clarity, we present the corresponding hyperparameters in Figure 1 and list the steps to reconstruct data with our network as follow: (a) Data processing. The scale of data has an effect on training. Hence, we

normalize the distance moduli of the sorted Pantheon data and re-arrange $\mu - z$ as sequences with the step number t = 4. (b) Building RNN. We build RNN with three layers, i.e. an input layer, a hidden layer and an output layer as described in Figure 1. The first two layers are constructed with the LSTM cells of 100 neurons. The redshifts $z_{<t>}$ and the corresponding distance moduli $\mu_{<t>}$ are the input and output vectors, respectively. We employ the Adam optimizer to minimize the cost function MSE and train the network 1000 times. (c) Building BNN. We set the dropout rate to 0 in the input layer to avoid the lost of information, and to 0.2 in the second layer as well as the output layer (Bonjean 2020; Mangena et al. 2020). We execute the trained network 1000 times to obtain the distribution of distance moduli

### 1.6.1 Training

We train the neural network using pantheon data. The pantheon data is split into train and test data in equal size randomly. 512 datapoints are used for training and remaiing for testing. The network architecture is described in previous section. We use meansquared error loss and adam optimizer, with early stopping technique to prevent overfitting. Dropout technique with $dropout_rate = 0.2$. The hyperparameters used are $batch_size = 10$, $learning_rate$ = 1e-3, $patience = 5$.

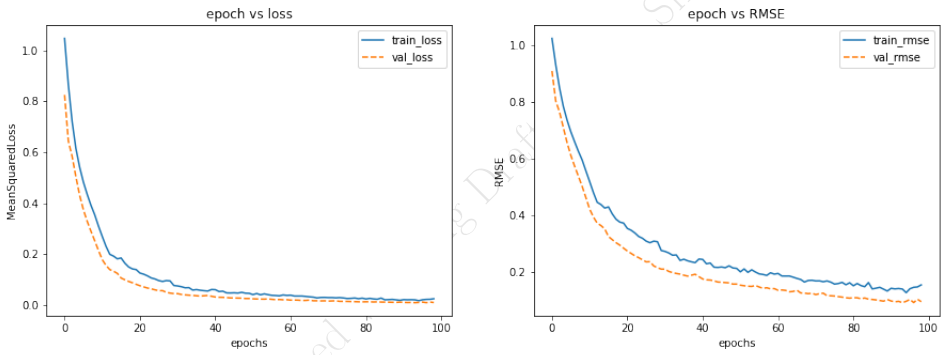


Figure 1.6: Loss curve

### 1.6.2 Testing redshift dependence of luminosity correlations

### 1.6.3 Calibrating distance modulus from $E_{peak} - E_{gamma}$ relation

### 1.6.4 Constraints on dark energy

## 1.7 Redoing analysis with Union Data

We redo all the analysis done for pantheon with union2.1 data and below are the results.

### 1.7.1 using Gaussian Processes

#### Training

The posterior drawn Gaussian process is shown below

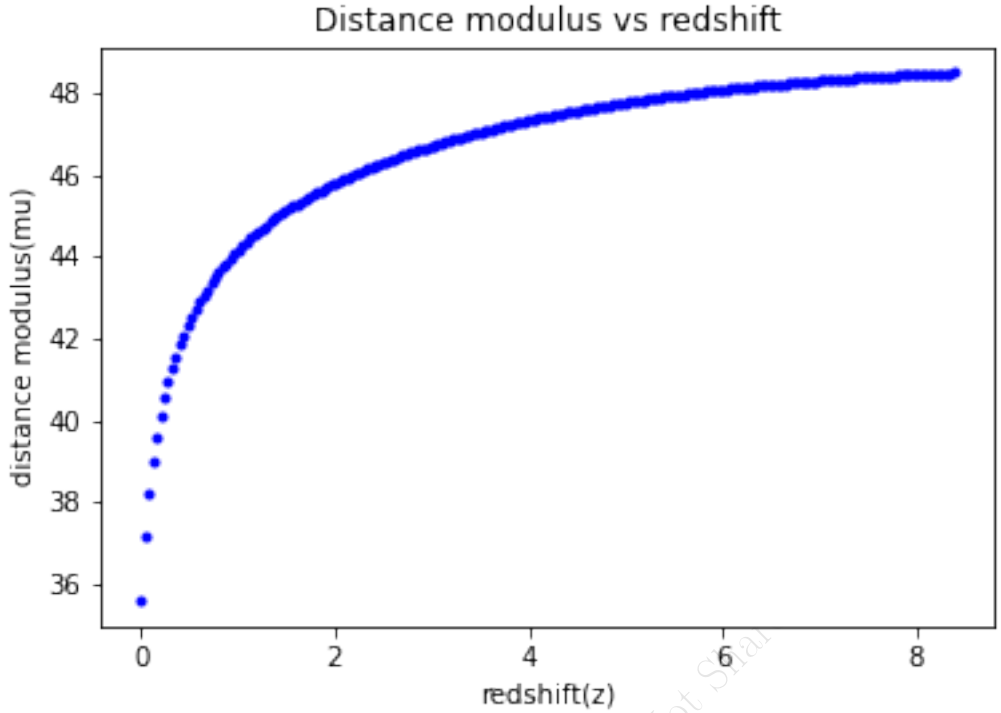


Figure 1.7: Loss curve



The error bars with predictions are shown belwo
Log Marginal Likelihood = -20.3
Score = 99.51

**Testing redshift dependence of luminosity correlations**

**Calibrating distance modulus from $E_{peak} - E_{gamma}$ relation**

**Constraints on the dark energy**

### 1.7.2 using Deep Learning

**Training**

**Testing redshift dependence of luminosity correlations**

**Calibrating distance modulus from $E_{peak} - E_{gamma}$ relation**

**Constraints on dark energy**

## 1.8 Conclusion

| Correlation | sample | N | a | $a_err$ | b | $b_err$ | $\sigma$ | $\sigma_{int}$ |
|---|---|---|---|---|---|---|---|---|
| $T_{lag} - L$ | low-z | 37 | 52.1 | 0.1 | -0.77 | 0.15 | 0.49 | 0.08 |
| | high-z | 32 | 52.37 | 0.07 | -0.6 | 0.12 | 0.29 | 0.07 |
| | All-z | 69 | 52.22 | 0.06 | -0.7 | 0.1 | 0.42 | 0.05 |
| $V - L$ | low-z | 47 | 52.12 | 0.25 | 0.65 | 0.36 | 0.91 | 0.13 |
| | high-z | 57 | 52.63 | 0.18 | 0.25 | 0.17 | 0.63 | 0.07 |
| | All-z | 104 | 52.34 | 0.13 | 0.46 | 0.14 | 0.75 | 0.07 |
| $E_{peak} - L$ | low-z | 50 | 51.89 | 0.09 | 1.43 | 0.18 | 0.59 | 0.07 |
| | high-z | 66 | 52.23 | 0.05 | 1.09 | 0.14 | 0.34 | 0.05 |
| | All-z | 116 | 52.05 | 0.05 | 1.35 | 0.12 | 0.5 | 0.04 |
| $E_{peak} - E_{\gamma}$ | low-z | 12 | 50.66 | 0.09 | 1.47 | 0.2 | 0.25 | 0.09 |
| | high-z | 12 | 50.53 | 0.13 | 1.37 | 0.43 | 0.39 | 0.16 |
| | All-z | 24 | 50.61 | 0.06 | 1.45 | 0.16 | 0.25 | 0.07 |
| $T_{RT} - L$ | low-z | 39 | 52.68 | 0.13 | -1.3 | 0.19 | 0.48 | 0.07 |
| | high-z | 40 | 52.61 | 0.09 | -0.74 | 0.17 | 0.39 | 0.06 |
| | All-z | 79 | 52.62 | 0.07 | -1.08 | 0.12 | 0.44 | 0.04 |
| $E_{peak} - E_{iso}$ | low-z | 40 | 52.57 | 0.1 | 1.55 | 0.2 | 0.6 | 0.08 |
| | high-z | 61 | 52.74 | 0.06 | 1.2 | 0.15 | 0.4 | 0.04 |
| | All-z | 101 | 52.65 | 0.05 | 1.42 | 0.12 | 0.49 | 0.04 |

Table 1.2: A test caption



| Correlation | sample | N | a | $a_err$ | b | $b_err$ | $\sigma$ | $\sigma_{int}$ |
|---|---|---|---|---|---|---|---|---|
| $T_{lag} - L$ | low-z | 37 | 52.13 | 0.11 | -0.79 | 0.16 | 0.53 | 0.08 |
| | high-z | 32 | 52.62 | 0.07 | -0.65 | 0.12 | 0.36 | 0.06 |
| | All-z | 69 | 52.36 | 0.07 | -0.77 | 0.11 | 0.5 | 0.05 |
| $V - L$ | low-z | 47 | 52.11 | 0.25 | 0.65 | 0.37 | 0.93 | 0.14 |
| | high-z | 57 | 52.83 | 0.16 | 0.34 | 0.15 | 0.62 | 0.07 |
| | All-z | 104 | 52.4 | 0.14 | 0.6 | 0.15 | 0.76 | 0.07 |
| $E_{peak} - L$ | low-z | 50 | 51.9 | 0.09 | 1.47 | 0.19 | 0.61 | 0.07 |
| | high-z | 66 | 52.52 | 0.06 | 1.13 | 0.15 | 0.41 | 0.04 |
| | All-z | 116 | 52.22 | 0.06 | 1.44 | 0.14 | 0.58 | 0.04 |
| $E_{peak} - E_{\gamma}$ | low-z | 12 | 50.65 | 0.08 | 1.56 | 0.19 | 0.24 | 0.09 |
| | high-z | 12 | 50.76 | 0.14 | 1.18 | 0.42 | 0.4 | 0.14 |
| | All-z | 24 | 50.7 | 0.06 | 1.48 | 0.17 | 0.27 | 0.07 |
| $T_{RT} - L$ | low-z | 39 | 52.71 | 0.13 | -1.34 | 0.19 | 0.51 | 0.07 |
| | high-z | 40 | 52.9 | 0.08 | -0.83 | 0.18 | 0.43 | 0.06 |
| | All-z | 79 | 52.8 | 0.08 | -1.23 | 0.13 | 0.49 | 0.05 |
| $E_{peak} - E_{iso}$ | low-z | 40 | 52.58 | 0.1 | 1.6 | 0.2 | 0.6 | 0.08 |
| | high-z | 61 | 53.03 | 0.06 | 1.28 | 0.14 | 0.39 | 0.04 |
| | All-z | 101 | 52.83 | 0.06 | 1.53 | 0.13 | 0.52 | 0.04 |

Table 1.3: A test caption

| Correlation | sample | N | a | $a_err$ | b | $b_err$ | $\sigma$ | $\sigma_{int}$ |
|---|---|---|---|---|---|---|---|---|
| $T_{lag} - L$ | low-z | 37 | 52.14 | 0.1 | -0.78 | 0.16 | 0.51 | 0.08 |
| | high-z | 32 | 52.18 | 0.08 | -0.51 | 0.13 | 0.36 | 0.07 |
| | All-z | 69 | 52.14 | 0.06 | -0.65 | 0.1 | 0.43 | 0.05 |
| $V - L$ | low-z | 47 | 52.14 | 0.25 | 0.65 | 0.37 | 0.92 | 0.14 |
| | high-z | 57 | 52.56 | 0.24 | 0.1 | 0.23 | 0.66 | 0.07 |
| | All-z | 104 | 52.33 | 0.14 | 0.32 | 0.15 | 0.79 | 0.07 |
| $E_{peak} - L$ | low-z | 50 | 51.92 | 0.09 | 1.46 | 0.18 | 0.6 | 0.07 |
| | high-z | 66 | 52.0 | 0.06 | 0.99 | 0.16 | 0.4 | 0.05 |
| | All-z | 116 | 51.95 | 0.05 | 1.28 | 0.12 | 0.5 | 0.04 |
| $E_{peak} - E_{\gamma}$ | low-z | 12 | 50.67 | 0.08 | 1.56 | 0.18 | 0.21 | 0.08 |
| | high-z | 12 | 50.36 | 0.16 | 1.57 | 0.5 | 0.45 | 0.18 |
| | All-z | 24 | 50.54 | 0.07 | 1.58 | 0.17 | 0.28 | 0.08 |
| $T_{RT} - L$ | low-z | 39 | 52.73 | 0.13 | -1.33 | 0.19 | 0.48 | 0.07 |
| | high-z | 40 | 52.39 | 0.09 | -0.63 | 0.18 | 0.43 | 0.06 |
| | All-z | 79 | 52.51 | 0.07 | -0.98 | 0.12 | 0.46 | 0.05 |
| $E_{peak} - E_{iso}$ | low-z | 40 | 52.6 | 0.1 | 1.6 | 0.2 | 0.59 | 0.08 |
| | high-z | 61 | 52.51 | 0.07 | 1.13 | 0.17 | 0.47 | 0.05 |
| | All-z | 101 | 52.53 | 0.06 | 1.36 | 0.13 | 0.52 | 0.04 |

