Cosmological Reconstructions with Neural Networks: From Data to Theory

Kostas Dialektopoulos
kdialekt@gmail.com

L-Università ta' Malta

NEB-21, Corfu
September 2, 2025

Outline

Cosmological tensions

- H_0 tension: $\sim 5\sigma$
- S_8 tension: $\sim 3\sigma$
- CMB anisotropy anomalies
- Cosmic dipoles
- Missing satellites
- Core-cusp problem
- Baryonic Tully-Fisher relation

What is a Reconstruction in Cosmology?

What?

Helps distinguish between different cosmological models (e.g., ΛCDM vs. alternatives)

How?

Two main approaches:

- **Parametric**: Assumes a functional form, reducing uncertainty.
- **Non-parametric**: Data-driven, avoiding model biases.

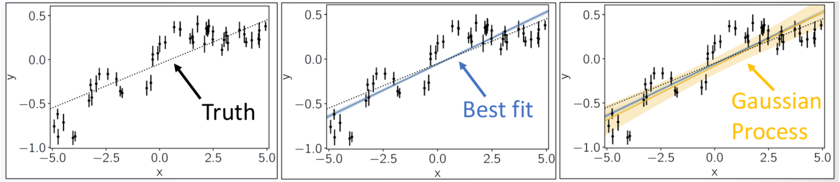
Parametric Reconstructions

- Phenomenological
  1. CPL: $w(z) = w_0 + w_a \frac{z}{1+z}$
  2. JBP: $w(z) = w_0 + w_a \frac{z}{(1+z)^2}$
  3. DM/DE interaction: $Q = \xi H \rho_{\mathrm{DM}}$
- Physically motivated
  1. Quintessence: $w(z) = \frac{\dot{\phi}^2/2 - V(\phi)}{\dot{\phi}^2/2 + V(\phi)}$
  2. Horndeski
  3. DM models: Warm, Self-interacting
- Model-independent
  1. Basis representation parametrizations: Taylor, Fourier, Padé
  2. Interpolation methods
  3. Cosmography

Non-Parametric Reconstructions

- **Gaussian Processes (GPs)**: Bayesian interpolation of cosmological functions.
- **Artificial Neural Networks (ANNs)**: Machine learning-based reconstructions.
- **Iterative Smoothing Method**: Refines data-driven reconstructions iteratively.
- **LOESS + SIMEX**: Non-parametric regression for trend analysis.

What are Gaussian processes?



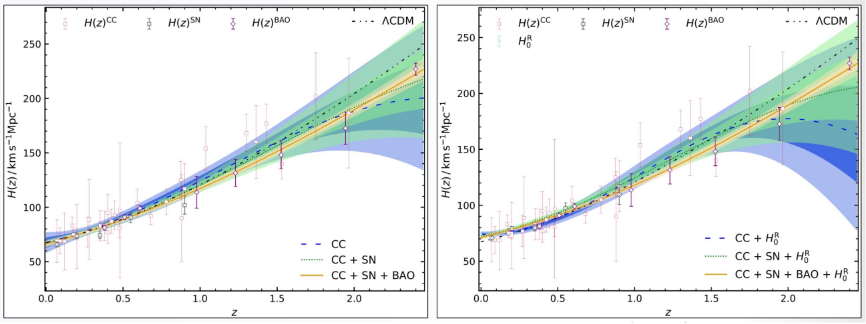
Definition: A GP is a stochastic (random) process where any finite subset is a **multivariant Gaussian distribution** with mean $\mu(x)$ and covariance $k(x, x')$.

Setting each $\mu(x)$ to zero, the **covariance function** can be used to learn the behavior that produced the data points.

Gaussian Process Regression

- The covariance function contains **non-physical hyperparameters** θ which define the distribution $k(\theta, x, x')$.
- Iterating over these values using Bayesian inference (or others) can produce better hyperparameters.
- The result is a **model independent reconstruction** (in physics) of the behabior of some parameter.
- This is superior to regular fitting because it is nonparametric and so **assumes no physical model** whatsoever.

