L-Università ta' Malta
Faculty of Science
Department of
Statistics &
Operations Research

# Estimation of the Hubble Constant using Gaussian Process Regression and Viable Alternatives

Samuel Zammit

**March 2023**

Supervisor: Dr David Suda
Co-Supervisors: Dr Fiona Sammut, Prof Jackson Levi-Said

A dissertation presented to the Department of Statistics & Operations Research, Faculty of Science in part fulfilment of the requirements for the degree of Master of Science at the University of Malta

## Abstract

Samuel Zammit, M.Sc.
Department of Statistics & Operations Research, March 2023
University of Malta

Several papers within the astrophysical literature are dedicated to obtaining accurate and reliable estimates for the Hubble constant $H_0$, using diverse data sources such as CC, SNIa, and BAO, and different methodologies. This results in estimates which do not agree with each other - the so-called '$H_0$ tension'. In this work, methods already established in literature for estimating $H_0$, such as Gaussian process regression (GPR) and Markov chain Monte Carlo (MCMC) methods based on the concordance Lambda Cold Dark Matter ($\Lambda CDM$) model, together with some novel approaches in the field, are assessed. The first novel approach makes use of non-parametric MCMC inference on the hyperparameters of a Gaussian process kernel, independently of any cosmological model such as $\Lambda CDM$. The second approach is Student's $t$-process regression (TPR), which is similar to GPR but makes use of the multivariate Student's $t$-distribution instead of the multivariate Gaussian distribution. TPR does not automatically assume Gaussianity of underlying observations and has the additional advantage of being a more generalised and flexible form of GPR. We also consider variants of GPR and TPR which account for heteroscedasticity within the data. A comparison of the novel and tried-and-tested approaches is made. In particular, the model-independent approaches investigated largely agree with predictions based on the $\Lambda CDM$ concordance model. Moreover, GPR is highly dependent on the prior specification, while TPR and the heteroscedastic variants of both GPR and TPR are more robust to this. TPR and both heteroscedastic models provide evidence for a Hubble constant value that is on the lower side, and is therefore closer to the Planck value of 67.4 km $s^{-1}$ $Mpc^{-1}$ than the Riess value of 74.22 km $s^{-1}$ $Mpc^{-1}$. Therefore, the novel approaches discussed in this dissertation may shed further light on the $H_0$ tension. A main challenge posed by these approaches is the use of small datasets for the model-independent approaches; further research can thus apply such approaches to larger datasets. Across all estimates obtained for $H_0$ in this work, the median value is $\hat{H}_0^{med} = 68.85 \pm 1.67$ km $s^{-1}$ $Mpc^{-1}$.

**Acknowledgements**

I would like to start by thanking my supervisors Dr David Suda and Dr Fiona Sammut from the Department of Statistics and Operations Research at the University of Malta, as well as Prof Jackson Levi Said from the Institute of Space Sciences and Astronomy at the same University, for supporting and guiding me throughout this dissertation. I would also like to thank my family and close friends for their constant support and love. In particular, I would like to thank my parents Vincent and Shirley, my brother Thomas, and his girlfriend Maria.

Parts of this research have been carried out using computational facilities procured through the European Regional Development Fund (ERDF), Project ERDF-080 'A Supercomputing Laboratory for the University of Malta', and in conjunction with the Department of Physics and the Institute of Space Sciences and Astronomy at the University of Malta. I would like to thank Mr Kristian Grixti from the Department of Physics at the University of Malta for making the necessary arrangements for me to be able to use this computational cluster for some of the more computationally-intensive parts of the application. Thanks are in order also to Ms Stephanie Dimech and Mr Martin Micallef from the Department of Statistics and Operations Research for allowing me to use the departmental computer for some of the calculations. Without these resources, the application part would have not been possible.

# Contents

# List of Figures

# List of Tables

# List of Notation

| | |
|---|---|
| Å | Ångström |
| $\mathbf{A}_i$ | Walker |
| $\overrightarrow{\mathbf{A}}$ | Ensemble of Walkers |
| $\overrightarrow{\mathbf{A}}_{[i]}$ | Complementary Ensemble of Walkers |
| $\alpha$ | Scale Mixture Hyperparameter |
| $c$ | Speed of Light |
| $C_h^{(t)}$ | Autocorrelation Function at Time $t$ |
| $\chi^2_{\hat{H}}$ | Chi-Squared Statistic for Estimated Hubble Parameter Value |
| $d$ | Distance |
| $d_L(z)$ | Luminosity Distance Value at Redshift $z$ |
| $\Delta$ | Delta / Difference |
| $\mathbb{\Delta}_r$ | $(r \times r)$ Matrix of Latent Variance Variables |
| $E(z)$ | Dimensionless Hubble Parameter Value at Redshift $z$ |
| $\epsilon$ | Noise Term |
| $\boldsymbol{\epsilon}$ | Vector of Noise Terms |
| $f\sigma_8$ | Amplitude of Density Fluctuations |
| $\Gamma(x)$ | Gamma Function |
| $GP$ | Gaussian Process |
| $H(z)$ | Hubble Parameter Value at Redshift $z$ |
| $H_{obs}(z)$ | Observed Value for Hubble Parameter at Redshift Value $z$ |
| $\hat{H}_{pred}(z)$ | Predicted Value for Hubble Parameter at Redshift Value $z$ |
| $H_0$ | Hubble Constant |
| $\hat{H}_0^{CM}$ | Camarena and Marra Prior for $H_0$ |
| $\hat{H}_0^{DES}$ | Dark Energy Survey Prior for $H_0$ |
| $\hat{H}_0^{HW}$ | H0LiCOW Prior for $H_0$ |

| | |
|---|---|
| $\hat{H}_0^P$ | Planck Prior for $H_0$ |
| $\hat{H}_0^R$ | Riess Prior for $H_0$ |
| $\hat{H}_0^{TRGB}$ | Tip of the Red-Giant Branch Prior for $H_0$ |
| $\hat{H}_0^{med}$ | Median Estimate for $H_0$ |
| $\mathbb{I}_p$ | $p$-Dimensional Identity Matrix |
| $IW_p$ | $p$-Variate Inverse-Wishart Distribution |
| $IWP$ | Inverse-Wishart Process |
| $k$ | Covariance Kernel / Kernel Function |
| $K_\nu(x)$ | Modified Bessel Function of the Second Kind of Order $\nu$ |
| $\mathbb{K}$ | Covariance Matrix |
| km $s^{-1}$ $Mpc^{-1}$ | Kilometres per Second per Megaparsec |
| $l$ | Length Scale Hyperparameter |
| $L$ | Number of Walkers |
| $\lambda$ | True (Empirical) Wavelength |
| $\lambda_{obs}$ | Observed Wavelength |
| $\Lambda$ | Likelihood Value |
| $\mathbb{\Lambda}_r$ | $(r \times r)$ Matrix of Latent Likelihood Variables |
| $m$ | Mean Function |
| $\mathbf{m}$ | Mean Vector |
| $\mu$ | Distance Modulus |
| $MVN_p$ | $p$-Variate Gaussian (Normal) Distribution |
| $MVT_p$ | $p$-Variate Student's $t$-Distribution |
| $n$ | Number of Observations |
| $\nu$ | Degrees of Freedom |
| $\Omega$ | Density Parameter |
| $\Omega_\Lambda$ | Effective Mass Density Parameter of Dark Energy |
| $\Omega_K$ | Spatial Curvature Parameter |
| $\Omega_M$ | Matter Density Parameter |
| $\Omega_R$ | Effective Mass Density Parameter of Relativistic Particles |
| $p$ | Number of Dimensions |
| $pc$ | Parsec |
| $\Pi(p)$ | Set of $(p \times p)$ Real-Valued Matrices |
| $\mathbb{P}$ | Probability |

| | |
|---|---|
| $q$ | Probability Density Function of Proposal Distribution |
| $\mathbb{R}^p$ | $p$-Dimensional Real Coordinate Space |
| $\sigma$ | Standard Error |
| $\sigma^2$ | Variance |
| $\hat{\sigma}_{\hat{H}_0}$ | Standard Error of Estimated Value for Hubble Constant |
| $\hat{\sigma}_H(z)$ | Standard Error of Observed Value for Hubble Parameter at Redshift Value $z$ |
| $\sigma_f$ | Amplitude Hyperparameter |
| $\mathbb{\Sigma}_{SN}$ | Covariance Matrix of SN Data |
| $\mathbb{\Sigma}_{SN_*}$ | Covariance Matrix of SN$*$ Data |
| $t$ | Time |
| $T$ | Number of Iterations |
| $\mathbb{T}$ | Transformation Matrix |
| $TP$ | Student's $t$-Process |
| $\boldsymbol{\tau}$ | Integrated Autocorrelation Time |
| $\boldsymbol{\theta}$ | Vector of Hyperparameters |
| $\mathbf{W}$ | Random Vector of Weights |
| $\mathbf{w}$ | Vector of Weights |
| $W_p$ | $p$-Variate Wishart Distribution |
| $X$ | Indexing Set |
| $\mathbf{x}$, $\mathbf{x}'$ | Observation / Input |
| $\mathbb{X}$ | Data Matrix of Inputs |
| $y$ | Output |
| $\mathbf{y}$ | Vector of Outputs |
| $z$ | Redshift Value |
| $\zeta$ | Projection Function |
| $\mathbb{Z}$ | Projection Matrix |
| $\sim$ | Is Distributed As |
| $\overset{\mathbb{T},\mathbf{c}}{\longmapsto}$ | Affine Transformation |

# List of Abbreviations

| | |
|---|---|
| BAO | Baryon Acoustic Oscillations |
| CANDELS | Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey |
| CC | Cosmic Chronometers |
| CM | Camarena and Marra |
| CLASH | Cluster Lensing and Supernova Survey with Hubble |
| CMB | Cosmic Microwave Background |
| DES | Dark Energy Survey |
| DIC | Deviance Information Criterion |
| ERDF | European Regional Development Fund |
| GaPP | Gaussian Processes in Python |
| GP | Gaussian Process |
| GPR | Gaussian Process Regression |
| H0LiCOW | $H_0$ Lenses in COSMOGRAIL's Wellspring |
| HRD | Hertzsprung-Russell Diagram |
| HST | Hubble Space Telescope |
| IAU | International Astronomical Union |
| IWP | Inverse-Wishart Process |
| $\Lambda CDM$ | Lambda Cold Dark Matter |
| LOESS | Locally Estimated Scatterplot Smoothing |
| Ly-$\alpha$ | Lyman-Alpha |

| | |
|---|---|
| MAP | Maximum *a Posteriori* |
| MCMC | Markov Chain Monte Carlo |
| MH | Metropolis-Hastings |
| ML | Maximum Likelihood |
| MVN | Multivariate Normal (Gaussian) Distribution |
| MVT | Multivariate Student's $t$-Distribution |
| P | Planck |
| R | Riess |
| RBF | Radial Basis Function |
| SciPy | Scientific Computing in Python |
| SDSS | Sloan Digital Sky Survey |
| SN | Supernovae |
| SNIa | Supernovae of Type Ia |
| TP | Student's $t$-Process |
| TPR | Student's $t$-Process Regression |
| TPRD | Student's $t$-Process Regression with Dependent Student's $t$-noise |
| TPRK | Student's $t$-Process Regression with noise term in the Kernel |
| TRGB | Tip of the Red-Giant Branch |
| WMAP | Wilkinson Microwave Anisotropy Probe |

# Chapter 1

# Introduction

Cosmology may broadly be defined as the study of the universe. Understanding cosmology can therefore lead scientists to understanding the structure and behaviour of the universe and its constituent parts. As one can expect from such a complex field of study, there are several open questions within this field. Some of these are related to very fundamental concepts, such as the large-scale structure of the universe as well as the values of core cosmological parameters. In particular, we shall mainly concern ourselves with one such parameter, namely the Hubble constant $H_0$.

The Hubble constant is the constant of proportionality that affects the linear relationship between the velocity and distance of receding galaxies – more details on this are given in Section 1.2. However, whereas the linear relationship, initially made famous by Hubble (1929), has been widely accepted within the astronomical community, the exact value of $H_0$ is still a matter of great debate, with different techniques for estimating this parameter producing different results that do not agree with each other. This problem is known as the '$H_0$ tension' (see e.g. Gómez-Valent, 2019), and is explained in more detail in Section 1.3. In this opening chapter, our aim is to introduce the $H_0$ tension problem and describe the different statistical methods which will be employed within this dissertation in an effort to try and resolve this tension. We will start this chapter by giving a general overview of cosmology, and in particular the use of statistical methods within this field.

## 1.1 Introduction to Cosmology from a Statistical Perspective

Cosmology is the study of the large-scale structure of the universe, with the word 'cosmology' derived from Greek 'kosmos' (meaning 'order' or 'harmony') and 'logia' (meaning 'study of'). Cosmology is a sub-branch of the wider field of study of astrophysics. In particular, cosmologists aim to understand what the universe is made of, whether it is finite or infinite, and whether it has a defined beginning and a projected end. The answers to such questions would enable us to provide answers to the pressing metaphysical questions that concern humanity: "where do we come from, what are we, where are we going?" (Ryden, 2017, p. 2). We also want to answer more practical questions like what kind of universe we live in and how can we use the answers to such questions to understand the laws of nature more precisely.

While providing answers to such questions and accurately modelling our universe might seem like an impossibly large task, a number of simple observations may instantly be made without much prior knowledge. For example, one can look at the night sky and confirm that it is dark, with only a few stars visible. This indicates that most of space is 'empty', or at least not directly visible. This is paradoxical, as for the hypothetical case of a static and homogenous universe one would expect the night sky to be completely illuminated. The observed darkness of the night sky is therefore one of the main pieces of evidence for a dynamic universe. This is known as Olbers' paradox, named after the German astronomer H. W. Olbers (1758–1840), or alternatively as the 'dark night sky' paradox. However, the widely-accepted Big Bang model explains the so-called paradox by positing an expanding universe. This expansion means that an observer on Earth would not be able to see most of the light emitted by celestial objects, since the wavelength of this light would have increased due to the effect of redshift (see Section 1.2) such that this wavelength ends up lying outside the range of visible light.

Cosmologists are also able to understand and explain the overall structure of the universe in a remarkably simple manner. This is because, it has been observed that when observed on a large scale, the universe is highly regular. In fact, the 'cosmological principle' (Ryden, 2017, pp. 11–14) states that, when viewed on a sufficiently large scale, the universe appears the same for all observers, irrespective of the observer's location in the universe or their orientation. In other words, the universe is said to be homogenous and isotropic, where the former relates to independence from position and the

latter relates to independence from orientation. Moreover, various approximations and statistical assumptions also play an important role to help explain the universe. For example, measurements in cosmology, such as the cosmic microwave background (CMB) radiation, appear to be Gaussian. In general, cosmologists often consider observations to be Gaussian for simplicity. In the case that observations are not truly Gaussian, one can resort to the Central Limit Theorem to justify use of the Gaussian distribution. However, an open question within cosmology is the extent to which this widespread normality assumption may be ignored while still being in concordance with observed measurements. Within this dissertation, we shall be considering the $H_0$ estimation problem from various different perspectives. In particular, we will consider Gaussian process regression (GPR) and Markov chain Monte Carlo (MCMC) sampling methods which make use of the Gaussian assumption, then we shall deviate from the normality assumption by considering Student's $t$-process regression (TPR). We also consider variants of GPR and TPR that account for any possible heteroscedasticity within the data. More details on each of the methods employed are given in Section 1.5.

The use of statistics in cosmology is a rather recent endeavour, largely spurred by the availability of high-quality data such as that obtained from observations from the Hubble Space Telescope (HST), the European Space Agency's Planck mission, and the Sloan Digital Sky Survey (SDSS). Verde (2010, p. 148) argues that cosmology and statistics are intrinsically linked, since any cosmological theory is probabilistic in nature. This has recently been referred to as 'precision cosmology' (see e.g. Fan, 2020).

The two main schools of thought in probability are the frequentist and Bayesian schools. Cosmologists tend to prefer the Bayesian approach to statistics (see e.g. Padilla et al., 2021). This is because cosmological experiments are different in nature to laboratory experiments in that cosmological observations are often difficult to replicate or repeat, and we have the limitation of a unique universe. Moreover, one cannot perturb the universe to see how it reacts. The statistical methods described within this dissertation are Bayesian in nature, and some of them have already been extensively used within the astronomical and cosmological literature.

The different types of statistical problems within cosmology include testing whether a given set of data supports a hypothesis, obtaining estimates for cosmological parameters, or comparing a number of models and determining their fit. Statistical techniques that are applied within cosmology include the use of Fisher matrix forecasts, particularly in CMB temperature data and baryon acoustic oscillations (BAO) (Verde, 2010),

the use of GPR and MCMC techniques, the use of censored/truncated data, and the application of classification methods and data mining techniques (see e.g. Feigelson and Babu, 2012).

In this dissertation, we will deal mainly with estimation of $H_0$, as described at the beginning of this chapter. We introduce this parameter in the next section; then, in Section 1.3, we discuss different methods by which estimates for $H_0$ have been obtained and reported in the literature. We also discuss the issue of $H_0$ tension, i.e. that different observations and techniques for estimating $H_0$ give wildly different results that do not concord with each other. Lastly, in Section 1.5, we describe a number of different statistical methods that will be used to deal with such discrepancies. Within this dissertation, we apply each of these methods and compare and contrast the results – the code used in each case is available at `https://github.com/samuelzammit/EstCosmoPar-GPR-TPR`.

## 1.2 Redshift, Hubble's Law, and $H_0$

A number of celestial objects, such as stars and galaxies, emit waves with frequencies throughout most of the electromagnetic spectrum. When we look at a particular star or galaxy through the visible light part of the electromagnetic spectrum, we are mainly detecting the light emitted by this object. However, the wavelength of this light is 'shifted' such that the observed wavelength is $\lambda_{obs}$ while the true (empirical) wavelength is $\lambda$. We say that the object has a redshift $z$, where

$$z := \frac{\lambda_{obs} - \lambda}{\lambda}.$$

Red-shifted, or 'reddish', light waves have $z > 0$, and therefore the observed wavelength $\lambda_{obs}$ is larger than the actual emitted wavelength $\lambda$. However, a number of celestial objects, such as the nearby Andromeda galaxy, are observed to have a negative redshift, and the light waves arriving at Earth from these objects are said to be blue-shifted. Reddish objects are receding away from Earth over time, while blueish ones are approaching it.

Hubble (1929) showed that, for celestial objects that are sufficiently far away from Earth, the redshift exhibited by a celestial object is linearly proportional to its distance from Earth. Moreover, such objects are receding from Earth at a rate that is linearly proportional to their distances from Earth. Thus, there is a linear relationship between

the recessional galaxy 'velocity' $v(d)$ and its relative distance $d$:

$$v(d) = H_0 d.$$

The Hubble constant $H_0$ is hence the predicted constant of proportionality in this relation, where the 0 in the subscript refers to zero redshift, i.e. $z = 0$. The Hubble parameter $H(z)$ in general may be expressed as a function of redshift. It is also possible for this quantity to be expressed in terms of the time $t$ as opposed to the redshift $z$, namely by defining the dimensionless 'cosmic scale factor' $a := \frac{1}{1+z}$, and then using the change of variable

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \frac{\mathrm{d}H}{\mathrm{d}z}\frac{\mathrm{d}z}{\mathrm{d}a}\frac{\mathrm{d}a}{\mathrm{d}t} = (1+z)^2\left(\frac{H}{1+z}\right)\frac{\mathrm{d}H}{\mathrm{d}z}.$$

At zero redshift, the recessional velocity is measured with respect to the current time and zero distance. However, at the redshift value $z$, the measurement $H(z)$ refers to a universe at a size $1/z$ relevant to the current size, i.e. as $z$ increases, a smaller – and earlier – universe is considered.

The original diagram used by Hubble (1929) to show the linear relation between recessional velocity and relative distance from Earth is reproduced in Figure 1.1. The $y$-axis shows the recessional velocity in km/s, while the $x$-axis represents the distance from Earth in parsecs, where 1 parsec (pc) is approximately 3.26 light years. Therefore, it stands to reason that the unit of $H_0$ is km $\mathrm{s}^{-1}$ $\mathrm{Mpc}^{-1}$. The black points and solid regression line represent the celestial objects considered by the author on an individual basis, while the white points and dotted regression line represent these objects as combined into groups by the author.

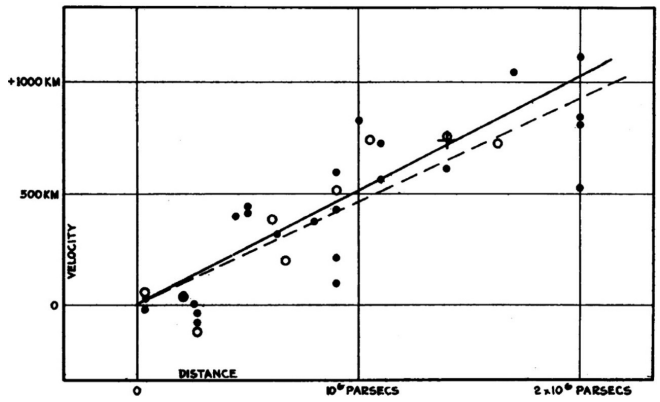



Figure 1.1: Linear relationship between velocity and distance (from Hubble, 1929)

Hubble's paper propelled the Big Bang theory to the prominent spot that it currently occupies as a cosmological model for the observable universe. However, previous work, such as that by Lemaître (1927), suggested this possible linear relation. In fact, the International Astronomical Union (IAU, 2018) has recognised Lemaître's contribution and suggested renaming Hubble's Law to the 'Hubble-Lemaître Law'. However, the value of $H_0$ itself has been the subject of uncertainty ever since Hubble's paper was published. Hubble himself estimated the value of $H_0$ to be around 500 km $s^{-1}$ $Mpc^{-1}$; however, his measurements are today known to have been off by a factor of 7. By the 1990s, the value for $H_0$ was estimated to lie in the range of between 50 and 100 km $s^{-1}$ $Mpc^{-1}$, while more recent work establishes the value of $H_0$ at around 70 km $s^{-1}$ $Mpc^{-1}$ (Prokopec, 2014, pp. 39–42). A review of recent estimates for $H_0$ can be found in the literature (di Valentino et al., 2021); some of these estimates are also described in Section 1.3 and will be used within the application part of this dissertation.

Use of the Hubble constant is not limited to Hubble's distance-velocity law. Note that $H_0$ is expressed in units of $\frac{\text{distance/time}}{\text{distance}}$, i.e. $\frac{1}{\text{time}}$. In fact, the Hubble constant also finds use in estimating the age of the universe, with a lower $H_0$ value suggesting an older universe (see e.g. Bahcall, 2015). In particular, if we assume that $H_0 = 70$ km $s^{-1}$ $Mpc^{-1}$and using the fact that $1Mpc \sim 3.0857 \times 10^{19}$ km, we get that $H_0 = 3.0857 \times 10^19 s^{-1}/70 = 4.4086 \times 10^{17} s^{-1}$, so that the age of the universe is approximately $4.4086 \times 10^{17}$ seconds, i.e. approximately 14 billion years. Small variations in the value of $H_0$ can dramatically change this figure; for example, a value of $H_0 = 65$ km $s^{-1}$ $Mpc^{-1}$will lead to an estimate closer to 15 billion years.

Additionally, this quantity was used to show that the expansion of the universe is in fact accelerating, not uniform as previously thought (see e.g. Perlmutter et al., 1999; Perlmutter and Schmidt, 2003; Riess et al., 1998). This eventually led to the 2011 Nobel Prize in Physics being awarded jointly to Riess, Perlmutter and Schmidt (see Royal Swedish Academy of Sciences, 2011).

Nowadays, there is an essentially universal consensus on the accelerating expansion of the universe, but the precise nature of this acceleration is still a matter of vigorous debate. In particular, while presently the 'standard' Lambda Cold Dark Matter ($\Lambda CDM$) model of Big Bang cosmology is widely accepted, there is no shortage of viable competing cosmological models such as the linearly varying deceleration parameter (Akarsu and Dereli, 2012), decaying dark matter (Pandey et al., 2020), and scale-invariant (Maeder, 2017) models. This is in no small part because the $\Lambda CDM$ model adequately explains

the basic observed structure of the universe (see e.g. Aghanim et al., 2020a), including the existence of CMB radiation as well as the large-scale structure within the distribution of galaxies. The $\Lambda CDM$ model also explains the observed abundances of simpler elements such as hydrogen, helium, and lithium within the universe, as well as the accelerating expansion of the universe as mentioned earlier in this section. Nevertheless, this model produces a number of theoretical challenges. In particular, the predicted values for some cosmological parameters such as $H_0$ using this model do not agree with more recent model-independent estimates and measurements (see e.g. Pandey et al., 2020). This separation, or tension, is quite a serious concern in cosmology, as it has the potential to 'spill over' into other cosmological parameters, calling into question the predictive power of the $\Lambda CDM$ model (Escamilla-Rivera et al., 2021). A more detailed description of the $H_0$ tension is given in Section 1.3. Some researchers, such as Guo et al. (2019), argue that looking beyond $\Lambda CDM$ may shine more light on the $H_0$ tension problem. Some work on this front has already been done (see e.g. Briffa et al., 2023). Section 1.4 will give more insight on some of the different estimates of $H_0$ present in the literature. Then, a number of techniques by which the $H_0$ tension may be mitigated are discussed in Section 1.5; we shall explore each of these techniques within this dissertation.

## 1.3 The $H_0$ Tension

Using CMB data and assuming the standard $\Lambda CDM$ model, the Planck collaboration (Aghanim et al., 2020b) reported an $H_0$ value of $\hat{H}_0^P = 67.4 \pm 0.5$ km s$^{-1}$ Mpc$^{-1}$, while the Dark Energy Survey (DES; Abbott et al., 2016) obtained a similar value of $\hat{H}_0^{DES} = 67.4^{+1.1}_{-1.2}$ km s$^{-1}$ Mpc$^{-1}$.

However, recent approaches suggest a different range of values from the $\Lambda CDM$ values – the so-called '$H_0$ tension' (see e.g. Gómez-Valent, 2019). For example, Riess et al. (2019) used long-period observations of Cepheids from the HST to obtain an $H_0$ estimate of $\hat{H}_0^R = 74.22 \pm 1.82$ km s$^{-1}$ Mpc$^{-1}$. Cepheids are often used as distance indicators since they are stars which brighten and dim periodically and predictably (see e.g. Feast and Walker, 1987), The use of the HST has allowed for better distance measurements compared with earlier, more primitive telescopes, and HST distance readings are suitably calibrated (see e.g. Tanvir et al., 1995). Similarly, Camarena and Marra (2020) used HST observations of Type Ia supernovae (SNIa) to obtain $\hat{H}_0^{CM} = 75.35 \pm 1.68$

km $s^{-1}$ $Mpc^{-1}$.

The H0LiCOW Collaboration (Wong et al., 2019) also used HST observations. However, the observations considered were those of strong gravitational lensing from quasars, which are distant cosmological structures produced by active black holes. The method for estimating $H_0$ in this case is based on "the measurement of the light variations in the lensed images of a distant source" (Courbin et al., 2002, pp. 19–20). This method was originally proposed by Refsdal (1964) and is hence widely known as Refsdal's method. Gravitational lensing occurs when light rays take different paths to produce multiple images and are therefore delayed by differing local potentials. In this case, the estimate obtained was $\hat{H}_0^{HW} = 73.3^{+1.7}_{-1.8}$ km $s^{-1}$ $Mpc^{-1}$.

The tip of the red-giant branch (TRGB) value $\hat{H}_0^{TRGB} = 69.8 \pm 1.9$ km $s^{-1}$ $Mpc^{-1}$ was obtained by Freedman et al. (2019) from a turning point in the evolutionary track of stars on the Hertzsprung–Russell diagram (HRD). This diagram, which is presented in Figure 1.2, is a plot of the stars' luminosity against their surface temperature. It is used in the study of stellar evolution; in particular, it may be used to roughly measure how far away a star cluster or galaxy is from Earth by comparing the apparent magnitudes of the stars under observation to the magnitudes of known reference stars.

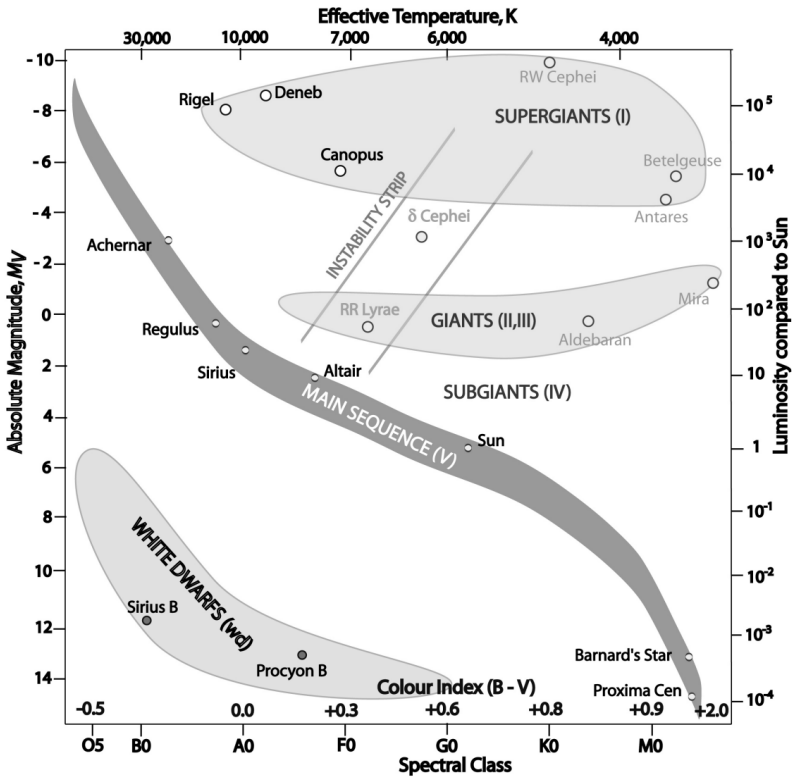



Figure 1.2: Hertzsprung-Russell diagram (from Hollow, 2006)

The life-cycle of a star, such as our Sun, consists of several stages. The material that eventually goes on to form a star typically starts out as a cloud of dust and gas, known as a nebula. Eventually, the dust and gas starts forming into protostars. If the protostar has sufficient mass, the dust and gas starts to burn hotter and hotter until the temperature is high enough for nuclear fusion of hydrogen into helium to occur. Then, the star is said to have entered its 'main sequence' phase. As an example, the Sun is about halfway through its main sequence phase.

The HRD is split into three main regions, with main-sequence stars running from the upper-left corner (i.e. high luminosity and temperature) to the lower-right corner (i.e. low luminosity and temperature). Giant and supergiant stars occupy the space above the main sequence, while white dwarfs occupy the space below. Once the main sequence phase of a star is over, the luminosity of the star starts increasing as it expands, while the surface temperature starts to decrease. At a certain point, helium particles within the star's core will start to undergo nuclear fusion into heavier elements such as carbon and iron, causing the temperature to increase quickly and consequently the star's evolutionary track to move sharply to the left. The result is a discontinuity, which is referred to as the TRGB, in the evolutionary track in the HRD for that star (Harpaz, 1993, pp. 103–110).

Figure 1.3, obtained from a meta-analysis by di Valentino et al. (2021), shows many different estimates of $H_0$ obtained using different techniques. The authors classify the different approaches to estimating $H_0$ into two main groups, namely the 'direct' and 'indirect' estimates. The 'direct' methods involve measuring directly the recessional velocities and distances of standard reference objects, or 'candles', such as supernovae and Cepheids. On the other hand, the 'indirect' estimates are obtained by analysing the large-scale structure of the universe at low or medium redshifts, or through measurements of the CMB radiation at higher redshifts (Odderskov et al., 2014).

Of particular note are the significant discrepancies between the indirect (e.g. CMB, Planck) and direct (e.g. TRGB, H0LiCOW) estimates of $H_0$. These discrepancies are sometimes as high as 4 standard deviations from each other – in such cases, we say that there is a $4\sigma$ tension between these estimates. Moreover, the distance in terms of standard deviations is set to increase in the near future as more precise estimates for $H_0$ are obtained, owing to the availability of more precise measurements (see e.g. Gómez-Valent, 2019).

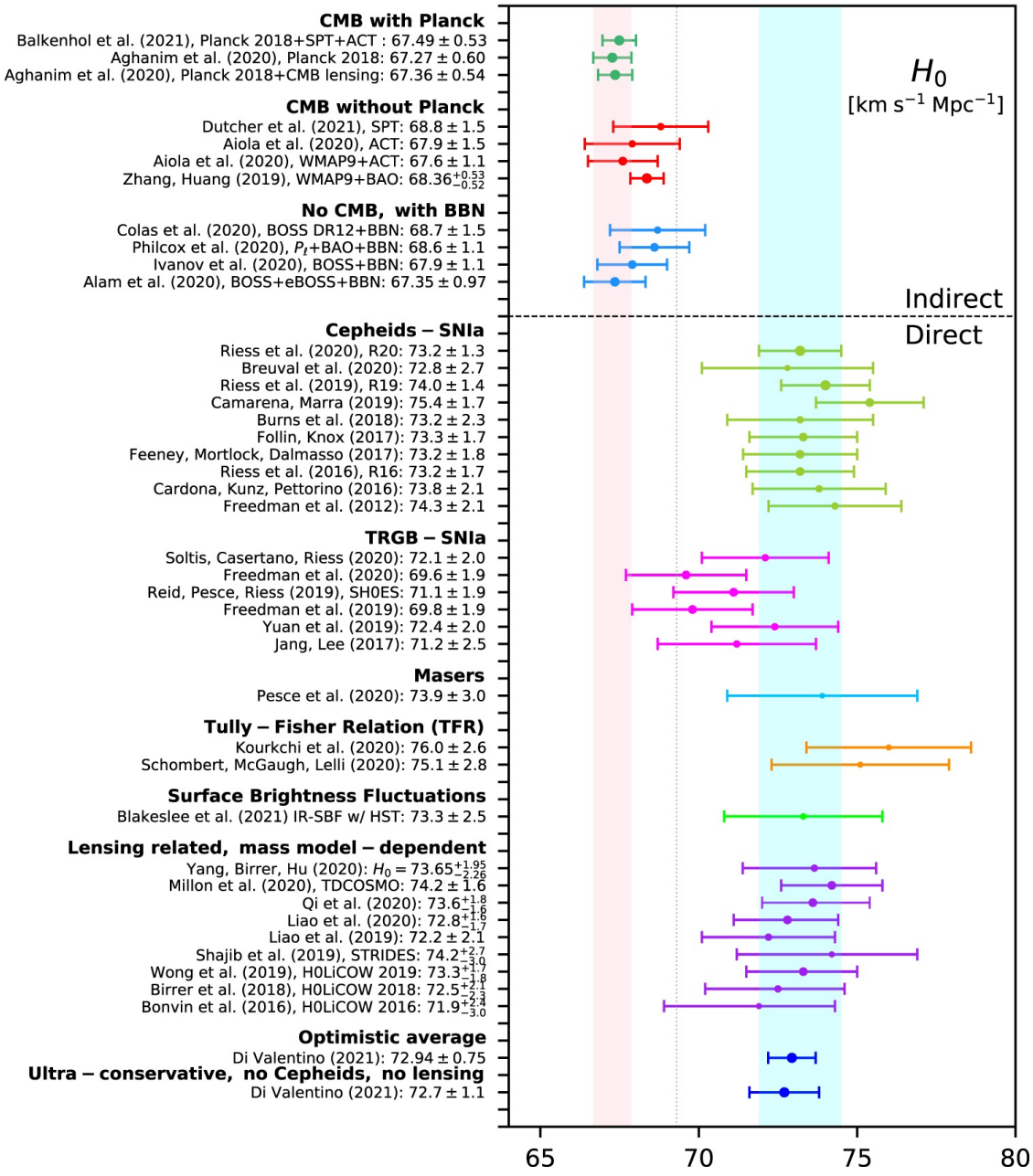



Figure 1.3: Some recent estimates of $H_0$ (from di Valentino et al., 2021), showing the tension between the 'indirect' and 'direct' measurements

## 1.4 Description of the Hubble Parameter Data

Throughout this dissertation we shall be concerned with three different data sources for $H(z)$. These are cosmic chronometers (CC), supernovae (SN) of Type Ia, and baryonic acoustic oscillations (BAO).

CC are used to obtain measurements for $H(z)$ using red-envelope galaxies (Stern et al., 2010), which are large passively-evolving galaxies that were formed early on in the universe's history, around 10 billion years ago. According to the CC approach, the expansion rate $H(z)$ may be measured by using the redshift-time derivative $\frac{\mathrm{d}z}{\mathrm{d}t}$ between these galaxies (Melia and Yennapureddy, 2018). These measurements are based on the observed 4000Å break in the passively-evolving galaxy spectra, where 1Å is equivalent to $1 \times 10^{-10}$m or 0.1nm, caused by the absorption of high-energy radiation from metals and by a scarcity of hot blue stars (Vogt, 2012). The amplitude of the absorption lines scales linearly with both the age and the metal abundance of the star. If the metallicity of these stars is known, one can calculate the age difference $\Delta t$ of two nearby galaxies proportionally to the difference of their 4000Å amplitude by measuring the slope. One can also determine the Hubble constant $H(z)$ given the redshift difference $\Delta z$ between the two galaxies using the relation

$$H(z) = -\frac{1}{1+z}\frac{\mathrm{d}z}{\mathrm{d}t} \approx -\frac{1}{1+z}\frac{\Delta z}{\Delta t}. \tag{1.1}$$

While several factors, such as star formation history and initial mass distribution, may affect the accuracy of the obtained readings, the 4000Å feature mainly depends on the age and metallicity of the host galaxies (Melia and Yennapureddy, 2018). The CC approach has the advantage of not relying on any cosmological models, since one is measuring only the redshift difference between 'local' galaxies. Moreover, it is ideal for obtaining $H(z)$ data at redshifts of $z \leq 2$ (Gómez-Valent and Amendola, 2018). For CC, the data in Table 4 of the study by Moresco et al. (2016) is used. To avoid the effects of cosmological model dependence and therefore retain independence from any cosmological models, only those points that are independent of BAO observations are considered.

For the SN data, the compressed Pantheon compilation data (Riess et al., 2018) is used throughout most of the dissertation. While the compressed dataset consists of only six points, these data points are effectively a compressed version of a set of around 1048 SN data points at $z < 1.5$ of the Pantheon compilation (Scolnic et al., 2018) together

with the 15 data points obtained by the HST at $z > 1.5$ as part of the CANDELS and CLASH Multi-Cycle Treasury programs (Riess et al., 2018). The values given in Riess et al. (2018) are values for $E(z) = H(z)/H_0$ instead of $H(z)$, but these values are still proportional to $H(z)$, since $H_0$ is a constant. Moreover, the compressed observations, except for one, have almost perfect Gaussian errors. The compressed data point for $z = 1.5$ in Table 6 of Riess et al. (2018) is excluded, since the error is non-Gaussian and therefore not valid for the GP approach. Therefore, the SN dataset in our case will only consist of five points. The full set of 1048 SN data points, which we will denote as SN$_*$, is used in conjunction with the $\Lambda CDM$ cosmological model within Section 5.2 of this dissertation. A more recent development in this field is the release of the 'Pantheon+' dataset (Scolnic et al., 2022), which combines the SN$_*$ dataset with other cosmological surveys.

The 1048 observations of the full Pantheon dataset (Scolnic et al., 2018) are given in terms of the distance modulus $\mu(z)$, but the observations within the compressed Pantheon dataset used by Riess et al. (2018) is given in terms of the 'dimensionless' Hubble parameter $E(z) = H(z)/H_0$. To obtain approximate $H(z)$ values corresponding to the given $E(z)$ values, we multiply the given values of $E(z)$ by 70 to obtain approximate $H(z)$ values. This quantity is dimensionless because $H(z)$ and $H_0$ have the same units, which cancel out. Therefore, within the compressed Pantheon dataset, the observations are given in terms of expansion data, and the nuisance parameter related to the distance modulus has been removed. This allows us to gain readings that are independent of any physical or cosmological model, but comes with the drawback that some information is lost. On the other hand, use of SN$_*$ requires the distance modulus values to be converted to $H(z)$ values, which can only be done given a cosmological model. Therefore, model-independent statistical methods, such as GPR and TPR, cannot be carried out using the full Pantheon dataset.

Converting the $\mu(z)$ values to $H(z)$ is quite an involved process that involves numerical integration. In particular, defining the luminosity distance $d_L(z)$ by

$$d_L(z) = c\,(1+z)\int_0^z \frac{1}{H(z')}\,\mathrm{d}z',$$

where $z'$ is a variable of integration and $c$ is the speed of light, then

$$\mu(z) = 5\log_{10}\left(d_L(z)\right) + M,$$

where $M$ is an arbitrary fiducial absolute magnitude (see e.g. Briffa et al., 2022). Such adjustments are not required for the CC data, since in this case the value of $H(z)$ can

be measured through direct observations, as in equation (1.1).