Table 1.4: A test caption

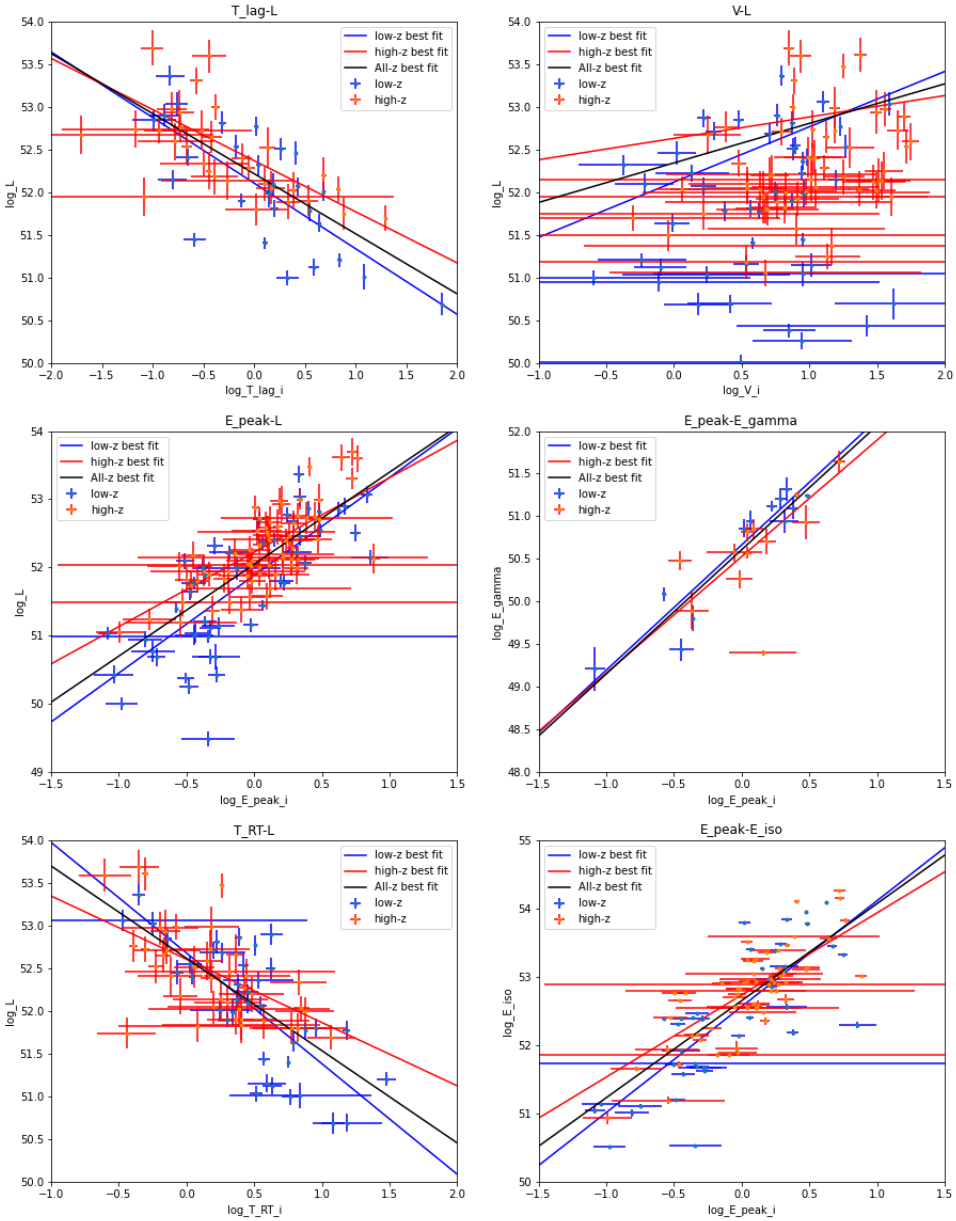


Figure 1.8: Luminsosity correlations best fit

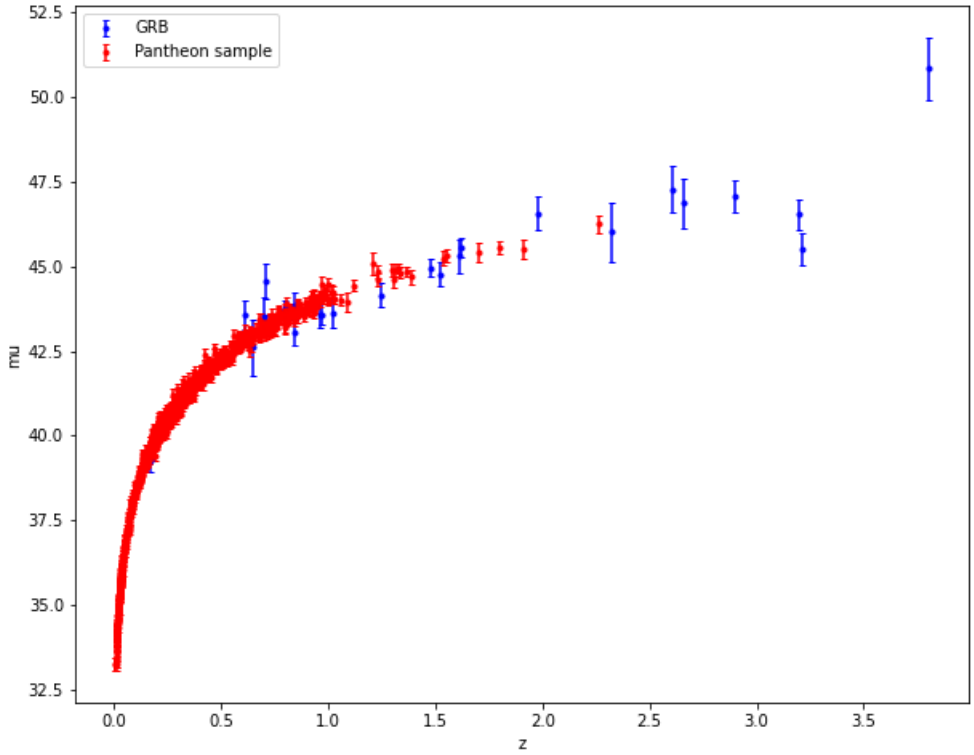


Figure 1.9: GRB Hubble Diagram

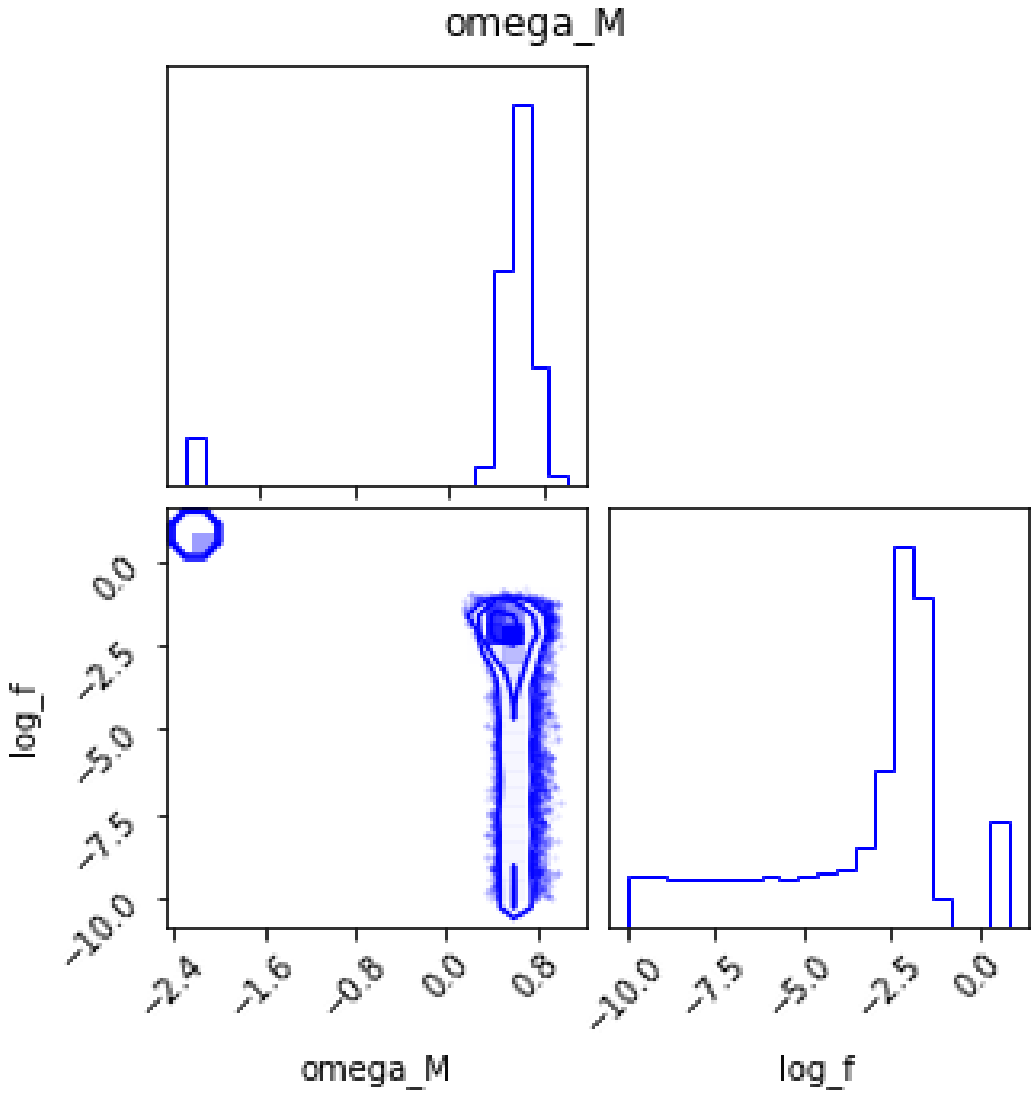


Figure 1.10: GRB Hubble Diagram

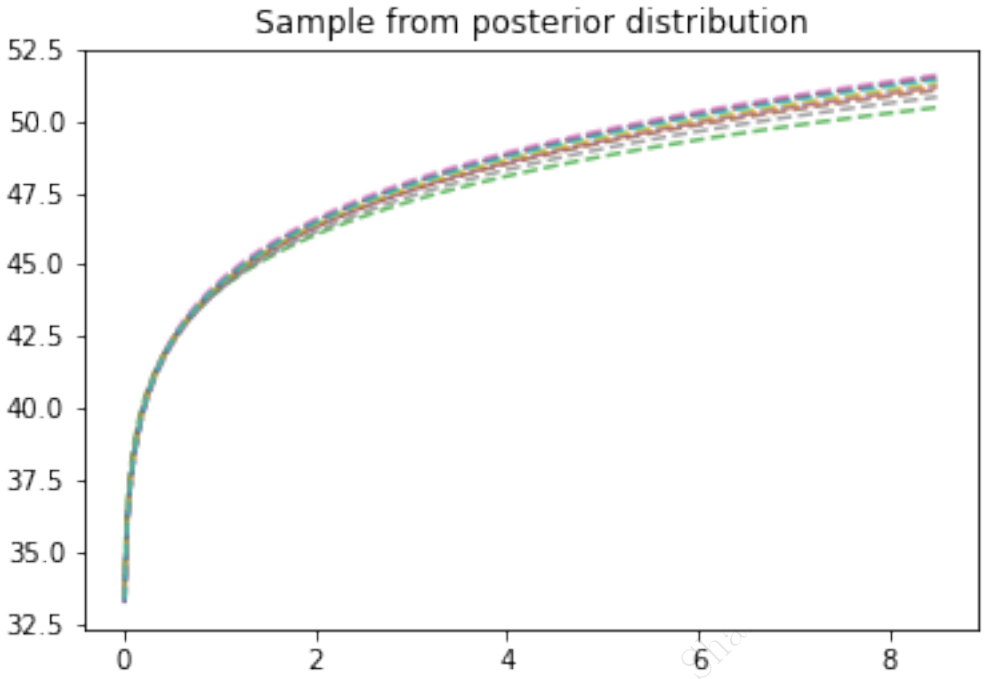


Figure 1.11: Posterior samples drawn from GP

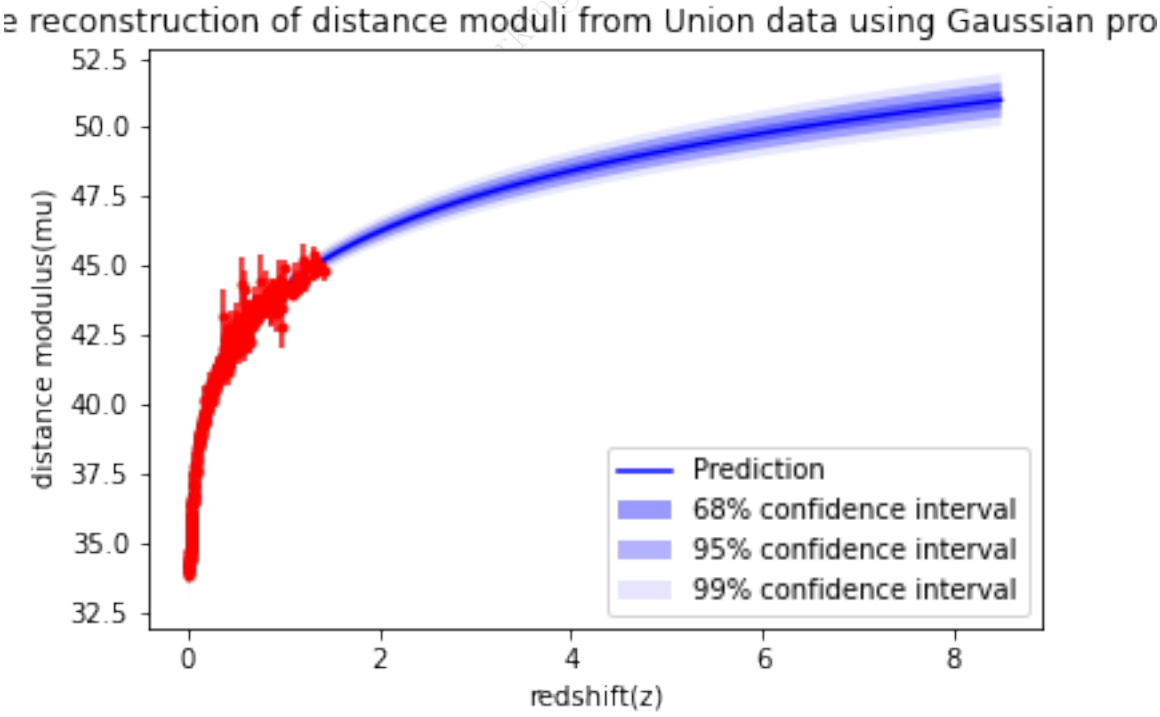


Figure 1.12: Reconstruction from Gaussian Processes

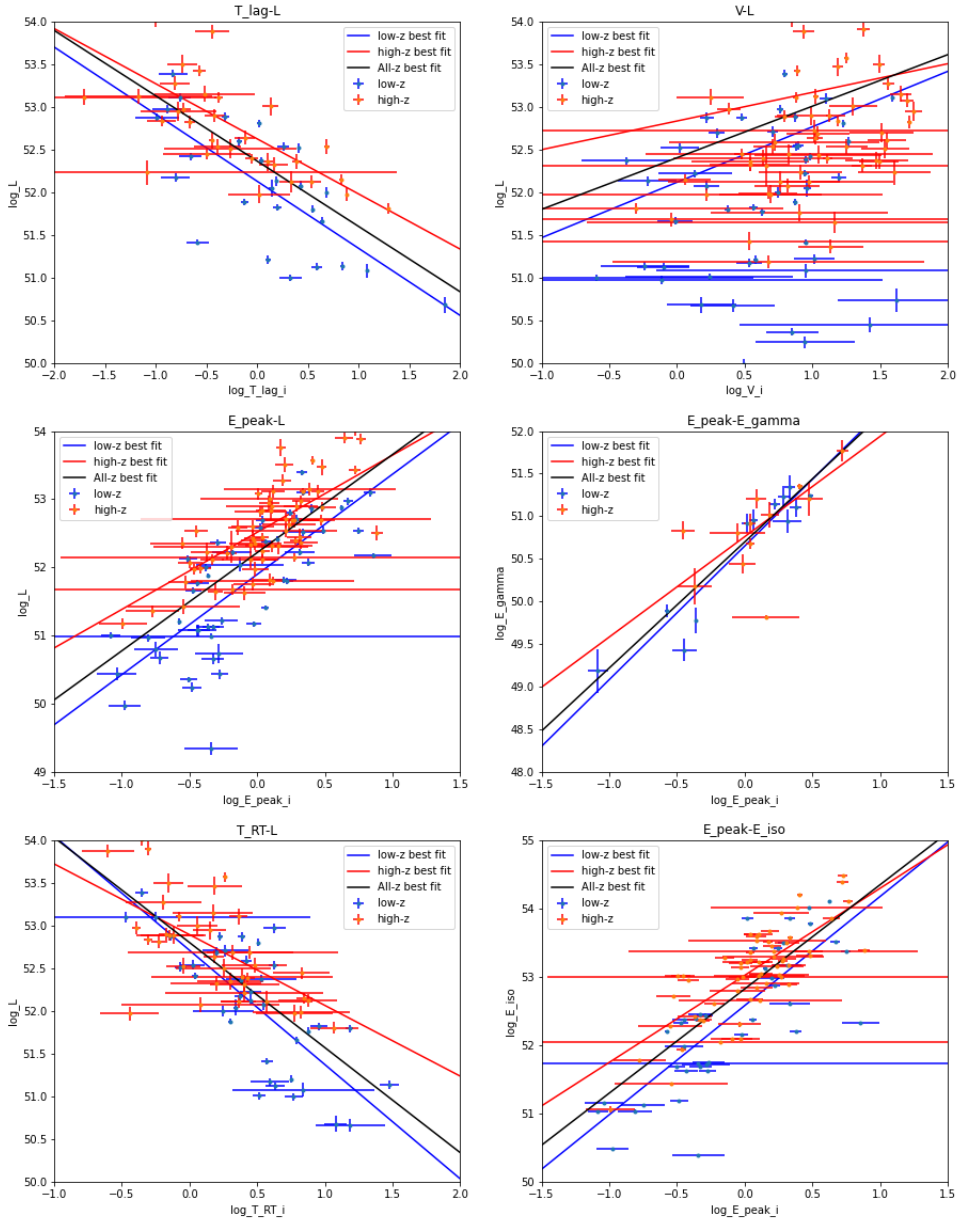


Figure 1.13: Luminsosity correlations best fit

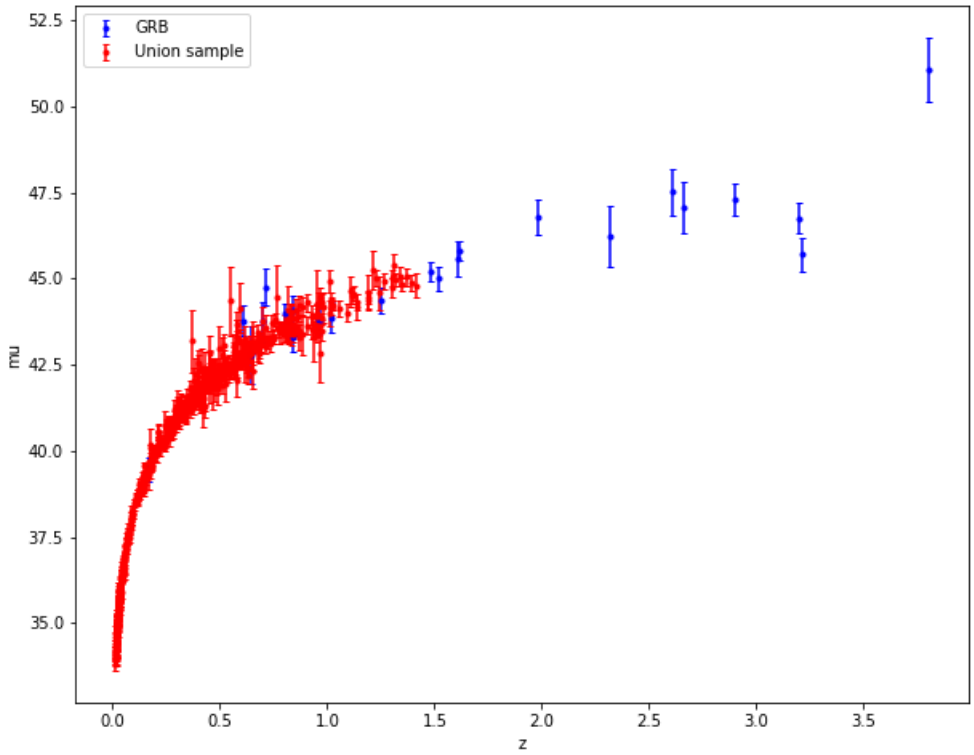


Figure 1.14: GRB Hubble Diagram

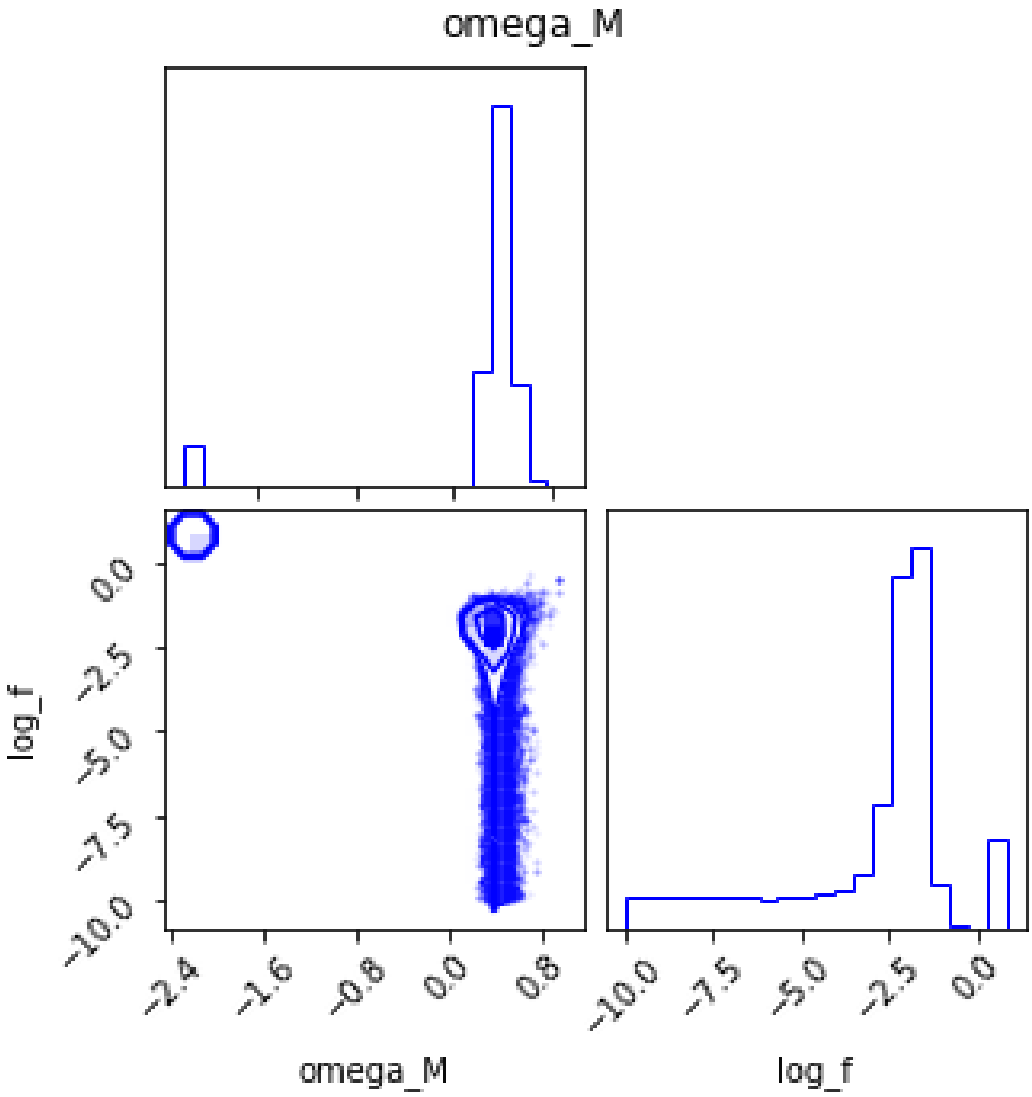


Figure 1.15: GRB Hubble Diagram

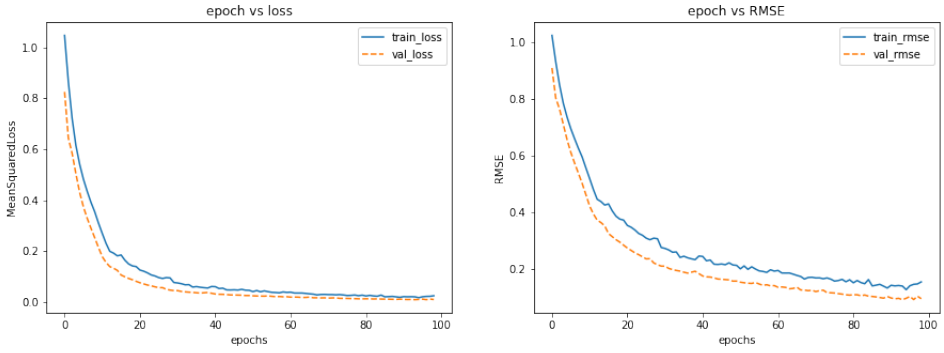


Figure 1.16: Loss curve

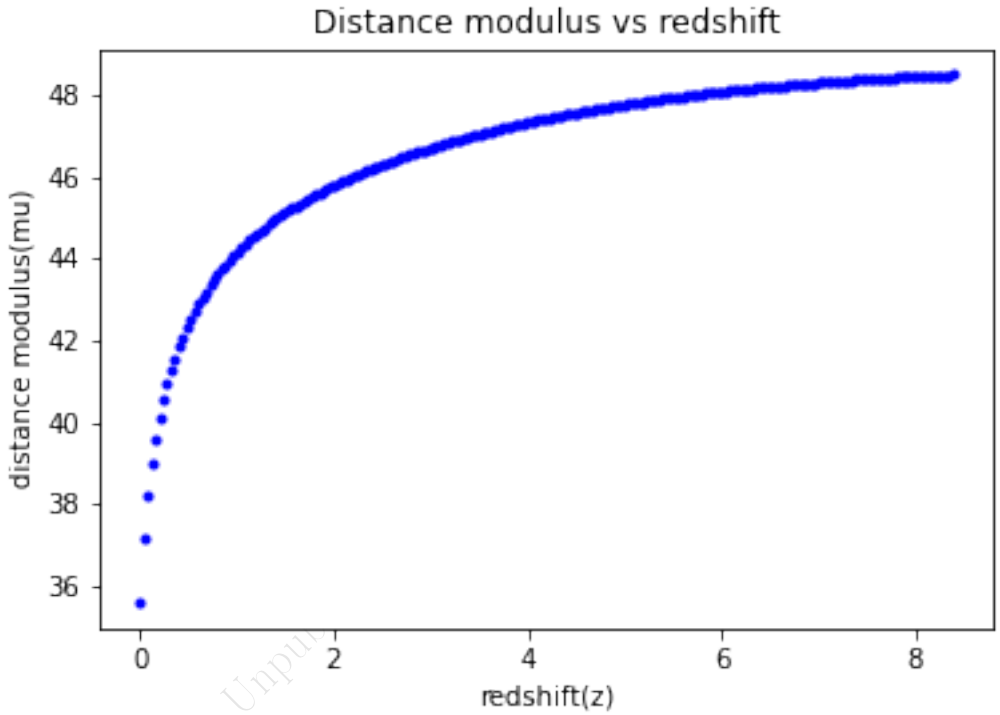


Figure 1.17: Loss curve

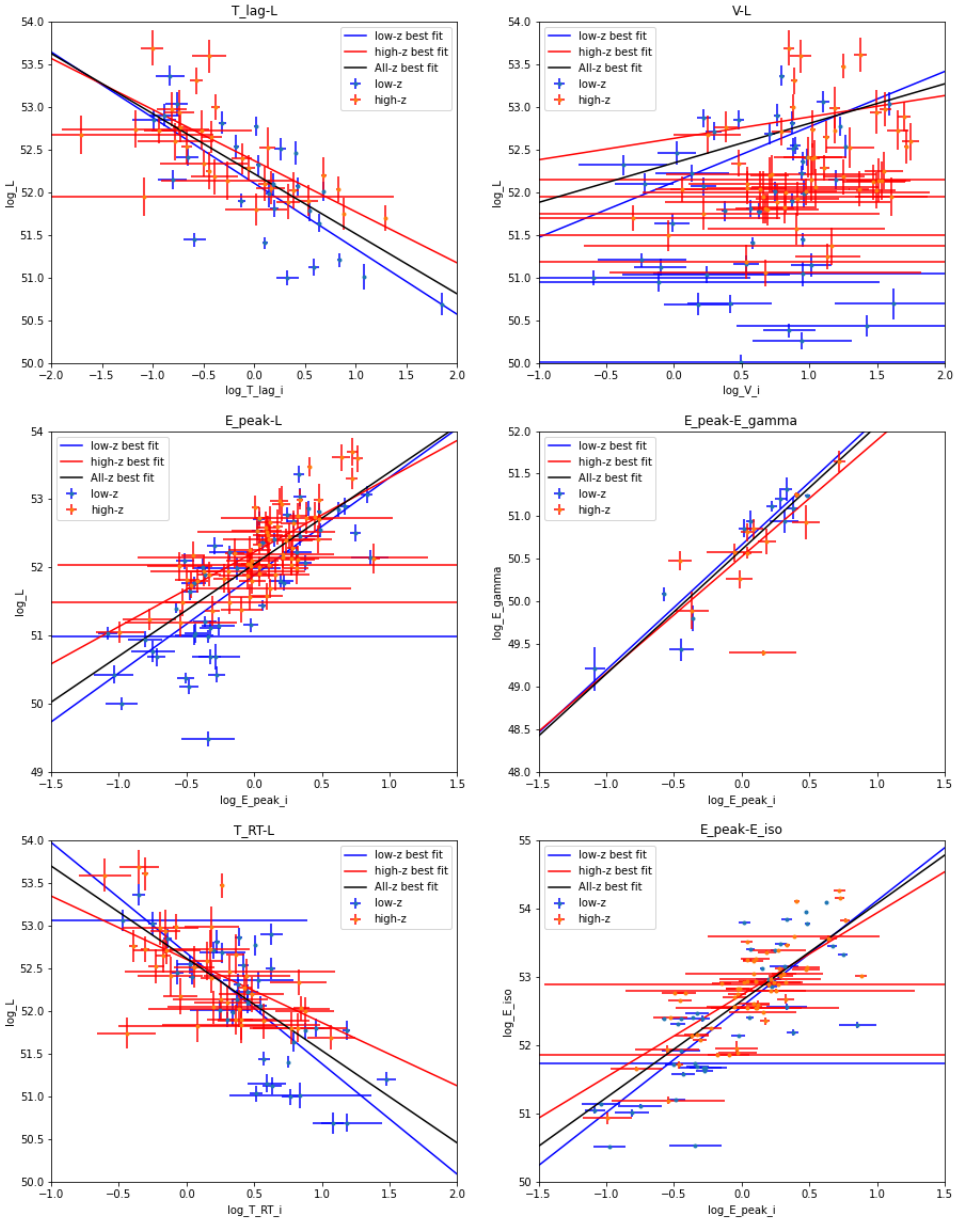


Figure 1.18: Luminsosity correlations best fit

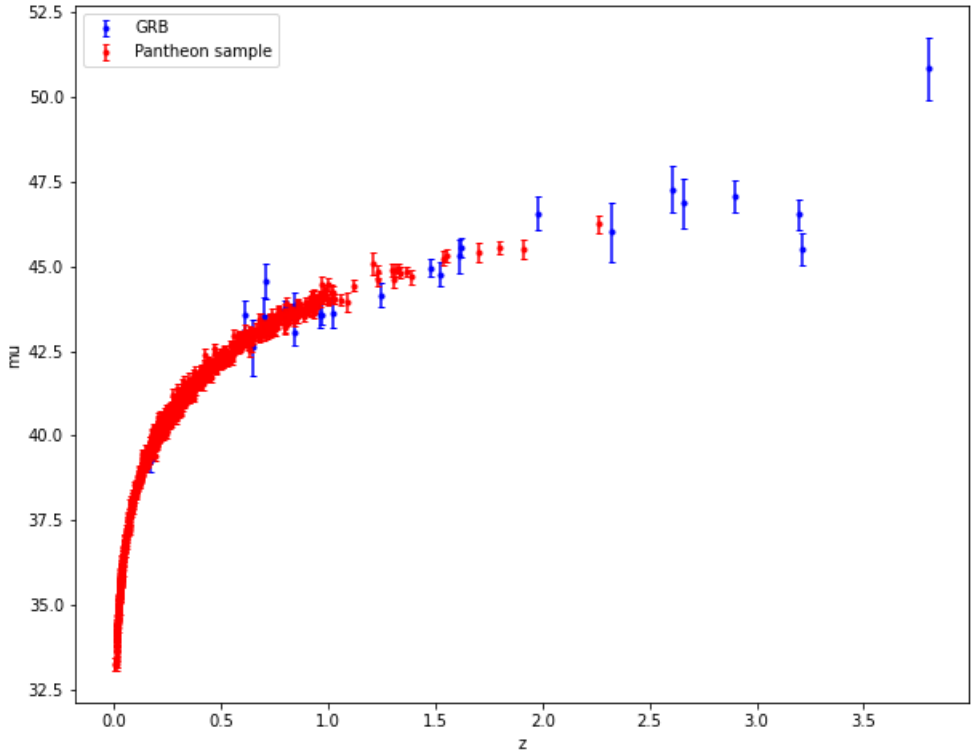


Figure 1.19: GRB Hubble Diagram

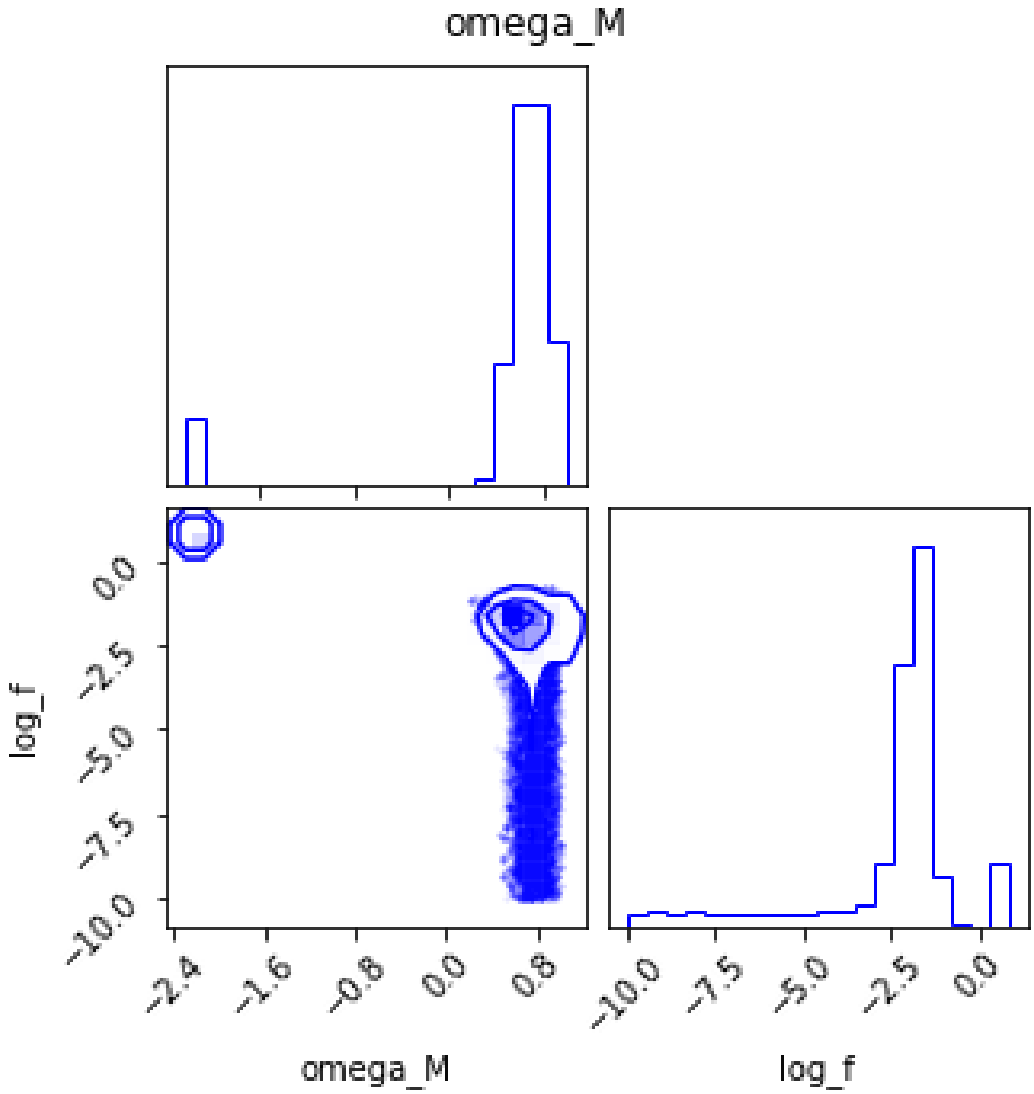


Figure 1.20: GRB Hubble Diagram

# References


[1] A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner et al. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The Astronomical Journal* 116, (1998) 1009.

[2] S. Perlmutter, G. Aldering, G. Goldhaber, R. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom et al. Measurements of $\Omega$ and $\Lambda$ from 42 high-redshift supernovae. *The Astrophysical Journal* 517, (1999) 565.

[3] B. E. Schaefer. The Hubble diagram to redshift$>$ 6 from 69 gamma-ray bursts. *The Astrophysical Journal* 660, (2007) 16.

[4] L.-X. Li. Variation of the Amati relation with cosmological redshift: a selection effect or an evolution effect? *Monthly Notices of the Royal Astronomical Society: Letters* 379, (2007) L55–L59.

[5] S. Basilakos and L. Perivolaropoulos. Testing gamma-ray bursts as standard candles. *Monthly Notices of the Royal Astronomical Society* 391, (2008) 411–419.

[6] F.-Y. Wang, S. Qi, and Z.-G. Dai. The updated luminosity correlations of gamma-ray bursts and cosmological implications. *Monthly Notices of the Royal Astronomical Society* 415, (2011) 3423–3433.

[7] L. Tang, X. Li, H.-N. Lin, and L. Liu. Model-independently calibrating the luminosity correlations of gamma-ray bursts using deep learning. *The Astrophysical Journal* 907, (2021) 121.

[8] D. M. Scolnic, D. Jones, A. Rest, Y. Pan, R. Chornock, R. Foley, M. Huber, R. Kessler, G. Narayan, A. Riess et al. The complete light-curve sample of spectroscopically confirmed SNe Ia from Pan-STARRS1 and cosmological constraints from the combined pantheon sample. *The Astrophysical Journal* 859, (2018) 101.