Squared Exponential H_0 GP (GaPP code: Seikel et al. 2012)



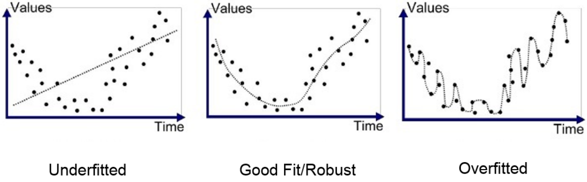
$H_0 = 67.539 \pm 4.772 \mathrm{km/s/Mpc}$ $H_0 = 73.782 \pm 1.374 \mathrm{km/s/Mpc}$

$H_0 = 67.001 \pm 1.653 \mathrm{km/s/Mpc}$ $H_0 = 72.022 \pm 1.076 \mathrm{km/s/Mpc}$

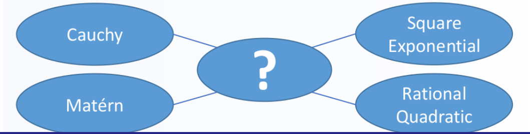
$H_0 = 66.197 \pm 1.464 \mathrm{km/s/Mpc}$ $H_0 = 71.180 \pm 1.025 \mathrm{km/s/Mpc}$

Open problems with GP reconstructions

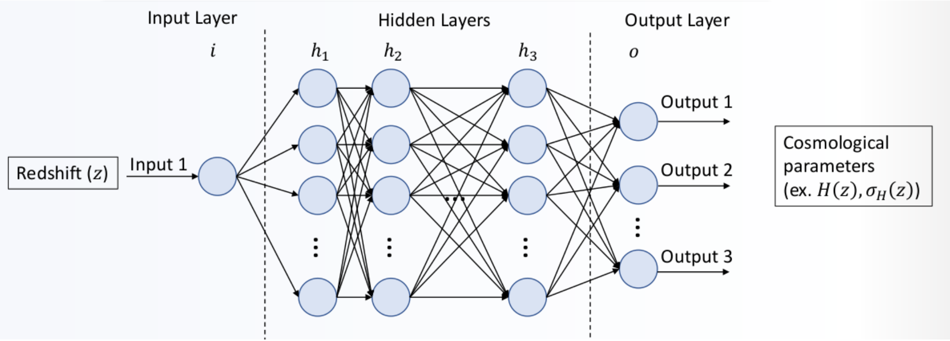
- **Overfitting**: GP is very prone to overfitting for small data sets, which is especially pronounced at the origin, i.e. Hubble constant



- **Kernel Selection Problem**: There is no natural kernel for cosmology

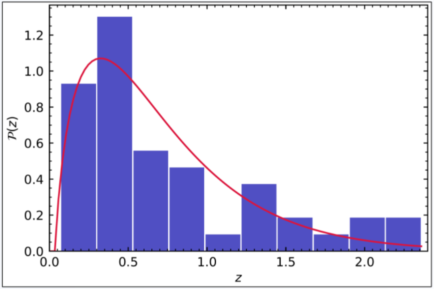


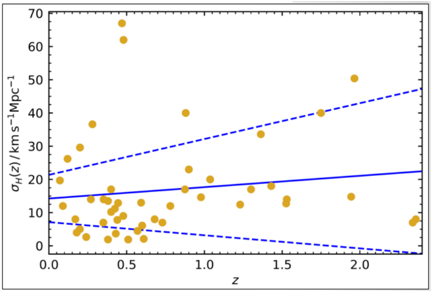
Artificial Neural Networks (ANN)



ReFANN code from Wang et al. (2020)

Training data for the ANN





$$P(z,\alpha,\lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} z^{\alpha-1} e^{-\lambda z}$$

Mean: $\sigma_H = 14.25 + 3.42z$
Upper error: $\sigma_H = 21.37 + 10.79z$
Lower error: $\sigma_H = 7.14 - 3.95z$