The BAO data points included are from the SDSS (Alam et al., 2017; Zhao et al., 2018). BAO observations are constructed from measurements of cosmic background radiation at high redshifts ($z \approx 1100$). At lower redshifts, BAO readings may be obtained by observing galaxy clusters (Cuceu et al., 2019). Another source of BAO measurements is from the Lyman-alpha (Ly-$\alpha$) forest (see e.g. du Mas des Bourboux et al., 2017); however, a significant tension of $2.5\sigma$ exists between galaxy BAO and Ly-$\alpha$ BAO. A similar tension exists between Ly-$\alpha$ BAO and predictions based on the widely-accepted $\Lambda CDM$ cosmological model, while galaxy BAO measurements are consistent with $\Lambda CDM$ predictions.

The mechanism behind obtaining BAO observations is more complicated than for both CC and SN. Essentially, BAO are fluctuations in the density of matter of the universe and provide a 'standard ruler' for measuring large distances within the universe in the same way that SN of known brightness provide a 'standard candle' (see e.g. Eisenstein, 2005; Perlmutter et al., 1999).

For each of the statistical approaches discussed throughout this dissertation, the different data sources will often be amalgamated. We will follow the convention in many astrophysical papers and use the + symbol to denote concatenation of datasets. In particular, for each approach taken throughout this dissertation, we shall consider the datasets CC, CC+SN, and CC+SN+BAO. The rationale behind concatenating the datasets in this way is twofold. In particular, CC acts as a 'base' dataset in that it contains direct measurements of $H(z)$ and has most of the data points at lower redshifts. Therefore, excluding the CC dataset from any of the analysis may lead to more uncertain estimates of $H_0$, since one would have to extrapolate further to arrive at $z = 0$. Adding in turn the SN and BAO data points then allows the CC data to be enriched with other measurements of $H(z)$, and the effect of adding these additional data points can be analysed. The datasets joined together also maintain independence from any cosmological model. Moreover, within Section 5.2, we replace SN with SN$_*$; thus, the datasets under consideration in this case are CC+SN$_*$ and CC+SN$_*$+BAO. For each dataset, we will consider each 'prior' $H_0$ value separately and also the priorless case.

The CC, SN, and BAO data points used throughout this dissertation are shown in Tables 1.1 through 1.3, whereas the full SN$_*$ dataset and its underlying $1048 \times 1048$ covariance matrix are publicly available on `https://github.com/dscolnic/Pantheon`.

| $z$ | $H(z)$ | $\sigma(H(z))$ | $z$ | $H(z)$ | $\sigma(H(z))$ | $z$ | $H(z)$ | $\sigma(H(z))$ |
|---|---|---|---|---|---|---|---|---|
| 0.07 | 69 | 19.6 | 0.3802 | 83 | 13.5 | 0.7812 | 105 | 12 |
| 0.09 | 69 | 12 | 0.4 | 95 | 17 | 0.8754 | 125 | 17 |
| 0.12 | 68.6 | 26.2 | 0.4004 | 77 | 10.2 | 0.88 | 90 | 40 |
| 0.17 | 83 | 8 | 0.4247 | 87.1 | 11.2 | 0.9 | 117 | 23 |
| 0.1791 | 75 | 4 | 0.4497 | 92.8 | 12.9 | 1.037 | 154 | 20 |
| 0.1993 | 75 | 5 | 0.47 | 89 | 49.6 | 1.3 | 168 | 17 |
| 0.2 | 72.9 | 29.6 | 0.4783 | 80.9 | 9 | 1.363 | 160 | 33.6 |
| 0.27 | 77 | 14 | 0.48 | 97 | 62 | 1.43 | 177 | 18 |
| 0.28 | 88.8 | 36.6 | 0.5929 | 104 | 13 | 1.53 | 140 | 14 |
| 0.3519 | 83 | 14 | 0.6797 | 92 | 8 | 1.75 | 202 | 40 |
| ... | ... | ... | ... | ... | ... | 1.965 | 186.5 | 50.4 |

Table 1.1: Hubble parameter data: CC

| $z$ | $E(z)$ | $\sigma(E(z))$ |
|---|---|---|
| 0.07 | 1.007 | 0.024 |
| 0.2 | 0.898 | 0.016 |
| 0.35 | 0.893 | 0.029 |
| 0.55 | 0.732 | 0.033 |
| 0.9 | 0.652 | 0.051 |

Table 1.2: Hubble parameter data: SN

| $z$ | $H(z)$ | $\sigma(H(z))$ | $z$ | $H(z)$ | $\sigma(H(z))$ |
|---|---|---|---|---|---|
| 0.24 | 79.69 | 2.65 | 0.6 | 87.9 | 6.1 |
| 0.35 | 84.4 | 7 | 0.73 | 97.3 | 7 |
| 0.43 | 86.45 | 3.68 | 2.3 | 224 | 8 |
| 0.44 | 82.6 | 7.8 | 2.34 | 222 | 7 |
| 0.57 | 92.4 | 4.5 | 2.36 | 226 | 8 |

Table 1.3: Hubble parameter data: BAO

## 1.5 Mitigating the $H_0$ Tension

In the previous section, we have seen that there is a significant tension between different measurements and estimates for $H_0$. For example, as shown in Figure 1.3, local measurements of the Hubble constant using Cepheids and Type Ia supernovae (see e.g. Camarena and Marra, 2020; Riess et al., 2019) seem to be more than 4 standard deviations apart from the estimates of the Hubble constant obtained from the $\Lambda CDM$ model fitted to the Planck CMB data (Aghanim et al., 2020b). Moreover, going beyond these basic models to include other data sources such as gravitational lensing (see e.g. Wong et al., 2019) will only complicate such discrepancies.

This dissertation aims to consider a number of statistical techniques which may be used to obtain estimates for $H_0$; these include GPR, MCMC approaches, TPR, and heteroscedastic approaches to GPR and TPR. On the cosmological front, a possible solution is to consider approaches other than the concordance $\Lambda CDM$ model (see e.g. Guo et al., 2019). Some of the statistical techniques considered in this dissertation, in particular GPR (Chapter 2), the non-parametric instance of MCMC (Section 3.4), and TPR (Chapter 4), as well as the heteroscedastic variants of GPR and TPR discussed in Section 5.1, have the advantage of being independent of any cosmological models. On the other hand, the MCMC approach involving the parameters of the $\Lambda CDM$ cosmological model (Sections 3.3 and 5.2) involves an explicit dependence on this model.

### 1.5.1 Gaussian Process Regression

A Gaussian process (GP) is a stochastic process; that is, a collection of random variables indexed by some set, from which any finite selection of random variables follows the multivariate normal distribution. GPs may therefore be thought of as a generalisation of the multivariate Gaussian distribution. In the same way that a multivariate Gaussian distribution can be characterised by its mean vector and variance-covariance matrix, a GP can be completely characterised by its mean and kernel functions. More detail on kernel functions is given in Section 2.2.

Although GPs have a long and storied history within statistics, they have not seen much in the way of applications within cosmology until quite recently. However, GP-based models are now commonly used in both regression and classification tasks, as

well as in some more specialised applications. The basic theory behind GPs may be attributed to Wiener (1949) and Kolmogorov (1941, as cited in Shiryayev, 1992), with the first applications of GPR to the prediction/regression problem appearing within the fields of geostatistics (see e.g. Matheron, 1973; Whittle, 1963) and meteorology (see e.g. Thompson, 1956). Another area in which GPR has been applied is in computer simulations (Sacks et al., 1989; Santner et al., 2003), in which the observations are synthetic and are often assumed to be free of noise.

The application of GP-based methods to astronomical data, which we shall be concerned with in this dissertation, is a more recent phenomenon. Nevertheless, GPR has been used within many different areas of astrophysics, such as the analysis of light curves of stars and active galactic nuclei (Brewer and Stello, 2009; Kelly et al., 2014), as well as measurement of the brightness of X-ray binary star systems (Uttley et al., 2005). GPs have also been used to model CMB radiation (Bond and Efstathiou, 1987) and in quantifying correlated noise in instruments (Gibson et al., 2012) and for the analysis of gravitational waves (see e.g. Belgacem et al., 2020). GPs have also more recently found use in the estimation of core cosmological parameters, such as $H_0$ and $f\sigma_8$, with the goal of understanding better the tension between the different estimates for these parameters. In particular, GPR has been widely used in the literature for smoothing Hubble data (see e.g. Gómez-Valent and Amendola, 2018; Li et al., 2020; Seikel and Clarkson, 2013; Seikel et al., 2012).

GPR, also known as kriging (after D. G. Krige), Wiener–Kolmogorov prediction, or Gaussian spatial modelling, is a model-independent supervised learning approach based on constraining a kernel function, which can then be used to reconstruct data, that is, to smoothen out observed data and simulate unobserved data points through interpolation/extrapolation. In this dissertation, the term GPR is preferred in order to emphasise that we are carrying out regression with the end goals of reconstruction and uncertainty quantification in mind. In particular, we will consider three different datasets of $H(z)$ for varying values of $z$ as described in Section 1.4 and each of the kernel functions described in Section 2.2, then reconstruct the $H(z)$-against-$z$ function and extrapolate this to $z = 0$ to obtain the estimate for $H_0$ in each case.

GPR may be viewed from differing perspectives. One can view it as a generalisation of the 'usual' regression to the infinite-dimensional case, i.e. a regression in function space as opposed to coordinate space. An alternative interpretation of GPR is as a Bayesian technique, where a GP prior and multivariate Gaussian likelihood function lead to a

GP posterior whose mean and kernel functions are suitably modified from the prior by the observed data. Gramacy (2020) mentions that this Bayesian reinterpretation of classical statistical techniques arose first alongside the surge of interest in Bayesian learning in the mid-to-late 1990s within the statistical learning community (see e.g. Williams and Rasmussen, 1996). Another example of this remodelling includes the interpretation of logistic regression as the single-layer perceptron. GPR is re-imagined within the Bayesian framework within Chapter 2; in this case, the GP acts as the 'prior', while the predictive distribution may be considered as the 'posterior'.

A main advantage of GPR is that it is mostly non-parametric, and therefore only a limited number of hyperparameters would need to be specified for good performance of the regression model. Other related non-parametric regression techniques, such as locally estimated scatterplot smoothing (LOESS), have also been applied to $H_0$ data (Escamilla-Rivera et al., 2021). While GP-based methods, as with any other statistical technique, have their limitations, they are very useful when there is a limited knowledge of the modelling context, in particular due to their model-independence and high degree of flexibility (Gramacy, 2020, p. 143).

One of the more noteworthy advantages is the simplicity by which this method may be applied, in part due to its reliance on the Gaussian distribution. Various statistical packages (Pedregosa et al., 2011; Seikel et al., 2012), which may apply GPR to a given set of data, are readily available on different software such as Python and R. Another advantage of using GPs is that the model uncertainty is known and directly captured, which is not always the case with other statistical learning methods such as neural networks (Gal and Ghahramani, 2016). Moreover, optimal estimation methods based on Gaussian models, such as maximum likelihood estimation, often result in mathematically tractable linear solutions. GP-based methods are also highly flexible and allow for detailed specification of the kernel function depending on prior belief. Indeed, a wide variety of kernel functions may be used, with some of these being described in Section 2.2. Products, summations, and compositions of these functions may also be used. Additionally, Gaussian distributions can be used to model the sum of independent random processes as a result of the Central Limit Theorem and are therefore highly suited to modelling many different classes of signals and noise. Non-Gaussian processes may also be approximated by a weighted mixture of Gaussian distributions (see e.g. Daemi et al., 2019).

However, as with any other statistical method, GPR also comes with its drawbacks.

Two main disadvantages are the cubic computational complexity which may present problems with large datasets and the tendency to overfit on the training data (Mohammed and Cawley, 2017). The overfitting problem is especially common with smaller datasets – such as those used within this dissertation – or when the prior knowledge of the covariance structure is limited. In order to alleviate the computational complexity problem, approximations of GPs (Liu et al., 2020) or sparse GPs (Ranganathan et al., 2010) may be used instead. To reduce overfitting, one can either try different initial hyperparameter values or increase the contribution of the random noise component. Research on extensions of GPs is ongoing. This includes sparse and minibatch GPs (Liu et al., 2020) which enable scalability on larger datasets, as well as deep and convolutional GPs (Damianou and Lawrence, 2013; Dutordoir et al., 2020) which allow for processing of high-dimensional and image data.

### 1.5.2 Markov Chain Monte Carlo-Based Models

MCMC comprises a class of algorithms that may be used to sample from a desired joint distribution $f(\mathbf{a})$, where $\mathbf{a} = (a_1, a_2, \ldots, a_p)$, in cases where sampling directly from this distribution is not feasible computationally. These methods work by constructing a Markov chain whose limiting distribution is the desired distribution.

One of the first MCMC methods to be developed, and arguably one of the most common MCMC methods in use even nowadays, is the Metropolis algorithm. The idea behind this algorithm is to sample values from a 'proposal' distribution instead of directly from the joint distribution, then accepting or rejecting each sampled value based on an acceptance function. This algorithm was first introduced by Metropolis et al. (1953) and generalised as the Metropolis-Hastings (MH) algorithm to non-symmetric proposal distributions by Hastings (1970), with further variations such as quasi-Monte Carlo methods (see e.g. Owen and Tribble, 2005) being developed in more recent years.

A special case of the MH algorithm, known as the Gibbs sampler, introduced by Geman and Geman (1984), is often regarded as a separate algorithm in its own right. Given a desired joint distribution on $p$-dimensional $\mathbf{a}$, the Gibbs sampler obtains a sampled value for each component $\mathbf{a}_i, i \in \{1, 2, \ldots, k\}$, where $k \leq p$, of $\mathbf{a}$ by sampling from the conditional distribution

$$f(\mathbf{a}_i | \mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{i-1}, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_k).$$

In this case, a 'component' may be either a single value or a small group of related values. The performance of the Gibbs sampler naturally depends on the ease of deriving and sampling from the conditional distributions for each $\mathbf{a}_i$. Therefore, it is not suitable when the conditional distributions are difficult to deal with. Variations, extensions and approximations to the Gibbs sampler have been developed over the years; these include the 'data augmentation' algorithm (Tanner and Wong, 1987), the 'hit-and-run' algorithm (Bélisle et al., 1993), and approximations based on the theory of acceptance functions (see e.g. Ritter and Tanner, 1992; Tierney, 1994).

In this dissertation, we shall mainly be considering a different family of MCMC methods, known as ensemble samplers. The term 'ensemble' relates to multiple samplers, or 'walkers', of the standard MCMC algorithms being run in parallel. Ensemble samplers that are invariant under affine transformations of the coordinate space have been proposed by Goodman and Weare (2010), and are easy to construct and implement. Such samplers are described in more detail in Section 3.2.

Within the context of cosmology, MCMC techniques such as the ensemble samplers have been used extensively, particularly in the estimation of the parameters of some pre-defined cosmological model given a set of data. In this case, the desired distribution often arises from Bayesian inference of model parameters, such as the angular power spectrum of the CMB radiation within the $\Lambda CDM$ model. This parameter may either be obtained from observations on the CMB or else predicted theoretically (Akeret et al., 2013). As an example, the parameters of the 'standard' $\Lambda CDM$ model would include the Hubble constant $H_0$ and the matter density parameter $\Omega_{M0}$, where the latter represents the present-day density of the universe due to baryonic and dark matter. Various software packages are available for this purpose (see e.g. Foreman-Mackey et al., 2013; Lewis, 2013).

In particular, within this dissertation, we shall be considering the use of MCMC techniques in both a parametric and a non-parametric setting. For the former, in Section 3.3, we consider the $\Lambda CDM$ cosmological model and use an ensemble sampler to obtain point estimates for the parameters $H_0$ and $\Omega_{M0}$. On the other hand, for the non-parametric setting, we consider a GPR structure and use an ensemble sampler to obtain estimates for the hyperparameters of the kernel functions involved and then obtain estimates for the Hubble constant. To our knowledge, the latter approach is novel since it has not yet been tried within cosmology.

### 1.5.3 Student's $t$-Process Regression

While GPR has been extensively used within the astrophysical literature, including to obtain estimates for the Hubble constant (see e.g. Bernal et al., 2016; Briffa et al., 2022), Student's $t$-processes (TPs) have not yet been used for this purpose to our knowledge. Within this dissertation, we explore the use of TPR to obtain estimates for $H_0$. As the name implies, TPs are highly similar to GPs, but with the multivariate Student's $t$-distribution replacing the multivariate Gaussian distribution. A TP is therefore characterised by its mean and kernel functions, as well as an additional degrees of freedom parameter.

The main drawbacks of GPR include the normality assumption taken on the observations, as well as the poor performance of GPR when outliers are present in the data and the tendency to overfit. Various alternative regression models have been proposed in order to address these disadvantages. The first step away from the 'vanilla' GPR model is to consider the GPR model with independent and identically-distributed Student's $t$ noise. Under such a model, the latent function $f$ is assumed to be a Gaussian process, but the noise vector $\boldsymbol{\mathcal{E}}$ is assigned a multivariate Student's $t$-distribution. However, such a model is not analytically tractable and therefore one would have to resort to computationally-intensive numerical methods. An alternate approach would be to assume that $f$ is a Student's $t$-process, that is, a TPR model, but TPR with either Gaussian or Student's-$t$ noise is still analytically intractable. Because of this, Williams and Rasmussen (2006) state that "the $t$-process is perhaps not as exciting as one might have hoped" (p. 194).

To overcome this problem, two extensions of the basic TPR model may be considered. The first approach (Shah et al., 2014; Zhang and Yeung, 2010) incorporates the noise within the kernel function by including an additional diagonal kernel term instead of considering independent Student's $t$-noise separately. This is known as the TPRK model. While the posterior distribution in this case is still not analytically tractable, at least we get that the TPRK model tends to a GPR model with independent Gaussian noise as $\nu \to \infty$, while behaving similarly to a GPR model with independent Student's $t$ noise (Shah et al., 2014).

The second approach (Tang et al., 2017) introduces dependent Student's $t$-noise, in that the variance of the noise is dependent on how well the corresponding noise-free model fits the data. This is referred to as the Student's $t$-process regression with dependent

Student's $t$ noise (TPRD) model, According to the authors, the TPRD model inherits all the advantages of GPR while giving a probabilistic interpretation of TPRK. Moreover, the three models can be shown to have the same predictive mean under some conditions imposed on the kernel, as shown in Appendix A. In this dissertation, we consider only the TPRD model. Analogously to the Gaussian case, in TPR we let the latent function $f$ be a TP and assign a multivariate Student's $t$-distributed likelihood function, leading to a TP posterior.

In the same way that the multivariate Student's $t$-distribution is a more general version of the multivariate Gaussian distribution owing to the additional degrees of freedom parameter, so is TPR a more general version of GPR. Moreover, as in GPR, the mean is often taken to be $\mathbf{0}$ within TPR. However, TPR may offer two possible advantages. Firstly, it does not assume normality of the underlying observations. Additionally, TPR has the capacity of being more flexible than GPR due to the additional degrees of freedom parameter. In the worst-case scenario, one can set the degrees of freedom parameter to a high value, in which case a TP will tend to a GP. This follows from the well-known fact that the Student's $t$-distribution converges to the Gaussian distribution as the degrees of freedom parameter approaches infinity.

As opposed to GPR, the main motivation behind using TPR is to deviate from the assumption of Gaussianity that is often taken on cosmological data. As in GPR, we consider the three data sources for $H_0$ described in Section 1.4 and the kernels as in Section 2.2, and we shall obtain a reconstruction of the $H(z)$-against-$z$ function and an estimate for $H_0$ in each case. The use of TPR within the literature is still quite limited, and there is a significant opportunity for future research on the applications within a wide variety of fields. In particular, we are not aware of any research within cosmology using TPR, while quite a lot of work has been done using GPR.

The next few chapters will describe and deal with each of the above-mentioned methods within the context of quantifying and examining the $H_0$ tension. In particular, Chapter 2 will deal with GPR while Chapter 3 will deal with the parametric and non-parametric MCMC approaches. Chapter 4 will deal with TPR. Then, within Chapter 5, we will describe some extensions to the models discussed within Chapters 2 to 4. In particular, we consider extensions to 'vanilla' GPR and TPR which account for heteroscedasticity within the data, and also revisit the $\Lambda CDM$ approach where we replace the compressed SN dataset with the full SN$_*$ dataset. Lastly, within the concluding Chapter 6, we present a comparison of the results obtained, state the limitations encountered and suggest areas for further research on this topic.

# Chapter 2

# Gaussian Process Regression

In this chapter, we provide an introduction to GPs and GPR, as well as apply GPR to obtain model-independent estimates for the Hubble constant. Firstly, in Sections 2.1 and 2.2, we define GPs and state some of the kernel functions which a GP may have. Then, in Section 2.3, we give the necessary theoretical background for GPR. In particular, we view GPR from two equivalent viewpoints, namely the 'weight-space' and the 'function-space' views.

In Section 2.4, we give a detailed description of our application of GPR to three sources of data, each containing measurements of $H(z)$ for different values of the redshift $z$, as described in Section 1.4. In our implementation of GPR for this purpose, we fit a GP regression model to each of the datasets in order to estimate the value of $H$ at $z = 0$, i.e. to obtain an estimate for $H_0$. We also consider the case where an additional data point, which represents an estimated value for $H_0$ based on prior literature, is added to each dataset. Six different values are considered for this purpose. These additional values are referred to as 'priors' because these values reflect our prior understanding of $H_0$ based on the literature, but they are not to be considered as priors in the Bayesian sense. The priorless case will also be considered. An estimate of $H_0$ is obtained for each combination of dataset, prior and kernel function.

We define some notational conventions that will be used throughout this dissertation. Vectors of random variables will be denoted by a bold uppercase symbol, for example $\mathbf{F}$, whereas realisations of such vectors are denoted by the corresponding bold lowercase symbol, such as $\mathbf{f}$. Similarly, matrices will be denoted in 'blackboard bold' font, for example $\mathbb{K}$. We consider $n$ points $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$, each of $p$ dimensions. The notation

$MVN_p$ and $MVT_p$ will be used to denote the $p$-variate Gaussian and Student's $t$-distributions. Hat notation, for example $\hat{H}_0$, will be used to denote an estimated value, while bar notation, such as $\bar{x}$, will be used to denote a mean value.

## 2.1 Defining Gaussian Processes

A GP is a type of stochastic process in which the distribution of any finite subset of the random variables making up this process is multivariate normally distributed. In particular, the distribution for each realisation of the process (i.e. the one-dimensional case) is Gaussian, making GPs useful in statistical modelling of data observed over time or space. In particular, we will carry out regression over GPs as introduced in Section 1.5.1 and as will be described in detail in Section 2.3. A GP is completely characterised by its mean and kernel functions; this characterisation of GPs allows for the following definition:

**Definition 2.1.1** (Gaussian process)**.** Let $X \subseteq \mathbb{R}^p$ be a nonempty indexing set containing elements of some dimension $p$. Also, let $k : X \times X \times \Theta \to \mathbb{R}$ be a positive definite kernel and $m : X \to \mathbb{R}$ be any real-valued function. Moreover, consider any combination $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ of size $n \in \mathbb{N}$ of elements in $X$, and let $\mathbf{f} := (f(\mathbf{x}_1), f(\mathbf{x}_2), ..., f(\mathbf{x}_n))^T$. Then, a random function $f : X \to \mathbb{R}$ is said to be a GP with mean function $m$ and covariance kernel $k$ if the random vector $\mathbf{F}$ follows the $n$-variate normal distribution with mean vector $\mathbf{m} := (m(\mathbf{x}_1), m(\mathbf{x}_2), ..., m(\mathbf{x}_n))^T$ and covariance matrix $\mathbb{K}$, where the $(i, j)^{\text{th}}$ entry of $\mathbb{K}$ is the kernel function $k(\mathbf{x}_i, \mathbf{x}_j)$.

Kernel functions are defined and described in more detail in Section 2.2. In the case that this kernel matrix is singular, the multivariate normal distribution is still well-defined. This may occur, for example, in the case of polynomial kernels, or when some of the points $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ are identical.

If $f$ is a GP with mean function $m$ and kernel function $k$, this is denoted by $f \sim GP(m, k)$ for notational convenience. Since there is a one-to-one correspondence between $f$ and pairs $(m, k)$ of mean and kernel functions, then specifying $m$ and $k$ implicitly corresponds to a 'unique' $GP(m, k)$. In other words, the mean and covariance structure of a GP determine all possible joint distributions of finite subsets of the random variables. If $m(\mathbf{x}) = 0 \ \forall \mathbf{x} \in X$, then the GP is said to be centered. The kernel function $k$ may also be written in terms of $f$ and $m$ using the standard definition of

covariance, i.e.:

$$\mathrm{Cov}(\mathbf{x}, \mathbf{x}') = \mathbb{E}_f \left[ (f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}')) \right], \tag{2.1}$$

where $f \sim GP(m, k)$ and $\mathbf{x}, \mathbf{x}' \in X$.

The most-well known example of a GP is the Brownian motion $\{W_t\}_{t \geq 0}$, while other examples include the Ornstein-Uhlenbeck process $\{Y_t\}_{t \in \mathbb{R}}$ and the Brownian bridge $\{W_t^o\}_{t \in [0,1]}$. These may be expressed in terms of their mean and covariance, namely

$$\begin{aligned}
\mathbb{E}[W_t] = 0 \ \forall t, \qquad & \mathrm{Cov}(W_{t_1}, W_{t_2}) = \mathbb{E}[W_{t_1} W_{t_2}] = \min(t_1, t_2); \\
\mathbb{E}[Y_t] = 0 \ \forall t, \qquad & \mathrm{Cov}(Y_{t_1}, Y_{t_2}) = \mathbb{E}[Y_{t_1} Y_{t_2}] = \exp\left(-|t_2 - t_1|\right); \\
\mathbb{E}[W_t^o] = 0 \ \forall t, \qquad & \mathrm{Cov}(W_{t_1}^o, W_{t_2}^o) = \mathbb{E}[W_{t_1}^o W_{t_2}^o] = \min(t_1, t_2) - t_1 t_2.
\end{aligned}$$

These three GPs are all centered and indexed by a unidimensional time component, but in general the indexing set is not always time and may in fact be multidimensional.

In the next section, we will mention some of the most common functions which are used as covariance kernels for GPs; combinations of these functions are also often considered.

## 2.2 Covariance Kernels

As specified in Definition 2.1.1, a GP may be defined by specifying its mean function and covariance kernel. This section will detail some common covariance kernels in use in conjunction with GPs. Firstly, however, we give a general definition for kernel functions:

**Definition 2.2.1** (Kernel function)**.** A kernel function $k : X \times X \times \Theta \to \mathbb{R}$ is a symmetric function for which

$$\sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j k(\mathbf{x}_i, \mathbf{x}_j; \boldsymbol{\theta}) \geq 0$$

holds for any $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n \in X$, $n \in \mathbb{N}$, $c_1, c_2, ..., c_n \in \mathbb{R}$. The vector $\boldsymbol{\theta} := (\theta_1, \theta_2, ..., \theta_a)^T$ for some $a \in \mathbb{N}$ contains the hyperparameters of the kernel, which allow for additional flexibility to the specification of the kernel function. The number of hyperparameters $a$ depends on the definition of the kernel. The set $\Theta$ contains all possible values for the hyperparameters and is therefore made up of vectors of length $a$.

Covariance kernels are kernel functions that serve as covariance functions for a GP. These covariance kernels mentioned in this section will be used for a GPR application with astronomical data (see Section 2.4). Only kernels which are available in the Python package `GaPP` (Seikel et al., 2012) will be considered. This list of kernels is by no means exhaustive; in general, kernels may be defined on any nonempty set and are not limited to Euclidean spaces. However, the functions available within `GaPP` offer a varied selection of different types of kernels.

The first kernel we will discuss is the square exponential kernel, which is defined by:

$$k(\mathbf{x}, \mathbf{x}'; \sigma_f, l) = \sigma_f^2 \exp\left[-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2l^2}\right]. \tag{2.2}$$

This kernel is also known as the Gaussian or the radial basis function (RBF) kernel and is widely used, both in GPR and in other supervised learning techniques such as support vector machines. The RBF kernel has some desirable properties such as it being easily differentiable and integrable, which allows for a wide class of related functions to also be available. Some approximations to the square exponential kernel, for example one based on the Fourier transform, have also been introduced and used in the literature (see e.g. Mueller, 2012).

For this kernel, the quantities $\sigma_f$ and $l$ are the hyperparameters, which allow for flexibility in the kernel specification. For example, a larger value for the length scale $l$ means that the function values change more slowly and is therefore suitable for capturing a long-term trend. On the other hand, $\sigma_f$ is a scale factor that determines the average distance of the function from its mean (see e.g. Duvenaud, 2014). Similarly, Jähnichen et al. (2018) consider the length scale as parametrising the 'memory effects' produced by the exponential decay.

A generalisation of the RBF kernel is known as the rational quadratic kernel. This kernel incorporates an additional hyperparameter $\alpha$, which allows this kernel to be equivalent to the addition of multiple RBF kernels with different length scales. The rational quadratic kernel with hyperparameters $\sigma_f$, $l$, and $\alpha$ is defined by:

$$k(\mathbf{x}, \mathbf{x}'; \sigma_f, l, \alpha) = \sigma_f^2 \left[1 + \frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\alpha l^2}\right]^{-\alpha}.$$

As $\alpha \to \infty$, this kernel becomes the 'standard' RBF kernel. It is important to note that while the RBF and rational quadratic kernels are two of the most popular kernel functions for use in GPR, they are not without their disadvantages. In particular, they

may lead to model misspecification especially when the function or its derivatives are discontinuous (Duvenaud, 2014).

Another widely-used class of kernels is known as the Matérn kernels. Apart from the length scale, a Matérn kernel has an additional hyperparameter $\nu > 0$, where typically $\nu = m + \frac{1}{2}$ for some non-negative integer $m$. In fact, the Matérn kernel with parameters $\nu, l > 0$ is the function defined by:

$$k(\mathbf{x}, \mathbf{x}'; \nu, l) = \frac{1}{2^{\nu-1}\Gamma(\nu)} \left[ \frac{\sqrt{2\nu}\|\mathbf{x} - \mathbf{x}'\|}{l} \right]^{\nu} K_{\nu} \left[ \frac{\sqrt{2\nu}\|\mathbf{x} - \mathbf{x}'\|}{l} \right],$$

where $\Gamma$ is the gamma function and $K_{\nu}$ is the modified Bessel function of the second kind of order $\nu$.

In the special case that $\nu = m + \frac{1}{2}$ for some non-negative integer $m$, then the expression for the Matérn kernel may be reduced to a simpler form (Williams and Rasmussen, 2006, p. 85):

$$k(\mathbf{x}, \mathbf{x}'; \nu, l) = \exp\left[ -\frac{\sqrt{2\nu}\|\mathbf{x} - \mathbf{x}'\|}{l} \right] \frac{m!}{(2m)!} \sum_{i=1}^{m} \frac{(m+1)!}{i!(m-1)!} \left[ \frac{\sqrt{8\nu}\|\mathbf{x} - \mathbf{x}'\|}{h} \right]^{m-i}.$$

In `GaPP`, Matérn functions for $\nu = \frac{3}{2}, \frac{5}{2}, \frac{7}{2}, \frac{9}{2}$ are implemented as kernel functions. The Matérn kernel function may only be used for the reconstruction of the $k^{\text{th}}$ derivative of the dependent function $f(\mathbf{x})$ if $k < \nu$. Therefore, reconstructions for $\nu = \frac{3}{2}$ are only available up to the first derivative, for $\nu = \frac{5}{2}$ these are available up to the second derivative, and so on.

The last kernel function we shall consider for this section is the Cauchy kernel, which has hyperparameters $\sigma_f$ and $l$ and is defined by:

$$k(\mathbf{x}, \mathbf{x}'; \sigma_f, l) = \sigma_f^2 \left[ \frac{l}{\|\mathbf{x} - \mathbf{x}'\|^2 + l^2} \right].$$

This kernel is similar to the square exponential kernel, but with a Cauchy density function in place of the Gaussian density function. Jähnichen et al. (2018) state that this kernel function may be compared with the square exponential kernel, but the temporal correlations for the Cauchy kernel decay more slowly than for the square exponential kernel.

It can be shown that the sum of two or more kernels is also a 'valid' kernel in that it satisfies Definition 2.2.1 (see e.g. Wu et al., 2019). In particular, the 'double' square

exponential kernel is the sum of two square exponential kernels and is available in `GaPP`:

$$k(\mathbf{x}, \mathbf{x}'; \sigma_{f_1}, l_1, \sigma_{f_2}, l_2) = \sigma_{f_1}^2 \exp\left[-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2l_1^2}\right] + \sigma_{f_2}^2 \exp\left[-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2l_2^2}\right].$$

Here, there are two pairs of hyperparameters, one for each square exponential kernel. This makes for four hyperparameters in total, allowing for greater flexibility in modelling through the use of different length scales. Similarly, the product of two or more kernels is also a kernel. In general, the `GaPP` package (Seikel et al., 2012) allows for the specification of arbitrary sums and products of the kernels discussed within this section.

Now that we have provided a brief overview of GPs as well as some of the covariance structures employed by these GPs, we will move on to describe the theory behind GPR. In particular, we will look at two equivalent ways on how to approach the regression problem; these are known as the weight-space view and the function-space view.

## 2.3 Theoretical Background on Gaussian Process Regression

GPs offer simplicity in model fitting and prediction. In particular, they may be used for data interpolation through GPR, in which the interpolated values are modelled by a GP with prespecified prior covariance.

In GPR, the observed data is taken to follow a finite-dimensional multivariate normal distribution, which we denote as $f \sim GP(m, k)$. For simplicity, the mean function $m(\mathbf{x})$ is often taken to be $\mathbf{0}$, so only the kernel function would need to be specified. The GPR technique is mathematically closely related to standard regression analysis and similarly results in a best linear unbiased estimator of $f$ (see e.g. Krige, 1979, p. 25). Even if the normality assumption is not always satisfied in practice, GPs are often used as basic building blocks for constructing more complex non-Gaussian models.

Another way of looking at GPR is as a Bayesian inference problem. The GP prior $f \sim GP(m, k)$, usually with $m(\mathbf{x}) = \mathbf{0}$ and with one of the covariance structures described in Section 2.2, is combined with a Gaussian likelihood function for each of the observed vectors. The combination of GP prior and Gaussian likelihood results in a Gaussian posterior distribution of $f(\mathbf{x})$. The hyperparameters present in the mean and covariance structure of the GP, as well as the posterior mean and covariance, are

unknown *a priori*; instead, they are inferred from the data. In the empirical Bayes framework, a point estimate of these hyperparameters is computed; this point estimate is then used within the standard GP for the purpose of reconstruction/'fitting' of the function (see e.g. Teckentrup, 2020). Regression using GPs is described in greater detail in Section 2.3.

The aim of regression is to simulate a 'test' output $y_*$ given the input $\mathbf{x}_*$, based on a training set of observations $(\mathbf{x}_i, y_i), i \in \{1, 2, ..., n\}$. For GPR, we assume that

$$y_i = f(\mathbf{x}_i) + \epsilon_i, \tag{2.3}$$

where $f$ is the unknown function that we want to reconstruct and to which we assign a GP prior, and $\epsilon_i$ is a Gaussian noise term. Moreover, the $\mathbf{x}_i$ may be combined into a $(p \times n)$ data matrix $\mathbb{X}$, where the $(i, j)^{\text{th}}$ entry of this data matrix is the $j^{\text{th}}$ component of $\mathbf{x}_i$. The $y_i$ may also be combined into a vector $\mathbf{y}$ of dimension $n$.

The probability of $y_*$ given the available data is given by:

$$\mathbb{P}(y_*|\mathbf{x}_*, \mathbb{X}, \mathbf{y}) = \int \mathbb{P}(y_*|f, \mathbf{x}_*)\ \mathbb{P}(f|\mathbb{X}, \mathbf{y})\ \mathrm{d}f.$$

The density functions $\mathbb{P}(y_*)$ and $\mathbb{P}(f)$ are not the same in general, but the notation $\mathbb{P}$ will be used in all cases for convenience.

### 2.3.1 Weight-Space View

GPR may be approached through two opposing views, namely the weight-space view and the function-space view. In the weight-space view, we start with the 'usual' linear regression equation, i.e.:

$$f(\mathbf{x}_i) = \mathbf{x}_i^T \mathbf{w}, \quad y_i = f(\mathbf{x}_i) + \epsilon_i, \tag{2.4}$$

where the $\epsilon_i$ are independent and identically-distributed as $N(0, \sigma_\epsilon^2)$. Moreover, $f$ is the function that we want to reconstruct, and $y_i$ is the observed response, which differs from $f(\mathbf{x}_i)$ by $\epsilon_i$.

Under the usual assumption of independence of all the $\mathbf{x}_i$ and $\epsilon_i$, the marginal likelihood is obtained by multiplication of the probabilities of each observed vector of values in the training set, i.e.:

$$\mathbb{P}(\mathbf{y}|\mathbb{X}, \mathbf{w}) = \prod_{i=1}^{n} \mathbb{P}(y_i|\mathbf{x}_i, \mathbf{w}) \tag{2.5}$$

so that

$$\mathbf{Y}|\mathbb{X}, \mathbf{W} \sim MVN_n(\mathbb{X}^T\mathbf{w}, \sigma_\epsilon^2\mathbb{I}).$$

The prior distribution over the weights will be taken to be Gaussian with zero mean vector and variance-covariance matrix $\mathbb{\Sigma}$, i.e. $\mathbf{W} \sim MVN_p(\mathbf{0}, \mathbb{\Sigma})$. Then,

$$\mathbb{P}(\mathbf{y}|\mathbb{X}) = \int \mathbb{P}(\mathbf{y}|\mathbb{X}, \mathbf{w})\ \mathbb{P}(\mathbf{w})\ \mathrm{d}\mathbf{w},$$

and the posterior distribution over the weights is then derived using:

$$\begin{aligned}
\mathbb{P}(\mathbf{w}|\mathbb{X}, \mathbf{y}) &= \frac{\mathbb{P}(\mathbf{y}|\mathbb{X}, \mathbf{w})\ \mathbb{P}(\mathbf{w})}{\mathbb{P}(\mathbf{y}|\mathbb{X})} \\
&\propto \exp\left[-\frac{1}{2\sigma_\epsilon^2}(\mathbf{y} - \mathbb{X}^T\mathbf{w})^T(\mathbf{y} - \mathbb{X}^T\mathbf{w})\right] \exp\left[-\frac{1}{2}\mathbf{w}^T\mathbb{\Sigma}^{-1}\mathbf{w}\right] \\
&\propto \exp\left[-\frac{1}{2}(\mathbf{w} - \bar{\mathbf{w}})^T\left(\frac{1}{\sigma_\epsilon^2}\mathbb{X}\mathbb{X}^T + \mathbb{\Sigma}^{-1}\right)(\mathbf{w} - \bar{\mathbf{w}})\right] \\
&= \exp\left[-\frac{1}{2}(\mathbf{w} - \bar{\mathbf{w}})^T\mathbb{A}(\mathbf{w} - \bar{\mathbf{w}})\right]
\end{aligned}$$

so that

$$\mathbf{W}|\mathbb{X}, \mathbf{Y} \sim MVN_p\left(\bar{\mathbf{w}}, \mathbb{A}^{-1}\right),$$

where $\mathbb{A} = \sigma_\epsilon^{-2}\mathbb{X}\mathbb{X}^T + \mathbb{\Sigma}^{-1}$ and $\bar{\mathbf{w}} = \sigma_\epsilon^{-2}\mathbb{A}^{-1}\mathbb{X}\mathbf{y}$. The posterior mean is the same as its mode, i.e. the maximum *a posteriori* (MAP) estimate. Care should be taken when using the MAP estimate with small samples, as will be the case in our application of GPR. This is because the prior distribution taken significantly influences the MAP estimate, and there is not enough observed data for the effect of the prior to be dominated by the likelihood. On the other hand, the maximum likelihood (ML) estimate considers the frequentist approach as opposed to the Bayesian approach, and is hence based only on the observed data without taking into consideration any prior distribution. However, as the number of samples goes to infinity, the MAP and ML estimates become equal (Tsun, 2020).