[9] N. Suzuki, D. Rubin, C. Lidman, G. Aldering, R. Amanullah, K. Barbary, L. Barrientos, J. Botyanszki, M. Brodwin, N. Connolly et al. The Hubble Space Telescope cluster supernova survey. V. Improving the dark-energy constraints above z$>$ 1 and building an early-type-hosted supernova sample. *The Astrophysical Journal* 746, (2012) 85.

[10] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Per-

rot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, (2011) 2825–2830.

[11] G. D'Agostini. Fits, and especially linear fits, with errors on both axes, extra variance of the data points and other complications. *arXiv preprint physics/0511182* .

[12] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman. emcee: The MCMC Hammer. *PASP* 125, (2013) 306–312.

[13] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems 2015. Software available from tensorflow.org.

# Chapter 2

# Model Comparison of Dark Energy models Using Deep Network

## 2.1 Introduction

[1] and [2] discovered that luminosity of Type Ia Supernovae are fainter than expected for decelerating Universe. This lead to conclusion that universe expansion is accelerating. Dark energy is proposed to account for this accelerating expansion, and its makes 73% of universe. Other observations from Cosmic Microwave Background(CMB)[**?**] and Baryon Acoustic Oscillations(BAO)[**?**] also supports this claim. The study of the nature of dark energy has become one of the most important issues in the field of fundamental physics. The simplest model for dark energy is ΛCDM, where Λ is the cosmological constant, which is equivalent to the quantum vacuum energy. For ΛCDM, the equation of state parameter is $w = -1$, so $p = -\rho$. ΛCDM model is very popular and accepted, since it can explain the current various astronomical observations quite well. But the cosmological constant has always been facing the severe theoretical challenges, such as the fine-tuning and coincidence problems. Hence other possible models are proposed. For example, a spatially homogeneous, slowly rolling scalar field can also provide a negative pressure, driving the cosmic acceleration[**?**]. More generally, one can phenomenologically characterize the property of dynamical dark energy through parametrizing $w$ of its equation of state (EoS) $p = -w\rho$, where $w$ is usually called the EoS parameter of dark energy. For example, the simplest parametrization model corresponds to the case of $w = constant$, and this cosmological model is sometimes called the $\omega$CDM model. A more physical and realistic situation is that $w$ is time variable, which is often probed by the so-called Chevalliear–Polarski–Linder (CPL) parametrization, $w(a) = w_0 + w_a(1 - a)$ [**?**].