Designing the ANN

- **Risk**: Optimizes the **number of hidden layers and neurons** in an ANN

$$\text{risk} = \sum_{i=1}^{N} (\text{Bias}_i^2 + \text{Variance}_i) = \sum_{i=1}^{N} \left([H_{\text{obs}}(z_i) - H_{\text{pred}}(z_i)]^2 + \sigma_H^2(z_i) \right)$$

- **Loss**: Balances the **number of iterations** a system needs to predict the observational data

1. Least absolute deviation (**L1**)

$$L1 = \sum_{i=1}^{N} |H_{\text{obs}}(z_i) - H_{\text{pred}}(z_i)|$$

2. Smoothed L1 (**SL1**)
3. Mean Square Error (**MSE**)

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (H_{\text{obs}}(z_i) - H_{\text{pred}}(z_i))^2$$

Designing the ANN

What we use here

$$\mathrm{L}_{\chi^2} = \sum_{i,j} \left[m_{\rm obs}(z_i) - m_{\rm pred}(z_i)\right]^{\mathsf{T}} \mathcal{C}_{ij}^{-1} \left[m_{\rm obs}(z_j) - m_{\rm pred}(z_j)\right] \, ,$$

where $\mathcal{C}_{ij}$ is the total noise covariance matrix of the data, which includes the statistical noise and systematics.

Building the ANN (**KD**, Levi Said et al. '21)

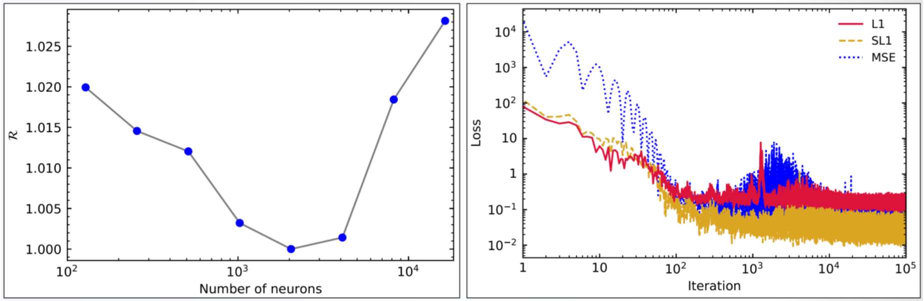


Figure: **Left:** Risk function for **one layer** (number of neurons 2^n, $n \in 7, ..., 14$),
Right: Evolution of L1, SL1 and MSE loss functions

Using the ANN (KD, Mukherjee et al. '23)

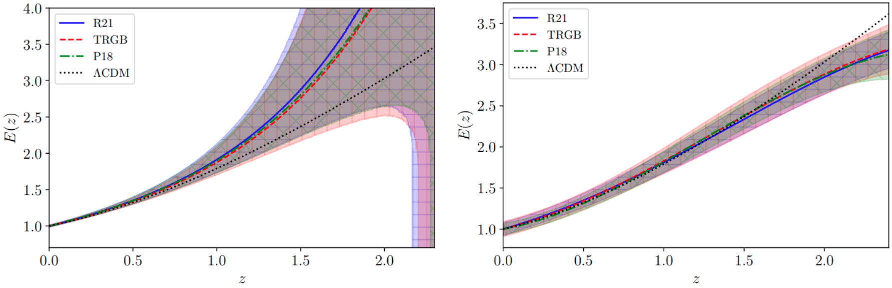


Figure: Reconstructed reduced Hubble parameter from the (i) Pantheon SN compilation (left) and (ii) combined CC+BAO Hubble data set (right), using ANNs.

Om diagnostics (Sahni, Shafieloo, Starobinsky '08) (**KD**, Mifsud et al. '21)

Distinguish ΛCDM from alternative dark energy and modified gravity models:

$$Om(z) = \frac{E^2(z) - 1}{(1+z)^3 - 1} .$$

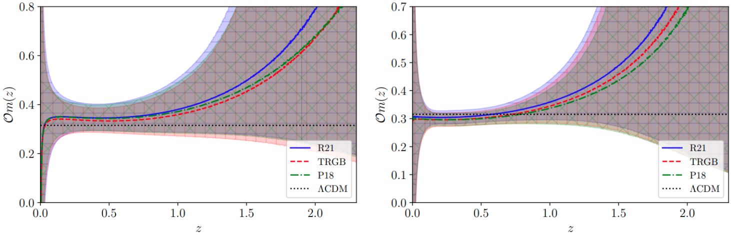


Figure: Reconstructed *Om* diagnostics using (i) ANNs (left) and (ii) GPs (right) from the Pantheon SN data for three different priors.