The distribution of $f_* := f(\mathbf{x}_*)$ for the test input $\mathbf{x}_*$ is obtained by the average of all possible linear models $F_*|\mathbf{X}_*, \mathbf{W}$ given the Gaussian posterior distribution $\mathbf{W}|\mathbb{X}, \mathbf{Y}$, i.e.:

$$\mathbb{P}(f_*|\mathbf{x}_*, \mathbb{X}, \mathbf{y}) = \int \mathbb{P}(f_*|\mathbf{x}_*, \mathbf{w})\mathbb{P}(\mathbf{w}|\mathbb{X}, \mathbf{y})\ \mathrm{d}\mathbf{w}. \tag{2.6}$$

The linear regression model may be extended by using a set of basis functions by projecting $\mathbf{x}$ into $\zeta(\mathbf{x})$ for some general function $\zeta$, for example for use in polynomial regression. Instead of equation (2.4), the regression function therefore becomes $f(\mathbf{x}) = \zeta(\mathbf{x})^T\mathbf{w}$,

with $\mathbf{w}$ having length equal to the dimension of the space in which $\mathbf{x}$ is projected. For example, polynomial regression may be implemented by projecting an input $\mathbf{x}$ into $\zeta(\mathbf{x}) = (\mathbf{1}, \mathbf{x}, \mathbf{x}^2, \mathbf{x}^3, ..., \mathbf{x}^n)^T$. In general, we define the kernel $k$ using:

$$k(\mathbf{x}, \mathbf{x}') := \zeta(\mathbf{x})^T \mathbb{\Sigma} \zeta(\mathbf{x}') = \psi(\mathbf{x}) \cdot \psi(\mathbf{x}'),$$

where $\psi(\mathbf{x}) := \mathbb{\Sigma}^{\frac{1}{2}} \zeta(\mathbf{x})$. The reformulation in terms of $\psi$ is known as the kernel trick and is advantageous in that the computation is made simpler (see e.g. Pedregosa et al., 2011).

If we define $\mathbb{Z}_{\mathbb{X}}$ to be made up of all the columns $\zeta(\mathbf{x})$ for $\mathbf{x}$ in the training set, then, instead of equation (2.5), the marginal likelihood equation becomes:

$$\mathbb{P}(\mathbf{y}|\mathbb{Z}_{\mathbb{X}}, \mathbf{w}) = \prod_{i=1}^{n} \mathbb{P}\big(y_i|\zeta(\mathbf{x}_i), \mathbf{w}\big) \tag{2.7}$$

and therefore $\mathbf{Y}|\mathbb{Z}_{\mathbb{X}}, \mathbf{W} \sim MVN_n\left(\mathbb{Z}_{\mathbb{X}}^T \mathbf{w}, \sigma_\epsilon^2 \mathbb{I}\right)$. Similarly, from equation (2.6), the posterior predictive distribution becomes:

$$\mathbb{P}(f_*|\zeta(\mathbf{x}_*), \mathbb{Z}_{\mathbb{X}}, \mathbf{y}) = \int \mathbb{P}(f_*|\zeta(\mathbf{x}_*), \mathbf{w}) \mathbb{P}(\mathbf{w}|\mathbb{Z}_{\mathbb{X}}, \mathbf{y}) \, \mathrm{d}\mathbf{w}$$

and hence $F_*|\zeta(\mathbf{x}_*), \mathbb{Z}_{\mathbb{X}}, \mathbf{Y} \sim N\left(\sigma_\epsilon^{-2} \zeta(\mathbf{x}_*)^T \mathbb{A}^{-1} \mathbb{Z}_{\mathbb{X}} \mathbf{y}, \zeta(\mathbf{x}_*)^T \mathbb{A}^{-1} \zeta(\mathbf{x}_*)\right)$. Inversion of the matrix $\mathbb{A}$ may be avoided through a suitable matrix decomposition.

### 2.3.2 Function-Space View

In the function-space view, the regression problem is approached as a regression over functions. Starting from a GP prior $f \sim GP(m, k)$, where

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$$

and $k(\mathbf{x}, \mathbf{x}')$ is defined as in equation (2.1), we derive the mean and covariance using:

$$\begin{aligned} \mathbb{E}[f(\mathbf{x})] &= \zeta(\mathbf{x})^T \mathbb{E}[\mathbf{w}] = m(\mathbf{x}), \\ \mathbb{E}[f(\mathbf{x}) f(\mathbf{x}')] &= \zeta(\mathbf{x})^T \mathbb{E}[\mathbf{w}\mathbf{w}^T] \zeta(\mathbf{x}') = \zeta(\mathbf{x})^T \mathbb{\Sigma} \zeta(\mathbf{x}'). \end{aligned}$$

The posterior distribution is then the joint multivariate Gaussian distribution:

$$\mathbf{F}|\mathbb{X} \sim MVN_n(\mathbf{m}, \mathbb{K}),$$

where $\mathbb{K}$ and $\mathbf{m}$ are as in Definition 2.1.1. As in the weight-space view, the mean vector is usually taken to be zero, i.e. $\mathbf{m} = \mathbf{0}$, for simplicity. This is because the GP is flexible enough to model the mean arbitrarily well.

Corresponding with the data matrix $\mathbb{X}$ made up of the observations in the training set, we define a similar matrix $\mathbb{X}_*$ which is made up of the test points. The dimensionality of this matrix is therefore $(p \times n_*)$, where $n_*$ is the number of test points. The kernel function is evaluated at $\mathbb{X}_*$, generating a random Gaussian vector $\mathbf{F}_*$ which is analogous to $\mathbf{F}$, but corresponding to the test points, i.e.:

$$\mathbf{F}_* \sim MVN_{n_*}(\mathbf{0}, \mathbb{K}_{**}), \tag{2.8}$$

where $\mathbb{K}_{**} := \mathbb{K}(\mathbb{X}_*, \mathbb{X}_*)$.

In GPR, given the 'training' and 'testing' data matrices $\mathbb{X}$ and $\mathbb{X}_*$, we want to predict $\mathbf{f}_* := (f(\mathbf{x}_{*1}), f(\mathbf{x}_{*2}), ..., f(\mathbf{x}_{*n_*}))^T$, i.e. the fitted function value at each of the test points $\mathbf{x}_{*1}, \mathbf{x}_{*2}, ..., \mathbf{x}_{*n_*}$. Observed data may either be considered as 'noiseless' or as 'noisy' data. In the noise-free case, GPR produces a function that perfectly interpolates the observed data points. Therefore, if we feed the fitted function some point $\mathbf{x}$ that was already present in the observed data, we would obtain as output $f(\mathbf{x})$ without any uncertainty surrounding this output. On the other hand, in the noisy case we do not expect a perfect interpolator, but expect that any predicted value contains a level of uncertainty. This level of uncertainty is quite low for points interpolated between two observed values, but the uncertainty increases rapidly for vectors outside the range of observed values. In other words, GPR performs very well for interpolation but less so for extrapolation beyond the range of observed values. The noiseless case is not often considered in practice, but may be appropriate in some limited cases, for example in computer simulations (Sacks et al., 1989).

For the noiseless case, i.e. the model $y_i = f(\mathbf{x}_i)$ for $i \in \{1, 2, ..., n\}$, then the joint distribution is of the form:

$$\begin{pmatrix} \mathbf{f} \\ \mathbf{f}_* \end{pmatrix} \sim MVN_{n+n_*} \left[ \begin{pmatrix} \mathbf{m} \\ \mathbf{m}_* \end{pmatrix}, \begin{pmatrix} \mathbb{K} & \mathbb{K}_* \\ \mathbb{K}_*^T & \mathbb{K}_{**} \end{pmatrix} \right], \tag{2.9}$$

where $\mathbb{K} := \mathbb{K}(\mathbb{X}, \mathbb{X})$ as in Definition 2.1.1, $\mathbb{K}_* := \mathbb{K}(\mathbb{X}, \mathbb{X}_*)$, and $\mathbb{K}_{**} = \mathbb{K}(\mathbb{X}_*, \mathbb{X}_*)$ as defined in equation (2.8). Moreover, the mean vector $\mathbf{m} := (m(\mathbf{x}_1), m(\mathbf{x}_2), ..., m(\mathbf{x}_n))^T$ as in Definition 2.1.1 and similarly $\mathbf{m}_* := (m(\mathbf{x}_{*1}), m(\mathbf{x}_{*2}), ..., m(\mathbf{x}_{*n_*}))^T$.

Hence, the posterior distribution of $\mathbf{F}_*$ has the form:

$$\mathbf{F}_* | \mathbb{X}_*, \mathbb{X}, \mathbf{F} \sim MVN_{n_*} \left( \mathbf{m}_* + \mathbb{K}_*^T \mathbb{K}^{-1} (\mathbf{f} - \mathbf{m}), \mathbb{K}_{**} - \mathbb{K}_*^T \mathbb{K}^{-1} \mathbb{K}_* \right). \tag{2.10}$$

Additionally, the mean may be taken to be $\mathbf{0}$ for simplicity as discussed earlier. Therefore, under the zero mean assumption, the posterior predictive mean becomes $\mathbb{K}_*^T\mathbb{K}^{-1}\mathbf{f}$. Hence, given some test input $\mathbf{x}_*$, then the posterior predictive distribution for $f_* := f(\mathbf{x}_*)$ may be written as:

$$F_*|\mathbb{X}_*, \mathbb{X}, \mathbf{F} \sim N\left(\mathbf{k}_*^T\mathbb{K}^{-1}\mathbf{f}, k_{**} - \mathbf{k}_*^T\mathbb{K}_y^{-1}\mathbf{k}_*\right), \tag{2.11}$$

where $\mathbf{k}_* = (k(\mathbf{x}_*, \mathbf{x}_1), ..., k(\mathbf{x}_*, \mathbf{x}_n))^T$ and $k_{**} = k(\mathbf{x}_*, \mathbf{x}_*)$.

In the noisy case, the observed function is assumed to be of the form $y_i = f(\mathbf{x}_i) + \epsilon_i$, where a random noise term $\epsilon_i \sim N(0, \sigma_\epsilon^2)$ has been included. Therefore, an additional term has to be included within the covariance of the observed responses, i.e.:

$$\mathrm{Cov}(Y_i, Y_j) = k(\mathbf{x}_i, \mathbf{x}_j) + \sigma_\epsilon^2\delta(i, j),$$

where $\delta(i, j)$ is the Kronecker delta. Hence,

$$\mathrm{Cov}(Y|\mathbb{X}) = \mathbb{K} + \sigma_\epsilon^2\mathbb{I}_N := \mathbb{K}_y,$$

The joint density function in this case is given by:

$$\begin{pmatrix}\mathbf{y}\\ \mathbf{f}_*\end{pmatrix} \sim MVN_{n+n_*}\left[\begin{pmatrix}\mathbf{m}\\ \mathbf{m}_*\end{pmatrix}, \begin{pmatrix}\mathbb{K}_y & \mathbb{K}_*\\ \mathbb{K}_*^T & \mathbb{K}_{**}\end{pmatrix}\right], \tag{2.12}$$

while the marginal likelihood is given by:

$$\mathbb{P}(\mathbf{y}|\mathbb{X}) = \int \mathbb{P}(\mathbf{y}|\mathbf{f}, \mathbb{X})\mathbb{P}(\mathbf{f}|\mathbb{X})\ \mathrm{d}\mathbf{f} \tag{2.13}$$

and hence the logarithm of this quantity may be written in terms of the kernel matrix as:

$$\log\mathbb{P}(\mathbf{y}|\mathbb{X}) = -\frac{1}{2}\mathbf{y}^T\mathbb{K}_y^{-1}\mathbf{y} - \frac{1}{2}\log|\mathbb{K}_y| - \frac{n}{2}\log 2\pi. \tag{2.14}$$

In the right-hand side, the first two terms represent the data fit and model complexity respectively, while the last term is simply a constant. Moreover, for the noisy case, the posterior predictive distribution is given by:

$$\mathbf{F}_*|\mathbb{X}_*, \mathbb{X}, \mathbf{F} \sim MVN_{n_*}\left(\mathbf{m}_* + \mathbb{K}_*^T\mathbb{K}_y^{-1}(\mathbf{y} - \mathbf{m}), \mathbb{K}_{**} - \mathbb{K}_*^T\mathbb{K}_y^{-1}\mathbb{K}_*\right). \tag{2.15}$$

Equations (2.12) and (2.15) are very similar to equations (2.9) and (2.10), except that we have $\mathbf{y} := f(\mathbf{x}) + \boldsymbol{\epsilon}$ instead of $\mathbf{f}$ and $\mathbb{K}_y$ instead of $\mathbb{K}$. In addition, for simplification, the

mean is often taken to be zero also in this case. Under this zero mean assumption, and assuming a single test point $\mathbf{x}_*$, then the posterior predictive mean becomes $\mathbf{k}_*^T \mathbb{K}_y^{-1} \mathbf{y}$, similarly to equation (2.11).

In order to avoid any computational issues arising from inversion of the matrix $\mathbb{K}_y$, a Cholesky decomposition of $\mathbb{K}_y$ may be carried out, i.e. $\mathbb{K}_y = \mathbb{L}\mathbb{L}^T$ for some $\mathbb{L}$.

The most naïve way of estimating the kernel parameters is to carry out an exhaustive search over some finite grid of values and minimising validation loss, but this is rather slow. An empirical Bayes approach, which aims to maximise the marginal log-likelihood of equation (2.14), may be considered instead. In order to maximise this quantity, we differentiate with respect to each hyperparameter and set the derivative to zero. Therefore, let the hyperparameters form a vector $\boldsymbol{\theta}$. It then follows that the derivative of the marginal log-likelihood with respect to each $\theta_i$ is given by:

$$\begin{aligned} \frac{\partial}{\partial \theta_i} \log \mathbb{P}(\mathbf{y}|\mathbb{X}) &= \frac{1}{2} \mathbf{y}^T \mathbb{K}_y^{-1} \frac{\partial \mathbb{K}_y}{\partial \theta_i} \mathbb{K}_y^{-1} \mathbf{y} - \frac{1}{2} \mathrm{Tr} \left[ \mathbb{K}_y^{-1} \frac{\partial \mathbb{K}_y}{\partial \theta_i} \right] \\ &= \frac{1}{2} \mathrm{Tr} \left[ (\boldsymbol{\beta}\boldsymbol{\beta}^T - \mathbb{K}_y^{-1}) \frac{\partial \mathbb{K}_y}{\partial \theta_i} \right], \end{aligned} \tag{2.16}$$

where $\mathbf{y}^T \mathbb{K}_y^{-1} \frac{\partial \mathbb{K}_y}{\partial \theta_i} \mathbb{K}_y^{-1} \mathbf{y}$ is a scalar and $\boldsymbol{\beta} = \mathbb{K}_y^{-1} \mathbf{y}$.

In Section 2.4, we discuss the application of GPR to the Hubble tension problem and present our results. An introduction to this problem was provided in Section 1.4. In particular, the GPR procedure is repeated for each combination of dataset, prior, and covariance kernel.

## 2.4 Application to $H_0$ Data

A number of GP regressions (i.e. regressing $H(z)$ against $z$) will be carried out on the different datasets as well as on the combined data. We are interested in predicting $H_0$, i.e. the value for $H$ at $z = 0$. This is unobservable and would need to be extrapolated from the regression. The datasets to be considered for this exercise are CC, CC+SN, CC+SN+BAO, where 'CC+SN' denotes concatenation of the CC and SN data points, while 'CC+SN+BAO' refers to the full data as shown in Tables 1.1 through 1.3. Moreover, each of the values $\hat{H}_0^R$, $\hat{H}_0^{HW}$, $\hat{H}_0^{TRGB}$, $\hat{H}_0^{CM}$, $\hat{H}_0^P$ and $\hat{H}_0^{DES}$, as defined in Section 1.3 and summarised in Table 2.1, will individually be considered as additional data points at $z = 0$. It is important to note the tension present between

these different estimates of $H_0$, as discussed in Section 1.3. In the cosmological and astrophysical spheres, these estimates for $H_0$ are generally referred to as 'priors', as they reflect our prior knowledge about $H_0$, i.e. before reconstructing the $H(z)$-against-$z$ function. However, such terminology is not to be confused with the usual statistical definition of prior, that being the prior distribution within a Bayesian context. The GP regression in our case is a Bayesian problem, where the GP prior $f \sim GP(m, k)$ is combined with a Gaussian likelihood to produce a Gaussian posterior distribution, as described in Section 2.3. The case in which none of these data points are included will also be considered; we shall refer to this latter case as the 'priorless' case.

| **Value** | $\hat{H}_0$ **(km s$^{-1}$ Mpc$^{-1}$)** | **Source** |
|---|---|---|
| $\hat{H}_0^R$ | $74.22 \pm 1.82$ | Riess et al. (2019) |
| $\hat{H}_0^{TRGB}$ | $69.8 \pm 1.9$ | Freedman et al. (2019) |
| $\hat{H}_0^{HW}$ | $73.3^{+1.7}_{-1.8}$ | Wong et al. (2019) |
| $\hat{H}_0^{CM}$ | $75.35 \pm 1.68$ | Camarena and Marra (2020) |
| $\hat{H}_0^P$ | $67.4 \pm 0.5$ | Aghanim et al. (2020b) |
| $\hat{H}_0^{DES}$ | $67.4^{+1.1}_{-1.2}$ | Abbott et al. (2016) |

Table 2.1: Values of different estimates for $H_0$ to be used as 'priors'

This analysis will be carried out using each of the kernel functions defined in Section 2.2. The Python package `GaPP` (Seikel et al., 2012), suitably modified for our case, will be used for this analysis.

We obtain a reconstruction of the relationship between $H(z)$ and $z$ that is independent of any physical or cosmological models through the minimisation of a chi-squared statistic that measures the discrepancies between the observed points at redshifts $z_1, z_2, ..., z_N$ (Escamilla-Rivera et al., 2021), as in equation (2.17):

$$\chi^2_{\hat{H}} = \sum_{i=1}^{N} \frac{\left[\hat{H}_{pred}(z_i) - H_{obs}(z_i)\right]^2}{\hat{\sigma}^2_H(z_i)}, \tag{2.17}$$

where $\hat{H}_{pred}(z_i)$ and $H_{obs}(z_i)$ are the reconstructed and observed values of $H$ at redshift $z_i$, and $\hat{\sigma}^2_H(z_i)$ is the variance of the observation $H_{obs}(z_i)$. This chi-squared statistic reaches a minimum value of 0 when the observed and reconstructed $H(z_i)$ values are equal for each redshift value $z_i$. The statistic in equation (2.17) is highly similar to that used in equation (3.12), and minimisation of this statistic is equivalent to estimating $H(z)$ using ML (see e.g. Browne, 1973).

The relationship between $H(z)$ and $z$ is expected to be non-linear. In particular, we expect the value of $H(z)$ to increase more rapidly as $z$ increases. This is due to the accelerating expansion of the Universe (Riess et al., 1998) and is in line with results already in the literature (see e.g. Briffa et al., 2022). The $\hat{H}_0$ value for a given combination of dataset, kernel, and 'prior' $H_0$ estimate is taken to be the extrapolated value of the reconstructed function at $z = 0$. Extrapolation is necessary for this case since it is not possible to take observations directly at $z = 0$.

Table 2.2 shows the $\hat{H}_0$ values obtained after running the regressions on the different datasets and using the square exponential kernel function. The first three rows contain the estimated values of $H_0$ for the case where no 'prior' values of $H_0$ were added for each of the three datasets considered. Each additional set of three rows similarly represents the regression carried out with each 'prior' $H_0$ estimate added in turn for each of the datasets. Additionally, the 'distance' between each reconstructed value for $H_0$ and the 'prior' estimates for $H_0$ is calculated in the respective rows. The distance between two $\hat{H}_0$ values $\hat{H}_{0,i}$ and $\hat{H}_{0,j}$ is defined by:

$$d(\hat{H}_{0,i}, \hat{H}_{0,j}) := \frac{\hat{H}_{0,i} - \hat{H}_{0,j}}{\sqrt{\hat{\sigma}^2_{\hat{H}_{0,i}} + \hat{\sigma}^2_{\hat{H}_{0,j}}}}, \tag{2.18}$$

where $\hat{\sigma}_{\hat{H}_{0,i}}$ and $\hat{\sigma}_{\hat{H}_{0,j}}$ represent the standard deviations of the two estimates. We shall refer to this distance measure as the '$\sigma$-distance'. Wherever relevant, significant discrepancies of more than 2 units are presented in bold. While this threshold of 2 units is arbitrary, this is comparable to the threshold of 1.96 used as the 95% confidence limit of the normal distribution. Similar tables for the other kernel functions are presented in Appendix C.

The main takeaway from these results is that the $H_0$ value supplied as a 'prior' greatly affects the final result, irrespective of the kernel function being used. While each prior results in a slightly different estimate for $\hat{H}_0$, the different kernels largely agree to within 1 unit of distance as defined in equation (2.18). The inclusion of $\hat{H}_0$ values as described in Table 2.1, however, generally results in higher estimated values for $H_0$, since these 'prior' values are obtained from late-Universe measurements. The exceptions to this are the $\hat{H}_0^P$ and $\hat{H}_0^{DES}$, since they are related to CMB radiation. Moreover, the inclusion of a 'prior' $H_0$ estimate led to narrower confidence bands than the priorless case, as can be seen in Table 2.2 and in Appendix C.

Another observation that may be made is that the $\hat{H}_0$ value for the priorless regression is very close, in terms of distance as calculated in equation (2.18), to the $\hat{H}_0^P$ and $\hat{H}_0^{DES}$

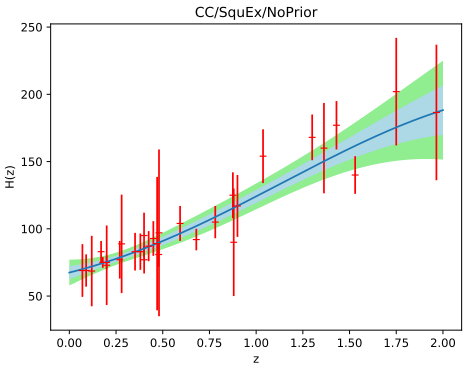
estimates. This is especially true for the CC dataset and the square exponential and double square exponential kernel functions. Furthermore, inclusion of the Planck or DES 'priors' in this regression understandably does not affect the regression much in these cases. This may be evidence in favour of a lower value of $H_0$. On the other hand, the largest discrepancies in terms of distance are encountered when adding the Riess, H0LiCoW, and CM 'priors'. The discrepancies between the results obtained using these two 'groups' of 'priors' may be attributed to the fact that the Planck and DES 'priors' assume the $\Lambda CDM$ model, while the Riess, H0LiCOW and CM 'priors' do not.

The dataset used, that is, CC or CC+SN or CC+SN+BAO, is also an important factor to consider, with the larger datasets understandably providing more confident predictions with lower errors. However, in all cases, care should be taken when interpreting results as the sample size is quite small, and also since GPs are prone to overfitting. Using the CC+SN data generally led to a higher value of $\hat{H}_0$ compared with when just the CC data was used; this may be since SN measurements are late-Universe. On the other hand, including also the BAO data usually led to a comparatively lower $\hat{H}_0$ value, possibly because the BAO data relates to the structure of galaxies and therefore represents an earlier, more distant Universe.

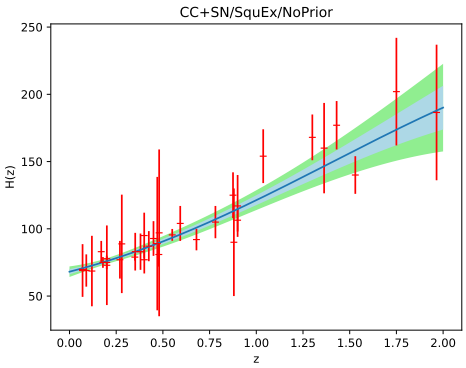
For the square exponential kernel function, reconstruction plots of $H(z)$ against $z$ are shown in Figures 2.1 through 2.7. The observed data points in each case and their corresponding errors are shown in each plot by red vertical lines. The reconstructed function is shown as a dark blue line and is surrounded by a light blue $1\sigma$ and light green $2\sigma$ confidence regions. As was deduced also from the respective Table 2.2, the uncertainty decreases considerably when considering the full CC+SN+BAO dataset when compared to CC and CC+SN, irrespective of the 'prior' $H_0$ estimate used. This is especially true for $z > 1.5$, owing to the fact that data points are available at higher values of $z$ within the BAO dataset. Similar plots for the other kernel functions are given in Appendix B. The estimates obtained using the other kernel functions are very similar to each other, which indicates that in this case the kernel used does not significantly affect the estimate obtained.

| **Dataset** | $\hat{H}_0$ (km s$^{-1}$ Mpc$^{-1}$) | $d(\hat{H}_0, \hat{H}_0^R)$ | $d(\hat{H}_0, \hat{H}_0^{TRGB})$ | $d(\hat{H}_0, \hat{H}_0^{HW})$ | $d(\hat{H}_0, \hat{H}_0^{CM})$ | $d(\hat{H}_0, \hat{H}_0^P)$ | $d(\hat{H}_0, \hat{H}_0^{DES})$ |
|---|---|---|---|---|---|---|---|
| CC | $67.448 \pm 4.753$ | $-1.3305$ | $-0.4595$ | $-1.1592$ | $-1.5674$ | 0.0100 | 0.0098 |
| CC+SN | $68.104 \pm 1.786$ | $\mathbf{-2.3983}$ | $-0.6503$ | $\mathbf{-2.1071}$ | $\mathbf{-2.9549}$ | 0.3796 | 0.3357 |
| CC+SN+BAO | $68.903 \pm 1.595$ | $\mathbf{-2.1971}$ | $-0.3617$ | $-1.8863$ | $\mathbf{-2.7830}$ | 0.8989 | 0.7755 |
| CC+$\hat{H}_0^R$ | $73.833 \pm 1.728$ | $-0.1543$ | 1.5701 | 0.2198 | $-0.6294$ | **3.5752** | **3.1399** |
| CC+SN+$\hat{H}_0^R$ | $71.646 \pm 1.334$ | $-1.1406$ | 0.7953 | $-0.7653$ | $-1.7265$ | **2.9810** | **2.4561** |
| CC+SN+BAO+$\hat{H}_0^R$ | $71.547 \pm 1.248$ | $-1.2112$ | 0.7684 | $-0.8312$ | $-1.8171$ | **3.0838** | **2.4924** |
| CC+$\hat{H}_0^{TRGB}$ | $69.602 \pm 1.773$ | $-1.8174$ | $-0.0761$ | $-1.5054$ | $\mathbf{-2.3531}$ | 1.1954 | 1.0555 |
| CC+SN+$\hat{H}_0^{TRGB}$ | $69.011 \pm 1.315$ | $\mathbf{-2.3197}$ | $-0.3414$ | $-1.9954$ | $\mathbf{-2.9710}$ | 1.1451 | 0.9397 |
| CC+SN+BAO+$\hat{H}_0^{TRGB}$ | $69.295 \pm 1.226$ | $\mathbf{-2.2443}$ | $-0.2232$ | $-1.9107$ | $\mathbf{-2.9113}$ | 1.4315 | 1.1507 |
| CC+$\hat{H}_0^{HW}$ | $72.998 \pm 1.622$ | $-0.5014$ | 1.2802 | $-0.1286$ | $-1.0074$ | **3.2988** | **2.8568** |
| CC+SN+$\hat{H}_0^{HW}$ | $71.269 \pm 1.278$ | $-1.3269$ | 0.6415 | $-0.9550$ | $-1.9333$ | **2.8190** | **2.2943** |
| CC+SN+BAO+$\hat{H}_0^{HW}$ | $71.225 \pm 1.202$ | $-1.3734$ | 0.6337 | $-0.9969$ | $-1.9972$ | **2.9382** | **2.3475** |
| CC+$\hat{H}_0^{CM}$ | $74.970 \pm 1.611$ | 0.3088 | **2.0758** | 0.7133 | $-0.1631$ | **4.4887** | **3.8813** |
| CC+SN+$\hat{H}_0^{CM}$ | $72.624 \pm 1.290$ | $-0.7154$ | 1.2296 | $-0.3168$ | $-1.2870$ | **3.7755** | **3.0812** |
| CC+SN+BAO+$\hat{H}_0^{CM}$ | $72.409 \pm 1.213$ | $-0.8282$ | 1.1572 | $-0.4269$ | $-1.4195$ | **3.8173** | **3.0586** |
| CC+$\hat{H}_0^P$ | $67.393 \pm 0.497$ | $\mathbf{-3.6188}$ | $-1.2258$ | $\mathbf{-3.3353}$ | $\mathbf{-4.5419}$ | $-0.0105$ | $-0.0061$ |
| CC+SN+$\hat{H}_0^P$ | $67.432 \pm 0.481$ | $\mathbf{-3.6064}$ | $-1.2085$ | $\mathbf{-3.3219}$ | $\mathbf{-4.5317}$ | 0.0454 | 0.0263 |
| CC+SN+BAO+$\hat{H}_0^P$ | $67.517 \pm 0.475$ | $\mathbf{-3.5635}$ | $-1.1657$ | $\mathbf{-3.2762}$ | $\mathbf{-4.4865}$ | 0.1695 | 0.0976 |
| CC+$\hat{H}_0^{DES}$ | $67.366 \pm 1.070$ | $\mathbf{-3.2464}$ | $-1.1162$ | $\mathbf{-2.9541}$ | $\mathbf{-4.0084}$ | $-0.0287$ | $-0.0221$ |
| CC+SN+$\hat{H}_0^{DES}$ | $67.533 \pm 0.931$ | $\mathbf{-3.2713}$ | $-1.0716$ | $\mathbf{-2.9756}$ | $\mathbf{-4.0701}$ | 0.1258 | 0.0922 |
| CC+SN+BAO+$\hat{H}_0^{DES}$ | $67.831 \pm 0.896$ | $\mathbf{-3.1497}$ | $-0.9373$ | $\mathbf{-2.8462}$ | $\mathbf{-3.9494}$ | 0.4204 | 0.3040 |

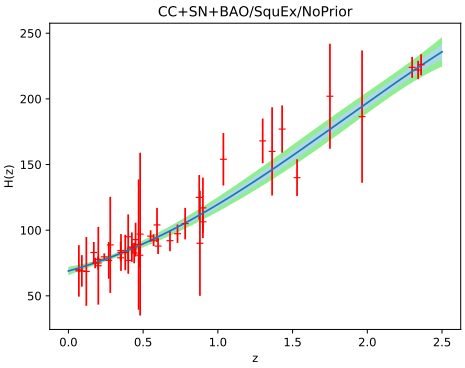
Table 2.2: GPR results with square exponential kernel function




(a) CC

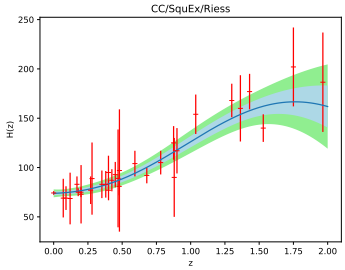



(b) CC+SN

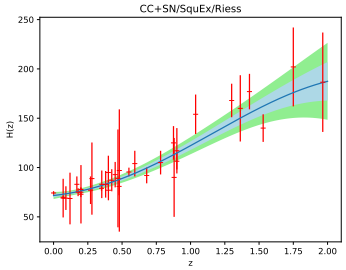



(c) CC+SN+BAO

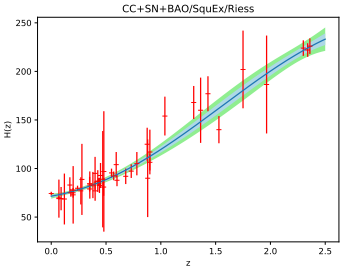
Figure 2.1: GPR reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and without adding any $H_0$ estimate to the data




(a) CC

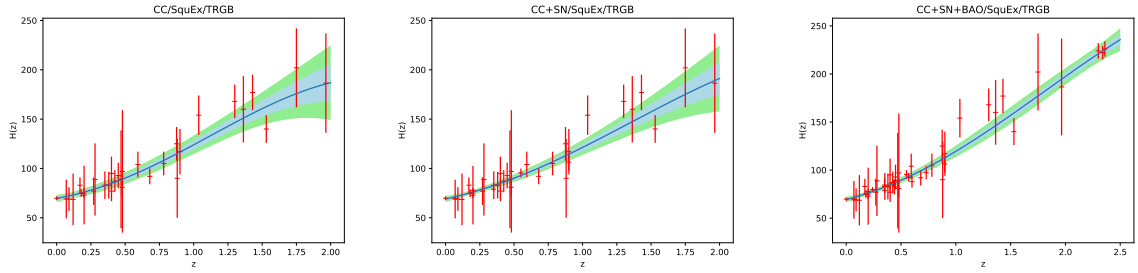



(b) CC+SN



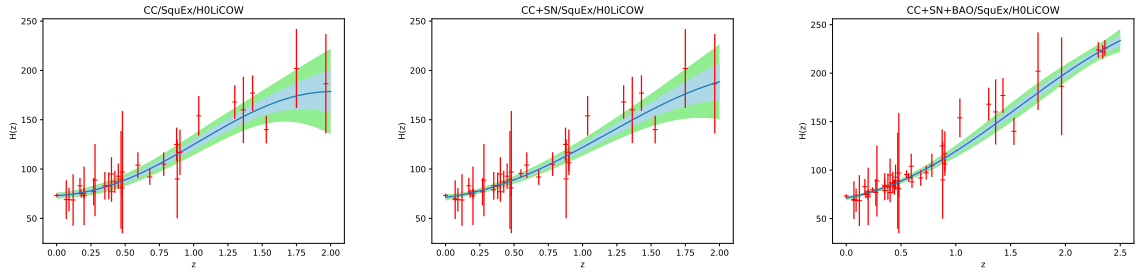

(c) CC+SN+BAO

Figure 2.2: GPR reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^R$



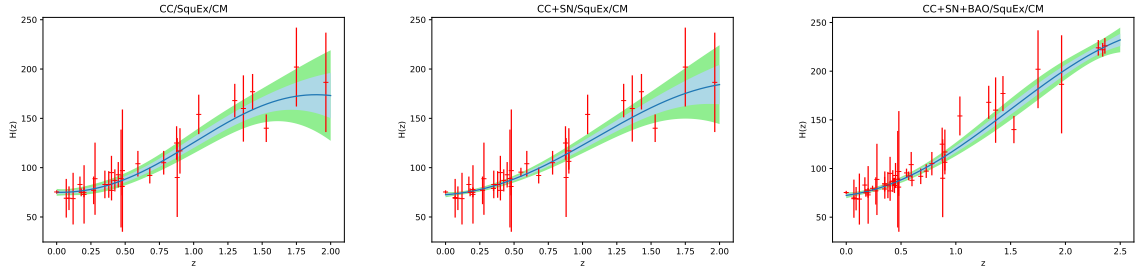

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure 2.3: GPR reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^{TRGB}$



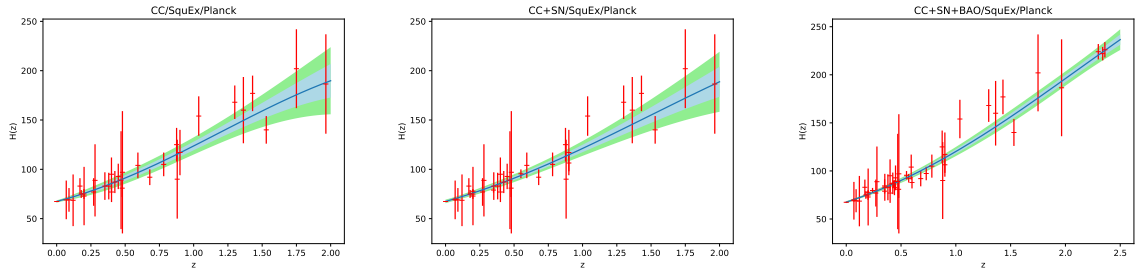

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure 2.4: GPR reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^{HW}$



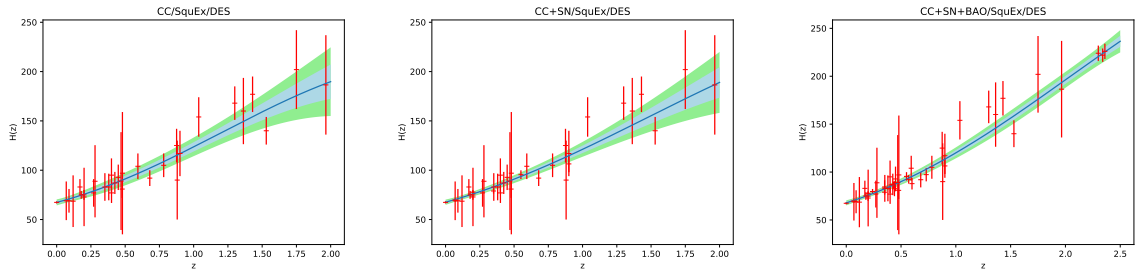

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure 2.5: GPR reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^{CM}$




(a) CC (b) CC+SN (c) CC+SN+BAO

Figure 2.6: GPR reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^P$




(a) CC (b) CC+SN (c) CC+SN+BAO

Figure 2.7: GPR reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^{DES}$

Given a particular dataset and 'prior' $\hat{H}_0$ value, the question of comparing the goodness of fit of the different kernel functions arises. This may be done through the use of a goodness of fit criterion, particularly the logarithm of the marginal likelihood of each model as defined in equation (2.7). One possible criticism of using this metric is that it is prone to selecting overfitted models, with the likelihood being maximised when the number of parameters is equal to the number of data points (Mohammed and Cawley, 2017). To mitigate this problem, other goodness of fit measures, such as the Deviance Information Criterion (DIC), have been devised and may be found in the literature; in particular, the DIC as proposed by Spiegelhalter et al. (2002) penalises more complex models by applying a penalty term on the number of parameters. However, in our case the only parameters involved are the hyperparameters of the kernel functions, and therefore this issue should not concern us.

Tables 2.3 to 2.5 show the marginal log-likelihood values obtained for each combination of kernel function, 'prior' $H_0$ value, with one table corresponding to each dataset. Two values obtained with a different prior and/or dataset are not comparable with each other; therefore, only the likelihood values within each column of a table may be compared separately. In each case, the kernel function that provides the best fit (i.e. the one with the highest log-likelihood value) is shown in bold.

The results in these tables confirm our earlier understanding that the different kernel functions do not provide results that are significantly different from each other. Nevertheless, some general observations may be made, for example that the double square exponential kernel function is the one that most often provides the best likelihood value. This may be since it contains two pairs of parameters and is therefore more flexible than, for example, the square exponential kernel.