## 2.2 Literature Survey

Since various dark energy models have been proposed, then the natural question is which model to select given the observational data. A variety of methods such as the $F-$test, Akaike information criterion (AIC), Mallows $C_p$, Bayesian information criterion (BIC), minimum description length (MDL), and Bayesian model averaging have been proposed to select a good or useful model in light of observations. [**?**] strongly recommends using Bayesian evidence to assign preferences to alternative

models since the evidence is the Bayesian's transportable quantity between models, and the popular easyto use AIC and BIC as well as MDL methods are all approximations to the Bayesian evidence[**?**]. The Bayesian evidence for model selection has been applied to the study of cosmology for a long time ([**?**] [**?**] [**?**], and recently a detailed study of Bayesian evidence for a large class of cosmological models taking into account around 21 different dark energy models has been performed by [**?**]. Although Bayesian evidence remains the preferred method compared with information criterions, a full Bayesian inference for model selection is very computationally expensive and often suffers from multi-modal posteriors and parameter degeneracies, which lead to a large time consumption to obtain the final result. Recently, deep learning models has been proposed for model selection. [**?**] proposed to use VAE−GAN model for both interpolation and model selection.

## 2.3 Observational Data

### 2.3.1 Union2.1

The observations are from the Union2.1 compilation [9] which contains 580 SNeIa, and $\boldsymbol{x}_{obs,\,\mathrm{real}}$ signify the measured distance moduli, $\Sigma_{\mathrm{obs}}$ represents the covariance of the distance moduli with systematics.



## 2.4 Methodology

### 2.4.1 VAE

A VAE[**?**] consists of two networks that encode a data sample x to a latent representation z and decode the latent representation back to data space, respectively:

$$z = Enc(x) = q(z|x), x' \approx Dec(z) = p(x|z) \tag{2.1}$$

The VAE regularizes the encoder by imposing a prior over the latent distribution $p(z)$. Typically $z \approx N(0, I)$ is chosen. The VAE loss is minus the sum of the expected log likelihood (the reconstruction error) and a prior regularization term:

$$L_{VAE} = -\mathbb{E}_q \left[ log \frac{p(x|z)p(z)}{q(z|x)} \right] = L_{like} + L_{prior} \tag{2.2}$$

with

$$L_{like} = -\mathbb{E} \left[ log \frac{p(x|z)}{} \right] \tag{2.3}$$

$$L_{prior} = -D_{KL} \left( q(z|x) || p(z) \right) \tag{2.4}$$

where $D_{KL}$ is the Kullback-Leibler divergence