H0 diagnostics (Krishnan, Colgáin, Sheikh-Jabbari, Yang '20) (**KD**, Mifsud et al. '21)

It is defined as

$$\mathbf{H0} = \frac{H(z)}{\sqrt{\Omega_{m0}(1+z)^3 + 1 - \Omega_{m0}}},$$

and its non-constancy suggests evidence for new physics beyond ΛCDM.

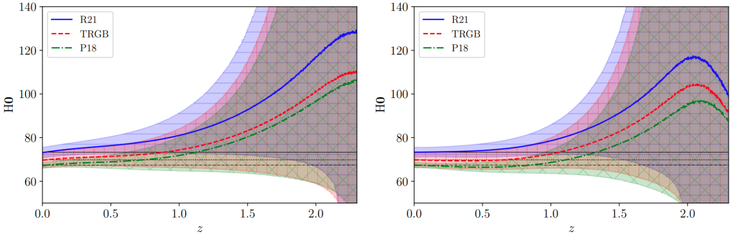


Figure: Reconstructed H_0 diagnostics using (i) ANNs (left) and (ii) GPs (right) from the Pantheon SN data for three different priors.

Constraining theories Arjona, Cardona, Nesseris '19

Example: Horndeski mapping:

$$G_2 = K(X),\ G_3 = G(X),\ G_4 = 1/2, \text{ and } G_5 = 0\,,$$

The action is given by:

$$S = \int \mathrm{d}^4x\sqrt{-g}\left(\frac{R}{2} - K(X) - G(X)\Box\phi\right) + S_{\rm mat}(\psi, g_{\mu\nu})\,.$$

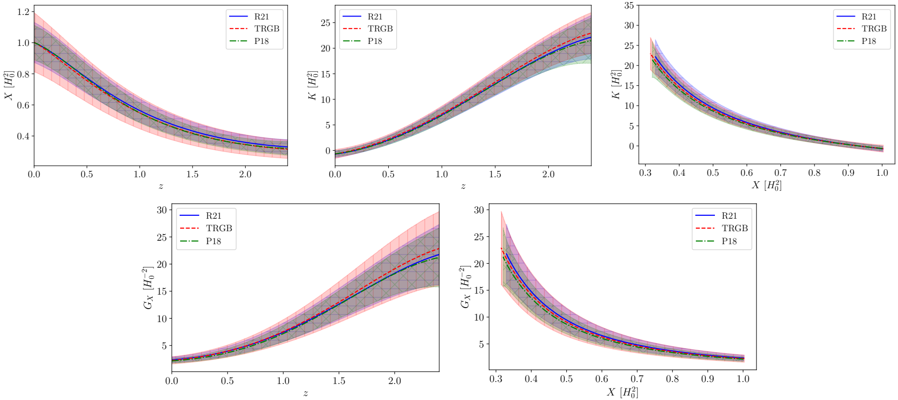
Cosmological equations (flat FLRW):

$$K(X) = -3H_0^2\left(1-\Omega_{m0}\right) + \frac{\mathcal{J}\sqrt{2X}H^2(X)}{H_0^2\Omega_{m0}} - \frac{\mathcal{J}\sqrt{2X}\left(1-\Omega_{m0}\right)}{\Omega_{m0}}\,,$$

and

$$G_X(X) = -\frac{2\mathcal{J}H'(X)}{3H_0^2\Omega_{m0}}\,.$$

Constraining theories of gravity



(KFD, Mukherjee, Levi Said, Mifsud '23)