Throughout the remainder of this dissertation, we consider other statistical methods and obtain further point estimates for the Hubble constant in each case, as mentioned in Section 1.5. Chapter 3 deals with parametric and non-parametric MCMC-based approaches to estimating the Hubble constant. In particular, the non-parametric approach of Section 3.4 is similar to the GPR approach discussed so far, except that the kernel hyperparameters are inferred using an MCMC sampler as opposed to through the minimisation of a likelihood function. On the other hand, Chapter 4 is concerned with TPR, which is similar to GPR but makes use of the multivariate Student's $t$-distribution as opposed to the multivariate Gaussian distribution. Also, in Section 5.1, we consider heteroscedastic versions of GPR and TPR.

| | None | $\hat{H}_0^R$ | $\hat{H}_0^{TRGB}$ | $\hat{H}_0^{HW}$ | $\hat{H}_0^{CM}$ | $\hat{H}_0^P$ | $\hat{H}_0^{DES}$ |
|---|---|---|---|---|---|---|---|
| **Square Exponential** | $-131.6806$ | $-135.6061$ | $-135.2328$ | $-135.5004$ | $-135.7448$ | $-135.0666$ | $-135.0880$ |
| **Double Square Exponential** | $-132.0972$ | $\mathbf{-134.9152}$ | $\mathbf{-134.8374}$ | $\mathbf{-134.8654}$ | $-139.4666$ | $-137.0009$ | $\mathbf{-134.8666}$ |
| **Rational Quadratic** | $-131.4817$ | $-135.0744$ | $-135.2290$ | $-135.4987$ | $\mathbf{-135.1578}$ | $-135.9966$ | $-134.8789$ |
| **Matérn** $\left(\nu = \frac{3}{2}\right)$ | $\mathbf{-131.3184}$ | $-135.3794$ | $-135.0255$ | $-135.2725$ | $-135.5250$ | $-134.9619$ | $-134.9764$ |
| **Matérn** $\left(\nu = \frac{5}{2}\right)$ | $-131.3869$ | $-135.3496$ | $-134.9757$ | $-135.2384$ | $-135.4997$ | $\mathbf{-134.8708}$ | $-134.8904$ |
| **Matérn** $\left(\nu = \frac{7}{2}\right)$ | $-131.4783$ | $-135.9491$ | $-135.0395$ | $-135.2978$ | $-135.5507$ | $-134.9161$ | $-134.9369$ |
| **Matérn** $\left(\nu = \frac{9}{2}\right)$ | $-131.5310$ | $-135.4482$ | $-135.0835$ | $-135.3415$ | $-135.5902$ | $-134.9494$ | $-134.9706$ |
| **Cauchy** | $-131.4132$ | $-135.2515$ | $-134.9564$ | $-135.1562$ | $-135.3797$ | $-134.8762$ | $-134.8969$ |

Table 2.3: Marginal log-likelihood values for each kernel function and prior using CC dataset

| | None | $\hat{H}_0^R$ | $\hat{H}_0^{TRGB}$ | $\hat{H}_0^{HW}$ | $\hat{H}_0^{CM}$ | $\hat{H}_0^P$ | $\hat{H}_0^{DES}$ |
|---|---|---|---|---|---|---|---|
| **Square Exponential** | $-147.1093$ | $-150.4788$ | $-148.2581$ | $-149.8828$ | $-151.6133$ | $-147.7594$ | $-147.8809$ |
| **Double Square Exponential** | $-146.9432$ | $\mathbf{-149.9851}$ | $\mathbf{-148.0194}$ | $\mathbf{-149.4278}$ | $\mathbf{-151.0232}$ | $-148.5970$ | $-150.1677$ |
| **Rational Quadratic** | $-147.1000$ | $-150.4775$ | $-148.1432$ | $-150.6641$ | $-151.6127$ | $-147.7590$ | $-149.4567$ |
| **Matérn** $\left(\nu = \frac{3}{2}\right)$ | $-146.9506$ | $-150.6652$ | $-148.3577$ | $-150.0882$ | $-151.7150$ | $-147.8004$ | $-147.8841$ |
| **Matérn** $\left(\nu = \frac{5}{2}\right)$ | $\mathbf{-146.9238}$ | $-150.4360$ | $-148.1220$ | $-149.8281$ | $-151.6043$ | $\mathbf{-147.6416}$ | $\mathbf{-147.7490}$ |
| **Matérn** $\left(\nu = \frac{7}{2}\right)$ | $-147.4102$ | $-150.3948$ | $-148.1309$ | $-149.7913$ | $-151.8288$ | $-147.6586$ | $-147.7725$ |
| **Matérn** $\left(\nu = \frac{9}{2}\right)$ | $-147.0008$ | $-150.3968$ | $-148.1532$ | $-149.7961$ | $-151.5472$ | $-147.6778$ | $-149.2782$ |
| **Cauchy** | $-146.9472$ | $-150.2450$ | $-148.0722$ | $-149.6495$ | $-151.3709$ | $-147.7967$ | $-147.7655$ |

Table 2.4: Marginal log-likelihood values for each kernel function and prior using CC+SN dataset

| | None | $\hat{H}_0^R$ | $\hat{H}_0^{TRGB}$ | $\hat{H}_0^{HW}$ | $\hat{H}_0^{CM}$ | $\hat{H}_0^P$ | $\hat{H}_0^{DES}$ |
|---|---|---|---|---|---|---|---|
| **Square Exponential** | $-177.7906$ | $-180.8676$ | $-178.7633$ | $-180.2532$ | $-182.0787$ | $-178.6273$ | $-178.7040$ |
| **Double Square Exponential** | $\mathbf{-177.6535}$ | $-184.1785$ | $\mathbf{-178.5955}$ | $-182.3232$ | $-184.7039$ | $-180.1961$ | $-178.6478$ |
| **Rational Quadratic** | $-177.7902$ | $-184.4148$ | $-178.6746$ | $\mathbf{-180.1456}$ | $-182.0782$ | $-178.6270$ | $-178.7027$ |
| **Matérn** $\left(\nu = \frac{3}{2}\right)$ | $-177.8056$ | $-181.6050$ | $-179.1653$ | $-180.9880$ | $-182.7901$ | $-178.6078$ | $-178.6977$ |
| **Matérn** $\left(\nu = \frac{5}{2}\right)$ | $-177.7030$ | $-181.0126$ | $-178.7637$ | $-180.3849$ | $-182.2596$ | $\mathbf{-178.5268}$ | $\mathbf{-178.6060}$ |
| **Matérn** $\left(\nu = \frac{7}{2}\right)$ | $-177.7206$ | $-180.8992$ | $-178.7297$ | $-180.2774$ | $-182.1325$ | $-178.5683$ | $-178.6420$ |
| **Matérn** $\left(\nu = \frac{9}{2}\right)$ | $-177.7313$ | $-180.8693$ | $-178.7258$ | $-180.2501$ | $-182.0957$ | $-178.5789$ | $-178.6529$ |
| **Cauchy** | $-177.6921$ | $\mathbf{-180.7615}$ | $-178.6747$ | $-180.1491$ | $\mathbf{-181.9729}$ | $-178.5788$ | $-178.6462$ |

Table 2.5: Marginal log-likelihood values for each kernel function and prior using CC+SN+BAO dataset

# Chapter 3

# Markov Chain Monte Carlo Methods

In this chapter, the use of Monte Carlo methods on the data described in Section 1.4 will be investigated, with the end goal in this case also being to obtain estimates for the Hubble constant $H_0$. As introduced in Section 1.5.2, MCMC methods are a broad class of algorithms that may be used in order to sample from a desired joint distribution $f(\mathbf{a}) = f(a_1, a_2, ..., a_p)$ in cases where directly sampling from this joint distribution is not possible or is computationally infeasible. In particular, we will apply the affine ensemble sampler due to Goodman and Weare (2010), which makes use of a number of parallel runs of 'standard' MCMC methods such as the MH algorithm and the Gibbs sampler. Here, we shall use MCMC methods so that, as in Chapter 2, we obtain reconstructions of the $H(z)$-against-$z$ function and obtain estimates for $H_0$. Two different MCMC approaches will be considered within this chapter. The first involves using the `mcmcdgp` module available in the Python package `GaPP`, which uses the affine ensemble sampler to perform inference on the hyperparameters of a specified covariance kernel of a GP. This therefore allows for a model-independent non-parametric reconstruction of the $H(z)$-against-$z$ function. Each of the kernel functions described in Section 2.2 will be considered in this case. In the second approach, we will consider reconstructions of $H_0$ and another parameter, namely the matter density parameter $\Omega_{M0}$, based on the $\Lambda CDM$ model. This will be done using the Python package `emcee` (Foreman-Mackey et al., 2013), which also implements the same affine ensemble sampler. In the latter case, the inference is carried out directly on the parameters $H_0$ and $\Omega_{M0}$ as opposed to on the kernel hyperparameters. In both cases, this analysis will be carried out on each of the three datasets, namely CC, CC+SN, and CC+SN+BAO, as in Chapter 2.

This chapter will be structured as follows. In Section 3.1, we will give a brief description of the two MCMC methods, namely the MH algorithm and the Gibbs sampler. This is followed by Section 3.2, in which a more detailed description of the ensemble affine sampler is given. Then, in Section 3.3, we use this sampler in a parametric setting to obtain estimates for the parameters of the $\Lambda CDM$ cosmological model. Lastly, in Section 3.4, we use the same sampler in a non-parametric GPR setting to obtain estimates for the kernel hyperparameters. The relevant results are presented in each case.

## 3.1 The Metropolis-Hastings Algorithm and the Gibbs Sampler

Both the MH algorithm and the Gibbs sampler may be used to sample from some desired joint distribution $f(\mathbf{a})$, but differ slightly in the approach they take. In particular, the MH algorithm samples proposal vectors $\mathbf{a}^*$ from a proposal distribution $Q$ which is similar to the desired distribution of the data but which is considered to be easier to sample from, and accepts or rejects each sampled vector accordingly. There is a clear advantage in using a simple proposal distribution as opposed to the more complex desired distribution, since we would be avoiding having to sample from 'difficult' distributions such as highly skewed distributions. The MH algorithm was originally developed as the Metropolis algorithm (Metropolis et al., 1953), and in its original form required the proposal distribution to be symmetric. However, the Metropolis algorithm was later generalised to allow for non-symmetric proposal distributions by Hastings (1970). On the other hand, the Gibbs sampler (Geman and Geman, 1984) obtains vectors from the desired joint distribution $f(\mathbf{a})$ by sampling values from the distribution of each component conditional on the values of the other components. Both sampling methods are iterative in nature and are usually run for a prespecified number of iterations $T$ starting from some initial state $\mathbf{a}^{(0)}$ within the state space of the Markov chain.

Suppose that we wish to sample from the joint distribution $f(\mathbf{a})$. The vector $\mathbf{a}$ may be split into components, i.e. $\mathbf{a} = (\mathbf{a}_1^T, \mathbf{a}_2^T, ..., \mathbf{a}_i^T)^T$. The components are not necessarily univariate; rather, each component may instead be a small group of related coordinates as explained in Section 1.5.2. Therefore, in cases when $\mathbf{a}$ is split into components, then $i \in \{1, 2, ..., k\}$ with $k \leq p$, and $k = p$ if and only if all the components are univariate. The proposal distribution $Q$ may either be defined on a 'local' basis, that is, defining a

$q_i(\mathbf{a}, \mathbf{a}_i^*)$ for each $i \in \{1, 2, ..., k\}$, where $\mathbf{a}_i^*$ is the $i^{\text{th}}$ component of $\mathbf{a}^*$. Alternatively, it may be defined on a 'global' basis, that is, defining a single $q(\mathbf{a}, \mathbf{a}^*)$.

The way in which the MH algorithm generates a new state $\mathbf{a}'$ from the previous state $\mathbf{a}$ is as follows:

1. Generate a candidate state $\mathbf{a}^*$, which is identical to $\mathbf{a}$ in all respects except for possibly the $i^{\text{th}}$ component.
2. Accept the proposal value, that is, set $\mathbf{a}' = \mathbf{a}^*$, with probability $A(\mathbf{a}, \mathbf{a}^*)$, otherwise let $\mathbf{a}' = \mathbf{a}$.

A common choice for the acceptance probability is the Metropolis acceptance function suitably modified to eliminate this symmetry assumption, that is,

$$A(\mathbf{a}, \mathbf{a}^*) = \min \left\{ 1, \ \frac{f(\mathbf{a}^*)}{f(\mathbf{a})} \frac{q_i(\mathbf{a}^*, \mathbf{a}_i)}{q_i(\mathbf{a}_i, \mathbf{a}^*)} \right\}, \tag{3.1}$$

where the proportion $\frac{q_i(\mathbf{a}^*, \mathbf{a}_i)}{q_i(\mathbf{a}_i, \mathbf{a}^*)}$ accounts for any asymmetry within the proposal distribution.

The performance of the MH algorithm depends greatly on the acceptance rate. For example, a low acceptance rate – which might arise due to misspecification of the proposal distribution – will result in wasted computation time.

Similarly to the MH algorithm, the Gibbs sampler is a method for sampling from $f(\mathbf{a})$ through the use of Markov chains. However, it differs from MH in that it makes use of conditional distributions. This sampling procedure, also known as the heatbath algorithm and named after the American scientist J. W. Gibbs, is originally due to Geman and Geman (1984). The Gibbs sampler works by generating a homogeneous Markov chain $\mathbf{a}^{(0)}, \mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \ldots, \mathbf{a}^{(T)}$ using an iterative process. The procedure for generating $\mathbf{a}^{(t)}$ from $\mathbf{a}^{(t-1)}$ is to sample each $\mathbf{a}_i^{(t)}$ from the conditional distribution $f(\mathbf{a}_i | \{\mathbf{a}_j^{(t-1)} : j \neq i\})$:

Sample $\mathbf{a}_1^{(t)}$ from $f\left(\mathbf{a}_1 \;\middle|\; \mathbf{a}_2^{(t-1)}, \mathbf{a}_3^{(t-1)}, \ldots, \mathbf{a}_k^{(t-1)}\right)$.

Sample $\mathbf{a}_2^{(t)}$ from $f\left(\mathbf{a}_2 \;\middle|\; \mathbf{a}_1^{(t-1)}, \mathbf{a}_3^{(t-1)}, \ldots, \mathbf{a}_k^{(t-1)}\right)$.

...

Sample $\mathbf{a}_i^{(t)}$ from $f\left(\mathbf{a}_i \;\middle|\; \mathbf{a}_1^{(t-1)}, \ldots, \mathbf{a}_{i-1}^{(t-1)}, \mathbf{a}_{i+1}^{(t-1)}, \ldots, \mathbf{a}_k^{(t-1)}\right)$.

...

Sample $\mathbf{a}_k^{(t)}$ from $f\left(\mathbf{a}_k \;\middle|\; \mathbf{a}_1^{(t-1)}, \mathbf{a}_2^{(t-1)}, \ldots, \mathbf{a}_{k-1}^{(t-1)}\right)$.

Sometimes, the Gibbs sampler is defined in a slightly different way in that $\mathbf{a}_i^{(t)}$ is sampled from the conditional distribution $f\left(\mathbf{a}_i \;\middle|\; \mathbf{a}_1^{(t)}, \ldots, \mathbf{a}_{i-1}^{(t)}, \mathbf{a}_{i+1}^{(t-1)}, \ldots, \mathbf{a}_k^{(t-1)}\right)$ instead.

Since the Gibbs sampler makes use of conditional distributions, efficient use of the Gibbs sampler naturally depends on the ability to derive and sample from the required conditional distributions. Additionally, the Gibbs sampler may be viewed as a special case of the MH algorithm (see e.g. Green and Han, 1992) for which the proposal distribution is $q_i(\mathbf{a}, \mathbf{a}_i^*) = f(\mathbf{a}_i^* | \{\mathbf{a}_j : j \neq i\})$, since we are sampling $\mathbf{a}_i$ from the conditional distribution $f(\mathbf{a}_i | \{\mathbf{a}_j : j \neq i\})$. In this case, the acceptance probability equates to 1, which means that the proposal value is always accepted for Gibbs sampling.

Both the MH algorithm and the Gibbs sampler are effective ways to sample from complex distributions, but their speed and ease of use depend greatly on the specification of the models. For example, the MH algorithm requires an appropriate specification of the proposal distribution in order to ensure a high acceptance rate, while the Gibbs sampler cannot be used if deriving the conditional distributions is not possible or if the conditional distributions are difficult to sample from (Brooks et al., 2011, p. 29).

## 3.2 Ensemble Affine Samplers

The ensemble affine sampler as proposed by Goodman and Weare (2010) solves the problem of 'difficult' badly-scaled distributions by carrying out affine transformations on the coordinates, transforming the complex sampling problem into a simpler one. Moreover, the 'ensemble' refers to a number of MCMC walkers executed in parallel. Using an ensemble of walkers in parallel can lead to faster convergence when compared to traditional 'single-particle' MCMC methods such as the MH algorithm and the Gibbs sampler consisting of a single walker. These single-particle MCMC methods may also be sensitive to scaling of distributions and may not work well with badly-scaled distributions (see e.g. Goodman and Weare, 2010, p. 76).

### 3.2.1 Affine Transformations and the Scaling Problem

Goodman and Weare (2010) state that if the single-particle methods perform well for some desired distribution $f(\mathbf{a})$ in $\mathbb{R}^p$, this does not necessarily mean that they will work well with the scaled distribution $f_\lambda(\mathbf{a}) := \lambda^{-p} f(\lambda \mathbf{a})$, especially if $\lambda$ is very large or very small. The authors propose to solve this scaling problem by carrying out affine transformations on the coordinates. An affine transformation is defined as an invertible mapping from $\mathbf{a} \in \mathbb{R}^p$ to $\mathbf{b} \in \mathbb{R}^p$ of the form

$$\mathbf{b} = \mathbb{T}\mathbf{a} + \mathbf{c}, \tag{3.2}$$

where $\mathbb{T}$ is a $(p \times p)$ transformation matrix, and $\mathbf{c}$ is a constant vector of dimension $p$. Therefore, if the joint probability density function of the random vector $\mathbf{A}$ is $f(\mathbf{a})$, that of $\mathbf{B} := \mathbb{T}\mathbf{A} + \mathbf{c}$ may be expressed as

$$f_{\mathbb{T},\mathbf{c}}(\mathbf{b}) = f_{\mathbb{T},\mathbf{c}}(\mathbb{T}\mathbf{a} + \mathbf{c}) \propto f(\mathbf{a}). \tag{3.3}$$

As a simple example in $\mathbb{R}^2$, consider the joint density function on $\mathbf{a} := (a_1, a_2)^T$ defined by:

$$f(\mathbf{a}) \propto \exp\left[\frac{-(a_1 - a_2)^2}{2\epsilon} - \frac{(a_1 + a_2)^2}{2}\right]. \tag{3.4}$$

For $\epsilon$ close to 1, obtaining samples from this joint density function should not produce any significant problems, since the function is unimodal with a clear maximum, as illustrated in Figure 3.1b.

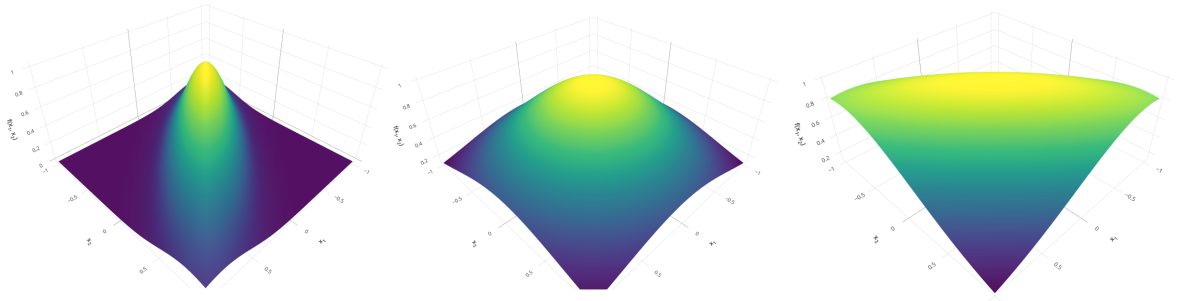
However, if the value of $\epsilon$ is significantly less than 1, then the function is almost flat except for a small region around the maximum at $\mathbf{a} = (0, 0)^T$. On the other hand, if $\epsilon$ is too large, then the function is almost flat around the maximum, as illustrated in the surface plots of Figures 3.1a and 3.1c. In both cases, single-particle MCMC methods or the 'usual' gradient-based optimisation methods may take a long time to achieve the maximum, and therefore obtaining samples from this joint density function would prove computationally expensive (Goodman and Weare, 2010, p. 66).

Let $\mathbf{b} := (b_1, b_2)^T$. Then, the affine transformation defined by $\mathbf{a} \mapsto \mathbf{b}$, where

$$(b_1, b_2)^T := \left[\frac{a_1 - a_2}{\sqrt{\epsilon}}, a_1 + a_2\right]^T$$

converts the challenging sampling problem described in equation (3.4) into the problem of sampling from

$$f_{\mathbb{T},\mathbf{c}}(\mathbf{b}) \propto \exp\left[\frac{-(b_1^2 + b_2^2)}{2}\right],$$



(a) $\epsilon = 0.1$ (b) $\epsilon = 1$ (c) $\epsilon = 10$

Figure 3.1: Surface plots of a skewed joint distribution in $\mathbb{R}^2$

which is easier to sample from since there is no saddle point. In this case, the function is unimodal with a clearly-defined maximum at $\mathbf{a} = (0, 0)^T$ that may be easily reached by either gradient optimisation methods or single-particle MCMC methods. A surface plot of this transformed density function is given in Figure 3.2.

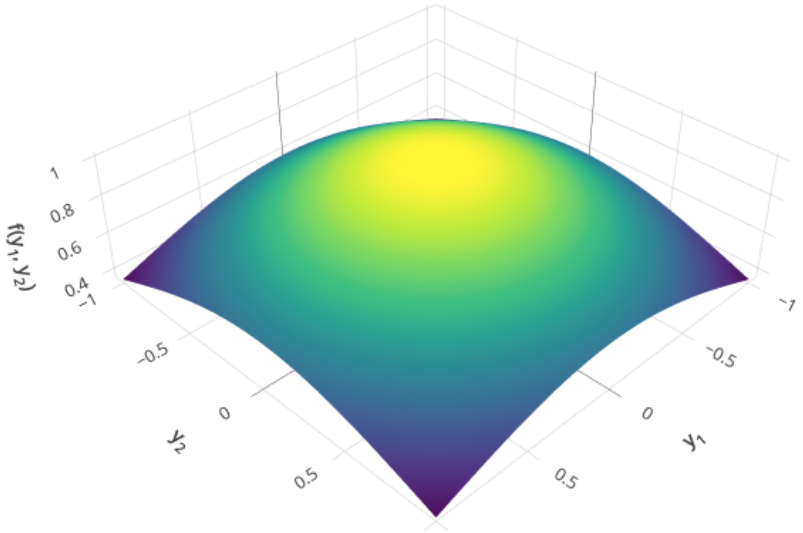



Figure 3.2: Surface plot of affine-transformed joint distribution in $\mathbb{R}^2$

The next section will provide a general overview and description of the ensemble sampling procedure.

### 3.2.2 Description of the Algorithm

Consider a general MCMC sampler for which the sampled value at some iteration $t$ depends only on that of the previous iteration and is independent of the sampled values at earlier iterations. As an example, the Gibbs sampler and the MH algorithm as

defined in Section 3.1 both follow this form. We denote this by:

$$\mathbf{a}^{(t)} = R\left(\mathbf{a}^{(t-1)}, \boldsymbol{\xi}^{(t-1)}, f\right),$$

where $\{\boldsymbol{\xi}^{(t)}\}_{t\in\mathbb{N}}$ is a sequence of independent and identically-distributed random variables and $f$ is the target distribution. Then, such an MCMC sampler is said to be affine invariant if

$$R\left(\mathbb{T}\mathbf{a} + \mathbf{c}, \boldsymbol{\xi}, f_{\mathbb{T},\mathbf{c}}\right) = \mathbb{T}R\left(\mathbf{a}, \boldsymbol{\xi}, f\right) + \mathbf{c} \tag{3.5}$$

for any $\mathbf{a}$ within the state space, any affine transformation $\mathbb{T}\mathbf{a} + \mathbf{c}$, for $f_{\mathbb{T},\mathbf{c}}$ as defined in equation (3.3), and for almost all $\boldsymbol{\xi}$. In other words, suppose that two Monte Carlo runs are carried out using such a sampler, starting with the same random seed so that the $\boldsymbol{\xi}^{(t)}$ will be identical for both runs. For the first run, we take as initial point $\mathbf{a}^{(0)}$ and density function $f$. However, for the second run, we consider the initial point $\mathbf{b}^{(0)} := \mathbb{T}\mathbf{a}^{(0)} + \mathbf{c}$ and density function $f_{\mathbb{T},\mathbf{c}}$. If the algorithm is indeed affine invariant, then the sequences $(\mathbf{a}^{(t)})_{t\in\mathbb{N}\cup 0}$ and $(\mathbf{b}^{(t)})_{t\in\mathbb{N}\cup 0}$ will satisfy $\mathbf{b}^{(t)} := \mathbb{T}\mathbf{a}^{(t)} + \mathbf{c}$.

An ensemble $\overrightarrow{\mathbf{A}}$ consists of $L$ random vectors known as 'walkers' $\{\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_L\}$, each of which is in $\mathbb{R}^p$. Hence, the ensemble can be thought of as being in $\mathbb{R}^p \times \mathbb{R}^L$. The idea behind the ensemble sampler is to independently sample each walker from $f$, i.e.:

$$g(\overrightarrow{\mathbf{A}}) = g(\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_L) := f(\mathbf{a}_1)f(\mathbf{a}_2)...f(\mathbf{a}_L). \tag{3.6}$$

The ensemble $\overrightarrow{\mathbf{A}}$ may be converted via affine transformation to $\overrightarrow{\mathbf{B}} := (\mathbf{B}_1, \mathbf{B}_2, ..., \mathbf{B}_L)$, where $\mathbf{B}_i := \mathbb{T}\mathbf{A}_i + \mathbf{c}$ for $i \in \{1, 2, ..., L\}$, by applying the affine transformation separately to each walker, i.e.:

$$(\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_L) \overset{\mathbb{T},\mathbf{c}}{\longmapsto} (\mathbb{T}\mathbf{A}_1 + \mathbf{c}, \mathbb{T}\mathbf{A}_2 + \mathbf{c}, ..., \mathbb{T}\mathbf{A}_L + \mathbf{c}) = (\mathbf{B}_1, \mathbf{B}_2, ..., \mathbf{B}_L), \tag{3.7}$$

where the notation $\overset{\mathbb{T},\mathbf{c}}{\longmapsto}$ denotes the affine transformation.

The term "ensemble MCMC algorithm" arises from the fact that the sampler is made up of a Markov chain on the state space of ensembles. The ensemble algorithm generates $\overrightarrow{\mathbf{A}}^{(0)}, \overrightarrow{\mathbf{A}}^{(1)}, \ldots, \overrightarrow{\mathbf{A}}^{(T)}$ in an iterative manner starting from the initial point $\overrightarrow{\mathbf{A}}^{(0)}$. Moreover, the product density $g$ of equation (3.6) may be preserved even if the individual walkers $\mathbf{A}_i^{(t)}$ are not independent of $t$, for example if $\mathbf{A}_i^{(t)}$ depends on $\mathbf{A}_j^{(t-1)}$ for some $i \neq j$.

The ensemble algorithm obtains $\overrightarrow{\mathbf{A}}^{(t)}$ from $\overrightarrow{\mathbf{A}}^{(t-1)}$ by updating one walker at a time, i.e. by cycling through each of the $L$ walkers in the ensemble. Each walker is typically

implemented as a Metropolis-Hastings Markov chain, although in theory any MCMC algorithm can be used (Foreman-Mackey et al., 2013). The walker $\mathbf{A}_i$ is updated based on the current positions of the other walkers, i.e. the walkers making up the complementary ensemble $\overrightarrow{\mathbf{A}}_{[i]} := \{\mathbf{A}_j : j \neq i\}$. If, for each $i$, the single-walker move from $\mathbf{A}_i^{(t-1)}$ to $\mathbf{A}_i^{(t)}$ preserves the conditional distribution of $\mathbf{A}_i$ given $\overrightarrow{\mathbf{A}}_{[i]}^{(t)}$, then the overall ensemble update from $\overrightarrow{\mathbf{A}}^{(t-1)}$ to $\overrightarrow{\mathbf{A}}^{(t)}$ preserves the product density $g$.

Goodman and Weare (2010) propose three different types of moves which one may wish to carry out on the individual walkers. The first type of move, and the one recommended by the authors, is the 'stretch move', where the walker $\mathbf{X}_i$ is updated using one complementary walker, that is, using $\mathbf{A}_j$ for some $i \neq j$. Starting from $\mathbf{A}_i^{(t-1)}$, a new value $\mathbf{A}_i^{(t)*}$ is proposed as follows:

$$\mathbf{A}_i^{(t)*} := \mathbf{A}_j^{(t-1)} + Y\left(\mathbf{A}_i^{(t-1)} - \mathbf{A}_j^{(t-1)}\right), \tag{3.8}$$

where $Y$ is a scaling variable. The stretch move described in equation (3.8) is similar in nature to the 'walk move' described by Christen and Fox (2007), but the former is affine invariant whilst the latter is not. Moreover, if the density $p$ of $Y$ satisfies $p\left(\frac{1}{y}\right) = yp(y)$, then the stretch move is symmetric in that one can move from $\mathbf{A}_i^{(t)*}$ to $\mathbf{A}_i^{(t-1)}$ just as easily as moving from $\mathbf{A}_i^{(t-1)}$ to $\mathbf{A}_i^{(t)*}$. The authors, as well as Christen and Fox (2007), suggest using the distribution

$$p(y) \propto \begin{cases} \frac{1}{\sqrt{y}} & \text{if } y \in \left[\frac{1}{\beta}, \beta\right], \\ 0 & \text{otherwise,} \end{cases}$$

where the parameter $\beta > 1$ may be adjusted. Assuming that the stretch step is symmetric, then the acceptance probability relating to the proposed value is given by:

$$A\left(\mathbf{A}_i^{(t-1)}, \mathbf{A}_i^{(t)*}\right) = \min\left\{1, \frac{1}{Y^{t-1}} \frac{f\left(\mathbf{A}_k^{(t)*}\right)}{f\left(\mathbf{A}_k^{(t-1)}\right)}\right\}.$$

As in the single-particle Metropolis-Hastings procedure, we set $\mathbf{A}_i^{(t)} = \mathbf{A}_i^{(t)*}$ if the proposal value is accepted, otherwise we set $\mathbf{A}_i^{(t)} = \mathbf{A}_i^{(t-1)}$. Additionally, for faster computation, the stretch move may be parallelised by splitting the full ensemble into two subsets and simultaneously updating all the walkers in one subset based only on the values of the walkers in the other subset, as described by Foreman-Mackey et al. (2013, p. 309).

The second type of move proposed by Goodman and Weare (2010) is the 'walk move'. This is highly similar to the stretch move, but it differs in that a subset $S$ of the

walkers, where $|S| \geq 2$ in the complementary ensemble $\overrightarrow{\mathbf{A}}_{[i]}^{(t-1)}$ are considered instead of only a single walker in order to update $\mathbf{A}_i^{(t-1)}$. The proposal value is then given by $\mathbf{A}_i^{(t)*} = \mathbf{A}_i^{(t-1)} + \mathbf{V}$, where $\mathbf{V}$ follows a multivariate normal distribution with mean vector $\mathbf{0}$ and variance-covariance structure equal to

$$\frac{1}{|S|} \sum_{\mathbf{A}_j \in S} \left(\mathbf{A}_j - \overline{\mathbf{A}}_S\right) \left(\mathbf{A}_j - \overline{\mathbf{A}}_S\right)^T,$$

where $\overline{\mathbf{A}}_S = \frac{1}{|S|} \sum_{\mathbf{A}_j \in S} \mathbf{A}_j$.

The third type of move considered is the 'replacement move'. Similarly to the walk move, a sub-ensemble $S$ is considered. However, for the replacement move, the proposal value is obtained by sampling from the distribution $f_S(\mathbf{a}|S)$, which is an estimate of the desired distribution $f(\mathbf{a})$ given the sub-ensemble $S$. This last type of move was only considered by Goodman and Weare (2010) for the case where $f$ is a Gaussian distribution.

Of the three types of moves described, the stretch move performed the best in the computational tests carried out by Goodman and Weare (2010), and it is the default option within `emcee`. However, Huijser et al. (2022) state that the stretch move should be used with caution with high-dimensional data. Other types of moves have also been implemented within `emcee` (Foreman-Mackey et al., 2021).

Goodman and Weare (2010) also describe how the ensemble methods may be compared with standard single-particle methods. Standard MCMC techniques are often employed to estimate expected values of the form:

$$E = E_f[h(\mathbf{A})] = \int_{\mathbb{R}^p} h(\mathbf{a}) f(\mathbf{a}) \, \mathrm{d}\mathbf{a},$$

where $h$ is some function of interest. Then, a natural single-particle estimator for $E$ is

$$\hat{E}_s = \frac{1}{T_s} \sum_{t=1}^{T_s} h\big(\mathbf{A}^{(t)}\big),$$

where $T_s$ is the number of iterations for which the method is run. Similarly, an ensemble $\overrightarrow{\mathbf{E}}^{(t)}$ with $L$ walkers and $N_e$ cycles (i.e. $LN_e$ iterations in total) would have corresponding estimator

$$\hat{E}_e = \frac{1}{T_e} \sum_{n=1}^{T_e} \left[ \frac{1}{L} \sum_{i=1}^{L} h\big(\mathbf{A}_i^{(t)}\big) \right].$$

In the case that $T_s = LT_e$, the single-particle and ensemble estimators do approximately equivalent computational work, depending on the sampling complexity of the individual

walkers. Foreman-Mackey et al. (2013, p. 310) suggest using an ensemble made up of hundreds of walkers and suggest to "go large" until one starts hitting computational issues, such as lack of computational memory or processing power.

In order to measure the performance of the ensemble sampler, we make use of the autocorrelation time, which is described by Foreman-Mackey et al. (2013, p. 309) as "a direct measure of the number of evaluations of the posterior [joint density] required to produce independent samples of the target density". Therefore, a lower autocorrelation time is desirable. To define and estimate the autocorrelation time, we start from the autocovariance function of a time series, i.e.:

$$C_h(t) = \lim_{t' \to \infty} \operatorname{Cov}\Big[h\big(\mathbf{A}^{(t+t')}\big), h\big(\mathbf{A}^{(t)}\big)\Big],$$

which measures the covariances between samples at a time lag $t$. In practice, given a Markov chain of length $N$, then this may be estimated using

$$\hat{C}_h(t) = \frac{1}{N-t} \sum_{n=1}^{N-t} \Big[h\big(\mathbf{A}^{(n+t)}\big) - E\Big]\Big[h\big(\mathbf{A}^{(n)}\big) - E\Big]$$

To ensure that the samples obtained are independent, the length of the Markov chain required is equivalent to the value of $t$ such that $C_h(t) \to 0$. Foreman-Mackey et al. (2013, p. 309) define the integrated autocorrelation time using

$$\tau_h = \sum_{t=-\infty}^{\infty} \frac{C_h(t)}{C_h(0)} = 1 + 2\sum_{t=1}^{\infty} \frac{C_h(t)}{C_h(0)}, \tag{3.9}$$

a natural estimate for which can be obtained using $\hat{C}_h(t)$ and $\hat{C}_h(0)$ in place of $C_h(t)$ and $C_h(0)$ respectively.

The authors propose using the autocorrelation time as a performance measure and suggest using the acceptance probability as a confirmatory test. If the acceptance probability is too low or too high, then the samples may not be representative of the target density. As a rule of thumb, they propose that the acceptance probability should ideally be between 0.2 and 0.5, and state that a low acceptance probability may be improved by increasing the number of walkers. They also state that parallelisation of chains within the ensemble method, i.e. updating the walkers in parallel rather than one-by-one, can be employed to further reduce the sampler variance. In particular, they propose to split the ensemble of $L$ walkers into two subsets $S_0 := \{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_{L/2}\}$ and $S_1 := \{\mathbf{A}_{L/2+1}, \mathbf{A}_{L/2+2}, \ldots, \mathbf{A}_L\}$. Then, all the walkers in $S_0$ are updated using the stretch move based on the values of all the walkers in $S_1$, and vice-versa. This also

has the effect of speeding up computation, since multiple walkers are being updated in parallel.

From the computational tests carried out by Goodman and Weare (2010), ensemble MCMC methods empirically offer significant performance improvements when compared to standard single-particle methods, with negligible added computational overhead. The ensemble sampler due to Goodman and Weare (2010) is one of the more widely adopted MCMC samplers, with several implementations available in different softwares. In particular, `emcee` (Foreman-Mackey et al., 2013) contains an open-source implementation of the ensemble sampler with computational optimisations such as the parallelisation of the chains. This toolkit is widely used within the astrophysics literature, for example as demonstrated for the $H_0$ tension problem (see e g. Bernal et al., 2016; Briffa et al., 2022), and also for the estimation of other cosmological parameters (see e.g. Oddo et al., 2021). Within Sections 3.3 and 3.4, we shall be using the ensemble sampler functionality available within the `emcee` package in order to obtain estimates for the Hubble constant on the data described in Section 1.4. In particular, the parametric MCMC approach considered in Section 3.3 makes use of `emcee` directly, while in Section 3.4 we make use of the `mcmcdgp` module within the `GaPP` package, which also makes use of the ensemble sampler within `emcee`.

## 3.3 Parametric MCMC Inference on Parameters of the $\Lambda CDM$ model

In this section, we shall consider the use of the ensemble sampler due to Goodman and Weare (2010) in order to obtain estimates for $H_0$ based on the concordance $\Lambda CDM$ cosmological model. In particular, we use the Python package `emcee` (Foreman-Mackey et al., 2013) and use the stretch move to update each walker in the ensemble.

In our case, the density of the Universe is expressed as a parameter $\Omega$ that is made up of three sub-quantities. The first is the mass density of matter (including both ordinary/baryonic matter as well as dark matter), and is denoted by $\Omega_M$. The second density parameter, denoted by $\Omega_R$, is the effective mass density of the relativistic particles, i.e. light and neutrinos, while the final density parameter $\Omega_\Lambda$ is the effective mass density of the dark energy:

$$\Omega = \Omega_M + \Omega_R + \Omega_\Lambda. \tag{3.10}$$

When discussing the values of these parameters at the current time, the subscript 0 is often added to each of these three parameters. From observations taken by the Wilkinson Microwave Anisotropy Probe (WMAP) and assuming the $\Lambda CDM$ model, it was found that $\Omega_{M0}$ is around 0.3, while $\Omega_{R0}$ is very close to zero and $\Omega_{\Lambda 0}$ is around 0.7 (Bennett et al., 2013, p. 129). Therefore, the total mass density as defined in equation (3.10) is close to 1, suggesting a Euclidean or 'flat' Universe. To this effect, the spatial curvature parameter at current time $\Omega_{K0} = 1 - \Omega_0$ is very close to zero. The dominance of the mass and dark energy densities compared to that of the relativistic particles at current time is contrasted with that at early times in the Universe. In fact, radiation was then dominant over both mass and dark energy (see e.g. Nave, 2016). The true values of the density parameters determine the shape and ultimate fate of the Universe, with current research mostly in favour of accelerating expansion (see e.g. Riess et al., 1998).

According to Nemiroff and Patla (2008), the first Friedmann equation (named after Friedmann, 1922) relates the different density parameters to the Hubble parameter and Hubble constant, namely:

$$\frac{H(z)^2}{H_0^2} = \Omega_{R0}(1+z)^4 + \Omega_{M0}(1+z)^3 + \Omega_{K0}(1+z)^2 + \Omega_{\Lambda 0},$$

where $\Omega_{K0} \approx 0$ so the third term may therefore be eliminated. Similarly, the term containing $\Omega_{R0}$ can also be eliminated, and $\Omega_{\Lambda 0}$ may alternatively be expressed as $1 - \Omega_{M0}$ from equation (3.10). Therefore, the model under consideration becomes:

$$H(z) = H_0 \sqrt{\Omega_{M0}(1+z)^3 + (1 - \Omega_{M0})}.$$

Virtually all modern estimates of $H_0$ lie in the range of $65 - 80$ km s$^{-1}$ Mpc$^{-1}$, and $\Omega_{M0}$ is known to be very close to 0.3 (see e.g. Hinshaw et al., 2013). Therefore, we conservatively set a mostly uninformative prior on the parameters, defined by:

$$\log \mathbb{P}(\boldsymbol{\theta}) = \begin{cases} 0 & \text{if } H_0 \in [50, 100] \text{ and } \Omega_{M0} \in [0.1, 0.5] \\ -\infty & \text{otherwise,} \end{cases} \tag{3.11}$$

where $\boldsymbol{\theta}$ is the vector of parameters $(H_0, \Omega_{M0})$. As initial points, we use $\hat{H}_0^{(0)} = 75$ km s$^{-1}$ Mpc$^{-1}$and $\hat{\Omega}_{M0}^{(0)} = 0.3$.

As in the GPR example in Chapter 2, we fit the $H(z)$-against-$z$ curve by comparing the predicted and observed values of $H(z)$. In particular, for each of the 31 CC data

points observed, our objective is analogous to weighted least squares estimation and is highly similar to equation (2.17). Therefore, the objective function is proportional to

$$-\sum_{i=1}^{31} \frac{\left[\hat{H}_{pred}(z_i) - H_{obs}(z_i)\right]^2}{\hat{\sigma}_H^2(z_i)}, \tag{3.12}$$

where $\hat{H}_{pred}(z_i)$ and $H_{obs}(z_i)$ are the predicted and observed values of $H$ at redshift $z_i$ and $\hat{\sigma}_H^2(z_i)$ is the variance of the observation $H_{obs}(z_i)$. The idea behind such a function is to minimise the distance between the predicted and observed values for $H$ at each of the redshifts $z_i$ present in the CC dataset. This quantity reaches a maximum of 0 when the predicted and observed values of $H$ for each $z_i$ are equal.

For the SN data, the objective function is similar but it differs in that we use $E(z) = H(z)/H_0$ values instead of using $H$ values directly, and additionally we need to incorporate the $5 \times 5$ covariance matrix $\mathbb{\Sigma}_{SN}$ of these readings (Briffa et al., 2022). Moreover, as described in Section 1.4, Riess et al. (2018) converted the original Pantheon SNIa dataset from brightness data to expansion data. Therefore, the original readings of the distance modulus $\mu$ were compressed and converted to $E$ values.

Therefore, suppose that the five values of $z$ form a vector $\mathbf{z} = (z_1, z_2, ..., z_5)^T$, and let the corresponding observed values of $\mathbf{E}$ at these $z$-values form a vector $\mathbf{E}_{obs}(\mathbf{z})$, and similarly for the predicted values. In this case, we have that the objective function is proportional to

$$-\left(\hat{\mathbf{E}}_{pred}(\mathbf{z}) - \mathbf{E}_{obs}(\mathbf{z})\right)^T \mathbb{\Sigma}_{SN}^{-1} \left(\hat{\mathbf{E}}_{pred}(\mathbf{z}) - \mathbf{E}_{obs}(\mathbf{z})\right). \tag{3.13}$$

For the BAO data, the function used is similar to that of the CC data, except that the summation is carried out over the 10 BAO data points instead of the 31 CC data points.