### 2.4.2 GAN

A GAN[?] consists of two networks: the generator network $Gen(z)$ maps latents $z$ to data space while the discriminator network assigns probability $y = Dis(x) \in [0, 1]$ that $x$ is an actual training sample and probability $1 - y$ that $x$ is generated by our model through $x = Gen(z)$ with $z \approx p(z)$. The GAN objective is to find the binary classifier that gives the best possible discrimination between true and generated data and simultaneously encouraging Gen to fit the true data distribution. We thus aim to maximize/minimize the binary cross entropy:

$$L_{GAN} = log(Dis(x)) + log(1 - Dis(Gen(z))) \tag{2.5}$$

with respect to $Dis/Gen$ with $x$ being a training sample and $z \approx p(z)$.

### 2.4.3 VAEGAN

[?] proposes a combination of VAE and GAN, that outperforms traditional VAEs. A property of GAN is that its discriminator network implicitly has to learn a rich similarity metric for inputs, so as to discriminate them from generated data. They exploit this observation so as to transfer the properties of input learned by the discriminator into a more abstract reconstruction error for the VAE. The end result will be a method that combines the advantage of GAN as a high quality generative model and VAE as a method that produces an encoder of data into the latent space $z$.

Specifically, since element-wise reconstruction errors are not adequate for images and other signals with invariances, we propose replacing the VAE reconstruction (expected log likelihood) error term from Eq. 3 with a reconstruction error expressed in the GAN discriminator. To achieve this, let $Dis_l(x)$ denote the hidden representation of the $l$th layer of the discriminator. We introduce a Gaussian observation model for $Dis_l(x)$ with mean $Dis_l(x')$ and identity covariance:

$$p(Dis_l(x)|z) = N(Dis_l(x)|Dis_l(x'), I) \tag{2.6}$$

where $x' \approx Dec(z)$ is the sample from the decoder of x. We can now replace the VAE error of Eq. 3 with

$$L_{llike}^{Dis_l} = -E_{q(z|x)}[logp(Dis_l(x)|z)] \tag{2.7}$$

We train our combined model with the triple criterion

$$L = L_{prior} + L_{llike}^{Dis_l} + L_{GAN} \tag{2.8}$$

Notably, we optimize the VAE wrt. $L_{GAN}$ which we regard as a style error in addition to the reconstruction error which can be interpreted as a content error using the terminology from Gatys et al. (2015). Moreover, since both Dec and Gen map from z to x, we share the parameters between the two (or in other words, we use Dec instead of Gen in Eq. 5). In practice, we have observed the devil in the details during development and training of this model. We therefore provide a list of practical considerations in this section. We refer to Fig. 2 and Alg. 1 for overviews of the training procedure. To be written