We can also compute the DE EoS as

$$w_\phi = \frac{-K + \sqrt{2X}\dot{X}G_X}{K - 2X(K_X + 3\sqrt{2X}HG_X)}$$

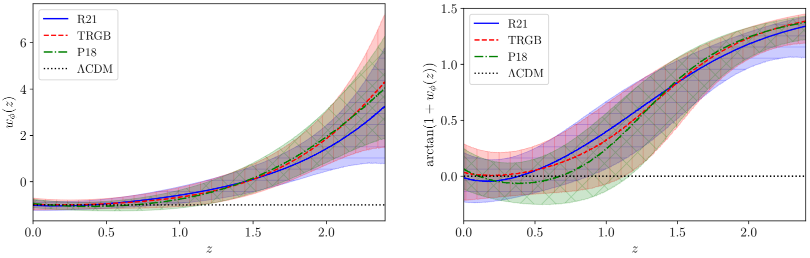


Figure: Plots for dark energy EoS $w_\phi(z)$ (left) and its compactified form $\arctan(1 + w_\phi(z))$ (right) considering R21, TRGB, and P18 H_0 priors. The shaded regions with '−', '|' and '×' hatches represent the 1σ confidence levels for the above priors respectively.

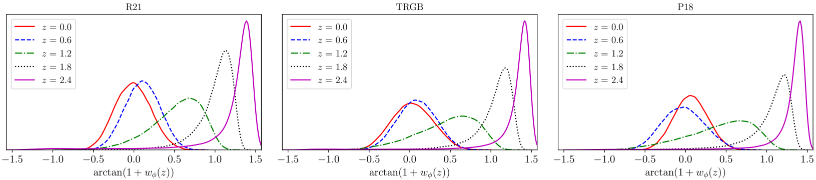


Figure: Plots showing the posteriors of probability distribution of the compactified dark energy EoS for the theory at some sample redshifts for the R21, TRGB, and P18 H_0 priors, respectively.

Observational Datasets

- **Pantheon+**: SNIa observations from 1701 light curves that represent 1550 distinct SNIa spanning the redshift range $z < 2.3$.
- **CC**: 32 $H(z)$ measurements, along with the full covariance matrix that includes systematic and calibration errors, as reported in *Moresco, et al, Astrop. J. 2020*.

ANN Training and Validation

1. We split the Pantheon+ dataset into training (70%) and validation (30%) sets and we train the network.
2. To incorporate the covariance matrix of the dataset, we minimize the χ^2 loss function.
3. The optimal network is one with two hidden layers and 128 neurons each.
4. The optimal network architecture is iterated over 500 times for random initialization of the hyperparameters along with the dropout effect. Out of these 500 samples, we compute the mean function and the respective uncertainties.

In a spatially flat Friedmann-Lemaître-Robertson-Walker universe, the luminosity distance is related to the Hubble parameter $H(z)$ at some redshift z, as,

$$d_L(z) = c(1+z)\int_0^z \frac{d\tilde{z}}{H(\tilde{z})}.$$

The observed luminosity of SNIa, from a specific redshift, is related to the apparent peak magnitude m via the following relation, independent of any physical model as,

$$m(z) - M_B = 5\log_{10}\left[\frac{d_L(z)}{1\ \mathrm{Mpc}}\right] + 25\,.$$

We can rewrite the luminosity distance as,

$$d_L(z) = 10^{\frac{1}{5}[m(z)-M_B-25]}.$$

and we can compute d_L', the first order derivative of d_L with respect to the redshift z as,

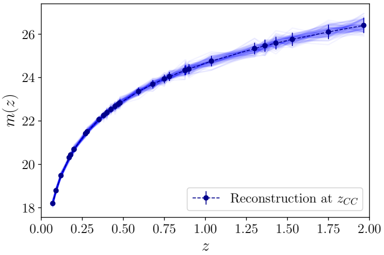
$$d_L'(z) = \frac{\log(10)}{5} 10^{-\frac{M_B}{5}} m'(z).$$

Combining them, we can express the Hubble parameter as,

$$H(z) = \frac{c(1+z)^2}{(1+z)d_L'(z) - d_L(z)}.$$

In this way, we can derive the Hubble parameter $H(z)$, from the Pantheon+ apparent magnitudes m and its corresponding derivatives m' employing specific values of M_B.