For all three data sources, since we are assuming a regression structure with independent Gaussian errors, then the minima of the least-squares objective functions are equivalent to the ML estimator (see e.g. Bazot et al., 2012), and the ML estimator is asymptotically consistent and efficient (Browne, 1973). In fact, the objective functions often directly referred to as 'likelihood functions' within the astrophysics literature (see e.g. Briffa et al., 2022). Estimation using the CC and BAO cases is analogous to ordinary least squares regression, while the SN data takes a weighted least squares structure since the SN error terms are correlated. Therefore, we denote the likelihood relating to each

dataset as $\Lambda_{CC}$, $\Lambda_{SN}$, and $\Lambda_{BAO}$ respectively. The overall log-likelihood when all the datasets are considered altogether is then given by the summation of the component log-likelihoods, i.e.:

$$\log \Lambda_{tot}(f|\boldsymbol{\theta}) = \log \Lambda_{CC}(f|\boldsymbol{\theta}) + \log \Lambda_{SN}(f|\boldsymbol{\theta}) + \log \Lambda_{BAO}(f|\boldsymbol{\theta}). \tag{3.14}$$

By Bayes' rule, the posterior distribution is

$$\mathbb{P}(\boldsymbol{\theta}|f) = \frac{\mathbb{P}(\boldsymbol{\theta})\Lambda_{tot}(f|\boldsymbol{\theta})}{\mathbb{P}(f)},$$

where the denominator is a constant. Therefore, the posterior log-probability is proportional to the summation of the log-prior and log-likelihood, i.e.:

$$\mathbb{P}(\boldsymbol{\theta}|f) \propto \mathbb{P}(\boldsymbol{\theta})\Lambda_{tot}(f|\boldsymbol{\theta}) \implies \log \mathbb{P}(\boldsymbol{\theta}|f) \propto \log \mathbb{P}(\boldsymbol{\theta}) + \log \Lambda_{tot}(f|\boldsymbol{\theta}).$$

In order to obtain the parameter estimates $\hat{H}_0$ and $\hat{\Omega}_{M0}$ for each combination of dataset and 'prior', we use the affine ensemble sampler as implemented in `emcee` with 5000 iterations and 200 walkers. Each walker is initialised close to the point $\boldsymbol{\theta}^{(0)} := \left(\hat{H}_0^{(0)}, \hat{\Omega}_{M0}^{(0)}\right)^T = (75, 0.3)^T$, and the parameter values are updated iteratively using the parallelised version of the stretch move described in Section 3.2.2 such that the log-likelihood function of equation (3.14) is maximised with respect to $\boldsymbol{\theta}$. As goodness of fit measures, the final log-likelihood and the estimated autocorrelation time $\hat{\boldsymbol{\tau}} := \left(\hat{\tau}_{\hat{H}_0}, \hat{\tau}_{\hat{\Omega}_{M0}}\right)^T$ will be used. The parameter estimates and goodness of fit measures in each case are presented in Table 3.1.

From this table, it is evident that both the dataset considered and the prior used (if any) have an effect on the estimated values for $H_0$ and $\Omega_{M0}$. For all the combinations of dataset and prior considered, however, the estimated value for $H_0$ lies between 67 and 74 km s$^{-1}$ Mpc$^{-1}$. These observations are very similar to those made in Section 2.4, i.e. when we made use of GPR on the same data and with the same prior values for $\hat{H}_0$.

For each of the prior values considered, the $\hat{H}_0$ value obtained using the CC dataset is generally lower than that obtained using the CC+SN dataset. The larger datasets provided more confident predictions for both $\hat{H}_0$ and $\hat{\Omega}_{M0}$, since the errors in the estimates were the lowest for the CC+SN+BAO data in each case. This is true for all the priors considered, but the smallest prediction error occurred when using the Planck prior.

| **Dataset** | **Parameter estimates and prediction errors** | | **Goodness of fit** | |
|---|---|---|---|---|
| | $\hat{H}_0$ (km s$^{-1}$ Mpc$^{-1}$) | $\hat{\Omega}_{M0}$ | $\hat{\boldsymbol{\tau}} = \left(\hat{\tau}_{\hat{H}_0}, \hat{\tau}_{\hat{\Omega}_{M0}}\right)^T$ | $\log \Lambda_{tot}$ |
| CC | $67.785^{3.176}_{3.161}$ | $0.327^{0.058}_{0.067}$ | $(31.526, 31.278)$ | $-8.246$ |
| CC+SN | $68.482^{3.133}_{3.078}$ | $0.311^{0.054}_{0.063}$ | $(31.764, 31.634)$ | $-9.579$ |
| CC+SN+BAO | $70.461^{1.789}_{1.773}$ | $0.254^{0.020}_{0.021}$ | $(32.512, 31.126)$ | $-11.421$ |
| CC+$\hat{H}_0^R$ | $72.505^{1.605}_{1.604}$ | $0.256^{0.033}_{0.036}$ | $(31.018, 31.223)$ | $-9.729$ |
| CC+SN+$\hat{H}_0^R$ | $72.636^{1.603}_{1.596}$ | $0.251^{0.033}_{0.035}$ | $(31.862, 31.503)$ | $-10.805$ |
| CC+SN+BAO+$\hat{H}_0^R$ | $72.220^{1.282}_{1.284}$ | $0.238^{0.015}_{0.016}$ | $(31.722, 30.866)$ | $-12.489$ |
| CC+$\hat{H}_0^{TRGB}$ | $69.245^{1.669}_{1.670}$ | $0.303^{0.038}_{0.042}$ | $(33.129, 31.458)$ | $-8.359$ |
| CC+SN+$\hat{H}_0^{TRGB}$ | $69.424^{1.661}_{1.671}$ | $0.296^{0.038}_{0.041}$ | $(33.029, 31.494)$ | $-9.660$ |
| CC+SN+BAO+$\hat{H}_0^{TRGB}$ | $70.161^{1.331}_{1.324}$ | $0.258^{0.016}_{0.017}$ | $(33.191, 31.268)$ | $-11.542$ |
| CC+$\hat{H}_0^{HW}$ | $71.981^{1.538}_{1.525}$ | $0.263^{0.033}_{0.036}$ | $(31.730, 31.264)$ | $-9.364$ |
| CC+SN+$\hat{H}_0^{HW}$ | $72.104^{1.526}_{1.522}$ | $0.258^{0.033}_{0.036}$ | $(31.383, 30.887)$ | $-10.454$ |
| CC+SN+BAO+$\hat{H}_0^{HW}$ | $71.885^{1.244}_{1.243}$ | $0.241^{0.015}_{0.016}$ | $(32.719, 31.123)$ | $-12.098$ |
| CC+$\hat{H}_0^{CM}$ | $73.547^{1.509}_{1.514}$ | $0.242^{0.031}_{0.034}$ | $(31.617, 31.640)$ | $-10.420$ |
| CC+SN+$\hat{H}_0^{CM}$ | $73.665^{1.513}_{1.500}$ | $0.238^{0.031}_{0.033}$ | $(32.028, 31.867)$ | $-11.406$ |
| CC+SN+BAO+$\hat{H}_0^{CM}$ | $72.945^{1.237}_{1.230}$ | $0.231^{0.014}_{0.015}$ | $(32.070, 31.371)$ | $-13.449$ |
| CC+$\hat{H}_0^{P}$ | $67.413^{0.515}_{0.509}$ | $0.333^{0.033}_{0.034}$ | $(46.791, 32.379)$ | $-8.594$ |
| CC+SN+$\hat{H}_0^{P}$ | $67.430^{0.511}_{0.511}$ | $0.328^{0.032}_{0.034}$ | $(45.231, 30.923)$ | $-10.086$ |
| CC+SN+BAO+$\hat{H}_0^{P}$ | $67.634^{0.497}_{0.497}$ | $0.285^{0.012}_{0.013}$ | $(49.550, 31.493)$ | $-13.281$ |
| CC+$\hat{H}_0^{DES}$ | $67.455^{1.076}_{1.070}$ | $0.332^{0.036}_{0.038}$ | $(36.991, 31.764)$ | $-8.350$ |
| CC+SN+$\hat{H}_0^{DES}$ | $67.527^{1.074}_{1.076}$ | $0.326^{0.035}_{0.037}$ | $(36.269, 31.402)$ | $-9.769$ |
| CC+SN+BAO+$\hat{H}_0^{DES}$ | $68.278^{0.963}_{0.964}$ | $0.277^{0.015}_{0.016}$ | $(36.622, 31.391)$ | $-12.753$ |

Table 3.1: Results of MCMC using $\Lambda CDM$ model

Another general trend identified within this analysis is that a smaller estimate for $H_0$ coincides with a larger value for $\hat{\Omega}_{M0}$. Therefore, a higher matter density is related to slower expansion of the Universe and to an older Universe. This makes sense as the attractive force of gravity pulls all matter together, so a higher matter density implies a stronger gravitational force that is opposing the expansion of the Universe. This finding is also consistent with other studies in the literature (see e.g. Briffa et al., 2022). A higher matter density is also associated with a higher metallicity, i.e. the abundance of metallic elements such as iron and nickel, within the universe. In the absence of such metallic elements, it is not possible for planets to be formed, except possibly for gas giants.

For each of the given datasets, the autocorrelation times are mostly comparable between the different priors, except for when the Planck prior was used. The estimated autocorrelation time for $\hat{H}_0$ is significantly higher when using the Planck prior; this may be since this prior has a much smaller error than the other priors, as seen in Table 2.1. For the MCMC run using the DES prior, this quantity is slightly higher than for the rest of the priors (excluding the Planck prior) but to a lesser extent than that obtained when using the Planck prior. On the other hand, the log-likelihood value is the highest for the priorless case, followed closely by the TRGB and DES priors. Note that the log-likelihood values should not be compared across different datasets, but instead, these values allow us to compare between the different priors given a particular dataset.

For each combination of dataset and 'prior', trace plots showing the evolution of the walkers at each iteration provided a visual confirmation that the desired convergence of the parameter estimates was achieved. An example of one of the trace plots obtained is given in Figure 3.3.

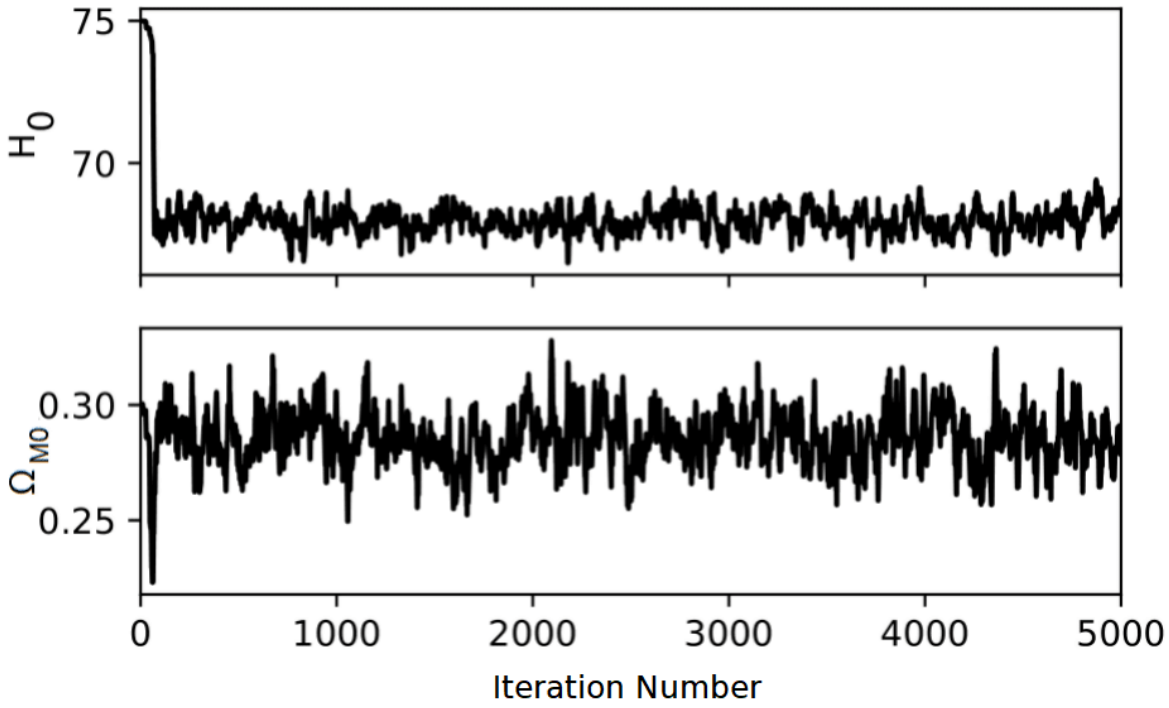
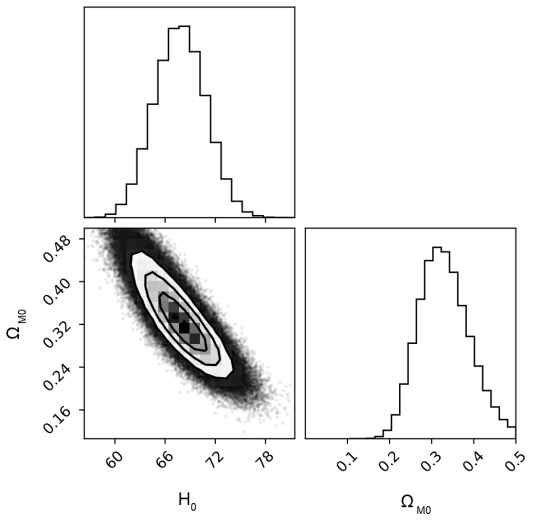


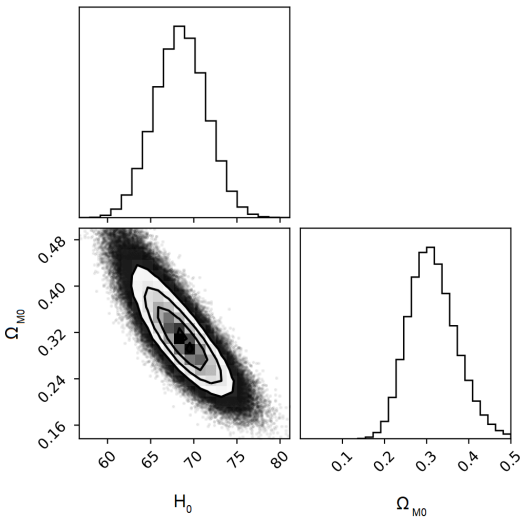

Figure 3.3: Example of trace plot for MCMC on parameters of $\Lambda CDM$ model

Additionally, corner plots showing the distribution of the parameters in each case were produced. The corner plots for each dataset for the priorless case are presented in Figure 3.4, while those for the cases where a 'prior' value was included are shown in Appendix D. These plots provide a visual confirmation of the points just mentioned,

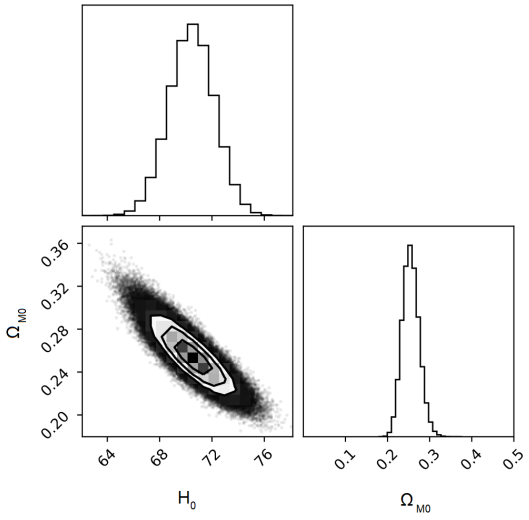
mainly the lower errors in the estimates when using the larger datasets and the decrease in $\hat{\Omega}_{M0}$ with increasing $\hat{H}_0$. The former can visually be seen from the distributions of the estimated values in each case, while the latter is visible from the space of the 'oval' in the bottom-left corner of each subplot.




(a) CC




(b) CC+SN




(c) CC+SN+BAO

Figure 3.4: Corner plots for MCMC on parameters of $\Lambda CDM$ model for priorless case

## 3.4 Non-Parametric MCMC Inference on Kernel Hyperparameters

MCMC inference may also be carried out independently from any cosmological model. For example, the Python package `GaPP` contains the module `mcmcdgp`, which allows for a non-parametric reconstruction of the $H(z)$-against-$z$ function through MCMC inference on the hyperparameters of the kernel function. This reconstruction is carried out in exactly the same way as in Section 2.4. In particular, we minimise the chi-squared statistic between the reconstructed/estimated and observed value of $H(z)$ in each case, as in equation (2.17). The only difference between the theoretical setup used here when compared to GPR is that the values for the kernel hyperparameters are obtained using the affine ensemble sampler described in Section 3.2.2 and as implemented in `emcee` instead of the usual minimisation of derivatives as in equation (2.16).

In this case, the `mcmcdgp` procedure in the `GaPP` package was employed using 100 walkers, with 1000 iterations of burn-in and a further 100000 iterations thereafter. As in Section 2.4, we consider each combination of dataset, kernel, and prior. In each case, the dimensionality of the problem is equal to the number of hyperparameters for that kernel. Therefore, for all kernels except the rational quadratic and the double square exponential kernels, each of the walkers will be 2-dimensional. For the rational quadratic, each of the walkers is 3-dimensional, while for the double square exponential kernel it is 4-dimensional. As initial points for the walkers, we assign $(10, 10)^T$, $(10, 10, 10)^T$ or $(10, 10, 10, 10)^T$ depending on the kernel. The stretch move is then used to update the walkers, and the values obtained are the hyperparameters of the respective kernel. Then, this kernel is used within a GPR context to reconstruct the $H(z)$-against-$z$ curve, and a point estimate of $H_0$ is obtained for each combination of dataset, kernel, and prior. The results will be presented in a similar manner to those in Table 2.2, and the $\sigma$-distance between each $H_0$ estimate and the 'priors' in Table 2.1 is calculated as in equation (2.18). The results corresponding to the square exponential kernel for each combination of 'prior' and dataset are presented in Table 3.2, while the corresponding tables for the other kernel functions are presented in Appendix F.

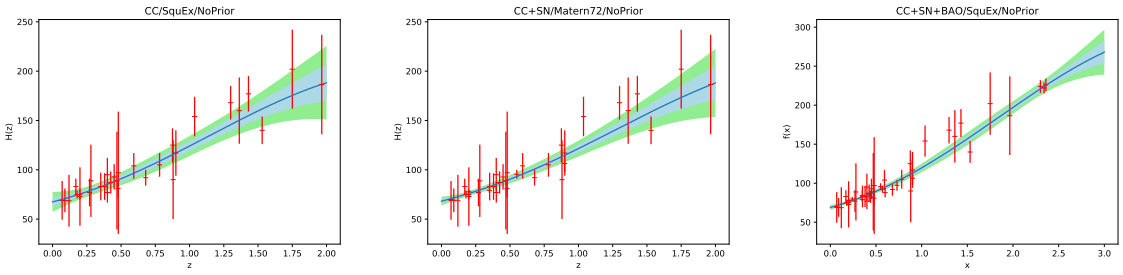
Due to the computational time required for the MCMC procedure, this research has been carried out using computational facilities procured through the European Regional Development Fund (ERDF), Project ERDF-080 'A Supercomputing Laboratory for the University of Malta', and in conjunction with the Department of Physics and the Institute of Space Sciences and Astronomy at the University of Malta.

| **Dataset** | $\hat{H}_0$ (km s$^{-1}$ Mpc$^{-1}$) | $d(\hat{H}_0, \hat{H}_0^R)$ | $d(\hat{H}_0, \hat{H}_0^{TRGB})$ | $d(\hat{H}_0, \hat{H}_0^{HW})$ | $d(\hat{H}_0, \hat{H}_0^{CM})$ | $d(\hat{H}_0, \hat{H}_0^P)$ | $d(\hat{H}_0, \hat{H}_0^{DES})$ |
|---|---|---|---|---|---|---|---|
| CC | $67.449 \pm 5.360$ | $-1.1962$ | $-0.4134$ | $-1.0406$ | $-1.4067$ | 0.0091 | 0.0090 |
| CC+SN | $68.104 \pm 3.191$ | $-1.6649$ | $-0.4566$ | $-1.4371$ | $\mathbf{-2.0094}$ | 0.2181 | 0.2087 |
| CC+SN+BAO | $68.903 \pm 2.545$ | $-1.6995$ | $-0.2825$ | $-1.4368$ | $\mathbf{-2.1143}$ | 0.5795 | 0.5421 |
| CC+$\hat{H}_0^R$ | $73.833 \pm 2.987$ | $-0.1107$ | 1.1391 | 0.1550 | $-0.4427$ | **2.1238** | **2.0207** |
| CC+SN+$\hat{H}_0^R$ | $71.646 \pm 1.779$ | $-1.0113$ | 0.7093 | $-0.6721$ | $-1.5136$ | **2.2977** | **2.0300** |
| CC+SN+BAO+$\hat{H}_0^R$ | $71.547 \pm 1.558$ | $-1.1157$ | 0.7109 | $-0.7602$ | $-1.6597$ | **2.5339** | **2.1740** |
| CC+$\hat{H}_0^{TRGB}$ | $69.602 \pm 3.144$ | $-1.2712$ | $-0.0538$ | $-1.0346$ | $-1.6125$ | 0.6918 | 0.6612 |
| CC+SN+$\hat{H}_0^{TRGB}$ | $69.011 \pm 1.730$ | $\mathbf{-2.0745}$ | $-0.3069$ | $-1.7683$ | $\mathbf{-2.6287}$ | 0.8949 | 0.7861 |
| CC+SN+BAO+$\hat{H}_0^{TRGB}$ | $69.295 \pm 1.503$ | $\mathbf{-2.0866}$ | $-0.2084$ | $-1.7650$ | $\mathbf{-2.6862}$ | 1.1966 | 1.0176 |
| CC+$\hat{H}_0^{HW}$ | $72.974 \pm 2.626$ | $-0.3900$ | 0.9792 | $-0.1043$ | $-0.7622$ | **2.0851** | 1.9578 |
| CC+SN+$\hat{H}_0^{HW}$ | $71.269 \pm 1.634$ | $-1.2066$ | 0.5863 | $-0.8614$ | $-1.7415$ | **2.2646** | 1.9645 |
| CC+SN+BAO+$\hat{H}_0^{HW}$ | $71.225 \pm 1.444$ | $-1.2892$ | 0.5969 | $-0.9304$ | $-1.8621$ | **2.5023** | **2.1066** |
| CC+$\hat{H}_0^{CM}$ | $74.970 \pm 2.594$ | 0.2368 | 1.6078 | 0.5385 | $-0.1228$ | **2.8652** | **2.6864** |
| CC+SN+$\hat{H}_0^{CM}$ | $72.624 \pm 1.664$ | $-0.6471$ | 1.1180 | $-0.2842$ | $-1.1527$ | **3.0059** | **2.6184** |
| CC+SN+BAO+$\hat{H}_0^{CM}$ | $72.408 \pm 1.472$ | $-0.7740$ | 1.0854 | $-0.3965$ | $-1.3171$ | **3.2226** | **2.7261** |
| CC+$\hat{H}_0^P$ | $67.393 \pm 0.247$ | $\mathbf{-3.7172}$ | $-1.2565$ | $\mathbf{-3.4388}$ | $\mathbf{-4.6861}$ | $-0.0133$ | $-0.0066$ |
| CC+SN+$\hat{H}_0^P$ | $67.432 \pm 0.231$ | $\mathbf{-3.7003}$ | $-1.2375$ | $\mathbf{-3.4206}$ | $\mathbf{-4.6695}$ | 0.0573 | 0.0281 |
| CC+SN+BAO+$\hat{H}_0^P$ | $67.517 \pm 0.226$ | $\mathbf{-3.6550}$ | $-1.1932$ | $\mathbf{-3.3722}$ | $\mathbf{-4.6210}$ | 0.2131 | 0.1041 |
| CC+$\hat{H}_0^{DES}$ | $67.391 \pm 1.147$ | $\mathbf{-3.1748}$ | $-1.0857$ | $\mathbf{-2.8818}$ | $\mathbf{-3.9131}$ | $-0.0075$ | $-0.0059$ |
| CC+SN+$\hat{H}_0^{DES}$ | $67.533 \pm 0.866$ | $\mathbf{-3.3175}$ | $-1.0856$ | $\mathbf{-3.0225}$ | $\mathbf{-4.1355}$ | 0.1329 | 0.0949 |
| CC+SN+BAO+$\hat{H}_0^{DES}$ | $67.833 \pm 0.802$ | $\mathbf{-3.2111}$ | $-0.9537$ | $\mathbf{-2.9082}$ | $\mathbf{-4.0375}$ | 0.4581 | 0.3181 |

Table 3.2: MCMC GPR Results with square exponential kernel function

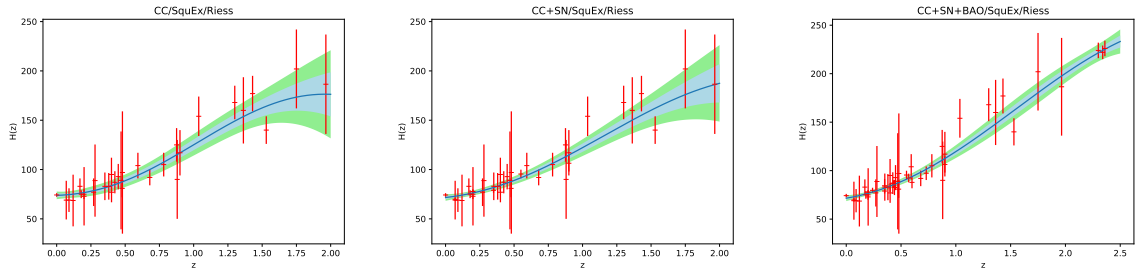
One main observation from Table 3.2 is that each estimated value of $H_0$ is highly dependent on the prior considered. The dataset considered also affects the estimate obtained in each case, but to a much lesser extent. Therefore, as was the case with the 'vanilla' GPR described in Chapter 2, the MCMC GPR produced highly similar results, once again showing the GPR is prone to overfitting and highly sensitive to the prior specification. This being said, the priorless case suggests a value for $H_0$ that is on the lower side, which was also the case for 'vanilla' GPR. Another observation that can be made from the results obtained is that the prediction error decreases as one considers larger datasets. Similarly, the prediction error is greatly decreased when adding a 'prior' value to the data, further confirming that inclusion of a prior greatly affects the estimates obtained.

In each case, the reconstruction plots for the $H(z)$-against-$z$ relationship were also produced as was done for the GPR analysis in Section 2.4. For the square exponential function, the relevant plots are shown in Figures 3.5 to 3.11, while those for the other kernel functions are shown in Appendix E. These plots again indicate similar results to those obtained in Section 2.4. In each case, the plot gives the reconstruction of $\hat{H}(z)$ against $z$, with the higher $\hat{H}(z)$ values as $z$ increases indicating higher distances at higher redshifts. Additionally, the increasing gradient in the plots indicates the accelerating expansion of the universe (see e.g. Riess et al., 1998). In other words, celestial objects at higher redshifts, i.e. at larger relative distances from Earth, are receding from Earth at a faster rate. For each plot, the extrapolated value at $z = 0$ is our estimate for $H_0$, as was the case in Section 2.4.



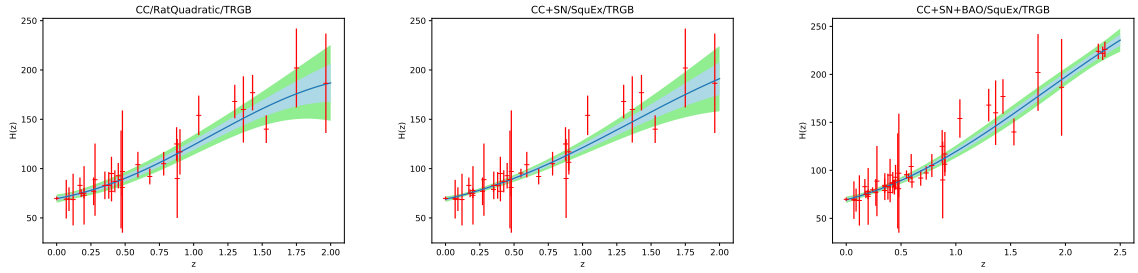

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure 3.5: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and without adding any $H_0$ estimate to the data



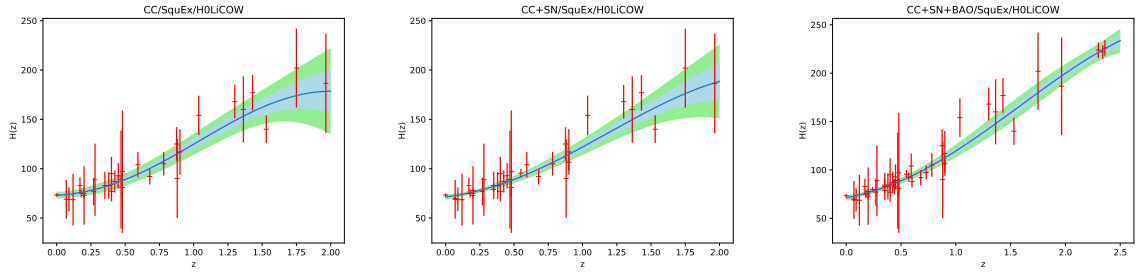

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure 3.6: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^R$



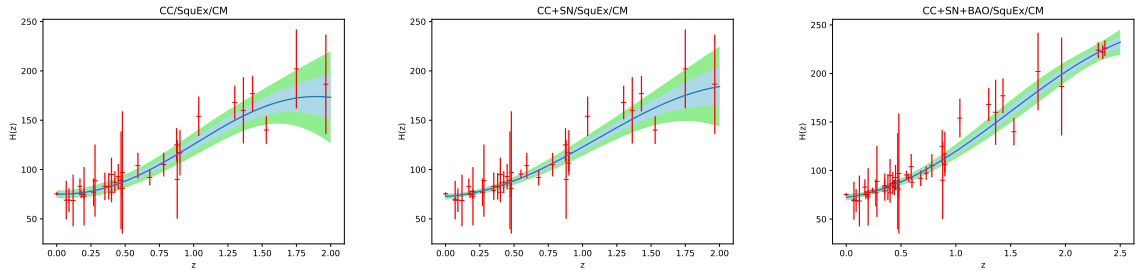

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure 3.7: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^{TRGB}$



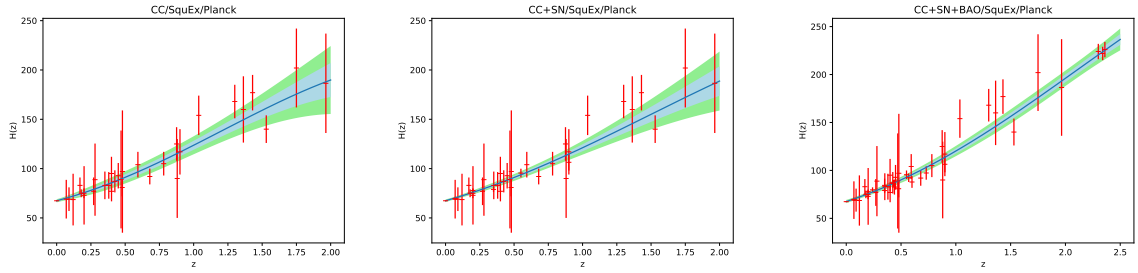

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure 3.8: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^{HW}$



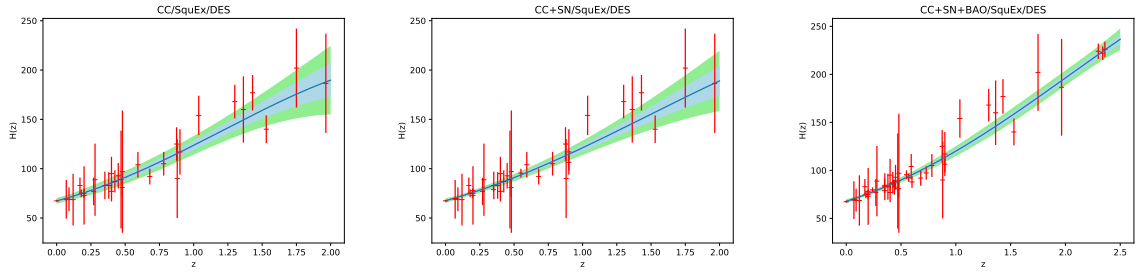

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure 3.9: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^{CM}$




(a) CC (b) CC+SN (c) CC+SN+BAO

Figure 3.10: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^{P}$




(a) CC (b) CC+SN (c) CC+SN+BAO

Figure 3.11: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^{DES}$

The final log-likelihood value is used as a goodness of fit measure, similarly to what was done in Section 2.4. Moreover, as a further metric, the acceptance probability and estimated autocorrelation times are also checked in each case. Values for the log-likelihood, acceptance probability, and autocorrelation times for each of the three datasets considered are given in Tables 3.3 through 3.11. In each case, the vector of estimated autocorrelation times contains an entry for each hyperparameter. For the double square exponential kernel, the vector $\hat{\boldsymbol{\tau}} := \left(\hat{\tau}(\hat{\sigma}_{f_1}), \hat{\tau}(\hat{l}_1), \hat{\tau}(\hat{\sigma}_{f_2}), \hat{\tau}(\hat{l}_2)\right)^T$, while this vector is similarly defined using $\left(\hat{\tau}(\hat{\sigma}_f), \hat{\tau}(\hat{l}), \hat{\tau}(\hat{\alpha})\right)^T$ for the rational quadratic kernel. For the remaining kernels, the vector $\hat{\tau}$ represents $\left(\hat{\tau}(\hat{\sigma}_f), \hat{\tau}(\hat{l})\right)^T$. The autocorrelation times are estimated as described in Section 3.2.2.

As was the case with the likelihood values in Section 2.4, these measures may only be compared between the different kernel functions given the same dataset and prior. As was the case with 'vanilla' GPR, there is no immediate indication that any one kernel function is preferable over another, further confirming that in this case the kernel used does not affect the results too much. Additionally, the acceptance probability values are mostly between 0.2 and 0.6, with a few notable exceptions where the acceptance probability was much lower. In some cases, the acceptance probability was less than 0.1, but no particular pattern can be identified as to why this is the case. Autocorrelation times are also generally comparable but in some cases slightly higher than others. There is no evidence in favour or against the use of any particular kernel function, but the autocorrelation times increase marginally as the size of the dataset increases. This means that more iterations are required to reach convergence of the parameter estimates for the larger datasets.

To conclude, the results obtained in this section are almost identical to those obtained using 'vanilla' GPR as seen in Chapter 2. However, a main issue with the non-parametric MCMC approach is the significant computational time and resources required. In fact, each combination of dataset, prior, and kernel required about 3 days to run on average. While the use of a dedicated computational platform allowed the different combinations to be run in parallel, this is still a striking difference when compared to the 'vanilla' GPR of Section 2.4. Using an ordinary laptop with a $6^{\text{th}}$ generation Intel i5 processor running at 2.3GHz and 8GB of DDR3 memory, each combination of dataset, prior, and kernel took only 10 seconds to run on average. Therefore, given that the two methods yielded virtually identical results, then this suggests that it may be more appropriate to consider 'vanilla' GPR in this case.