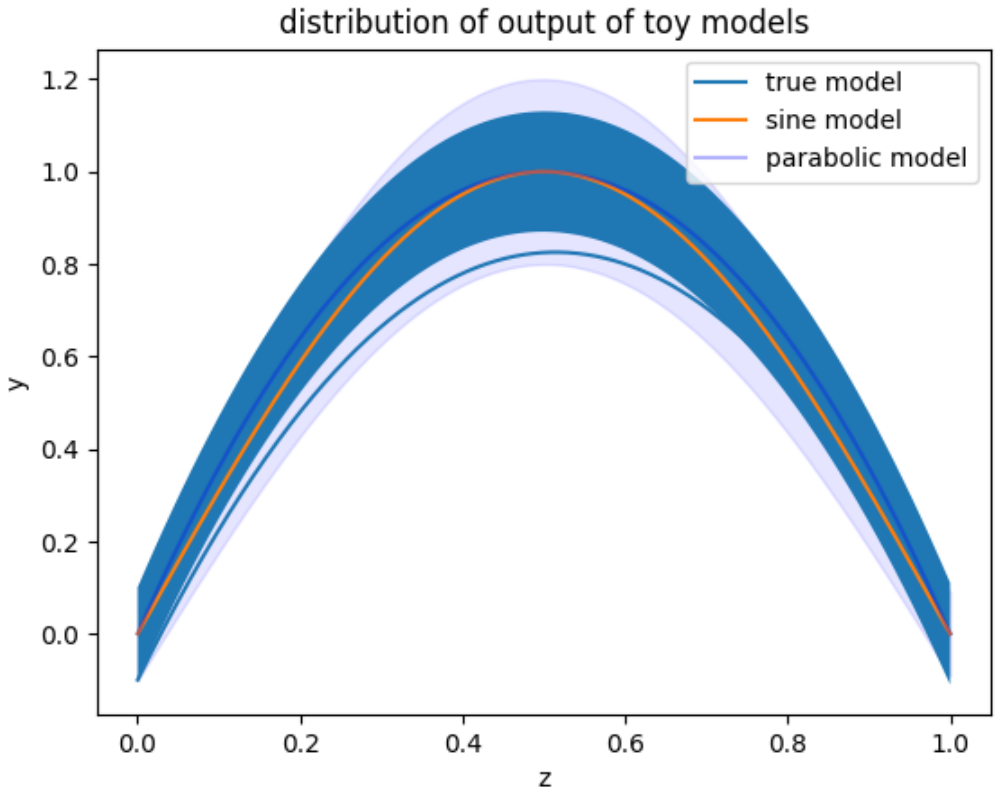


Figure 2.1: Toy models



## 2.5 Test on toy model

This section creates two toy models to test the data reconstruction and model comparison ability of the network.

Model 1,

$$y = Az^2 + (-A + B)z + C$$
$$\text{where, } A \sim \mathcal{N}(-4, 0.1), B \sim \mathcal{N}(0, 0.01), C \sim \mathcal{N}(0, 0.1)$$

Model 2,

$$y = A\sin(\omega z) + C$$
$$\text{where, } A \sim \mathcal{N}(1, 0.1), \omega \sim \mathcal{N}(\pi, 0.01), C \sim \mathcal{N}(0, 0.1)$$

Model 1 and Model 2 have similar distributions as shown in Figure 2 Given the observations $\boldsymbol{x}_{\text{obs,real}}$ which are generated by the underlying model $y_{\text{true}} = -3.5z^2 + 3.6z - 0.1$ on $\boldsymbol{z}_{\text{obs}} = \{z_1, z_2, \cdots, z_{580}\}$ with an error matrix $\Sigma_{obs}$, we would like to fit the two toy models to the observations to tell which one is most probable to be the true model, and interpolate the data with the model at $\boldsymbol{z}^* = \{z_1^*, \cdots, z_M^*\}$, for example, $\boldsymbol{z}^*$ even staying in the interval $[0, 1]$ with $M = 1468$.

First we concatenate and sort $\boldsymbol{z}$ and $\boldsymbol{z}^*$, and call the new one $\boldsymbol{z}$. Then sample $\{A_i, B_i, C_i, \omega_i\}$ from the priors of the toy models and generate the training samples $\boldsymbol{x}_i = M_k(\boldsymbol{z} \mid A_i, B_i, C_i, \omega_i)$ (Note that which set of parameters should be used depends on the toy model). Here 12800 samples

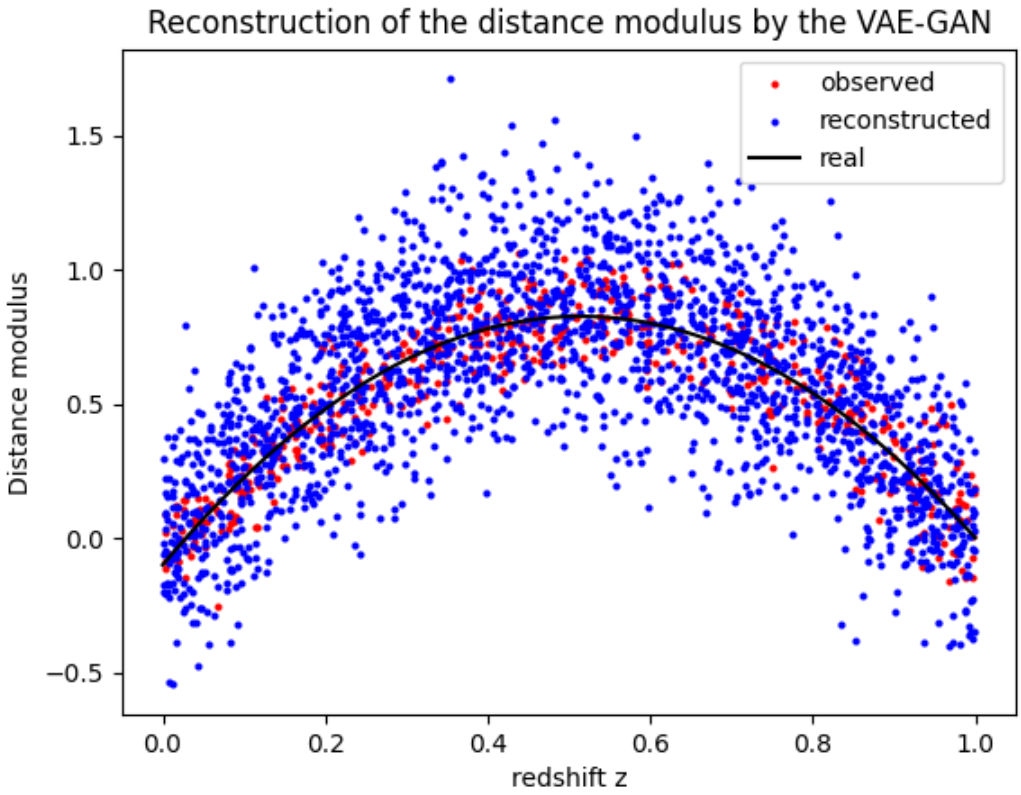


Figure 2.2: Reconstruction

for each model are generated as the training dataset. Finally, the training set $\{\boldsymbol{x}\}_{i=1}^{25600}$ together with the observation error $\Sigma_{obs}$ is fed into the network. Once the training converges, one can put the observations $\boldsymbol{x}_{obs,real}$ into the network to tell which toy model is most probable and get the interpolation, see Figure 3 In this task, the discriminator has a classification accuracy of almost 1. It assigns a probability of 97% to the parabolic model (Model 1), which is indeed the case.

## 2.6 Dark enrgy models

We consider three popular dark energy models to test out VAE-GAN network for model selection and interpolation.

### 2.6.1 ΛCDM

The equation of state parameter is

$$\omega(z) = -1 \tag{2.9}$$

### 2.6.2 ωCDM

$$\omega(z) = \omega_{DE} \tag{2.10}$$

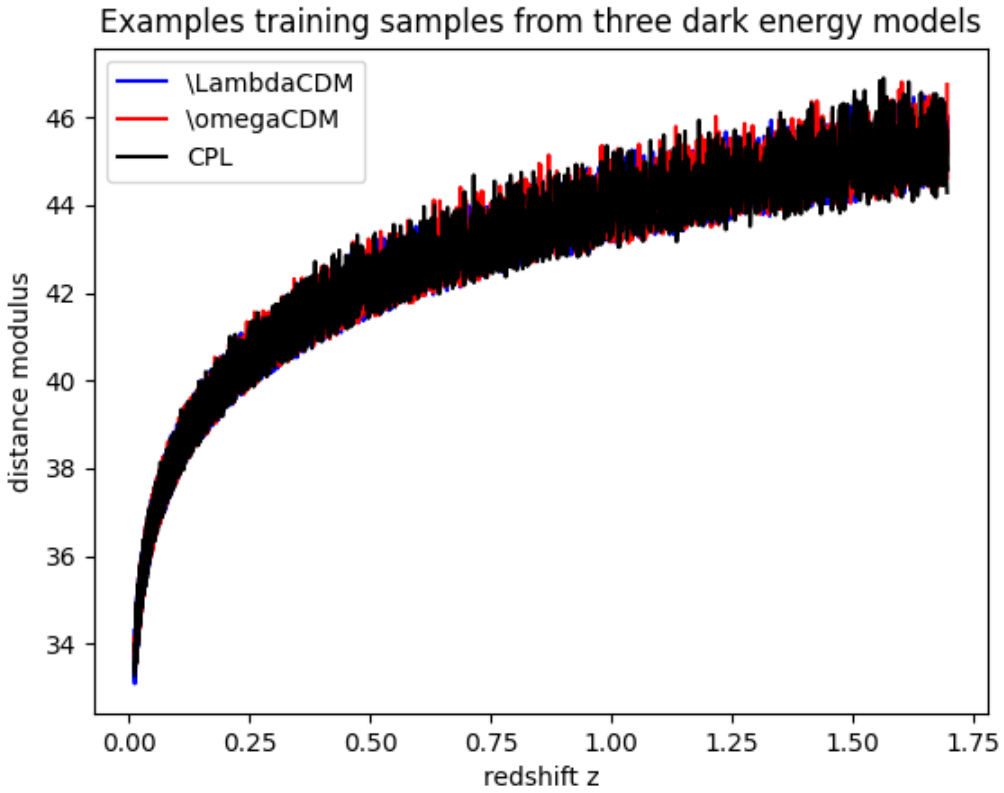


Figure 2.3: Samples from dark energy models

### 2.6.3 CPL

$$\omega(z) = \omega_0 + \omega_a \frac{z}{1+z} \tag{2.11}$$

### 2.6.4 Distance Modulus

We evaluate these models at redshifts $z_{obs}$ given by Union2.1 data and randomly sampled redshifts between $(0.8min(z_{obs}) - 1.2 * max(z_{obs})$. The expansion rate of a spatially flat FRW universe is determined by the matter and dark energy,

$$H^2(z) = H_0^2 \left\{ \Omega_{m0}(1+z)^3 + (1-\Omega_{m0}) \exp\left[ 3 \int \frac{1+\omega\left(z'\right)}{1+z'} dz' \right] \right\}$$

The luminosity distance is closely related to the Hubble expansion rate (Eq.12), and the distance modulus is given by Eq13

$$D_L(z) = c(1+z) \int_0^z dz' \frac{1}{H\left(z'\right)}$$

$$\mu(z) = 5 \log_{10} D_L(z) + 25$$

For each dark energy model, 12800 samples are generated at the redshift $\boldsymbol{z} = \mathrm{sort}\left\{\boldsymbol{z}_{\mathrm{obs}}, \boldsymbol{z}^*\right\}$, given the priors of the parameters as,

$$\begin{aligned} \Omega_{m0} &\sim \mathcal{U}(0.1, 0.9) \\ H_0 &\sim \mathcal{U}(50, 90) \\ \omega_{DE} &\sim \mathcal{U}(-1.8, -0.4) \\ \omega_0 &\sim \mathcal{U}(-1.9, -0.4) \\ \omega_a &\sim \mathcal{U}(-4.0, 4.0) \end{aligned}$$

$\boldsymbol{z}^*$ has 1468 elements evenly located in the interval, $[0.8 \min\left(\boldsymbol{z}_{obs}\right), 1.2 \max\left(\boldsymbol{z}_{obs}\right)]$. The $12800 \times 3$ samples