Constraints on M_B



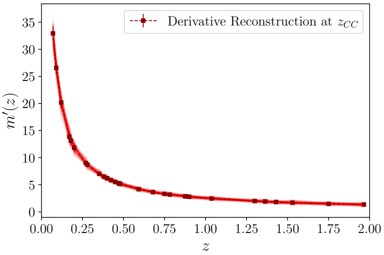


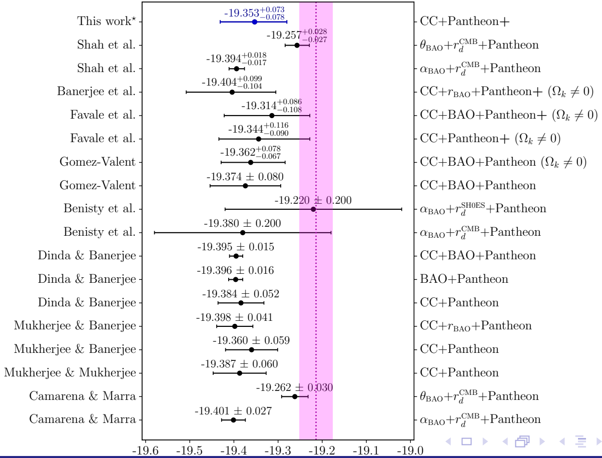
Figure: ANN reconstruction of the Pantheon+ apparent magnitudes $m(z)$ (left panel) and its corresponding derivatives $m'(z)$ [right panel] at the CC redshifts (z_{CC}).

Constraints on M_B [Mukherjee, KD, Said, Mifsud, JCAP 2024]

Reference	Methodology	Datasets	M_B
Camarena & Marra[144]	Cosmography	$\alpha_{\rm BAO} + r_d^{\rm CMB}$ + Pantheon	-19.401 ± 0.027
		$\theta_{\rm BAO} + r_d^{\rm CMB}$ + Pantheon	-19.262 ± 0.030
Mukherjee & Mukherjee[118]	**Gaussian Process**	**CC + Pantheon**	$\mathbf{-19.387 \pm 0.060}$
Mukherjee & Banerjee[61]	**Gaussian Process**	**CC + Pantheon**	$\mathbf{-19.360 \pm 0.059}$
		CC + $r_{\rm BAO}$ + Pantheon	-19.398 ± 0.041
Dinda & Banerjee[115]	**Gaussian Process**	**CC + Pantheon**	$\mathbf{-19.384 \pm 0.052}$
		BAO + Pantheon	-19.396 ± 0.016
		CC + BAO + Pantheon	-19.395 ± 0.015
Benisty *et al.*[68]	Neural Networks	$\alpha_{\rm BAO} + r_d^{\rm CMB}$ + Pantheon	-19.38 ± 0.20
		$\alpha_{\rm BAO} + r_d^{\rm SH0ES}$ + Pantheon	-19.22 ± 0.20
Gómez-Valent[145]	Index of Inconsistency	CC + BAO + Pantheon	-19.374 ± 0.080
		CC + BAO + Pantheon ($\Omega_k \neq 0$)	$-19.362^{+0.078}_{-0.067}$
Favale *et al.*[122]	Gaussian Process	CC + Pantheon+	$-19.344^{+0.116}_{-0.090}$
		CC + BAO + Pantheon+	$-19.314^{+0.086}_{-0.108}$
		CC + SH0ES + Pantheon+	$-19.252^{+0.024}_{-0.036}$
		CC + BAO + SH0ES + Pantheon+	$-19.252^{+0.024}_{-0.036}$
Banerjee *et al.*[119]	Gaussian Process	CC + $r_{\rm BAO}$ + Pantheon+ ($\Omega_k \neq 0$)	$-19.404^{+0.099}_{-0.104}$
Shah *et al.*[111]	Neural Networks	Pantheon + $\alpha_{\rm BAO} + r_d^{\rm CMB}$	$-19.394^{+0.018}_{-0.017}$
		Pantheon + $\theta_{\rm BAO} + r_d^{\rm CMB}$	$-19.257^{+0.028}_{-0.027}$
This work*	**Neural Networks**	**CC & Pantheon+**	$\mathbf{-19.353^{+0.073}_{-0.078}}$