For all the methods discussed so far within both Chapters 2 and 3, we have assumed that the underlying observations are Gaussian in nature. This normality assumption is required for GPR from the definition of Gaussian processes. In Chapter 4, however, we will be doing away with this normality assumption by considering Student's $t$-processes instead of Gaussian processes within the regression context. Since the multivariate Student's $t$-distribution is a generalisation of the multivariate Gaussian distribution, then in general we expect TPR to perform at least as well as GPR while offering greater flexibility and robustness.

| | None | $\hat{H}_0^R$ | $\hat{H}_0^{TRGB}$ | $\hat{H}_0^{HW}$ | $\hat{H}_0^{CM}$ | $\hat{H}_0^P$ | $\hat{H}_0^{DES}$ |
|---|---|---|---|---|---|---|---|
| **Square Exponential** | −145.630 | −149.247 | −149.413 | **−149.415** | −151.655 | −149.584 | −149.340 |
| **Double Square Exponential** | −145.622 | **−147.730** | −149.512 | −151.643 | −150.293 | −149.094 | −149.440 |
| **Rational Quadratic** | −146.154 | −148.220 | **−148.399** | −149.709 | −151.169 | −149.167 | −150.204 |
| **Matérn** ($\nu = \frac{3}{2}$) | −145.620 | −151.629 | −150.398 | −151.915 | −150.234 | −149.036 | −148.808 |
| **Matérn** ($\nu = \frac{5}{2}$) | **−144.502** | −153.384 | −149.420 | −151.511 | −151.802 | −148.608 | **−148.298** |
| **Matérn** ($\nu = \frac{7}{2}$) | −144.701 | −152.194 | −149.413 | −153.026 | −152.098 | −148.841 | −149.468 |
| **Matérn** ($\nu = \frac{9}{2}$) | −144.856 | −153.856 | −149.823 | −152.521 | −153.298 | **−148.542** | −149.249 |
| **Cauchy** | −145.744 | −153.469 | −150.126 | −151.340 | **−150.145** | −150.022 | −150.609 |

Table 3.3: Marginal log-likelihood values for each kernel function and prior using CC dataset

| | None | $\hat{H}_0^R$ | $\hat{H}_0^{TRGB}$ | $\hat{H}_0^{HW}$ | $\hat{H}_0^{CM}$ | $\hat{H}_0^P$ | $\hat{H}_0^{DES}$ |
|---|---|---|---|---|---|---|---|
| **Square Exponential** | −160.869 | **−166.368** | −163.911 | **−165.694** | −164.770 | −161.257 | −161.172 |
| **Double Square Exponential** | −161.549 | −166.600 | −165.029 | −166.605 | −166.938 | −161.784 | −161.566 |
| **Rational Quadratic** | **−158.219** | −169.466 | −166.098 | −167.333 | −165.160 | −162.579 | **−161.120** |
| **Matérn** ($\nu = \frac{3}{2}$) | −160.131 | −166.534 | −163.756 | −168.940 | −168.286 | −163.047 | −161.298 |
| **Matérn** ($\nu = \frac{5}{2}$) | −159.651 | −168.842 | −163.221 | −168.333 | −168.165 | −161.420 | −161.183 |
| **Matérn** ($\nu = \frac{7}{2}$) | −160.570 | −168.596 | −163.055 | −166.845 | **−163.973** | **−160.893** | −161.551 |
| **Matérn** ($\nu = \frac{9}{2}$) | −159.917 | −168.593 | **−162.181** | −169.976 | −168.020 | −161.044 | −161.430 |
| **Cauchy** | −159.252 | −172.694 | −165.762 | −169.338 | −168.778 | −161.853 | −162.232 |

Table 3.4: Marginal log-likelihood values for each kernel function and prior using CC+SN dataset

| | None | $\hat{H}_0^R$ | $\hat{H}_0^{TRGB}$ | $\hat{H}_0^{HW}$ | $\hat{H}_0^{CM}$ | $\hat{H}_0^P$ | $\hat{H}_0^{DES}$ |
|---|---|---|---|---|---|---|---|
| **Square Exponential** | −188.231 | −192.663 | −192.834 | −189.314 | −192.477 | −189.307 | **−188.027** |
| **Double Square Exponential** | −189.780 | −189.144 | −191.836 | −191.062 | −192.736 | −191.390 | −192.324 |
| **Rational Quadratic** | −189.919 | −191.848 | **−189.740** | −191.569 | **−189.188** | −191.160 | −189.010 |
| **Matérn** $(\nu = \frac{3}{2})$ | −189.263 | −192.484 | −192.544 | −191.324 | −192.110 | −190.396 | −190.942 |
| **Matérn** $(\nu = \frac{5}{2})$ | −189.330 | −191.211 | −190.030 | −190.779 | −193.811 | **−187.304** | −191.775 |
| **Matérn** $(\nu = \frac{7}{2})$ | −188.305 | −192.893 | −191.309 | **−188.550** | −189.422 | −187.359 | −188.775 |
| **Matérn** $(\nu = \frac{9}{2})$ | **−187.941** | **−188.527** | −190.828 | −191.191 | −191.529 | −188.927 | −192.002 |
| **Cauchy** | −188.003 | −189.939 | −192.017 | −188.546 | −191.397 | −191.064 | −191.086 |

Table 3.5: Marginal log-likelihood values for each kernel function and prior using CC+SN+BAO dataset

| | None | $\hat{H}_0^R$ | $\hat{H}_0^{TRGB}$ | $\hat{H}_0^{HW}$ | $\hat{H}_0^{CM}$ | $\hat{H}_0^P$ | $\hat{H}_0^{DES}$ |
|---|---|---|---|---|---|---|---|
| **Square Exponential** | 0.6224 | 0.3061 | 0.5021 | 0.5444 | 0.5396 | 0.5195 | 0.5128 |
| **Double Square Exponential** | 0.3250 | 0.0582 | 0.1872 | 0.0888 | 0.0629 | 0.1968 | 0.2773 |
| **Rational Quadratic** | 0.3644 | 0.0983 | 0.4027 | 0.2684 | 0.1830 | 0.2990 | 0.3174 |
| **Matérn** $(\nu = \frac{3}{2})$ | 0.5948 | 0.5507 | 0.5930 | 0.6217 | 0.1671 | 0.6196 | 0.4710 |
| **Matérn** $(\nu = \frac{5}{2})$ | 0.5454 | 0.4931 | 0.4956 | 0.5002 | 0.5809 | 0.5202 | 0.6236 |
| **Matérn** $(\nu = \frac{7}{2})$ | 0.6200 | 0.3775 | 0.5140 | 0.6436 | 0.5738 | 0.5572 | 0.6352 |
| **Matérn** $(\nu = \frac{9}{2})$ | 0.5041 | 0.5994 | 0.5965 | 0.5376 | 0.5802 | 0.4041 | 0.6037 |
| **Cauchy** | 0.5259 | 0.5771 | 0.5940 | 0.5244 | 0.6136 | 0.5679 | 0.5201 |

Table 3.6: Acceptance probability values for each kernel function and prior using CC dataset

| | None | $\hat{H}_0^R$ | $\hat{H}_0^{TRGB}$ | $\hat{H}_0^{HW}$ | $\hat{H}_0^{CM}$ | $\hat{H}_0^P$ | $\hat{H}_0^{DES}$ |
|---|---|---|---|---|---|---|---|
| **Square Exponential** | 0.4360 | 0.4942 | 0.6102 | 0.5265 | 0.5856 | 0.5371 | 0.5937 |
| **Double Square Exponential** | 0.3932 | 0.0503 | 0.4022 | 0.0892 | 0.3751 | 0.2432 | 0.4178 |
| **Rational Quadratic** | 0.4402 | 0.2825 | 0.3102 | 0.3264 | 0.2965 | 0.3401 | 0.4430 |
| **Matérn** $(\nu = \frac{3}{2})$ | 0.5473 | 0.5596 | 0.5641 | 0.3754 | 0.1765 | 0.5565 | 0.6181 |
| **Matérn** $(\nu = \frac{5}{2})$ | 0.5854 | 0.4890 | 0.4911 | 0.6010 | 0.5731 | 0.5465 | 0.5482 |
| **Matérn** $(\nu = \frac{7}{2})$ | 0.5960 | 0.6078 | 0.5230 | 0.6031 | 0.4409 | 0.6906 | 0.4853 |
| **Matérn** $(\nu = \frac{9}{2})$ | 0.5280 | 0.5746 | 0.4928 | 0.5202 | 0.4722 | 0.6114 | 0.6201 |
| **Cauchy** | 0.5995 | 0.5367 | 0.6060 | 0.5941 | 0.5167 | 0.6096 | 0.5174 |

Table 3.7: Acceptance probability values for each kernel function and prior using CC+SN dataset

| | None | $\hat{H}_0^R$ | $\hat{H}_0^{TRGB}$ | $\hat{H}_0^{HW}$ | $\hat{H}_0^{CM}$ | $\hat{H}_0^P$ | $\hat{H}_0^{DES}$ |
|---|---|---|---|---|---|---|---|
| **Square Exponential** | 0.5550 | 0.5769 | 0.5848 | 0.5934 | 0.5401 | 0.6130 | 0.6131 |
| **Double Square Exponential** | 0.2029 | 0.2569 | 0.4272 | 0.2684 | 0.2099 | 0.3024 | 0.2962 |
| **Rational Quadratic** | 0.4727 | 0.4543 | 0.4211 | 0.4756 | 0.4642 | 0.4599 | 0.6263 |
| **Matérn** $(\nu = \frac{3}{2})$ | 0.5763 | 0.5904 | 0.5940 | 0.6078 | 0.5935 | 0.5542 | 0.6408 |
| **Matérn** $(\nu = \frac{5}{2})$ | 0.6137 | 0.5068 | 0.5758 | 0.5844 | 0.5812 | 0.5295 | 0.5956 |
| **Matérn** $(\nu = \frac{7}{2})$ | 0.6021 | 0.5915 | 0.5610 | 0.5838 | 0.6023 | 0.6043 | 0.5760 |
| **Matérn** $(\nu = \frac{9}{2})$ | 0.5260 | 0.5794 | 0.6181 | 0.4785 | 0.5764 | 0.5938 | 0.5195 |
| **Cauchy** | 0.5428 | 0.5955 | 0.5310 | 0.4632 | 0.4839 | 0.5152 | 0.4826 |

Table 3.8: Acceptance probability values for each kernel function and prior using CC+SN+BAO dataset

| | None | $\hat{H}_0^R$ | $\hat{H}_0^{TRGB}$ | $\hat{H}_0^{HW}$ | $\hat{H}_0^{CM}$ | $\hat{H}_0^P$ | $\hat{H}_0^{DES}$ |
|---|---|---|---|---|---|---|---|
| **Square Exponential** | $\begin{pmatrix}16.70\\13.79\end{pmatrix}$ | $\begin{pmatrix}17.16\\15.38\end{pmatrix}$ | $\begin{pmatrix}17.33\\14.97\end{pmatrix}$ | $\begin{pmatrix}15.65\\13.29\end{pmatrix}$ | $\begin{pmatrix}10.83\\7.86\end{pmatrix}$ | $\begin{pmatrix}18.57\\14.57\end{pmatrix}$ | $\begin{pmatrix}19.66\\16.88\end{pmatrix}$ |
| **Double Square Exponential** | $\begin{pmatrix}22.18\\21.59\\24.50\\25.99\end{pmatrix}$ | $\begin{pmatrix}23.12\\22.73\\23.94\\25.68\end{pmatrix}$ | $\begin{pmatrix}23.33\\22.64\\23.14\\25.35\end{pmatrix}$ | $\begin{pmatrix}23.52\\22.53\\23.02\\24.29\end{pmatrix}$ | $\begin{pmatrix}23.14\\22.13\\22.96\\24.36\end{pmatrix}$ | $\begin{pmatrix}22.86\\22.62\\23.42\\24.76\end{pmatrix}$ | $\begin{pmatrix}23.03\\21.45\\23.15\\24.94\end{pmatrix}$ |
| **Rational Quadratic** | $\begin{pmatrix}22.91\\25.38\\22.80\end{pmatrix}$ | $\begin{pmatrix}23.67\\25.19\\23.41\end{pmatrix}$ | $\begin{pmatrix}23.42\\25.02\\23.53\end{pmatrix}$ | $\begin{pmatrix}23.71\\24.59\\23.64\end{pmatrix}$ | $\begin{pmatrix}22.95\\24.16\\22.62\end{pmatrix}$ | $\begin{pmatrix}23.16\\24.05\\23.17\end{pmatrix}$ | $\begin{pmatrix}22.94\\24.52\\23.51\end{pmatrix}$ |
| **Matérn** $(\nu = \frac{3}{2})$ | $\begin{pmatrix}17.03\\14.19\end{pmatrix}$ | $\begin{pmatrix}14.75\\10.72\end{pmatrix}$ | $\begin{pmatrix}15.56\\14.60\end{pmatrix}$ | $\begin{pmatrix}16.09\\12.76\end{pmatrix}$ | $\begin{pmatrix}17.14\\15.11\end{pmatrix}$ | $\begin{pmatrix}17.90\\16.97\end{pmatrix}$ | $\begin{pmatrix}14.85\\16.21\end{pmatrix}$ |
| **Matérn** $(\nu = \frac{5}{2})$ | $\begin{pmatrix}17.94\\15.68\end{pmatrix}$ | $\begin{pmatrix}20.70\\19.52\end{pmatrix}$ | $\begin{pmatrix}18.22\\19.59\end{pmatrix}$ | $\begin{pmatrix}20.07\\21.06\end{pmatrix}$ | $\begin{pmatrix}22.12\\21.68\end{pmatrix}$ | $\begin{pmatrix}19.85\\18.32\end{pmatrix}$ | $\begin{pmatrix}18.70\\16.91\end{pmatrix}$ |
| **Matérn** $(\nu = \frac{7}{2})$ | $\begin{pmatrix}15.58\\15.47\end{pmatrix}$ | $\begin{pmatrix}15.96\\15.98\end{pmatrix}$ | $\begin{pmatrix}16.25\\14.57\end{pmatrix}$ | $\begin{pmatrix}22.37\\19.20\end{pmatrix}$ | $\begin{pmatrix}16.59\\14.50\end{pmatrix}$ | $\begin{pmatrix}20.54\\16.17\end{pmatrix}$ | $\begin{pmatrix}17.15\\13.75\end{pmatrix}$ |
| **Matérn** $(\nu = \frac{9}{2})$ | $\begin{pmatrix}17.09\\17.48\end{pmatrix}$ | $\begin{pmatrix}20.38\\22.05\end{pmatrix}$ | $\begin{pmatrix}22.87\\21.78\end{pmatrix}$ | $\begin{pmatrix}17.00\\17.08\end{pmatrix}$ | $\begin{pmatrix}17.86\\19.83\end{pmatrix}$ | $\begin{pmatrix}13.44\\12.40\end{pmatrix}$ | $\begin{pmatrix}15.03\\19.43\end{pmatrix}$ |
| **Cauchy** | $\begin{pmatrix}18.36\\18.45\end{pmatrix}$ | $\begin{pmatrix}16.10\\16.67\end{pmatrix}$ | $\begin{pmatrix}20.65\\22.19\end{pmatrix}$ | $\begin{pmatrix}22.68\\23.50\end{pmatrix}$ | $\begin{pmatrix}22.73\\23.25\end{pmatrix}$ | $\begin{pmatrix}16.36\\15.15\end{pmatrix}$ | $\begin{pmatrix}19.65\\18.76\end{pmatrix}$ |

Table 3.9: Estimated autocorrelation times for each prior and kernel function for CC dataset

| | None | $\hat{H}_0^R$ | $\hat{H}_0^{TRGB}$ | $\hat{H}_0^{HW}$ | $\hat{H}_0^{CM}$ | $\hat{H}_0^P$ | $\hat{H}_0^{DES}$ |
|---|---|---|---|---|---|---|---|
| **Square Exponential** | $\begin{pmatrix}18.19\\16.12\end{pmatrix}$ | $\begin{pmatrix}20.49\\21.09\end{pmatrix}$ | $\begin{pmatrix}23.50\\24.79\end{pmatrix}$ | $\begin{pmatrix}23.34\\19.16\end{pmatrix}$ | $\begin{pmatrix}18.43\\18.36\end{pmatrix}$ | $\begin{pmatrix}21.86\\25.75\end{pmatrix}$ | $\begin{pmatrix}23.84\\23.41\end{pmatrix}$ |
| **Double Square Exponential** | $\begin{pmatrix}23.45\\22.97\\22.78\\17.74\end{pmatrix}$ | $\begin{pmatrix}23.77\\22.26\\22.73\\18.65\end{pmatrix}$ | $\begin{pmatrix}18.49\\18.13\\18.28\\18.72\end{pmatrix}$ | $\begin{pmatrix}19.78\\20.31\\20.46\\19.58\end{pmatrix}$ | $\begin{pmatrix}25.69\\20.11\\21.30\\19.11\end{pmatrix}$ | $\begin{pmatrix}20.55\\22.15\\27.39\\20.50\end{pmatrix}$ | $\begin{pmatrix}23.16\\24.18\\23.65\\21.90\end{pmatrix}$ |
| **Rational Quadratic** | $\begin{pmatrix}24.11\\20.56\\23.33\end{pmatrix}$ | $\begin{pmatrix}20.03\\18.96\\21.91\end{pmatrix}$ | $\begin{pmatrix}21.25\\20.78\\22.14\end{pmatrix}$ | $\begin{pmatrix}22.62\\19.37\\23.60\end{pmatrix}$ | $\begin{pmatrix}27.88\\21.30\\22.97\end{pmatrix}$ | $\begin{pmatrix}21.54\\20.57\\23.69\end{pmatrix}$ | $\begin{pmatrix}19.69\\21.42\\22.73\end{pmatrix}$ |
| **Matérn** $(\nu = \frac{3}{2})$ | $\begin{pmatrix}28.81\\28.12\end{pmatrix}$ | $\begin{pmatrix}14.60\\14.48\end{pmatrix}$ | $\begin{pmatrix}26.99\\26.96\end{pmatrix}$ | $\begin{pmatrix}21.47\\21.11\end{pmatrix}$ | $\begin{pmatrix}26.31\\24.58\end{pmatrix}$ | $\begin{pmatrix}22.91\\23.09\end{pmatrix}$ | $\begin{pmatrix}18.04\\15.22\end{pmatrix}$ |
| **Matérn** $(\nu = \frac{5}{2})$ | $\begin{pmatrix}19.75\\17.17\end{pmatrix}$ | $\begin{pmatrix}22.75\\20.54\end{pmatrix}$ | $\begin{pmatrix}22.69\\24.36\end{pmatrix}$ | $\begin{pmatrix}20.55\\27.70\end{pmatrix}$ | $\begin{pmatrix}21.45\\18.72\end{pmatrix}$ | $\begin{pmatrix}22.06\\16.58\end{pmatrix}$ | $\begin{pmatrix}12.14\\11.35\end{pmatrix}$ |
| **Matérn** $(\nu = \frac{7}{2})$ | $\begin{pmatrix}26.72\\23.23\end{pmatrix}$ | $\begin{pmatrix}9.97\\8.73\end{pmatrix}$ | $\begin{pmatrix}16.24\\14.05\end{pmatrix}$ | $\begin{pmatrix}25.99\\17.08\end{pmatrix}$ | $\begin{pmatrix}21.02\\17.38\end{pmatrix}$ | $\begin{pmatrix}22.94\\19.83\end{pmatrix}$ | $\begin{pmatrix}9.58\\7.64\end{pmatrix}$ |
| **Matérn** $(\nu = \frac{9}{2})$ | $\begin{pmatrix}21.05\\27.67\end{pmatrix}$ | $\begin{pmatrix}22.72\\22.00\end{pmatrix}$ | $\begin{pmatrix}21.98\\19.91\end{pmatrix}$ | $\begin{pmatrix}18.55\\18.51\end{pmatrix}$ | $\begin{pmatrix}17.11\\14.28\end{pmatrix}$ | $\begin{pmatrix}19.62\\15.74\end{pmatrix}$ | $\begin{pmatrix}17.31\\16.51\end{pmatrix}$ |
| **Cauchy** | $\begin{pmatrix}23.66\\23.17\end{pmatrix}$ | $\begin{pmatrix}27.19\\28.00\end{pmatrix}$ | $\begin{pmatrix}25.37\\25.28\end{pmatrix}$ | $\begin{pmatrix}29.91\\28.19\end{pmatrix}$ | $\begin{pmatrix}27.54\\20.79\end{pmatrix}$ | $\begin{pmatrix}29.55\\29.89\end{pmatrix}$ | $\begin{pmatrix}31.58\\29.69\end{pmatrix}$ |

Table 3.10: Estimated autocorrelation times for each prior and kernel function for CC+SN dataset

| | None | $\hat{H}_0^R$ | $\hat{H}_0^{TRGB}$ | $\hat{H}_0^{HW}$ | $\hat{H}_0^{CM}$ | $\hat{H}_0^P$ | $\hat{H}_0^{DES}$ |
|---|---|---|---|---|---|---|---|
| **Square Exponential** | $\begin{pmatrix}23.88\\18.47\end{pmatrix}$ | $\begin{pmatrix}20.30\\15.77\end{pmatrix}$ | $\begin{pmatrix}26.39\\26.92\end{pmatrix}$ | $\begin{pmatrix}27.78\\21.11\end{pmatrix}$ | $\begin{pmatrix}31.26\\32.26\end{pmatrix}$ | $\begin{pmatrix}25.01\\29.68\end{pmatrix}$ | $\begin{pmatrix}22.36\\21.52\end{pmatrix}$ |
| **Double Square Exponential** | $\begin{pmatrix}22.99\\22.89\\23.18\\19.41\end{pmatrix}$ | $\begin{pmatrix}25.81\\26.51\\20.25\\18.08\end{pmatrix}$ | $\begin{pmatrix}33.85\\34.76\\34.16\\33.37\end{pmatrix}$ | $\begin{pmatrix}19.60\\21.79\\18.32\\13.73\end{pmatrix}$ | $\begin{pmatrix}15.86\\17.95\\21.02\\22.12\end{pmatrix}$ | $\begin{pmatrix}15.36\\17.41\\18.35\\18.73\end{pmatrix}$ | $\begin{pmatrix}21.20\\26.07\\21.83\\15.56\end{pmatrix}$ |
| **Rational Quadratic** | $\begin{pmatrix}24.47\\23.33\\25.07\end{pmatrix}$ | $\begin{pmatrix}27.95\\30.31\\33.84\end{pmatrix}$ | $\begin{pmatrix}28.73\\33.25\\34.89\end{pmatrix}$ | $\begin{pmatrix}26.06\\23.83\\64.70\end{pmatrix}$ | $\begin{pmatrix}32.12\\27.05\\24.90\end{pmatrix}$ | $\begin{pmatrix}24.44\\26.01\\22.78\end{pmatrix}$ | $\begin{pmatrix}28.48\\27.22\\28.92\end{pmatrix}$ |
| **Matérn** $(\nu = \frac{3}{2})$ | $\begin{pmatrix}21.51\\21.53\end{pmatrix}$ | $\begin{pmatrix}24.01\\21.89\end{pmatrix}$ | $\begin{pmatrix}24.12\\25.94\end{pmatrix}$ | $\begin{pmatrix}31.56\\35.03\end{pmatrix}$ | $\begin{pmatrix}34.79\\35.79\end{pmatrix}$ | $\begin{pmatrix}27.62\\34.88\end{pmatrix}$ | $\begin{pmatrix}33.36\\34.80\end{pmatrix}$ |
| **Matérn** $(\nu = \frac{5}{2})$ | $\begin{pmatrix}26.13\\25.75\end{pmatrix}$ | $\begin{pmatrix}31.81\\32.61\end{pmatrix}$ | $\begin{pmatrix}13.95\\14.61\end{pmatrix}$ | $\begin{pmatrix}28.59\\29.69\end{pmatrix}$ | $\begin{pmatrix}31.31\\27.81\end{pmatrix}$ | $\begin{pmatrix}26.52\\34.18\end{pmatrix}$ | $\begin{pmatrix}34.90\\36.84\end{pmatrix}$ |
| **Matérn** $(\nu = \frac{7}{2})$ | $\begin{pmatrix}26.34\\27.38\end{pmatrix}$ | $\begin{pmatrix}33.29\\30.28\end{pmatrix}$ | $\begin{pmatrix}26.76\\31.01\end{pmatrix}$ | $\begin{pmatrix}27.15\\26.07\end{pmatrix}$ | $\begin{pmatrix}33.81\\30.93\end{pmatrix}$ | $\begin{pmatrix}28.55\\27.40\end{pmatrix}$ | $\begin{pmatrix}31.79\\34.65\end{pmatrix}$ |
| **Matérn** $(\nu = \frac{9}{2})$ | $\begin{pmatrix}24.17\\27.28\end{pmatrix}$ | $\begin{pmatrix}11.60\\10.02\end{pmatrix}$ | $\begin{pmatrix}21.09\\23.06\end{pmatrix}$ | $\begin{pmatrix}21.11\\26.04\end{pmatrix}$ | $\begin{pmatrix}14.89\\12.77\end{pmatrix}$ | $\begin{pmatrix}34.01\\30.59\end{pmatrix}$ | $\begin{pmatrix}20.42\\17.45\end{pmatrix}$ |
| **Cauchy** | $\begin{pmatrix}24.60\\16.44\end{pmatrix}$ | $\begin{pmatrix}20.40\\19.31\end{pmatrix}$ | $\begin{pmatrix}31.44\\30.65\end{pmatrix}$ | $\begin{pmatrix}25.19\\23.90\end{pmatrix}$ | $\begin{pmatrix}12.21\\12.28\end{pmatrix}$ | $\begin{pmatrix}19.55\\22.77\end{pmatrix}$ | $\begin{pmatrix}14.89\\20.62\end{pmatrix}$ |

Table 3.11: Estimated autocorrelation times for each prior and kernel function for CC+SN+BAO dataset

# Chapter 4

# Student's $t$-Process Regression

In this chapter, we shall introduce the Student's $t$-process (TP) and the Student's $t$-process regression (TPR) model in a manner which is analogous to GPs and GPR as described in Chapter 2. Just as the defining characteristic of a GP is that any finite sample of random variables follows the multivariate normal distribution, that of a TP is that any such finite sample follows the multivariate Student's $t$-distribution. A TP is defined in a similar way to a GP, except that an additional degrees of freedom parameter $\nu$ must be specified. Two possible advantages of TPR when compared to GPR are (a) the additional flexibility granted by the additional degrees of freedom parameter of the $t$-distribution and (b) the better performance on data with outliers due to the comparatively heavier tails of the Student's $t$-distribution (see e.g. Kirkby et al., 2019, p. 2).

In Section 4.1, we start by providing the necessary preliminaries for introducing and constructing Student's $t$-processes. In particular, we first define the multivariate Student's $t$-distribution, and then we provide two ways of constructing a TP. This is followed by the derivation of some desirable properties of the TP, such as its convergence to a GP as the parameter $\nu \to \infty$. Then, in Section 4.3, we describe the theory behind TPR. In particular, we consider the TPRD model due to Tang et al. (2017), which considers the noise within the regression model as dependent multivariate Student's $t$-distributed. Lastly, in Section 4.4, we carry out TPR on the $H_0$ data sources described in Section 1.4, as was done with GPR in Section 2.4. As is the case throughout the rest of the dissertation, our aim here is to obtain different estimates for $H_0$ through the use of different datasets, kernel functions, and 'prior' $H_0$ estimates.

## 4.1 Defining and Constructing Student's $t$-Processes

We start from the definition of the univariate Student's $t$-distribution:

**Definition 4.1.1** (Student's $t$-distribution)**.** The random variable $X$ is said to follow a (univariate) Student's $t$-distribution with mean $m \in \mathbb{R}$ and degrees of freedom parameter equal to $\nu > 2$ if it has probability density function given by:

$$\mathbb{P}(x) = \frac{\Gamma(\frac{\nu+1}{2})}{(\nu\pi)^{\frac{1}{2}}\Gamma\left(\frac{\nu}{2}\right)}\left[1 + \frac{(x-m)^2}{\nu}\right]^{-\frac{\nu+1}{2}}.$$

The degrees of freedom parameter controls the heaviness of the tails. In particular, a lower value of $\nu$ results in heavier tails. The Student's $t$-distribution assigns comparatively higher probabilities for values of $x$ far away from the mean when compared to the Gaussian, but this difference decreases as $\nu$ is increased. At the limit $\nu \to \infty$, then the Student's $t$-distribution becomes equivalent to the Gaussian distribution.

The multivariate Student's $t$-distribution is defined similarly to the corresponding univariate distribution. Therefore, we may define it by specifying the joint probability density function as follows:

**Definition 4.1.2** (Multivariate Student's $t$-distribution)**.** The random vector $\mathbf{X}$ is multivariate Student's $t$-distributed with mean function $\mathbf{m}$, positive definite $(p \times p)$ scale matrix $\mathbb{K} \in \Pi(p)$, and degrees of freedom parameter $\nu > 2$ if it has joint probability density function

$$\mathbb{P}(\mathbf{x}) = \frac{\Gamma(\frac{\nu+p}{2})}{(\nu\pi)^{\frac{p}{2}}\Gamma\left(\frac{\nu}{2}\right)}|\mathbb{K}|^{-\frac{1}{2}} \times \left[1 + \frac{(\mathbf{x}-\mathbf{m})^T\mathbb{K}^{-1}(\mathbf{x}-\mathbf{m})}{\nu}\right]^{-\frac{\nu+p}{2}}, \tag{4.1}$$

where $p$ is the number of dimensions (see e.g. Tang et al., 2017). We denote this as $\mathbf{X} \sim MVT_p(\mathbf{m}, \mathbb{K}, \nu)$.

The conditional distribution for this distribution can be expressed analytically. In particular, given $\mathbf{X} \sim MVT_p(\mathbf{m}, \mathbb{K}, \nu)$ partitioned into $\mathbf{X}_1$ and $\mathbf{X}_2$ of lengths $p_1$ and $p_2$ respectively, then:

$$\mathbf{X}_2|\mathbf{X}_1 \sim MVT_{p_2}\left[\widetilde{\mathbf{m}}_2,\ \frac{\beta_1+\nu}{p_1+\nu}\,\widetilde{\mathbb{K}}_{22},\ \nu+p_1\right] \tag{4.2}$$

(Shah et al., 2014), where:

$$\begin{aligned}\widetilde{\mathbf{m}}_2 &= \mathbb{K}_{21}\mathbb{K}_{11}^{-1}(\mathbf{x}_1 - \mathbf{m}_1) - \mathbf{m}_2,\\ \beta_1 &= (\mathbf{x}_1 - \mathbf{m}_1)^T\mathbb{K}_{11}^{-1}(\mathbf{x}_1 - \mathbf{m}_1),\\ \widetilde{\mathbb{K}}_{22} &= \mathbb{K}_{22} - \mathbb{K}_{21}\mathbb{K}_{11}^{-1}\mathbb{K}_{12},\end{aligned}$$

$\mathbf{m}_1$ and $\mathbf{m}_2$ are the mean vectors corresponding to $\mathbf{X}_1$ and $\mathbf{X}_2$ respectively, and $\mathbb{K}_{ij}$ for $i, j \in \{1, 2\}$ is the relevant $(p_i \times p_j)$ submatrix of $\mathbb{K}$. The derivation of equation (4.2) is given in Theorem A.1 within Appendix A.

Moreover, as mentioned in the introduction of this chapter, a TP is defined in a similar manner to a GP, that is, as a stochastic process from which any finite selection of random variables is jointly multivariate Student's $t$-distributed. A formal definition of a TP is provided in Definition 4.1.3.

**Definition 4.1.3** (Student's $t$-process)**.** Let $X \subseteq \mathbb{R}^p$ be a nonempty indexing set containing elements of some dimension $p$, and define the constant $\nu > 2$. Also, let $k : X \times X \times \Theta \to \mathbb{R}$ be a positive definite kernel and $m : X \to \mathbb{R}$ be a real-valued function. Moreover, consider any combination $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ of size $n \in \mathbb{N}$ of elements in $X$, and let $\mathbf{f} := (f(\mathbf{x}_1), f(\mathbf{x}_2), ..., f(\mathbf{x}_n))^T$. Then, a random function $f : X \to \mathbb{R}$ is said to be a TP with parameters $m, k, \nu$ if the random vector $\mathbf{F}$ is distributed as $MVT_n(\mathbf{m}, \mathbb{K}, \nu)$. In other words, $\mathbf{F}$ has the $n$-variate Student's $t$-distribution with mean vector $\mathbf{m} := (m(\mathbf{x}_1), m(\mathbf{x}_2), ..., m(\mathbf{x}_n))^T$, and for which the $(n \times n)$ scale matrix $\mathbb{K}$ has $(i, j)^{\text{th}}$ entry equal to $k(\mathbf{x}_i, \mathbf{x}_j)$. Similar to the notation for Gaussian processes, we denote a TP using $f \sim TP(m, k, \nu)$.

A TP may be constructed using the inverse-Wishart and multivariate Gaussian distributions. In Definition 4.1.4, we define the Wishart and inverse-Wishart distributions, which are multivariate versions of the gamma and inverse-gamma distributions. The former distributions are defined over the set $\Pi(p)$ of $(p \times p)$ positive-definite real matrices instead of over the positive real numbers.

**Definition 4.1.4** (Wishart and inverse-Wishart distributions)**.** The matrix $\overline{\mathbb{\Sigma}} \in \Pi(p)$ is Wishart-distributed with parameters $\mathbb{K} \in \Pi(p)$ and $\nu > p - 1$ if it has joint probability density function

$$\mathbb{P}(\overline{\mathbb{\Sigma}}) = \left[|\mathbb{K}|^{\frac{\nu}{2}} 2^{\frac{\nu p}{2}} \Gamma_p(\nu/2)\right]^{-1} |\overline{\mathbb{\Sigma}}|^{\frac{\nu-p-1}{2}} \exp\left[-\frac{1}{2}\mathrm{Tr}(\mathbb{K}^{-1}\overline{\mathbb{\Sigma}})\right],$$

where $\Gamma_p := \int_{\mathcal{S}} \det(S)^{x-\frac{p+1}{2}} \exp\big(\mathrm{Tr}(S)\big)\, \mathrm{d}S$ is a multivariate version of the gamma function. In such a case, we write $\overline{\mathbb{\Sigma}} \sim W_p(\mathbb{K}, \nu)$, and it follows that $\overline{\mathbb{\Sigma}}^{-1} \sim W_p(\mathbb{K}^{-1}, \nu)$.

Similarly, $\mathbb{\Sigma}$ is inverse-Wishart-distributed with parameters $\mathbb{K} \in \Pi(p)$ and $\nu \in \mathbb{R}_+$, denoted $\mathbb{\Sigma} \sim IW_p(\mathbb{K}, \nu)$, if it has joint probability density function

$$\mathbb{P}(\mathbb{\Sigma}) = \frac{|\mathbb{K}|^{\frac{\nu+p-1}{2}}}{2^{\frac{(\nu+p-1)p}{2}}\Gamma_p\left(\frac{\nu+p-1}{2}\right)}|\mathbb{\Sigma}|^{-\frac{\nu+2p}{2}}\exp\left[-\frac{1}{2}\mathrm{Tr}(\mathbb{K}\mathbb{\Sigma}^{-1})\right].$$

Dawid (1981) shows that any $(m \times m)$ principal submatrix $\mathbb{\Sigma}_m$ of $\mathbb{\Sigma}$ would be distributed as $IW_m(\nu, \mathbb{K}_m)$, where $\mathbb{K}_m$ is the corresponding $(m \times m)$ principal submatrix of $\mathbb{K}$.

Similarly to a GP, one might want to define a 'Wishart process', that is, a stochastic process from which any finite sample of random matrices is Wishart-distributed, but the restriction $\nu > p - 1$ on the degrees of freedom parameter prevents us from constructing Wishart marginals of arbitrary size. However, the inverse-Wishart distribution is not restricted by such a condition. Therefore an inverse-Wishart process $IWP(k, \nu)$ can easily be defined, where the joint distribution of any $n$ random matrices sampled from this process is inverse-Wishart-distributed with kernel function $k$ and $\nu$ degrees of freedom. Shah et al. (2014) give the definition of an IWP:

**Definition 4.1.5** (Inverse-Wishart process)**.** Let $X \subseteq \mathbb{R}^p$ be a nonempty input space, and let $k : X \times X \times \Theta \to \mathbb{R}$ be a kernel function. Then, $\sigma$ is an IWP on with kernel function $k$ and $\nu$ degrees of freedom, denoted $\sigma \sim IWP(k, \nu)$, if, for any finite collection of elements $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ in $X$, the joint covariance $\mathbb{\Sigma} := \sigma(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n) \sim IW_n(\nu, \mathbb{K})$, where $\mathbb{K} \in \Pi(n)$ and the $(i, j)^{\text{th}}$ entry of $\mathbb{K}$ is equal to $k(\mathbf{x}_i, \mathbf{x}_j)$.

We now state two theorems that may be used to construct Student's $t$-processes; these theorems further show that GPs and TPs are highly related. Since the Student's $t$-distribution may be considered as a mixture of Gaussian distributions (see e.g. Cook, 2009), we may construct a Student's $t$-process by placing an IWP prior on the kernel function, as proposed by Shah et al. (2014).

**Theorem 4.1.6.** *Let $\mathbf{x} \in X \subseteq \mathbb{R}^p$, $k : X \times X \times \Theta \to \mathbb{R}$ be a kernel, $m : X \to \mathbb{R}$ be a mean function, and $\nu > 2$. Then, the generative process*

$$\begin{aligned} \sigma &\sim IWP(k, \nu) \\ x|\sigma &\sim GP(m, \nu\mathbb{\Sigma}). \end{aligned} \tag{4.3}$$

*results in the marginal distribution $\mathbf{X} \sim MVT_p(\mathbf{m}, \mathbb{K}, \nu)$.*

An alternate method for constructing a TP that avoids use of the inverse-Wishart distribution is also given by Yu et al. (2007):

**Theorem 4.1.7.** *Suppose that* $\mathbf{x} \in X \subseteq \mathbb{R}^p$, $\mathbb{K} \in \Pi(p)$, $m : X \to \mathbb{R}^p$, $\nu > 2$, *and* $\rho > 0$. *Then, using the generative process*

$$\begin{aligned} R^{-1} &\sim Gamma\left(\frac{\nu}{2}, \frac{\rho}{2}\right) \\ \mathbf{X}|R &\sim MVN_p\left(\mathbf{m}, \frac{\nu r}{\rho}\mathbb{K}\right) \end{aligned} \tag{4.4}$$

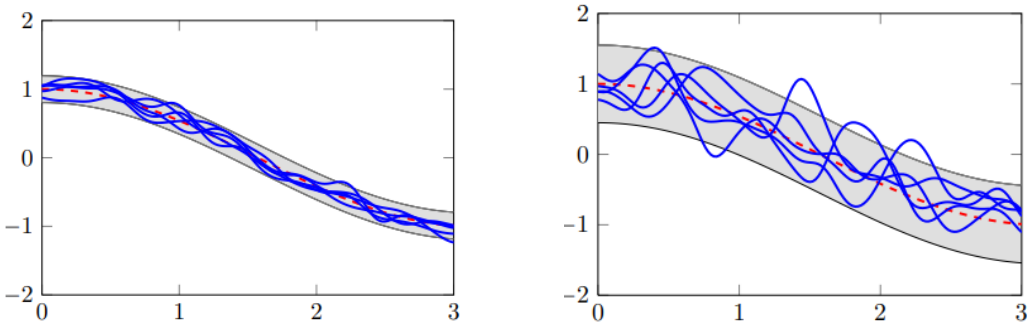
*we get the marginal distribution* $\mathbf{X} \sim MVT_p(\mathbf{m}, \mathbb{K}, \nu)$.

Therefore, the two generative processes of Theorems 4.1.6 and 4.1.7 are identical in that they produce the same marginal distribution for $\mathbf{X}$. The proof of this result is given in Theorem A.2 within Appendix A.

## 4.2 Properties of Student's $t$-Processes

Many of the properties of TPs follow from those of GPs, since the Student's $t$-distribution is closely related to the Gaussian distribution. In particular, given some fixed $\mathbf{m}$ and $\mathbb{K}$, then a TP becomes equivalent to the corresponding GP at the limit $\nu \to \infty$. This follows immediately from the fact that the multivariate Student's $t$-distribution converges to the multivariate Gaussian distribution at this limit (see e.g. Cassidy, 2016), and from Definitions 2.1.1 and 4.1.3.

Moreover, the additional 'degrees of freedom' parameter $\nu$ within the TP controls the degree of dependence between jointly $t$-distributed variables (Shah et al., 2014) as well as the heaviness of the tails, with smaller values of $\nu$ corresponding to heavier tails. In particular, the tails of the MVT distribution decay at a polynomial rate. On the other hand, those of the multivariate Gaussian distribution decay at an exponential rate, and thus the MVT distribution has comparatively heavier tails (see e.g. Kirkby et al., 2019, p. 2). This indicates that TPs are in general more robust to outliers than GPs, since the GP forces samples to be closer to the mean, as can be seen in Figure 4.1.



(a) GP (b) TP with $\nu = 5$

Figure 4.1: Samples (in blue) taken from a Gaussian process and a Student's $t$-process with $\nu = 5$; grey shaded areas are the 95% confidence regions (from Shah et al., 2014)

However, the behaviour of a TP becomes closer to that of a GP as we increase $\nu$. Additionally, from equation (4.2), we get that the posterior predictive distribution of a TP converges to that of a GP as $\nu \to \infty$. From this result, we also get

$$\mathbb{E}[\mathbf{X}_2|\mathbf{X}_1 = \mathbf{x}_1] = \widetilde{\mathbf{m}}_2 \tag{4.5a}$$

and

$$\mathrm{Cov}[\mathbf{X}_2|\mathbf{X}_1 = \mathbf{x}_1] = \frac{\beta_1 + \nu}{p_1 + \nu}\,\widetilde{\mathbb{K}}_{22}. \tag{4.5b}$$

Therefore, assuming that the same kernel with the same hyperparameters is used, the posterior predictive mean of a TP has the same form as in a GP, since equation (4.5a) is independent of $\nu$. However, the same cannot be said for the predictive covariance of equation (4.5b). In fact, $\beta_1 = (\mathbf{x}_1 - \mathbf{m}_1)^T \mathbb{K}_{11}^{-1}(\mathbf{x}_1 - \mathbf{m}_1)$ implies that the predictive covariance for a TP depends on the observed values. This allows for increased flexibility in using a TP when compared with GPs, as in the latter case the predictive covariance is independent of the training observations.

Both GPs and TPs are also said to be elliptical processes, in that the joint distribution of any finite set of random variables sampled from such a process is elliptically symmetric. This follows from the elliptical symmetry of both the multivariate Gaussian distribution and the multivariate Student's $t$-distribution (see e.g. Hult and Lindskog, 2002). Intuitively, this means that both distributions are unimodal and are 'smooth' in that the likelihood of a point is decreasing as the distance from this mode increases. This smoothness property makes elliptical processes desirable for multivariate modelling applications.

While elliptical processes are desirable for multivariate modelling applications such as

regression, not all elliptical processes have defined densities. In fact, it can be shown that the only two elliptical processes with a defined density function are the GP and the TP (Shah et al., 2014, p. 881), with the TP being a more flexible version of a GP. Therefore, the TP is an ideal candidate for such tasks. The next section will in fact describe the TPR framework, which is analogous to GPR as described in Section 2.3. When compared to GPR, however, TPR is more robust to outliers and is also less prone to overfitting (see e.g. Shah et al., 2014; Zhang and Yeung, 2010). Therefore, TPR can help mitigate some of the problems of GPR while additionally making it possible to model data which might not adhere to the normality assumption.

## 4.3 Theory of Student's $t$-Process Regression

Within the TPR framework, we assume the same basic model as in equation (2.3), i.e. $y_i = f(\mathbf{x}_i) + \epsilon_i$ for $i \in \{1, 2, ..., n\}$, which may be vectorised as $\mathbf{y} = f(\mathbb{X}) + \boldsymbol{\epsilon}$. However, in this case, we assign $f$ a Student's $t$-process prior $f \sim TP(m, k, \nu)$. Therefore, by Definition 4.1.3, the vector $\mathbf{f} := (f(\mathbf{x}_1), f(\mathbf{x}_2), ..., f(\mathbf{x}_n))^T$ is distributed as $\mathbf{F} \sim MVT_n(\mathbf{m}, \mathbb{K}, \nu)$, i.e.:

$$\mathbb{P}(\mathbf{f}) = \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{(\nu\pi)^{\frac{n}{2}}\Gamma\left(\frac{\nu}{2}\right)}|\mathbb{K}|^{-\frac{1}{2}}\left[1 + \frac{1}{\nu}(\mathbf{f}-\mathbf{m})^T\mathbb{K}^{-1}(\mathbf{f}-\mathbf{m})\right]^{-\frac{\nu+n}{2}}. \tag{4.6}$$

As was the case in GPR, we have that $\mathbf{m} := (m(\mathbf{x}_1), m(\mathbf{x}_2), ..., m(\mathbf{x}_n))^T$ and that $\mathbb{K}$ is the $(n \times n)$ covariance matrix with $(i,j)^{\text{th}}$ term equal to $k(\mathbf{x}_i, \mathbf{x}_j)$. Also as in GPR, the mean vector $\mathbf{m}$ is often set to $\mathbf{0}$ for simplicity, as the TP is flexible enough that specification of a prior mean is not required.