## 2.7 Conclusion

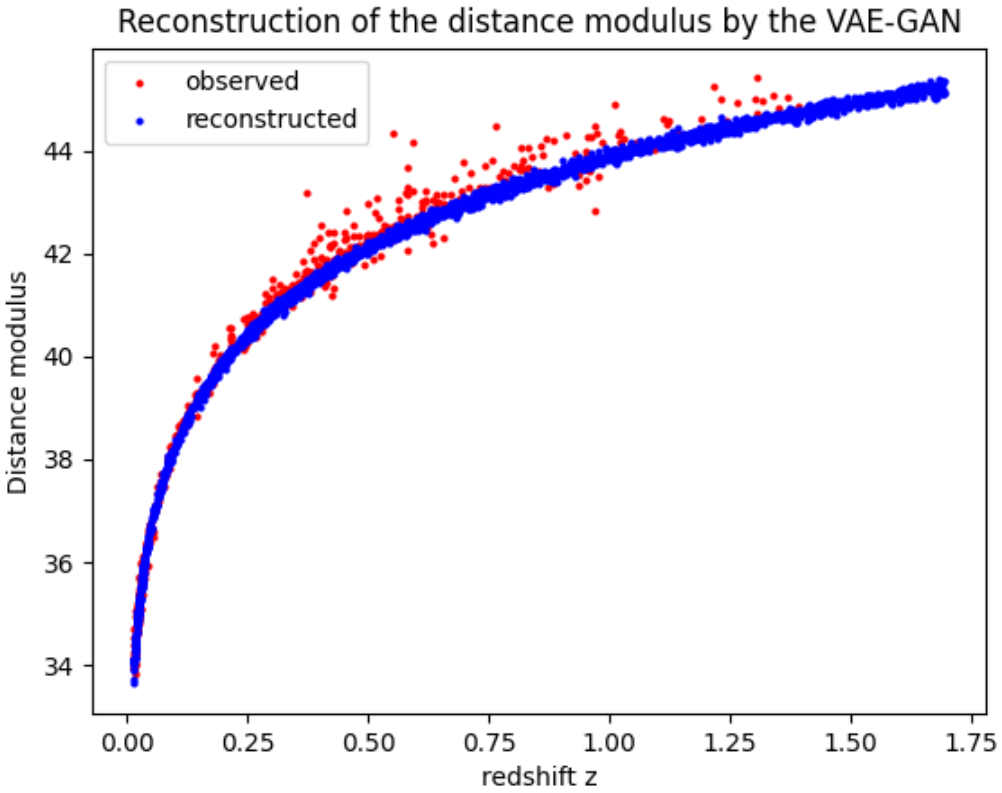


Figure 2.4: Reconstruction

# References


[1] A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner et al. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The Astronomical Journal* 116, (1998) 1009.

[2] S. Perlmutter, G. Aldering, G. Goldhaber, R. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom et al. Measurements of $\Omega$ and $\Lambda$ from 42 high-redshift supernovae. *The Astrophysical Journal* 517, (1999) 565.

[3] B. E. Schaefer. The Hubble diagram to redshift$>$ 6 from 69 gamma-ray bursts. *The Astrophysical Journal* 660, (2007) 16.

[4] L.-X. Li. Variation of the Amati relation with cosmological redshift: a selection effect or an evolution effect? *Monthly Notices of the Royal Astronomical Society: Letters* 379, (2007) L55–L59.

[5] S. Basilakos and L. Perivolaropoulos. Testing gamma-ray bursts as standard candles. *Monthly Notices of the Royal Astronomical Society* 391, (2008) 411–419.

[6] F.-Y. Wang, S. Qi, and Z.-G. Dai. The updated luminosity correlations of gamma-ray bursts and cosmological implications. *Monthly Notices of the Royal Astronomical Society* 415, (2011) 3423–3433.

[7] L. Tang, X. Li, H.-N. Lin, and L. Liu. Model-independently calibrating the luminosity correlations of gamma-ray bursts using deep learning. *The Astrophysical Journal* 907, (2021) 121.

[8] D. M. Scolnic, D. Jones, A. Rest, Y. Pan, R. Chornock, R. Foley, M. Huber, R. Kessler, G. Narayan, A. Riess et al. The complete light-curve sample of spectroscopically confirmed SNe Ia from Pan-STARRS1 and cosmological constraints from the combined pantheon sample. *The Astrophysical Journal* 859, (2018) 101.

[9] N. Suzuki, D. Rubin, C. Lidman, G. Aldering, R. Amanullah, K. Barbary, L. Barrientos, J. Botyanszki, M. Brodwin, N. Connolly et al. The Hubble Space Telescope cluster supernova survey. V. Improving the dark-energy constraints above z$>$ 1 and building an early-type-hosted supernova sample. *The Astrophysical Journal* 746, (2012) 85.

[10] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Per-

rot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, (2011) 2825–2830.

[11] G. D'Agostini. Fits, and especially linear fits, with errors on both axes, extra variance of the data points and other complications. *arXiv preprint physics/0511182* .

[12] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman. emcee: The MCMC Hammer. *PASP* 125, (2013) 306–312.

[13] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems 2015. Software available from tensorflow.org.

# Chapter 3

# Photometric redshift estimation using Symbolic Regression

## 3.1 Introduction

Large scale structure cosmology and extragalactic astronomy rely heavily on accurate estimate of the redshift of objects under study. For example the reconstruction of the two point correlation function for galaxies is critical to understand the history of structure formation in the Universe and probe theories beyond ΛCDM. Unfortunately it is a very time consuming and expensive task to obtain spectroscopic data for the millions of observed galaxies. It has therefore long been a challenge to estimate the redshift of galaxies using the much easier to obtain photometric data.



## 3.2 Literature Survey

The estimation of redshifts from photometric data has been an industry for some years in astronomy culminating in the production of MegaZ-LRG [**?**], a database of photometric redshifts of 1 million luminous red galaxies in the range $0.4 < z < 0.7$ and the 2MPZ database[**?**] for $z < 0.3$. The aim is to model the spectroscopic redshift using photometric redshift estimator, $z_{phot}$(u, g, r, i, z), where u, g, r, i, z are the standard photometric magnitudes. There have been two main approaches to the problem: template based methods ([**?**] [**?**] [**?**] [**?**] [**?**] [**?**] [**?**] [**?**] [**?**] [**?**]) and machine learning/empirical methods ([**?**] [**?**] [**?**] [**?**] [**?**] [**?**] [**?**]). For comparisons of the various codes see ([**?**] [**?**] [**?**]). One of the best performing codes is ANNz[**?**] which is based on artificial neural networks and was used in creating the MegaZ and 2MPZ databases.[**?**] proposes non-linear regression with genetic optimization.

## 3.3 Observation Data

The data in this study are drawn from SDSS Data Release 17 [**?**]. The SDSS I-III uses a 4 meter telescope at Apache Point Observatory in New Mexico and has CCD wide field photometry in 5 bands (u, g, r, i, z [**?**] [**?**]), and an expansive spectroscopic follow up program [**?**] covering $\pi$ radians of the northern sky. The SDSS collaboration has obtained approximately 2 million galaxy spectra using

dual fiberfed spectrographs. An automated photometric pipeline performed object classification to a magnitude of $r \approx 22$ and measured photometric properties of more than 100 million galaxies. The complete data sample, and many derived catalogs such as the photometric redshift estimates, are publicly available through the CasJobs server[**?**] [1].

### 3.3.1 SDSS DR17 photometry

The SDSS is well suited to the analysis presented in this paper due to the enormous number of photometrically selected galaxies with spectroscopic redshifts to use as training, cross-validation and test samples. We select 1,958,727 galaxies from CasJobs with both spectroscopic redshifts and photometric properties. In detail we run the following MySQL query in the DR17 schema:

```
-- Goto http://skyserver.sdss.org/casjobs/, create an account run the following sql query
-- http://skyserver.sdss.org/dr17/SearchTools/sql cannot be used to bulk data (only 500000)
-- SQL qeury
-- =========
-- only select galaxies that have a photometric galaxy classification type = 3,
-- and spectscopic redshifts, r band magnitudes, -- and radii greater than 0
-- make a magnitude error cut of < 0.3 (in all 5 bands) to ensure that you don't get junk obje
-- dered_ is simplified mag, corrected for extinction: modelMag - extinction



SELECT
    q.dered_u as u, q.dered_g as g, q.dered_r as r,
    q.dered_i as i, q.dered_z as z, q.modelMagErr_u as u_err,
    q.modelMagErr_g as g_err, q.modelMagErr_r as r_err,
    q.modelMagErr_i as i_err, q.modelMagErr_z as z_err,
    s.z AS specz, s.zerr AS specz_err,
    p.z AS photoz, p.zerr AS photoz_err
INTO mydb.specPhotoDR10v2 FROM
SpecPhotoAll AS s JOIN photoObjAll AS q ON s.objid=q.objid
AND q.dered_u>0
AND q.dered_g>0
AND q.dered_r>0
AND q.dered_z>0
AND q.dered_i>0
AND q.expAB_r>0
AND q.modelMagErr_u < 0.3
AND q.modelMagErr_g < 0.3
AND q.modelMagErr_r < 0.3
AND q.modelMagErr_i < 0.3
AND q.modelMagErr_z < 0.3
```

[1] http://skyserver.sdss.org/casjobs/

```
AND q.type=3
AND s.z > 0
--AND s.zerr > -0.3 AND s.zerr < 0.3
--AND q.petroRad_u > 0 -- has no effect
--AND q.petroRad_g > 0
--AND q.petroRad_r > 0
--AND q.petroRad_i > 0
--AND q.petroRad_z > 0
AND q.CLEAN=1  -- Clean photometry flag
-- (1=clean, 0=unclean)
AND s.zWarning = 0  -- Bitmask of warning
-- vaules; 0 means all
-- is well
LEFT OUTER JOIN Photoz AS p ON s.objid=p.objid
```

We apply the SDSS extinction corrections to the psf and fiber magnitudes, and further only select galaxies that have a photometric galaxy classification type = 3, have spectroscopic redshifts, r band magnitudes, and radii greater than zero. This reduces the sample size to 1,922,231 galaxies.



## 3.4 Methodology

### 3.4.1 Symbolic Regression

Symbolic regression (SR) is a novel machine-learning technique that approximates the relation between an input and an output through analytic mathematical formulae ([?] [?] [?] [?] [?] [?] [?]). The advantage of using SR over other ML regression models like RF or deep neural networks is that it provides analytic expressions that can be readily generalized and that facilitate understanding the underlying physics. Furthermore, SR is shown to outperform other ML models when the size of dataset is small[?].

## 3.5 Photometric redshift estimation

For our symbolic regression we rely on PySR[?]. It uses genetic programming to find a symbolic expression for a numerically defined function in terms of pre-defined variables. The population consists of symbolic expressions, visualized as a tree and consisting of nodes with an operator function or an operand. We use the operators for addition, subtraction, multiplication. The tree population evolves when new individuals are created and old ones are discarded. To breed the next generation, several mutation operators can be applied, for instance exchanging, adding or deleting nodes of the parent tree. The hyperparameter populations = 30 defines the number of populations and is per default set to the number of processors used (procs). The number of individuals per populations is given by npop = 1000. As the figure of merit for the PySR algorithm we take the mean squared

error between the data points $t_i(x, z|\theta)$ and the functional description $g_i$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(g_i\left(x\right) - t_i(x, z|\theta)\right) \tag{3.1}$$

and balance it with the function's complexity, defined as

$$complexity = \#nodes \tag{3.2}$$

For the PySR score value, not to be confused with the statistics version of the optimal observable defined in, the parameter parsimony balances the two conditions,

$$score = \frac{MSE}{baseline} + parsimony \times complexity \tag{3.3}$$

The normalization factor baseline is the MSE between the data and the constant unit function. The hyperparameter *maxsize* restricts the complexity to a maximum value. We adjust this value depending on the difficulty of the regression task taking 50 as a starting point and increase (decrease) it if the required complexity is larger (smaller). Additionally we can restrict the complexity of specific operators to obtain a more readable result. We set the maximal complexity of square to 5 and cube to 3. Note that in some instances we choose to not extract the score, but the score scaled by a constant, to improve the numerics with an order-one function. Simulated annealing allows us to search for a global optimum in a high-dimensional space while preventing the algorithm from being stuck in a local optimum. A mutation is accepted with the probability

$$p = exp\left(-\frac{score_{new} - score_{old}}{alpha \times T}\right) \tag{3.4}$$

The parameter T is referred to as temperature. It linearly decreases with each cycle or generation, starting with 1 in the first cycle and 0 in the last. The hyperparameter ncyclesperiterations = 200 sets the amount of cycles. We choose alpha = 1. If the new function describes the data better than the reference tree, $score_{new}$ $score_{old}$, the exponent has a positive sign and the new function is accepted. If the new sore is larger than the old score, the acceptance of the new function is given by p and hence exponentially suppressed. We use this default PySR form for our simple example and discuss a bettersuited form for our application in Sec. 3. The hyperparameter niterations = 300 defines the number of iterations of a full simulated annealing process. After each iteration the best formulas are compared to the hall of fame (HoF). For each complexity the best equation is chosen and saved in the output file. An equation of higher complexity is only added if its MSE is smaller than for previous formulas. Equations from different populations or the hall of fame can migrate to other populations. This process is affected by the parameters fractionReplaced = 0.5 and fractionReplacedHof = 0.2.