Figure: Comparison between the model-independent constraints on M_B obtained in this work vs those present in the literature.

Constraints on M_B [Mukherjee, KD, Said, Mifsud, JCAP 2024]



Cumulative binning method/Redshift layer binning with $\bar{z}$

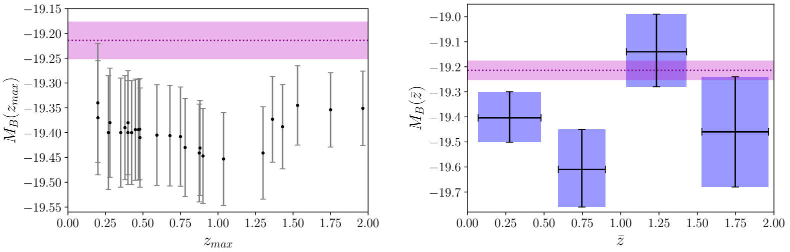


Figure: Predictions of the supernovae absolute magnitudes: $M_B(z_{\rm max})$ by adopting cumulative binning, where $M_B(z_{\rm max})$ is the derived value of M_B by considering CC $H(z)$ data up to $z_{\rm max}$ (left panel), and $M_B(\bar{z})$ by adopting the redshift layer binning, where $M_B(\bar{z})$ is the derived value of M_B by considering CC $H(z)$ data within that redshift layer with a mean redshift $\bar{z}$ (right panel). The purple region corresponds to the $1-\sigma$ model independent constraint $M_B = -19.214 \pm 0.037$, as inferred from the Pantheon and SH0ES data sets.

Calibration of GRBs [Mukherjee, Dainotti, KD, Said, Mifsud 2024]

- Model-independent calibration of the 2D and 3D Dainotti relations using an MCMC approach driven by ANNs.
- **Key Findings:**
  - ANN-based calibration significantly reduces the scatter in the GRB calibration relations compared to GPs.
  - The method avoids issues like kernel function dependence and overfitting in low redshift.
  - Improved calibration of GRBs as high-redshift distance indicators.
- **Implications:**
  - GRBs can be used as reliable tools for extending the cosmic distance ladder.
  - The method offers a robust alternative for addressing the current tensions in cosmological parameter estimations.

Take-home messages

- **Cosmological tensions are there!** Either them being systematics or new physics, we have to acknowledge them and try to resolve them.
- **Reconstruction techniques help probe cosmic evolution** without assuming a specific model, making them crucial for testing ΛCDM and alternative theories.

Take-home messages

- **Parametric methods provide controlled, interpretable models**, while **non-parametric methods offer flexibility** to extract insights directly from data.
- **Different reconstruction approaches (Gaussian Processes, Neural Networks, Interpolation, etc.)** allow us to cross-check results and uncover hidden trends in cosmic expansion and structure growth.
- By **improving** reconstruction techniques and **using** new datasets (e.g., DESI, JWST, Euclid), we can **refine** our understanding of dark energy, the Hubble tension, and fundamental physics.

Take-home messages

- **Parametric methods provide controlled, interpretable models**, while **non-parametric methods offer flexibility** to extract insights directly from data.
- **Different reconstruction approaches (Gaussian Processes, Neural Networks, Interpolation, etc.)** allow us to cross-check results and uncover hidden trends in cosmic expansion and structure growth.
- By **improving** reconstruction techniques and **using** new datasets (e.g., DESI, JWST, Euclid), we can **refine** our understanding of dark energy, the Hubble tension, and fundamental physics.

Thank You!