On letting $\mathbf{m} = \mathbf{0}$, equation (4.6) becomes slightly simpler, i.e:

$$\mathbb{P}(\mathbf{f}) = \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{(\nu\pi)^{\frac{n}{2}}\Gamma\left(\frac{\nu}{2}\right)}|\mathbb{K}|^{-\frac{1}{2}}\left[1 + \frac{1}{\nu}\mathbf{f}^T\mathbb{K}^{-1}\mathbf{f}\right]^{-\frac{\nu+n}{2}}. \tag{4.7}$$

In the TPRD model (Tang et al., 2017), we let the noise vector $\boldsymbol{\epsilon} := (\epsilon_1, \epsilon_2, ..., \epsilon_n)^T$ follow an $n$-dimensional MVT distribution with correlation matrix dependent on how well the noise-free model $\mathbf{y} = f(\mathbb{X})$ fits to the data. In particular, the noise vector is distributed as:

$$\boldsymbol{\mathcal{E}}|\beta \sim MVT_n\left[\mathbf{m},\ \left(1 + \frac{1}{\nu}(\mathbf{f}-\mathbf{m})^T\mathbb{K}^{-1}(\mathbf{f}-\mathbf{m})\right)\frac{1}{\beta}\mathbb{I}_n,\ \nu+n\right], \tag{4.8}$$

where $\beta$ is a constant scaling factor and $\mathbb{I}_n$ is the $(n \times n)$ identity matrix. The quantity $\frac{1}{\beta}\mathbb{I}_n$ thus ensures that the covariance matrix of the noise distribution is indeed a matrix, as is required from the definition of the multivariate Student's $t$-distribution. Under the assumption that $\mathbf{m} = \mathbf{0}$, then the expression (4.8) may be rewritten as:

$$\boldsymbol{\mathcal{E}}|\beta \sim MVT_n\left[\mathbf{0},\ \left(1 + \frac{1}{\nu}\mathbf{f}^T\mathbb{K}^{-1}\mathbf{f}\right)\frac{1}{\beta}\mathbb{I}_n,\ \nu + n\right], \tag{4.9}$$

and then the density function of the noise in equation (4.9) can be written down as:

$$\mathbb{P}(\boldsymbol{\epsilon}|\beta) = \frac{\Gamma(\frac{\nu+2n}{2})}{((\nu+n)\pi)^{\frac{n}{2}}\Gamma(\frac{\nu+n}{2})} \times \frac{\beta^{\frac{n}{2}}}{\left(1 + \frac{1}{\nu}\mathbf{f}^T\mathbb{K}^{-1}\mathbf{f}\right)^{\frac{n}{2}}} \times \left[1 + \frac{\beta}{\nu+n}\ \frac{\boldsymbol{\epsilon}^T\boldsymbol{\epsilon}}{1 + \frac{1}{\nu}\mathbf{f}^T\mathbb{K}^{-1}\mathbf{f}}\right]^{-\frac{\nu+2n}{2}}. \tag{4.10}$$

In equation (4.10), the term $\frac{1}{\nu}\mathbf{f}^T\mathbb{K}^{-1}\mathbf{f}$ is the data-fit term. This can be compared with the GPR case, for example with the marginal log-likelihood of equation (2.14). In particular, if the noise-free model fits the data well, then this term would be small, since the negative marginal log-likelihood would also be small. Additionally, under the TPRD model, the degrees of freedom parameter of the distribution of the noise terms is $\nu + n$. This means that for large datasets, or if $\nu$ is sufficiently large, the distribution of $\boldsymbol{\mathcal{E}}$ can be approximated by a multivariate Gaussian distribution with a diagonal covariance matrix, implying that the noise terms $\epsilon_i$ would be approximated by independent and identically-distributed Gaussian random variables. As a result, TPRD is essentially a more robust and generalised version of GPR.

Based on equations (4.6) and (4.8), we obtain the likelihood derived from the multivariate Student's $t$-distribution:

$$\mathbf{Y}|\mathbf{F}, \beta \sim MVT_n\left[\mathbf{f} + \mathbf{m},\ \left(1 + \frac{1}{\nu}(\mathbf{f} - \mathbf{m})^T\mathbb{K}^{-1}(\mathbf{f} - \mathbf{m})\right)\frac{1}{\beta}\mathbb{I},\ \nu + n\right]. \tag{4.11}$$

Therefore, assuming that $\mathbf{m} = \mathbf{0}$, then:

$$\mathbb{P}(\mathbf{y}|\mathbf{f}, \beta) = \frac{\Gamma(\frac{\nu+2n}{2})}{((\nu+n)\pi)^{\frac{n}{2}}\Gamma(\frac{\nu+n}{2})} \times \frac{\beta^{\frac{n}{2}}}{\left(1 + \frac{1}{\nu}\mathbf{f}^T\mathbb{K}^{-1}\mathbf{f}\right)^{\frac{n}{2}}} \times \left[1 + \frac{\beta}{\nu+n}\ \frac{(\mathbf{y}-\mathbf{f})^T(\mathbf{y}-\mathbf{f})}{1 + \frac{1}{\nu}\mathbf{f}^T\mathbb{K}^{-1}\mathbf{f}}\right]^{-\frac{\nu+2n}{2}}. \tag{4.12}$$

The joint distribution of $\mathbf{y}$ and $\mathbf{f}$ is given by the multiplication of equations (4.7) and (4.12), i.e.:

$$\begin{aligned}\mathbb{P}(\mathbf{y},\mathbf{f}|\beta) &\propto \left[1+\frac{1}{\nu}\mathbf{f}^T\mathbb{K}^{-1}\mathbf{f}+\frac{\beta}{\nu+n}(\mathbf{y}-\mathbf{f})^T(\mathbf{y}-\mathbf{f})\right]^{-\frac{\nu+2n}{2}} \\ &\propto \left[1+\frac{\beta}{\nu+n}\mathbf{y}^T\left(\mathbb{I}-\frac{\beta}{\nu+n}\mathbb{A}^{-1}\right)\mathbf{y}+(\mathbf{f}-\bar{\mathbf{f}})^T\mathbb{A}(\mathbf{f}-\bar{\mathbf{f}})\right]^{-\frac{\nu+2n}{2}}, \end{aligned} \tag{4.13}$$

where $\mathbb{A}=\frac{1}{\nu}\mathbb{K}^{-1}+\frac{\beta}{\nu+n}\mathbb{I}$ and $\bar{\mathbf{f}}=\frac{\beta}{\nu+n}\mathbb{A}^{-1}\mathbf{y}$.

Then, we obtain the marginal likelihood by integrating out $\mathbf{f}$ in equation (4.13), i.e.

$$\mathbb{P}(\mathbf{y}|\boldsymbol{\theta},\mathbb{K})\propto\left[1+\frac{\beta}{\nu+n}\mathbf{y}^T\left(\mathbb{I}-\frac{\beta}{\nu+n}\mathbb{A}^{-1}\right)\mathbf{y}\right]^{-\frac{\nu+n}{2}}$$

and in fact:

$$\mathbb{P}(\mathbf{y}|\boldsymbol{\theta},\mathbb{K})=\frac{\Gamma(\frac{\nu+n}{2})}{(\nu\pi)^{\frac{n}{2}}\Gamma\left(\frac{\nu}{2}\right)}|\mathbb{\Sigma}|^{-1/2}\left[1+\frac{1}{\nu}\mathbf{y}^T\mathbb{\Sigma}^{-1}\mathbf{y}\right]^{-\frac{\nu+n}{2}}, \tag{4.14}$$

where the term $\frac{\Gamma(\frac{\nu+n}{2})}{(\nu\pi)^{\frac{n}{2}}\Gamma\left(\frac{\nu}{2}\right)}$ is a normalising constant and $\mathbb{\Sigma}=\mathbb{K}+\frac{\nu+n}{\nu}\frac{1}{\beta}\mathbb{I}$, i.e. the correlation matrix of $\mathbf{Y}$. Consequently, $\mathbb{\Sigma}^{-1}=\frac{\nu\beta}{\nu+n}\left[\mathbb{I}-\frac{\beta}{\nu+n}\mathbb{A}^{-1}\right]$.

Therefore, the marginal log-likelihood for the TPRD model is given by:

$$\log\mathbb{P}(\mathbf{y}|\boldsymbol{\theta},\mathbb{K})=-\frac{\nu+n}{2}\log\left[1+\frac{1}{\nu}\mathbf{y}^T\mathbb{\Sigma}^{-1}\mathbf{y}\right]-\frac{1}{2}\log|\mathbb{\Sigma}|+\log\left[\frac{\Gamma(\frac{\nu+n}{2})}{(\nu\pi)^{\frac{n}{2}}\Gamma\left(\frac{\nu}{2}\right)}\right]. \tag{4.15}$$

Comparing with equation (2.14) for the GPR case, the log-likelihood in this case is highly similar. In fact, the first term is related to the data-fit term $\mathbf{y}^T\mathbb{\Sigma}^{-1}\mathbf{y}$, while the second and third terms are a model complexity penalty and a normalisation constant respectively. The main difference from GPR in this case is that the first term in the TPRD marginal log-likelihood is a logarithmic, rather than linear, function of $\mathbf{y}^T\mathbb{\Sigma}^{-1}\mathbf{y}$. This provides further evidence in favour of the robustness of TPRD, as any outliers in $\mathbf{y}$ would disturb the marginal log-likelihood for TPRD less than the equivalent in GPR.

Moreover, comparing with the result of Shah et al. (2014), who used the TPRK model, we get that the marginal log-likelihood of TPRK is equivalent to that of TPRD, with the only difference arising since the authors use a slightly different definition for the multivariate Student's $t$-distribution as outlined in Definition 4.1.2. Therefore, TPRD can be said to give a "probabilistic interpretation" to TPRK (Tang et al., 2017, p. 85).

The derivative of the marginal log-likelihood as defined in equation (4.15) with respect to some hyperparameter $\theta_i$ of the kernel is given in equation (4.16):

$$\frac{\partial}{\partial \theta_i} \log \mathbb{P}(\mathbf{y}|\nu, \mathbb{K}) = \frac{1}{2}\text{Tr}\left[\left(\frac{\nu+n}{\nu+\beta-2}\boldsymbol{\alpha}\boldsymbol{\alpha}^T - \mathbb{K}^{-1}\right)\frac{\partial \mathbb{\Sigma}}{\partial \theta_i}\right] \tag{4.16}$$

where $\boldsymbol{\alpha} = \mathbb{\Sigma}^{-1}(\mathbf{y}-\mathbf{m})$. The hyperparameters of the kernel may each be inferred using cross-validation or numerical optimisation techniques, such as gradient-based optimisation or MCMC methods, similar to what was done in Sections 2.4 and 3.4 for GPR, and as will be done in Section 4.4 for TPR.

The ideal value of the degrees of freedom parameter $\nu$ can be inferred similarly, by setting the derivative

$$\begin{aligned}\frac{\partial}{\partial \nu} \log \mathbb{P}(\mathbf{y}|\nu, \mathbb{K}) = &-\frac{n}{2(\nu-2)} + \psi\left(\frac{\nu+n}{2}\right) - \psi\left(\frac{\nu}{2}\right) \\ &- \frac{1}{2}\log\left[1+\frac{\beta}{\nu-2}\right] + \frac{(\nu+n)\beta}{2(\nu-2)^2+2\beta(\nu-2)}\end{aligned} \tag{4.17}$$

to 0, where $\psi(x) := \frac{\mathrm{d}}{\mathrm{d}x}\log(\Gamma(x))$ is the digamma function. Such estimates inherit all the properties of maximum likelihood estimators, such as consistency and asymptotic normality.

After inferring the optimal kernel hyperparameter values and value for $\nu$, predictions can be made using the conditional distribution result of equation (4.2). Indeed, suppose that we have the test observation $\mathbf{x}_*$. Then, the posterior predictive probability attributed to some function value $f_*$ is given by:

$$F_*|\mathbf{X}_*, \mathbb{X}, \mathbf{Y} \sim t\left[\mathbf{k}_*^T\mathbb{\Sigma}^{-1}\mathbf{y},\ \frac{\nu+\mathbf{y}^T\mathbb{\Sigma}^{-1}\mathbf{y}}{\nu+n}\left(k_{**}-\mathbf{k}_*^T\mathbb{\Sigma}^{-1}\mathbf{k}_*\right),\ \nu+n\right] \tag{4.18}$$

where $\mathbf{k}_* = (k(\mathbf{x}_*, \mathbf{x}_1), k(\mathbf{x}_*, \mathbf{x}_2), ..., k(\mathbf{x}_*, \mathbf{x}_n))^T$ and $k_{**} = k(\mathbf{x}_*, \mathbf{x}_*)$, as defined in equations (2.9) and (2.12).

Comparing the result of equation (4.18) with that of GPR in equation (2.15), we get that the predictive means of TPRD and GPR have the same form. Therefore, given the same kernel function and hyperparameter values, the two predictive means would be equal. Additionally, the maximum likelihood estimates for the predictive mean in both cases inherit all the properties of maximum likelihood estimators, such as consistency and asymptotic normality. However, the key difference between the two approaches is that the predictive variance for TPRD is explicitly dependent on $\mathbf{y}$, whereas for GPR this quantity is independent of $\mathbf{y}$. Since GPR, TPRK, and TPRD all make use of

kernels, their performances naturally depend on the nature of the kernel used. However, if the kernel used has the $\beta$-property as specified in Definition 4.3.1, the predictive means resulting from maximum likelihood estimation for the three approaches are equal. Moreover, empirical results by Tang et al. (2017) on a variety of datasets show that TPRD is more robust than GPR in practice, even if the kernel used does not have the $\beta$-property.

Broadly speaking, a kernel is said to satisfy the $\beta$-property if one of the hyperparameters can be written down as a free factor:

**Definition 4.3.1** ($\beta$-property)**.** A kernel function $k(\mathbf{x}, \mathbf{x}'; \boldsymbol{\theta})$ with hyperparameter vector $\boldsymbol{\theta} := (\theta_1, \theta_2, ..., \theta_a)^T$ is said to have the $\beta$-property if it can be written down in the form

$$k(\mathbf{x}, \mathbf{x}'; \theta_1, \theta_2, ..., \theta_a) = g(\theta_1)k(\mathbf{x}, \mathbf{x}'; 1, \theta_2', ..., \theta_a'),$$

where each $\theta_i'$ for $i \in \{2, 3, ..., a\}$ corresponds one-to-one with some $\theta_i$ given some $\theta_1$, and $g(\theta_1)$ is a one-to-one function of $\theta_1$ with range $(0, \infty)$.

As examples of kernels that satisfy this definition, we have the square exponential, rational quadratic and Cauchy kernels described in Section 2.2. This is because the hyperparameter $\sigma_f$ can be 'extracted' as a free factor from the equation of these kernels.

We now show that, given that the kernel used satisfies this property, then maximum likelihood estimation of the kernel hyperparameters as well as the quantity $\sigma := \frac{\nu+n}{\nu}\frac{1}{\beta}$ are independent of $\nu$.

**Theorem 4.3.2.** *If the kernel $k$ used in TPRD has the $\beta$-property, then the maximum marginal likelihood estimate of the hyperparameters $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_a)^T$ of the kernel as well as that of $\sigma := \frac{\nu+n}{\nu}\frac{1}{\beta}$ are independent of $\nu$.*

*Proof.* From equation (4.14), the marginal likelihood of TPRD is:

$$\mathbb{P}(\mathbf{y}|\boldsymbol{\theta}, \mathbb{K}) \propto |\mathbb{\Sigma}|^{-1/2}\left[1 + \frac{1}{\nu}\mathbf{y}^T\mathbb{\Sigma}^{-1}\mathbf{y}\right]^{-\frac{\nu+n}{2}}, \tag{4.19}$$

where $\mathbb{\Sigma} := \mathbb{K} + \sigma\mathbb{I}$. Assuming that $k$ has the $\beta$-property, then $\mathbb{\Sigma}$ may be re-written as:

$$\mathbb{\Sigma} = \mathbb{K} + \sigma^2\mathbb{I} = g(\theta_1)\mathbb{K}' + \sigma^2\mathbb{I} = \sigma^2\mathbb{\Sigma}',$$

where $\mathbb{K}$ is the matrix with $(i, j)^{\text{th}}$ entry equal to $k(\mathbf{x}_i, \mathbf{x}_j)$ with respect to the hyperparameters $\theta_1, \theta_2, ..., \theta_a$ and similarly $\mathbb{K}'$ is a rewriting of this matrix with respect to the hyperparameters $1, \theta_2', ..., \theta_a'$. Additionally, $\mathbb{\Sigma}' := \lambda\mathbb{K}' + \mathbb{I}$ with $\lambda = \frac{g(\theta_1)}{\sigma^2}$.

Since $g(\theta_1)$ is an injective function of $\theta_1$, then the parameters $\sigma, \lambda, \theta'_2, ..., \theta'_a$ can be mapped on a one-to-one basis to $\boldsymbol{\theta}$. Therefore, maximising the likelihood $\mathbb{P}(\mathbf{y}|\boldsymbol{\theta}, \mathbb{K})$ is the same as maximising $\mathbb{P}(\mathbf{y}|\sigma, \lambda, \theta'_2, ..., \theta'_a)$, or equivalently minimising

$$h(\mathbf{y}|\sigma, \lambda, \theta'_2, ..., \theta'_a) := \frac{1}{\alpha}\Big[\mathbb{P}(\mathbf{y}|\sigma, \lambda, \theta'_2, ..., \theta'_a)\Big]^{-\frac{2}{\nu+n}}, \tag{4.20}$$

where the normalisation constant $\alpha = \frac{\Gamma(\frac{\nu+n}{2})}{(\nu\pi)^{\frac{n}{2}}\Gamma(\frac{\nu}{2})}$ is irrelevant for maximisation and/or minimisation purposes.

We can rewrite the quantity of equation (4.20) as

$$\begin{aligned} h(\mathbf{y}|\sigma, \lambda, \theta'_2, ..., \theta'_a) &= \frac{1}{\alpha}\Big[\mathbb{P}(\mathbf{y}|\sigma, \lambda, \theta'_2, ..., \theta'_a)\Big]^{-\frac{2}{\nu+n}} \\ &= \left[\frac{1}{|\mathbb{\Sigma}|^{\frac{1}{2}}}\left(1 + \frac{1}{\nu}\mathbf{y}^T\mathbb{\Sigma}\mathbf{y}\right)^{-\frac{\nu+n}{2}}\right]^{-\frac{2}{\nu+n}} \\ &= |\mathbb{\Sigma}|^{\frac{1}{\nu+n}}\left[1 + \frac{1}{\nu}\mathbf{y}^T\mathbb{\Sigma}^{-1}\mathbf{y}\right] \\ &= \sigma^{\frac{2n}{\nu+n}}|\mathbb{\Sigma}'|^{\frac{1}{\nu+n}}\left[1 + \frac{1}{\nu\sigma^2}\mathbf{y}^T\mathbb{\Sigma}^{-1}\mathbf{y}\right], \end{aligned} \tag{4.21}$$

To obtain the maximum likelihood estimate of $\sigma$, we can differentiate equation (4.21) with respect to $\sigma$, giving:

$$\frac{\partial h(\mathbf{y}|\sigma, \lambda, \theta'_2, ..., \theta'_a)}{\partial\sigma} = \frac{2n}{\nu+n}\sigma^{\frac{2n}{\nu+n}-1}\ |\mathbb{\Sigma}'|^{\frac{1}{\nu+n}} - \frac{2\sigma^{\frac{2n}{\nu+n}-3}|\mathbb{\Sigma}'|^{\frac{1}{\nu+n}}}{\nu+n}\mathbf{y}^T\mathbb{\Sigma}'^{-1}\mathbf{y}. \tag{4.22}$$

Setting equation (4.22) to 0, we can derive the maximum likelihood estimate $\hat{\sigma}$ for $\sigma$ as:

$$\begin{aligned} & \frac{2\hat{\sigma}^{\frac{2n}{\nu+n}-3}|\mathbb{\Sigma}'|^{\frac{1}{\nu+n}}}{\nu+n}\mathbf{y}^T\mathbb{\Sigma}'^{-1}\mathbf{y} = \frac{2n}{\nu+n}\hat{\sigma}^{\frac{2n}{\nu+n}-1}\ |\mathbb{\Sigma}'|^{\frac{1}{\nu+n}} \\ \implies \quad & 2\hat{\sigma}^{\frac{2n}{\nu+n}-3} = 2n\hat{\sigma}^{\frac{2n}{\nu+n}-1} \\ \implies \quad & \hat{\sigma}^{\frac{2n}{\nu+n}-3}\left[\mathbf{y}^T\mathbb{\Sigma}'^{-1}\mathbf{y} - n\hat{\sigma}^2\right] = 0 \\ \implies \quad & \mathbf{y}^T\mathbb{\Sigma}'^{-1}\mathbf{y} = n\hat{\sigma}^2 \end{aligned}$$

and hence $\hat{\sigma} = \left[\frac{\mathbf{y}^T\mathbb{\Sigma}^{-1}\mathbf{y}}{n}\right]^{\frac{1}{2}}$ is independent of $\nu$. Additionally, from this definition of $\sigma$

and equation (4.21), we get that:

$$\begin{aligned} h(\mathbf{y}|\hat{\sigma},\lambda,\theta_2',...,\theta_a') &= \hat{\sigma}^{\frac{2n}{\nu+n}}|\mathbb{\Sigma}'|^{\frac{1}{\nu+n}}\left[1+\frac{\mathbf{y}^T\mathbb{\Sigma}^{-1}\mathbf{y}}{\nu\hat{\sigma}^2}\right] \\ &= \left[\frac{\mathbf{y}^T\mathbb{\Sigma}^{-1}\mathbf{y}}{n}\right]^{\frac{n}{\nu+n}}|\mathbb{\Sigma}'|^{\frac{1}{\nu+n}}\left[1+\frac{\mathbf{y}^T\mathbb{\Sigma}^{-1}\mathbf{y}}{\left(\frac{\nu(\mathbf{y}^T\mathbb{\Sigma}^{-1}\mathbf{y})}{n}\right)}\right] \\ &= \left[\frac{\mathbf{y}^T\mathbb{\Sigma}^{-1}\mathbf{y}}{n}\right]^{\frac{n}{\nu+n}}|\mathbb{\Sigma}'|^{\frac{1}{\nu+n}}\left[1+\frac{n}{\nu}\right]. \end{aligned} \tag{4.23}$$

From equation (4.23), minimising $h(\mathbf{y}|\hat{\sigma},\lambda,\theta_2',...,\theta_a')$ with respect to any of the remaining hyperparameters $\lambda,\theta_2',...,\theta_a'$ is equivalent to minimising

$$\left[\frac{\mathbf{y}^T\mathbb{\Sigma}^{-1}\mathbf{y}}{n}\right]^{\frac{n}{\nu+n}}|\mathbb{\Sigma}'|^{\frac{1}{\nu+n}}.$$

Therefore, if we define $h'(\mathbf{y}|\hat{\sigma},\lambda,\theta_2',...,\theta_a')$ by:

$$h'(\mathbf{y}|\hat{\sigma},\lambda,\theta_2',...,\theta_a') := \left[\frac{\mathbf{y}^T\mathbb{\Sigma}^{-1}\mathbf{y}}{n}\right]^{n}|\mathbb{\Sigma}'|,$$

then minimising $h(\mathbf{y}|\hat{\sigma},\lambda,\theta_2',...,\theta_a')$ with respect to any one of the hyperparameters $\lambda,\theta_2',...,\theta_a'$ is equivalent to minimising $h'(\mathbf{y}|\hat{\sigma},\lambda,\theta_2',...,\theta_a')$ with respect to that parameter. Moreover, since $h'$ is independent of $\nu$, then the solution of

$$\frac{\partial h'(\mathbf{y}|\hat{\sigma},\lambda,\theta_2',...,\theta_a')}{\partial\phi}=0$$

is independent of $\nu$ for any $\phi\in\{\lambda,\theta_2',...,\theta_a'\}$.

Moreover, since by the $\beta$-property there is a one-to-one mapping between the hyperparameters $\sigma,\theta_1,\theta_2,...,\theta_a$ and $\sigma,\lambda,\theta_2',...,\theta_a'$, then the maximum likelihood estimates of $\sigma,\theta_1,\theta_2,...,\theta_a$ are also independent of $\nu$. □

As a corollary of Theorems A.1 and 4.3.2, we get the following:

**Corollary 4.3.3.** *If the kernel $k$ used in TPRD has the $\beta$-property, then the predictive mean of TPRD is the same as that of GPR.*

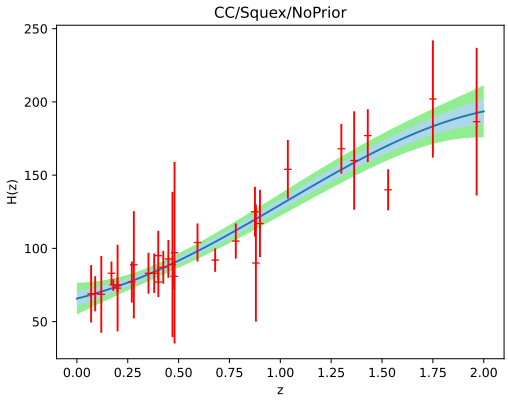
*Proof.* From Theorem A.1, the predictive mean of TPRD does not explicitly include $\nu$, but it is dependent on the hyperparameters $\boldsymbol{\theta}$ and $\sigma$, which may in general depend on $\nu$. However, if the kernel used has the $\beta$-property, then by Theorem 4.3.2 we get that these hyperparameters are independent of $\nu$, which implies that the predictive mean of TPRD is indeed independent of $\nu$. Since TPRD is in fact a special case of GPR with $\nu\to\infty$, then the predictive means of TPRD and GPR are equivalent. □

An intuitive justification for Corollary 4.3.3 is that the free factor $g(\theta_1)$ of the kernel neutralises the difference between the two marginal log-likelihoods of equations (2.14) and (4.15). While the two predictive means are equal for $\beta$-property kernels, Tang et al. (2017) still argue in favour of TPRD over GPR for multiple reasons, mainly due to the increased flexibility and robustness of the former when compared to the latter.

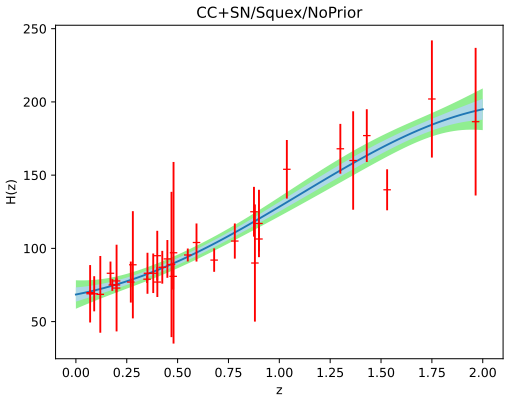
## 4.4 Application to $H_0$ Data

In this section, we apply the TPRD model to the Hubble data as described in Section 1.4. The procedure taken for such a task is equivalent to that taken for the GPR application in Section 2.4. In particular, we consider the datasets CC, CC+SN, and CC+SN+BAO and each of the kernel functions defined in Section 2.2. Moreover, for each combination of dataset and kernel, we consider each of the 'prior' values in Table 2.1 separately and consider also the priorless case. For the regression, we use the class `StudentTProcessRegressionModel` within the Python package `TensorFlow` (Abadi et al., 2015). The kernels described in Section 2.2 are all available within this package, except for the Cauchy kernel, which had to be defined manually.

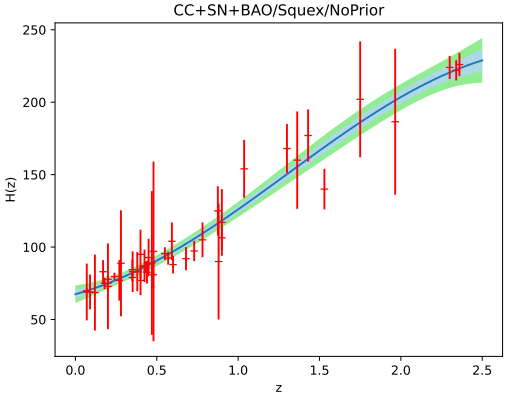
As in Section 2.4, we maximise the same objective function of equation (2.17). As was the case in GPR, such an objective function reaches a minimum value of 0 when the reconstructed and observed values $\hat{H}_{pred}(z_i)$ and $H_{obs}(z_i)$ are equal at each redshift value $z_i$, with the idea behind using such a function being to incentivise the reconstructed and observed values to be close to each other. Then, to obtain the estimates of $H_0$, we extrapolate the reconstructed function to $z = 0$. Moreover, we calculate the distance between each estimated value of $H_0$ and each of the 'prior' $H_0$ values of Table 2.1 as in equation (2.18). Significant distances of more than 2 units are shown in bold within the results tables. For the square exponential kernel function, reconstruction plots of $H(z)$ against $z$ are shown in Figures 4.2 through 4.8. Similarly to the corresponding plots in Sections 2.4 and 3.4, the observed data points and their errors are shown in each plot by red vertical lines. The reconstructed function is shown as a dark blue line and is surrounded by the $1\sigma$ and $2\sigma$ confidence regions, shaded light blue and light green respectively. Similar plots for the other kernel functions are given in Appendix G.




(a) CC

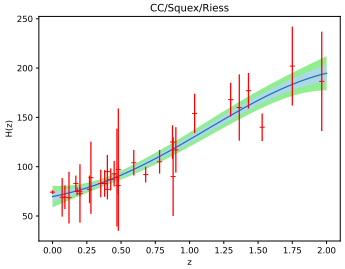



(b) CC+SN

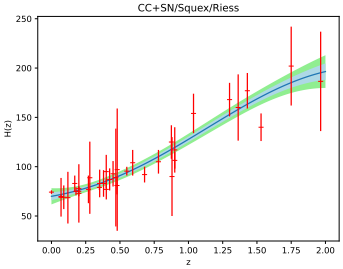



(c) CC+SN+BAO

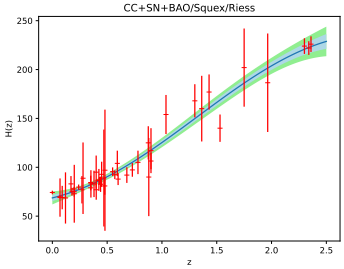
Figure 4.2: TPR reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and without adding any $H_0$ estimate to the data




(a) CC

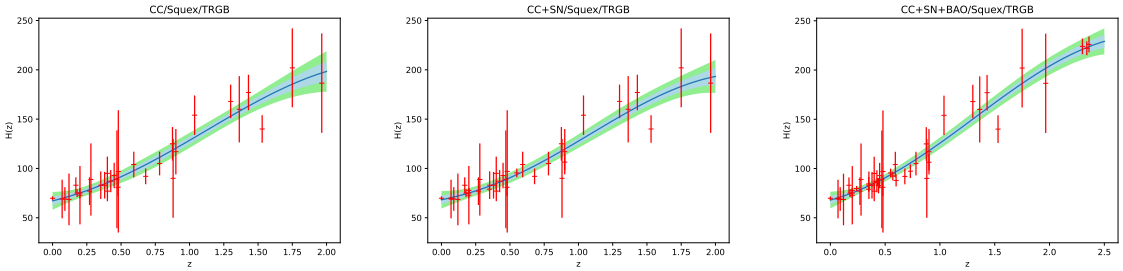



(b) CC+SN



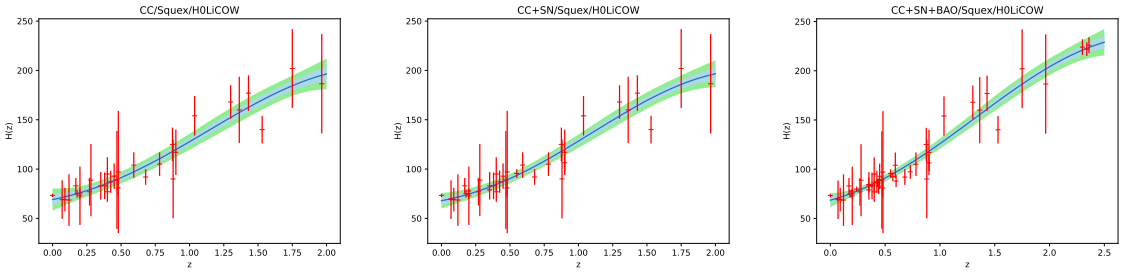

(c) CC+SN+BAO

Figure 4.3: TPR reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^R$



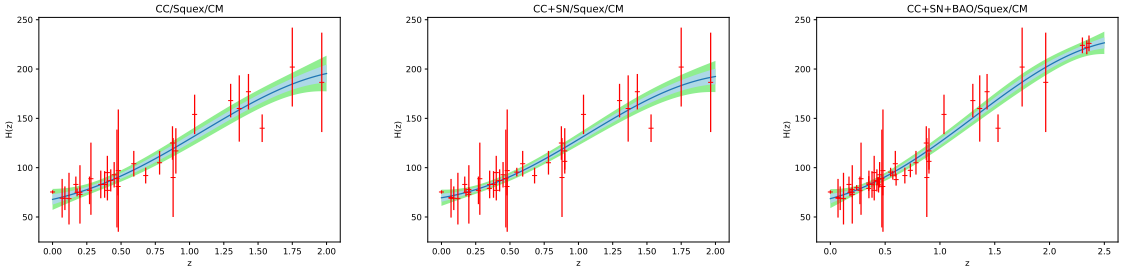

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure 4.4: TPR reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^{TRGB}$



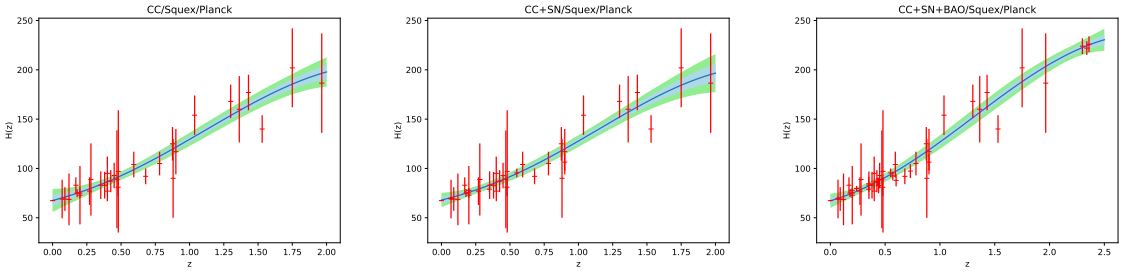

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure 4.5: TPR reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^{HW}$



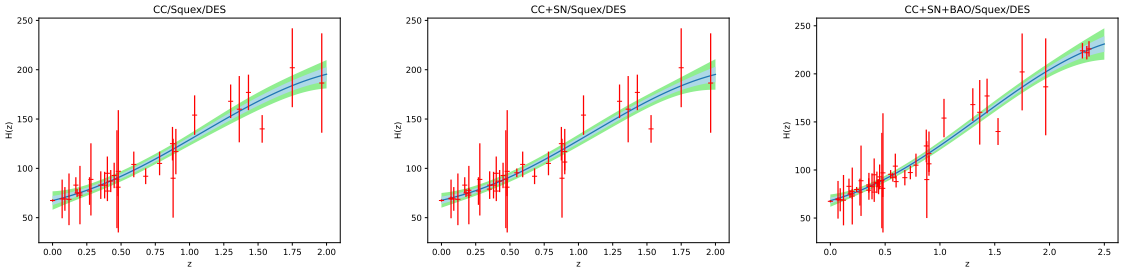

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure 4.6: TPR reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^{CM}$




(a) CC (b) CC+SN (c) CC+SN+BAO

Figure 4.7: TPR reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^P$




(a) CC (b) CC+SN (c) CC+SN+BAO

Figure 4.8: TPR reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^{DES}$

Table 4.1 shows the results for all datasets and priors, and for the square exponential kernel function. As was the case with GPR, the 'prior' value supplied and the dataset both affect the estimated $H_0$ values, but to a lesser extent than in GPR. This is a significant finding – as expected, the $H_0$ estimates obtained using TPR are more robust to changes in the prior specification than those obtained using GPR. In other words, whereas GPR has a tendency of overfitting on the data, TPR does not exhibit this behaviour. Similar results for the other kernel functions are given in Appendix H.

When compared to GPR, the estimates obtained using TPR are generally on the lower side. In fact, from Table 4.1, one can see that all the $H_0$ estimates obtained using the square exponential kernel are below 70km $s^{-1}$ $Mpc^{-1}$. Therefore, TPR seems to provide evidence in favour of lower values for the Hubble constant. This is a crucial finding, as it suggests that the value of the Hubble constant is more in line with the value obtained by Aghanim et al. (2020b) than that of Riess et al. (2019). Moreover, comparing with

Table 2.2, the estimates for the priorless case as well as when using the Planck and DES priors are highly similar for both GPR and TPR.

Another observation that can be made is that the $H_0$ estimates obtained using TPR have larger standard errors than those for GPR. This is because TPR in general produces more conservative estimates, especially when the degrees of freedom are low. In fact, the optimal value for the degrees of freedom was between 3 and 4 in most cases. Nevertheless, it seems that the confidence regions obtained using TPR are narrower than for GPR at higher redshifts, for example when comparing Figures 2.1 and 4.2. However, in estimating the Hubble constant, we are interested in the prediction error at $z = 0$, and the larger errors obtained for these estimates is a principal drawback of TPR.

Moreover, as opposed to GPR, in this case the choice of kernel may affect the quality of the estimates obtained. In particular, the standard errors of the estimates obtained using the rational quadratic kernel are significantly larger than those for the other kernels, as seen in Table H.2. While the larger prediction errors may compromise the meaningfulness of the estimates obtained, this is not necessarily a drawback of TPs. In fact, it may also be seen as an advantage of TPR when compared to GPR, as TPs do not have the tendency to overfit on the data. In an ideal situation, given the availability of larger datasets, TPR would be able to obtain accurate estimates for $H_0$ while avoiding the overfitting problems of GPR. Future work with regards to estimation of the Hubble constant may include the application of TPR to larger datasets.

| **Dataset** | $\hat{H}_0$ (km s$^{-1}$ Mpc$^{-1}$) | $d(\hat{H}_0, \hat{H}_0^R)$ | $d(\hat{H}_0, \hat{H}_0^{TRGB})$ | $d(\hat{H}_0, \hat{H}_0^{HW})$ | $d(\hat{H}_0, \hat{H}_0^{CM})$ | $d(\hat{H}_0, \hat{H}_0^P)$ | $d(\hat{H}_0, \hat{H}_0^{DES})$ |
|---|---|---|---|---|---|---|---|
| CC | $65.710 \pm 5.426$ | $-1.4869$ | $-0.7114$ | $-1.3348$ | $-1.6971$ | $-0.3102$ | $-0.3053$ |
| CC+SN | $68.492 \pm 4.858$ | $-1.1041$ | $-0.2508$ | $-0.9341$ | $-1.3341$ | 0.2236 | 0.2192 |
| CC+SN+BAO | $67.444 \pm 3.033$ | $-1.9159$ | $-0.6584$ | $-1.6845$ | $\mathbf{-2.2805}$ | 0.0142 | 0.0135 |
| CC+$\hat{H}_0^R$ | $69.701 \pm 5.485$ | $-0.7819$ | $-0.0170$ | $-0.6267$ | $-0.9847$ | 0.4179 | 0.4114 |
| CC+SN+$\hat{H}_0^R$ | $69.972 \pm 4.104$ | $-0.9463$ | 0.0380 | $-0.7492$ | $-1.2128$ | 0.6222 | 0.6054 |
| CC+SN+BAO+$\hat{H}_0^R$ | $68.657 \pm 3.330$ | $-1.4658$ | $-0.2980$ | $-1.2417$ | $-1.7944$ | 0.3734 | 0.3585 |
| CC+$\hat{H}_0^{TRGB}$ | $67.153 \pm 4.525$ | $-1.4491$ | $-0.5395$ | $-1.2718$ | $-1.6984$ | $-0.0544$ | $-0.0531$ |
| CC+SN+$\hat{H}_0^{TRGB}$ | $68.243 \pm 4.449$ | $-1.2434$ | $-0.3218$ | $-1.0618$ | $-1.4944$ | 0.1883 | 0.1839 |
| CC+SN+BAO+$\hat{H}_0^{TRGB}$ | $68.003 \pm 4.231$ | $-1.3500$ | $-0.3875$ | $-1.1618$ | $-1.6141$ | 0.1415 | 0.1379 |
| CC+$\hat{H}_0^{HW}$ | $69.053 \pm 5.656$ | $-0.8696$ | $-0.1252$ | $-0.7191$ | $-1.0672$ | 0.2911 | 0.2869 |
| CC+SN+$\hat{H}_0^{HW}$ | $67.896 \pm 3.943$ | $-1.4561$ | $-0.4350$ | $-1.2584$ | $-1.7390$ | 0.1248 | 0.1211 |
| CC+SN+BAO+$\hat{H}_0^{HW}$ | $68.523 \pm 3.526$ | $-1.4357$ | $-0.3189$ | $-1.2203$ | $-1.7479$ | 0.3152 | 0.3039 |
| CC+$\hat{H}_0^{CM}$ | $67.835 \pm 5.323$ | $-1.1349$ | $-0.3476$ | $-0.9779$ | $-1.3462$ | 0.0814 | 0.0800 |
| CC+SN+$\hat{H}_0^{CM}$ | $69.491 \pm 4.123$ | $-1.0493$ | $-0.0680$ | $-0.8541$ | $-1.3160$ | 0.5036 | 0.4901 |
| CC+SN+BAO+$\hat{H}_0^{CM}$ | $68.595 \pm 4.820$ | $-1.0918$ | $-0.2326$ | $-0.9206$ | $-1.3234$ | 0.2467 | 0.2418 |
| CC+$\hat{H}_0^P$ | $67.519 \pm 5.829$ | $-1.0974$ | $-0.3721$ | $-0.9521$ | $-1.2909$ | 0.0203 | 0.0200 |
| CC+SN+$\hat{H}_0^P$ | $67.956 \pm 3.659$ | $-1.5327$ | $-0.4473$ | $-1.3245$ | $-1.8364$ | 0.1504 | 0.1454 |
| CC+SN+BAO+$\hat{H}_0^P$ | $67.182 \pm 3.600$ | $-1.7446$ | $-0.6431$ | $-1.5366$ | $\mathbf{-2.0559}$ | $-0.0600$ | $-0.0579$ |
| CC+$\hat{H}_0^{DES}$ | $67.371 \pm 4.672$ | $-1.3660$ | $-0.4816$ | $-1.1926$ | $-1.6071$ | $-0.0062$ | $-0.0061$ |
| CC+SN+$\hat{H}_0^{DES}$ | $67.594 \pm 3.721$ | $-1.5994$ | $-0.5278$ | $-1.3945$ | $-1.8995$ | 0.0518 | 0.0501 |
| CC+SN+BAO+$\hat{H}_0^{DES}$ | $68.098 \pm 3.204$ | $-1.6615$ | $-0.4569$ | $-1.4343$ | $\mathbf{-2.0047}$ | 0.2154 | 0.2062 |

Table 4.1: TPR results with square exponential kernel function

The statistical methods discussed so far can be roughly divided into two groups, namely parametric and non-parametric methods. The former includes the MCMC approach discussed in Section 3.3, where we considered the ensemble sampler to obtain parameter estimates for the parameters of the $\Lambda CDM$ cosmological model, which is frequently referred to as the 'standard' or 'concordance' model of Big Bang cosmology (see e.g. Aghanim et al., 2020a). While this model seems to adequately explain the large-scale structure of the universe, such as the presence of CMB radiation, a myriad of competing cosmological models exist (Akarsu and Dereli, 2012; Maeder, 2017; Pandey et al., 2020, see e.g.).