## 3.6 Conclusion

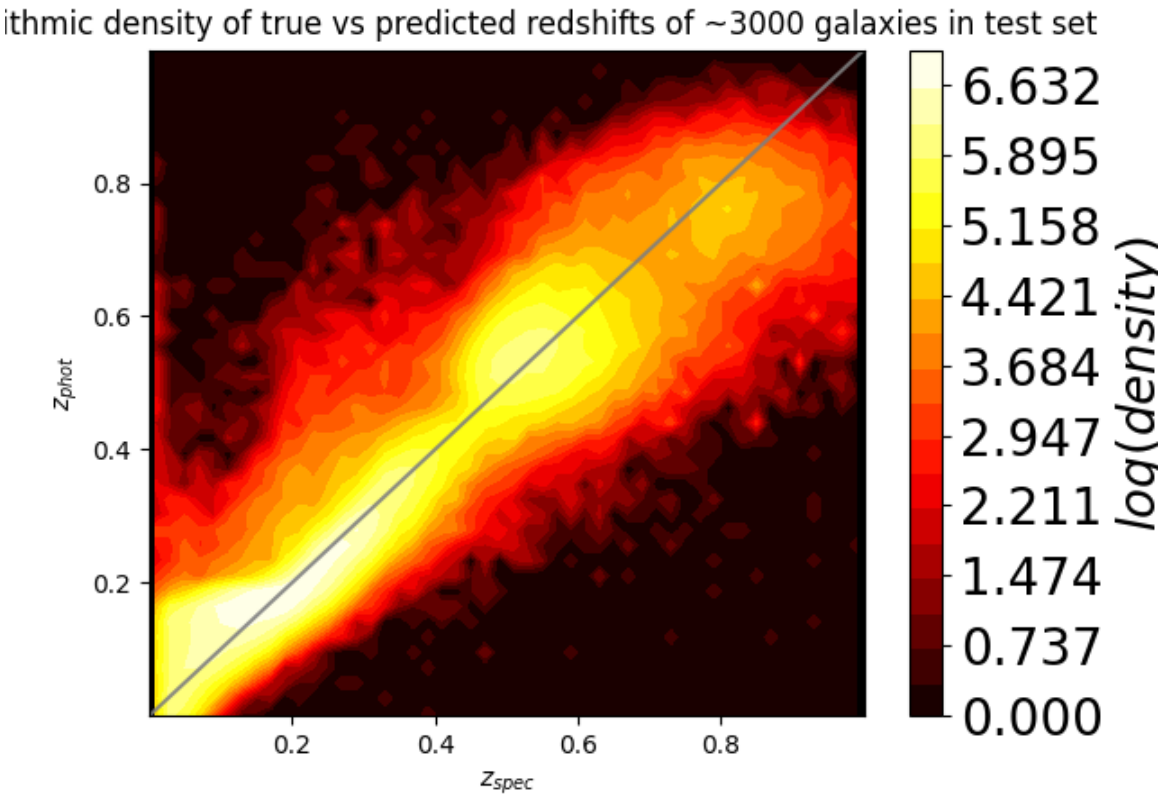


Figure 3.1: Photometric redshift prediction and errorbars for a representative subsample of 300 galaxies. The errorbars are due to errors in the photometric data and so depend on the particular model chosen for $z_{phot}$

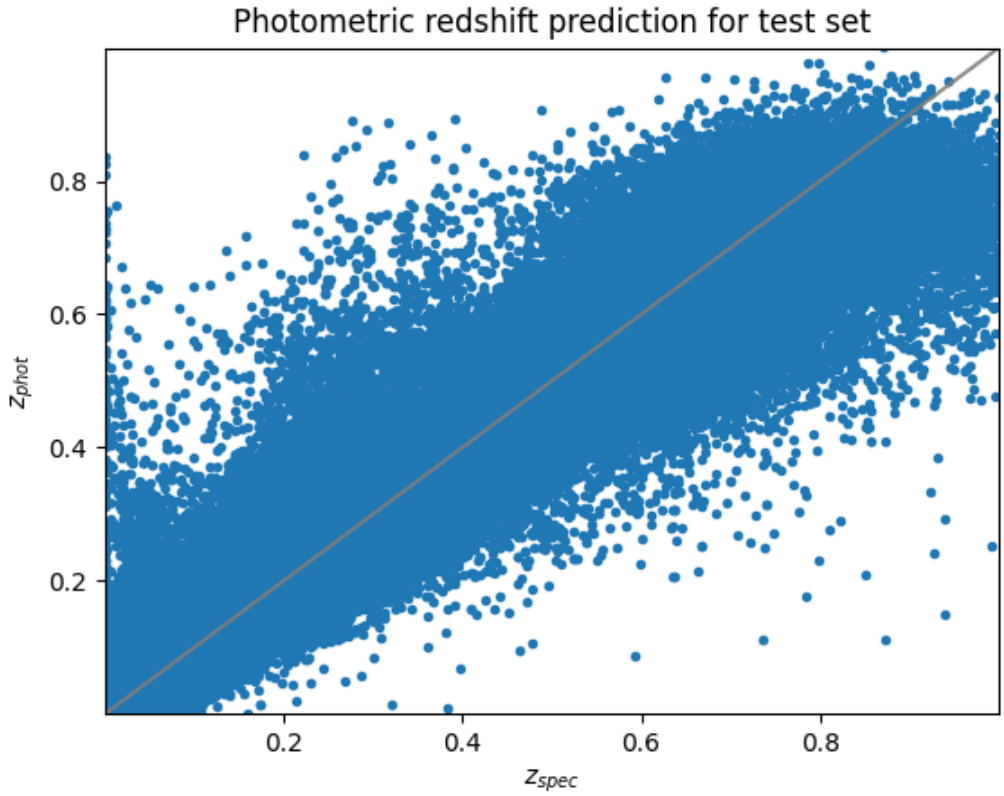


Figure 3.2: Predictions for test data

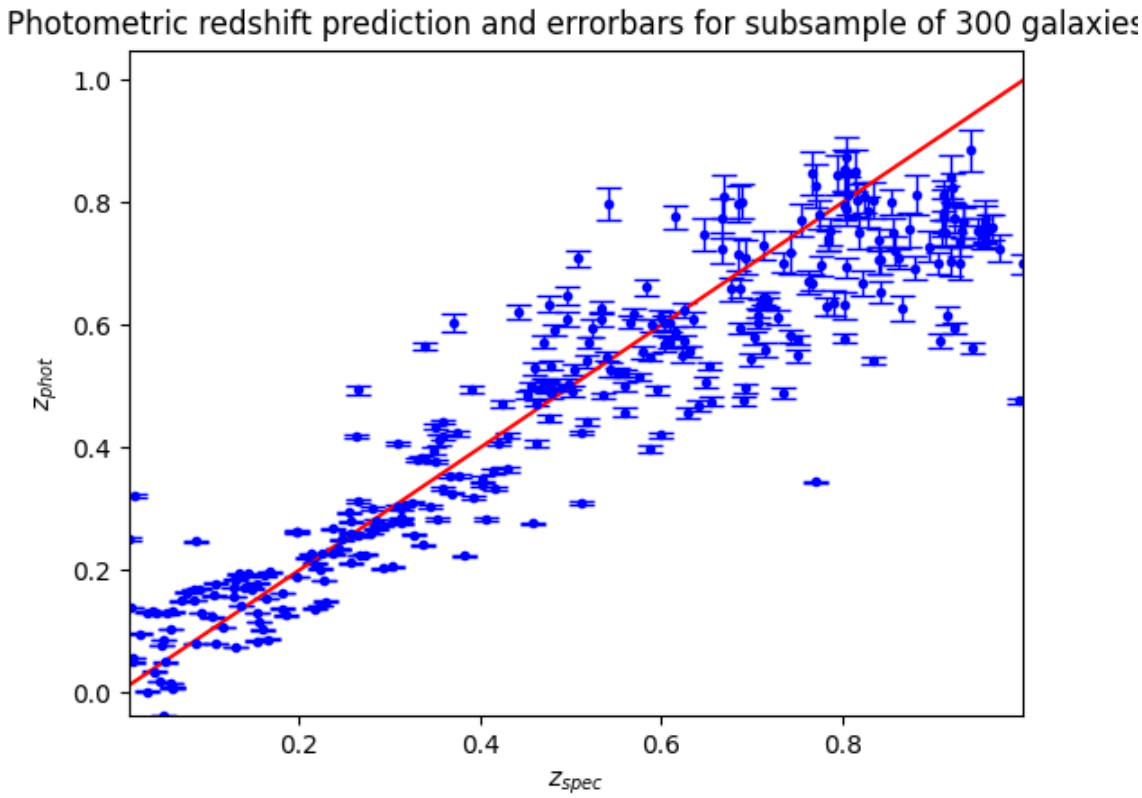


Figure 3.3: Photometric redshift prediction and errorbars for a representative subsample of 300 galaxies. The errorbars are due to errors in the photometric data and so depend on the particular model chosen for $z_{phot}$

# References


[1] A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner et al. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The Astronomical Journal* 116, (1998) 1009.

[2] S. Perlmutter, G. Aldering, G. Goldhaber, R. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom et al. Measurements of $\Omega$ and $\Lambda$ from 42 high-redshift supernovae. *The Astrophysical Journal* 517, (1999) 565.

[3] B. E. Schaefer. The Hubble diagram to redshift$>$ 6 from 69 gamma-ray bursts. *The Astrophysical Journal* 660, (2007) 16.

[4] L.-X. Li. Variation of the Amati relation with cosmological redshift: a selection effect or an evolution effect? *Monthly Notices of the Royal Astronomical Society: Letters* 379, (2007) L55–L59.

[5] S. Basilakos and L. Perivolaropoulos. Testing gamma-ray bursts as standard candles. *Monthly Notices of the Royal Astronomical Society* 391, (2008) 411–419.

[6] F.-Y. Wang, S. Qi, and Z.-G. Dai. The updated luminosity correlations of gamma-ray bursts and cosmological implications. *Monthly Notices of the Royal Astronomical Society* 415, (2011) 3423–3433.

[7] L. Tang, X. Li, H.-N. Lin, and L. Liu. Model-independently calibrating the luminosity correlations of gamma-ray bursts using deep learning. *The Astrophysical Journal* 907, (2021) 121.

[8] D. M. Scolnic, D. Jones, A. Rest, Y. Pan, R. Chornock, R. Foley, M. Huber, R. Kessler, G. Narayan, A. Riess et al. The complete light-curve sample of spectroscopically confirmed SNe Ia from Pan-STARRS1 and cosmological constraints from the combined pantheon sample. *The Astrophysical Journal* 859, (2018) 101.

[9] N. Suzuki, D. Rubin, C. Lidman, G. Aldering, R. Amanullah, K. Barbary, L. Barrientos, J. Botyanszki, M. Brodwin, N. Connolly et al. The Hubble Space Telescope cluster supernova survey. V. Improving the dark-energy constraints above z$>$ 1 and building an early-type-hosted supernova sample. *The Astrophysical Journal* 746, (2012) 85.

[10] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Per-

rot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, (2011) 2825–2830.

[11] G. D'Agostini. Fits, and especially linear fits, with errors on both axes, extra variance of the data points and other complications. *arXiv preprint physics/0511182* .

[12] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman. emcee: The MCMC Hammer. *PASP* 125, (2013) 306–312.

[13] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems 2015. Software available from tensorflow.org.