On the other hand, the non-parametric methods are arguably more interesting as they are independent of the specification of any cosmological model, and any Hubble constant estimates obtained from such methods are based on the observed data alone. The non-parametric methods discussed include GPR, TPR, and a non-parametric MCMC approach where the ensemble sampler was used to obtain values for the kernel hyper-parameters within a GPR context. These three methods were discussed in Sections 2.4, 4.4, and 3.4 respectively.

In Chapter 5, we consider extensions to both the parametric and non-parametric approaches undertaken so far for the purposes of estimating the Hubble constant. In particular, we extend the non-parametric GPR and TPR approaches of Chapters 2 and 4 respectively to similar approaches which allow for heteroscedasticity (Binois and Gramacy, 2021; Binois et al., 2018). We also consider an extension to the parametric MCMC $\Lambda CDM$ model discussed in Section 3.3; in this case, we replace the compressed SN dataset used so far with the full Panthon dataset (Scolnic et al., 2018), which contains 1048 data points. It should be noted that a main limitation of this dataset is that it can only be applied given some cosmological model, and that therefore one cannot use it for the model-independent methods. Then, in the concluding Chapter 6, we compare and contrast the different estimates obtained for $H_0$.

# Chapter 5

# Further Extensions

In this chapter, we discuss some extensions and amendments to the models described in the previous chapters. In particular, in Section 5.1, we carry out heteroscedastic GPR and TPR. These methods are similar to the homoscedastic cases described in Chapters 2 and 4, except that the error terms are allowed to differ for each vector of observations instead of assuming independent and identically-distributed error terms. This allows for greater flexibility in modelling when compared to the homoscedastic case. Then, in Section 5.2, we will re-run the MCMC $\Lambda CDM$ analysis of Section 3.3 with the full Pantheon dataset (Scolnic et al., 2018) replacing the 'compressed' Pantheon dataset used so far. The motivation behind this change is two-fold. First, it allows us to compare the results obtained using the compressed SN dataset to the full Pantheon dataset. Moreover, more accurate estimates may be achieved when considering the additional observations available. The heteroscedastic and amended $\Lambda CDM$ approaches will be compared and contrasted with the methods discussed within Chapters 2, 3, and 4.

## 5.1 Heteroscedastic Gaussian Process Regression and Student's $t$-Process Regression

In this section, we shall explore the use of heteroscedastic GPR and TPR as described by Binois and Gramacy (2021), who provide an implementation of this within the `R` package `hetGP`. The idea behind heteroscedastic GP and TP modelling is to simultaneously model the mean and variance, while allowing the error terms $\epsilon_i$ to have a different

variance for each element. In other words, given a vector of observations $(\mathbf{x}_i, y_i)$ for $i \in \{1, 2, ..., n\}$ and $f \sim GP(m, k)$ or $f \sim TP(m, k, \nu)$, the basic model $y_i = f(\mathbf{x}_i) + \epsilon_i$ shown in equation (2.3) is assigned error terms $\epsilon_i \sim N(0, \sigma_i^2)$ for each $i$ instead of assuming that the error terms are independent and identically-distributed for all $i$.

This `R` package also allows for some optimisations based on the 'stochastic kriging' (SK) predictor (Ankenman et al., 2010), which lead to large computational savings especially when a large number of observations are available at each $\mathbf{x}_i$. This assumption of repeated readings is often prohibitive when modelling real-world data. In fact, for the data described in Section 1.4, no repeated readings are present. While replication can still help speed up the computation, use of the SK predictor does not inherently require a set minimum dataset size or numbers of replicates.

The log-likelihood for the SK predictor is defined by:

$$\log \Lambda = c - \frac{1}{2} \sum_{i=1}^{n} \left[ (a_i - 1) \log \lambda_i + \log a_i \right] - \frac{1}{2} \log |\mathbb{K}_n| \tag{5.1}$$

(Binois et al., 2018, p. 10).

The SK predictor already incorporates some in-sample heteroscedasticity by independently calculating the moments for each element; therefore, a different variance value will be assigned to each of the sampled elements. However, this is not extended out-of-sample and is therefore not useful for heteroscedastic interpolation. Ankenman et al. (2010) suggest to incorporate out-of-sample heteroscedasticity fitting another GP or TP on the variances of the sample points to obtain a smoothed variance for use in interpolation. However, Binois and Gramacy (2021) go a step further and propose introducing latent variables under a GP or TP prior and performing MCMC inference on their joint distribution in order to obtain this smoothed variance. Let $\mathbb{\Delta}_r = \text{Diag}(\delta_1, \delta_2, ..., \delta_r)$ be made up of the latent variance variables corresponding to the $r \leq n$ unique observation points. We assign a GP or TP prior to this matrix, i.e.:

$$\mathbb{\Delta}_r \sim MVN_r(\beta_0, \mathbb{K}_{(g)}^{-1}),$$

or analogously:

$$\mathbb{\Delta}_r \sim MVT_r(\beta_0, \mathbb{K}_{(g)}^{-1}, \nu_{(g)}),$$

where $\mathbb{K}_{(g)} := \mathbb{C}_{(g)} + g\mathbb{A}_r^{-1}$. Here, the subscript $(g)$ is used to distinguish this process from the one on the latent function $f$. As for $\beta_0$, the natural estimator to use is $\hat{\beta}_0 = \mathbb{\Delta}_r^T \mathbb{K}_{(g)}^{-1} \mathbb{\Delta}_r (\mathbf{1}_r^T \mathbb{K}_{(g)}^{-1} \mathbf{1}_r)^{-1}$. As described by Binois et al. (2018), one can opt to use $\log(\mathbb{\Delta}_r)$ in place of $\mathbb{\Delta}_r$ to ensure that $\delta_i$ remains positive for each $i \in \{1, 2, \ldots, r\}$.

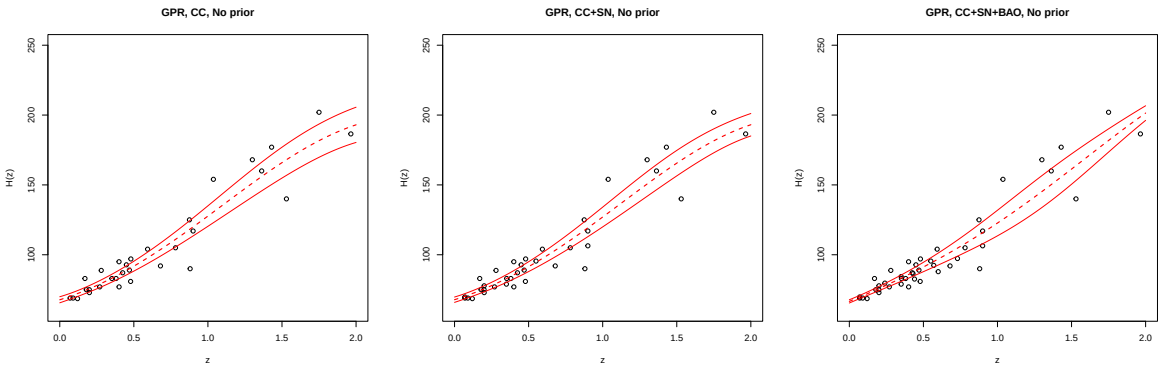
In heteroscedastic GPR, we define a joint log-likelihood function over both GPs, i.e.:

$$\log \mathbb{P}(\mathbf{y}|\mathbb{K}_r, \mathbb{K}_{(g)}) = c - \frac{1}{2}\sum_{i=1}^{r}[(a_i - 1)\log \lambda_i + \log a_i] - \frac{1}{2}\log|\mathbb{K}_r| - \frac{1}{2}\log|\mathbb{K}_{(g)}|, \quad (5.2)$$

and this likelihood may be optimised numerically with respect to any unknown parameter by differentiating it with respect to that parameter and setting the derivative to zero. For TPR, we can use a joint log-likelihood function over both TPs. Binois et al. (2018) state that, in the event that replication is present, the maximum likelihood estimates obtained using the $r$ unique points are equivalent to the full-$n$ estimates.

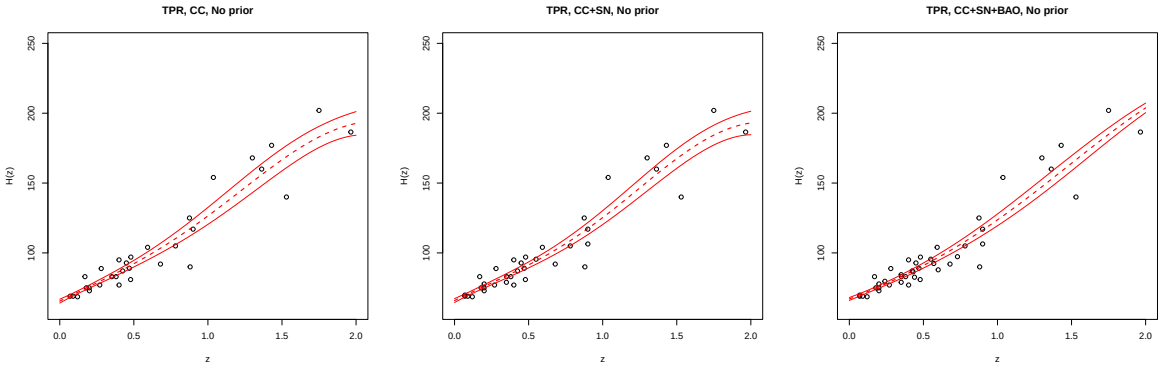
For the application of these techniques on the $H_0$ data of Tables 1.1 through 1.3, we use the `R` package `hetGP`, in particular the subroutines `mleHetGP` and `mleHetTP` for heteroscedastic GPR and TPR. In each case, we supply only $z$ and $H(z)$. As a consequence of this, the estimates obtained using the Planck and DES priors will be identical given some dataset and kernel, since these two priors have the same value of 67.4 km s$^{-1}$ Mpc$^{-1}$. We present only the results using the Gaussian kernel of equation (2.2) since the other kernels gave virtually identical results.

In Figures 5.1 and 5.2, we present the reconstruction of the $H(z)$-against-$z$ curve for the heteroscedastic GPR and TPR approaches respectively, for the case where none of the 'priors' were added to the data. Similar plots for the other priors are shown in Appendix I.




(a) CC (b) CC+SN (c) CC+SN+BAO

Figure 5.1: Heteroscedastic GPR reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and without adding any $H_0$ estimate to the data




(a) CC (b) CC+SN (c) CC+SN+BAO

Figure 5.2: Heteroscedastic TPR reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and without adding any $H_0$ estimate to the data

The plots indicate that lower prediction errors, i.e. more confident estimates for $H_0$, were achieved with the heteroscedastic approaches. The figures also show that a lower prediction error is achieved by adding the SN and BAO data to the analysis, as was seen in the previous chapters. Also as seen in previous chapters, the prediction uncertainties are much lower for low values of $z$ than for the higher values. This is mainly because more readings are available at low $z$ than at high $z$ in our data.

The estimated values obtained using heteroscedastic GPR and TPR, which are presented in Tables 5.1 and 5.2, may be compared to the GPR and and TPR results obtained in Tables 2.2 and 4.1. However, one of the immediate differences when compared to Table 2.2 is that heteroscedastic GPR is robust to the specification of the prior, whereas 'vanilla' GPR was not. Additionally, the estimates obtained for $H_0$ using heteroscedastic GPR and TPR are on the lower side irrespective of the prior used, with only one of the estimates exceeding 71 km $s^{-1}$ $Mpc^{-1}$.

The estimates obtained for $H_0$ are generally more in agreement with the Planck and DES prior values than the others considered. This is in agreement with 'vanilla' TPR as can be seen in Table 4.1, and this provides further evidence in favour of a lower value for $H_0$. When comparing to 'vanilla' TPR, the lower prediction errors obtained lead to larger tensions, particularly with the priors $\hat{H}_0^R$, $\hat{H}_0^{HW}$, and $\hat{H}_0^{CM}$. However, only one of the estimates obtained using heteroscedastic TPR has a greater than $2\sigma$ tension with $\hat{H}_0^P$, and none of the estimates in Tables 5.1 and 5.2 are more than $2\sigma$ apart from $\hat{H}_0^{DES}$. As with the other methods explored in this dissertation, the use of a larger dataset, i.e. CC+SN or CC+SN+BAO when compared to CC alone, also led to lower prediction errors throughout.

| **Dataset** | $\hat{H}_0$ (km s$^{-1}$ Mpc$^{-1}$) | $d(\hat{H}_0, \hat{H}_0^R)$ | $d(\hat{H}_0, \hat{H}_0^{TRGB})$ | $d(\hat{H}_0, \hat{H}_0^{HW})$ | $d(\hat{H}_0, \hat{H}_0^{CM})$ | $d(\hat{H}_0, \hat{H}_0^P)$ | $d(\hat{H}_0, \hat{H}_0^{DES})$ |
|---|---|---|---|---|---|---|---|
| CC | $67.768 \pm 2.917$ | 1.8766 | $-0.5836$ | $-1.6263$ | $\mathbf{-2.2525}$ | 0.1245 | 0.1179 |
| CC+SN | $67.827 \pm 2.429$ | $\mathbf{-2.1065}$ | $-0.6398$ | $-1.8283$ | $\mathbf{-2.5475}$ | 0.1723 | 0.1590 |
| CC+SN+BAO | $66.636 \pm 1.718$ | $\mathbf{-3.0301}$ | $-1.2352$ | $\mathbf{-2.7173}$ | $\mathbf{-2.6262}$ | $-0.4271$ | $-0.3697$ |
| CC+$\hat{H}_0^R$ | $70.351 \pm 2.374$ | $-1.2933$ | 0.1812 | $-0.9998$ | $-1.7187$ | 1.2161 | 1.1185 |
| CC+SN+$\hat{H}_0^R$ | $70.062 \pm 2.154$ | $-1.4745$ | 0.0913 | $-1.1667$ | $-1.9358$ | 1.2039 | 1.0903 |
| CC+SN+BAO+$\hat{H}_0^R$ | $70.284 \pm 1.471$ | $-1.6820$ | 0.2014 | $-1.3194$ | $\mathbf{-2.2688}$ | 1.8564 | 1.5446 |
| CC+$\hat{H}_0^{TRGB}$ | $68.558 \pm 2.202$ | $-1.9820$ | $-0.4272$ | $-1.6860$ | $\mathbf{-2.4522}$ | 0.5126 | 0.4659 |
| CC+SN+$\hat{H}_0^{TRGB}$ | $68.436 \pm 1.803$ | $\mathbf{-2.2574}$ | $-0.5206$ | $-1.9355$ | $\mathbf{-2.8052}$ | 0.5538 | 0.4846 |
| CC+SN+BAO+$\hat{H}_0^{TRGB}$ | $68.255 \pm 1.389$ | $\mathbf{-2.6054}$ | $-0.6564$ | $\mathbf{-2.2580}$ | $\mathbf{-3.2548}$ | 0.5793 | 0.4742 |
| CC+$\hat{H}_0^{HW}$ | $70.010 \pm 2.239$ | $-1.4592$ | 0.0714 | $-1.1579$ | $-1.9079$ | 1.1376 | 1.0368 |
| CC+SN+$\hat{H}_0^{HW}$ | $69.482 \pm 2.206$ | $-1.6569$ | $-0.1093$ | $-1.3560$ | $\mathbf{-2.1164}$ | 0.9205 | 0.8369 |
| CC+SN+BAO+$\hat{H}_0^{HW}$ | $69.734 \pm 1.451$ | $-1.9277$ | $-0.0277$ | $-1.5690$ | $\mathbf{-2.5304}$ | 1.5210 | 1.2607 |
| CC+$\hat{H}_0^{CM}$ | $70.699 \pm 2.391$ | $-1.1718$ | 0.2945 | $-0.8778$ | $-1.5917$ | 1.3509 | 1.2437 |
| CC+SN+$\hat{H}_0^{CM}$ | $70.451 \pm 2.117$ | $-1.3499$ | 0.2288 | $-1.0371$ | $-1.8125$ | 1.4022 | 1.2661 |
| CC+SN+BAO+$\hat{H}_0^{CM}$ | $70.744 \pm 1.611$ | $-1.4303$ | 0.3790 | $-1.0747$ | $-1.9791$ | 1.9829 | 1.6897 |
| CC+$\hat{H}_0^P$ | $67.471 \pm 2.227$ | $\mathbf{-2.3467}$ | $-0.7957$ | $\mathbf{-2.0582}$ | $\mathbf{-2.8245}$ | 0.0309 | 0.0282 |
| CC+SN+$\hat{H}_0^P$ | $67.570 \pm 1.885$ | $\mathbf{-2.5380}$ | $-0.8333$ | $\mathbf{-2.2279}$ | $\mathbf{-3.0813}$ | 0.0870 | 0.0769 |
| CC+SN+BAO+$\hat{H}_0^P$ | $67.008 \pm 1.238$ | $\mathbf{-3.2764}$ | $-1.2312$ | $\mathbf{-2.9352}$ | $\mathbf{-3.9973}$ | $-0.2937$ | $-0.2321$ |
| CC+$\hat{H}_0^{DES}$ | $67.471 \pm 2.227$ | $\mathbf{-2.3467}$ | $-0.7957$ | $\mathbf{-2.0582}$ | $\mathbf{-2.8245}$ | 0.0309 | 0.0282 |
| CC+SN+$\hat{H}_0^{DES}$ | $67.570 \pm 1.885$ | $\mathbf{-2.5380}$ | $-0.8333$ | $\mathbf{-2.2279}$ | $\mathbf{-3.0813}$ | 0.0870 | 0.0769 |
| CC+SN+BAO+$\hat{H}_0^{DES}$ | $67.008 \pm 1.238$ | $\mathbf{-3.2764}$ | $-1.2312$ | $\mathbf{-2.9352}$ | $\mathbf{-3.9973}$ | $-0.2937$ | $-0.2321$ |

Table 5.1: Heteroscedastic GPR results using square exponential kernel function

| **Dataset** | $\hat{H}_0$ (km s$^{-1}$ Mpc$^{-1}$) | $d(\hat{H}_0, \hat{H}_0^R)$ | $d(\hat{H}_0, \hat{H}_0^{TRGB})$ | $d(\hat{H}_0, \hat{H}_0^{HW})$ | $d(\hat{H}_0, \hat{H}_0^{CM})$ | $d(\hat{H}_0, \hat{H}_0^P)$ | $d(\hat{H}_0, \hat{H}_0^{DES})$ |
|---|---|---|---|---|---|---|---|
| CC | $65.482 \pm 2.575$ | $\mathbf{-2.7714}$ | $-1.3494$ | $\mathbf{-2.5114}$ | $\mathbf{-3.2099}$ | $-0.7313$ | $-0.6802$ |
| CC+SN | $65.888 \pm 2.477$ | $\mathbf{-2.7106}$ | $-1.2530$ | $\mathbf{-2.4438}$ | $\mathbf{-3.1613}$ | $-0.5982$ | $-0.5535$ |
| CC+SN+BAO | $67.027 \pm 1.861$ | $\mathbf{-2.7634}$ | $-1.0427$ | $\mathbf{-2.4557}$ | $\mathbf{-3.3198}$ | $-0.1936$ | $-0.1706$ |
| CC+$\hat{H}_0^R$ | $70.214 \pm 2.076$ | $-1.4512$ | $0.1471$ | $-1.1367$ | $-1.9233$ | $1.3180$ | $1.1858$ |
| CC+SN+$\hat{H}_0^R$ | $69.455 \pm 2.062$ | $-1.7322$ | $-0.1229$ | $-1.4214$ | $\mathbf{-2.2160}$ | $0.9686$ | $0.8705$ |
| CC+SN+BAO+$\hat{H}_0^R$ | $69.569 \pm 1.827$ | $-1.8036$ | $-0.0876$ | $-1.4748$ | $\mathbf{-2.3292}$ | $1.1452$ | $1.0048$ |
| CC+$\hat{H}_0^{TRGB}$ | $68.002 \pm 2.070$ | $\mathbf{-2.2557}$ | $-0.6397$ | $-1.9543$ | $\mathbf{-2.7559}$ | $0.2828$ | $0.2544$ |
| CC+SN+$\hat{H}_0^{TRGB}$ | $67.718 \pm 2.085$ | $\mathbf{-2.3494}$ | $-0.7381$ | $\mathbf{-2.0507}$ | $\mathbf{-2.8503}$ | $0.1482$ | $0.1335$ |
| CC+SN+BAO+$\hat{H}_0^{TRGB}$ | $68.079 \pm 1.460$ | $\mathbf{-2.6319}$ | $-0.7182$ | $\mathbf{-2.2908}$ | $\mathbf{-3.2667}$ | $0.4400$ | $0.3654$ |
| CC+$\hat{H}_0^{HW}$ | $69.729 \pm 2.044$ | $-1.6409$ | $-0.0254$ | $-1.3271$ | $\mathbf{-2.1245}$ | $1.1069$ | $0.9931$ |
| CC+SN+$\hat{H}_0^{HW}$ | $69.226 \pm 2.038$ | $-1.8274$ | $-0.2059$ | $-1.5164$ | $\mathbf{-2.3183}$ | $0.8702$ | $0.7804$ |
| CC+SN+BAO+$\hat{H}_0^{HW}$ | $69.293 \pm 1.714$ | $-1.9707$ | $-0.1982$ | $-1.6358$ | $\mathbf{-2.5236}$ | $1.0599$ | $0.9169$ |
| CC+$\hat{H}_0^{CM}$ | $71.226 \pm 1.663$ | $-1.2143$ | $0.5649$ | $-0.8589$ | $-1.7444$ | $\mathbf{2.2034}$ | $1.8924$ |
| CC+SN+$\hat{H}_0^{CM}$ | $69.935 \pm 2.167$ | $-1.5141$ | $0.0469$ | $-1.2080$ | $-1.9748$ | $1.1399$ | $1.0333$ |
| CC+SN+BAO+$\hat{H}_0^{CM}$ | $69.945 \pm 1.950$ | $-1.6028$ | $0.0532$ | $-1.2806$ | $\mathbf{-2.1001}$ | $1.2642$ | $1.1242$ |
| CC+$\hat{H}_0^P$ | $66.631 \pm 1.913$ | $\mathbf{-2.8738}$ | $-1.1751$ | $\mathbf{-2.5719}$ | $\mathbf{-3.4242}$ | $-0.3887$ | $-0.3443$ |
| CC+SN+$\hat{H}_0^P$ | $66.683 \pm 1.703$ | $\mathbf{-3.0235}$ | $-1.2214$ | $\mathbf{-2.7095}$ | $\mathbf{-3.6226}$ | $-0.4038$ | $-0.3488$ |
| CC+SN+BAO+$\hat{H}_0^P$ | $67.148 \pm 1.403$ | $\mathbf{-3.0779}$ | $-1.1231$ | $\mathbf{-2.7433}$ | $\mathbf{-3.7479}$ | $-0.1695$ | $-0.1391$ |
| CC+$\hat{H}_0^{DES}$ | $66.631 \pm 1.913$ | $\mathbf{-2.8738}$ | $-1.1751$ | $\mathbf{-2.5719}$ | $\mathbf{-3.4242}$ | $-0.3887$ | $-0.3443$ |
| CC+SN+$\hat{H}_0^{DES}$ | $66.683 \pm 1.703$ | $\mathbf{-3.0235}$ | $-1.2214$ | $\mathbf{-2.7095}$ | $\mathbf{-3.6226}$ | $-0.4038$ | $-0.3488$ |
| CC+SN+BAO+$\hat{H}_0^{DES}$ | $67.148 \pm 1.403$ | $\mathbf{-3.0779}$ | $-1.1231$ | $\mathbf{-2.7433}$ | $\mathbf{-3.7479}$ | $-0.1695$ | $-0.1391$ |

Table 5.2: Heteroscedastic TPR results using square exponential kernel function

## 5.2 Incorporating the Full Pantheon Dataset within the $\Lambda CDM$ model

In this section, we re-run the $\Lambda CDM$ analysis of Section 3.3 using the full Pantheon dataset (Scolnic et al., 2018) in place of the 'compressed' SN dataset used previously. Going forward, we shall denote the full Pantheon dataset as SN$*$. We recall that the compressed Pantheon dataset consists of six observations and we have removed one of these since it had a non-Gaussian error. On the other hand, SN$*$ consists of 1048 observations and therefore one would also have to consider the corresponding $1048 \times 1048$ covariance matrix $\mathbb{\Sigma}_{SN*}$. This process is consequently significantly more computationally-intensive when compared to the compressed dataset. As mentioned in Section 1.4, the data and underlying covariance matrix are available publicly (Scolnic et al., 2018).

The readings in the SN$*$ dataset are given as distance moduli $\mu$ as opposed to $H$. The distance modulus can be expressed as the difference of the apparent magnitude $m$ and absolute magnitude $M$, i.e. $\mu = m - M$. The former is a measure of the brightness of a celestial object as observed from Earth, while the latter is defined to be equivalent to the apparent magnitude that the object would have at a standard reference distance of exactly 10 parsecs (see e.g. Zeilik et al., 1992, p. 227).

In our analysis, we will be considering $M$ as a nuisance parameter, as is done by Briffa et al. (2022, p. 532). Given some redshift value $z_i$, hyperparameter values $\boldsymbol{\theta} = (H_0, \Omega_{M0}, M)^T$, and the observed distance modulus $\mu(z_i, \boldsymbol{\theta})$, we can calculate the luminosity distance $d_L(z_i, \boldsymbol{\theta})$ using:

$$\mu(z_i, \boldsymbol{\theta}) = 5 \log_{10} \left[ d_L(z_i, \boldsymbol{\theta}) \right] + M, \tag{5.3}$$

and we can link this to the Hubble parameter through the relation:

$$d_L(z_i, \boldsymbol{\theta}) = c \, (1 + z_i) \int_0^{z_i} \frac{1}{H(z', \boldsymbol{\theta})} \, \mathrm{d}z', \tag{5.4}$$

where this integral has to be calculated numerically. For this, we make use of the function `integrate` within the Python package `scipy` (Virtanen et al., 2020).

Similarly to Section 3.3, we obtain an estimate for the Hubble constant $H_0$ by optimising a likelihood function. We consider a different likelihood function for each of the datasets. In particular, for the CC data we shall use equation (3.12) as our likelihood, which has

the effect of minimising the 'distance' between the observed and predicted values of $\hat{H}(z_i)$ at each redshift value $z_i$ observed within the CC dataset. We also make use of an objective function of the same form for the BAO data.

For the SN$*$ data, we use a similar form of the likelihood function that however has to incorporate also the covariance matrix $\mathbb{\Sigma}_{SN*}$. Defining $k(\boldsymbol{\theta})$ by

$$k(\boldsymbol{\theta}) = [\hat{\mu}(z_i, \boldsymbol{\theta}) - \mu_{obs}(z_i)]^T \, \mathbb{\Sigma}_{SN*}^{-1},$$

then the log-likelihood function corresponding to the SN$*$ data is proportional to

$$\big[\hat{\mu}(z_i, \boldsymbol{\theta}) - \mu_{obs}(z_i)\big]^T \mathbb{\Sigma}_{SN*}^{-1} \big[\hat{\mu}(z_i, \boldsymbol{\theta}) - \mu_{obs}(z_i)\big] + \ln\left[\frac{S}{2\pi}\right] - \frac{k^2(\boldsymbol{\theta})}{S}, \tag{5.5}$$

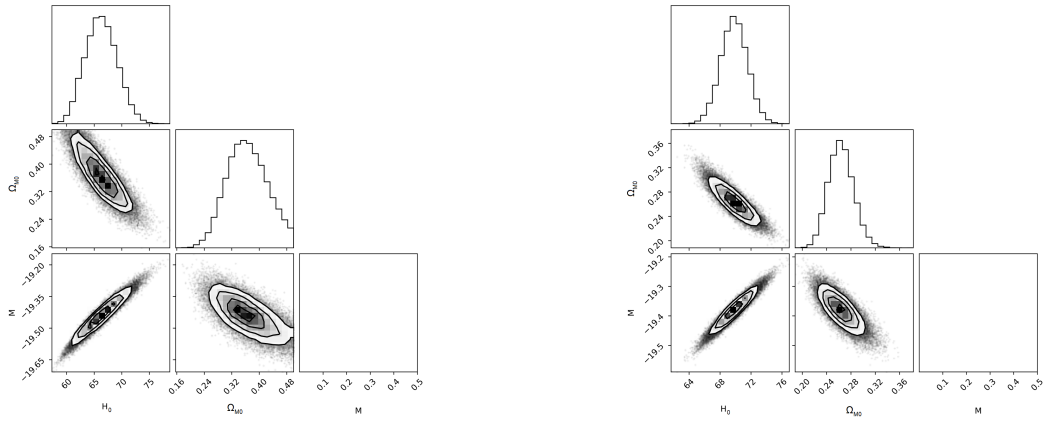
where $S$ is the sum of all the components of $\mathbb{\Sigma}_{SN*}^{-1}$ and the 'extra' terms $\ln\left[\frac{S}{2\pi}\right] - \frac{k^2(\boldsymbol{\theta})}{S}$ are a correction for host-galaxy properties (see e.g. Conley et al., 2010, pp. 9–12). To avoid any possible numerical issues with inversion of $\mathbb{\Sigma}_{SN*}$, we first carry out a Cholesky decomposition of this matrix.

As in Section 3.3, we use the affine ensemble sampler as implemented in `emcee` (Foreman-Mackey et al., 2013), and employ 200 walkers for 5000 iterations. For this, a specialised computer at the Department of Statistics & Operations Research at the University of Malta was used. This computer has an $11^{\text{th}}$ generation Intel i9 processor with 8 cores running at 5.3GHz, 64GB of 2666MHz DDR4 memory, and a 24GB NVIDIA GeForce RTX 3090 graphics card. We consider the datasets CC+SN$*$ and CC+SN$*$+BAO. Moreover, for both of these cases, we shall consider each of the priors mentioned in Table 2.1 one-by-one, and also the priorless case. Going through all these combinations took about 2 days of computational time, which represents a significant increase over the computational resources required when using the smaller datasets in Sections 2.4 and 4.4. The analysis for CC remains the same as in Section 3.3.

The results of this analysis, including parameter estimates and goodness of fit measures, are presented in Table 5.3. Within this table, we also make a comparison to the previously-obtained results using the compressed Pantheon dataset, i.e. the results presented in Table 3.1. The estimates obtained using the CC+SN$*$ and CC+SN$*$+BAO data are presented in Table 5.3. These are similar to those obtained using the CC+SN and CC+SN+BAO data, which indicates that the compressed SN dataset is comparable to the full Pantheon dataset. The compressed dataset also has the advantage of being much simpler to implement than the full dataset. In other words, the compressed SN dataset is effective in condensing the full Pantheon dataset while having

only a handful of data points. However, using SN$*$ in place of SN led to a slightly lower estimate for the Hubble constant in each case, but in each case the difference is within the margin of error of the two estimates. When comparing the results obtained to those obtained using the model-independent GPR and TPR, i.e. the results of Sections 2.4 and 4.4, it can be seen that similar results were obtained. This indicates that the estimates obtained using the $\Lambda CDM$ cosmological model are comparable with the estimates obtained without specifying any cosmological model. Moreover, comparing our results to cosmological models other than $\Lambda CDM$, for example the $f(T)$ or $f(T, B)$ modified gravity models (Briffa et al., 2022, 2023), similar values for $H_0$, $\Omega_{M0}$, and $M$ were obtained throughout, except for one of the $f(T)$ models. As for the goodness of fit measures, there is no clear indication that any one prior is more appropriate to consider than any other. In fact, given the same dataset and comparing likelihood values across the different priors, the values obtained are similar to each other. The estimated autocorrelation times $\hat{\boldsymbol{\tau}}$ shown in the penultimate column of Table 5.3 are also comparable, except for the cases with the Planck and DES priors. A similar discrepancy with the autocorrelation times was also noted in Table 3.1. However, all the estimated autocorrelation times are small enough that 5000 iterations are enough to achieve convergence of the parameter estimates in each case.

The corner plots corresponding to the parameter inference for the priorless case are shown in Figure 5.3. For both the CC+SN$*$ and CC+SN$*$+BAO datasets, we can see that a higher value for $\hat{\Omega}_{M0}$ increases as $\hat{H}_0$ decreases, which is consistent with our findings in Section 3.3. Additionally, the value of $\hat{M}$ increases slightly as $\hat{H}$ increases. The plots for the other priors are similar and can be found in Appendix J.




(a) CC+SN$*$ (b) CC+SN$*$+BAO

Figure 5.3: Corner plots for MCMC on parameters of $\Lambda CDM$ model for priorless case using the full Pantheon dataset

| **Dataset** | **Parameter estimates and prediction errors** | | | **Comparison (SN; Table 3.1)** | | **Goodness of fit** | |
|---|---|---|---|---|---|---|---|
| | $\hat{H}_0$ (km s$^{-1}$ Mpc$^{-1}$) | $\hat{\Omega}_{M0}$ | $\hat{M}$ | $\hat{H}_0$ (km s$^{-1}$ Mpc$^{-1}$) | $\hat{\Omega}_{M0}$ | $\hat{\boldsymbol{\tau}} = \left(\hat{\tau}_{\hat{H}_0}, \hat{\tau}_{\hat{\Omega}_{M0}}, \hat{\tau}_{\hat{M}}\right)^T$ | $\log \Lambda_{tot}$ |
| CC | | | | $67.785^{3.176}_{3.161}$ | $0.327^{0.058}_{0.067}$ | | |
| CC+SN$_*$ | $66.228^{2.989}_{3.094}$ | $0.363^{0.059}_{0.066}$ | $-19.439^{0.074}_{0.075}$ | $68.482^{3.133}_{3.078}$ | $0.311^{0.054}_{0.063}$ | $(47.563, 41.615, 31.384)$ | $-524.283$ |
| CC+SN$_*$+BAO | $69.820^{1.781}_{1.775}$ | $0.263^{0.021}_{0.022}$ | $-19.379^{0.047}_{0.046}$ | $70.461^{1.789}_{1.773}$ | $0.254^{0.020}_{0.021}$ | $(47.266, 37.126, 30.998)$ | $-529.434$ |
| CC+$\hat{H}_0^R$ | | | | $72.505^{1.605}_{1.604}$ | $0.256^{0.033}_{0.036}$ | | |
| CC+SN$_*$+$\hat{H}_0^R$ | $72.017^{1.597}_{1.580}$ | $0.273^{0.034}_{0.036}$ | $-19.307^{0.040}_{0.039}$ | $72.636^{1.603}_{1.596}$ | $0.251^{0.033}_{0.035}$ | $(39.903, 36.379, 52.813)$ | $-528.978$ |
| CC+SN$_*$+BAO+$\hat{H}_0^R$ | $71.911^{1.299}_{1.285}$ | $0.242^{0.015}_{0.016}$ | $-19.328^{0.036}_{0.034}$ | $72.220^{1.282}_{1.284}$ | $0.238^{0.015}_{0.016}$ | $(38.371, 35.367, 57.733)$ | $-534.006$ |
| CC+$\hat{H}_0^{TRGB}$ | | | | $69.245^{1.669}_{1.670}$ | $0.303^{0.038}_{0.042}$ | | |
| CC+SN$_*$+$\hat{H}_0^{TRGB}$ | $68.762^{1.676}_{1.671}$ | $0.302^{0.039}_{0.043}$ | $-19.380^{0.044}_{0.043}$ | $69.424^{1.661}_{1.671}$ | $0.296^{0.038}_{0.041}$ | $(49.374, 37.595, 39.669)$ | $-526.728$ |
| CC+SN$_*$+BAO+$\hat{H}_0^{TRGB}$ | $69.850^{1.323}_{1.309}$ | $0.263^{0.016}_{0.018}$ | $-19.379^{0.035}_{0.036}$ | $70.161^{1.331}_{1.324}$ | $0.258^{0.016}_{0.017}$ | $(48.161, 35.078, 36.966)$ | $-529.742$ |
| CC+$\hat{H}_0^{HW}$ | | | | $71.981^{1.538}_{1.525}$ | $0.263^{0.033}_{0.036}$ | | |
| CC+SN$_*$+$\hat{H}_0^{HW}$ | $71.546^{1.517}_{1.524}$ | $0.279^{0.033}_{0.036}$ | $-19.317^{0.040}_{0.039}$ | $72.104^{1.526}_{1.522}$ | $0.258^{0.033}_{0.036}$ | $(45.797, 38.099, 36.398)$ | $-524.919$ |
| CC+SN$_*$+BAO+$\hat{H}_0^{HW}$ | $71.606^{1.241}_{1.248}$ | $0.245^{0.015}_{0.016}$ | $-19.335^{0.034}_{0.033}$ | $71.885^{1.244}_{1.243}$ | $0.241^{0.015}_{0.016}$ | $(44.360, 33.321, 35.612)$ | $-528.314$ |
| CC+$\hat{H}_0^{CM}$ | | | | $73.547^{1.509}_{1.514}$ | $0.242^{0.031}_{0.034}$ | | |
| CC+SN$_*$+$\hat{H}_0^{CM}$ | $73.103^{1.514}_{1.503}$ | $0.259^{0.032}_{0.034}$ | $-19.282^{0.039}_{0.038}$ | $73.665^{1.513}_{1.500}$ | $0.238^{0.031}_{0.033}$ | $(38.912, 36.405, 40.150)$ | $-527.425$ |
| CC+SN$_*$+BAO+$\hat{H}_0^{CM}$ | $72.651^{1.234}_{1.231}$ | $0.235^{0.014}_{0.015}$ | $-19.310^{0.033}_{0.032}$ | $72.945^{1.237}_{1.230}$ | $0.231^{0.014}_{0.015}$ | $(41.210, 44.569, 40.711)$ | $-531.064$ |
| CC+$\hat{H}_0^{P}$ | | | | $67.413^{0.515}_{0.509}$ | $0.333^{0.033}_{0.034}$ | | |
| CC+SN$_*$+$\hat{H}_0^{P}$ | $67.359^{0.509}_{0.508}$ | $0.343^{0.032}_{0.034}$ | $-19.412^{0.024}_{0.023}$ | $67.430^{0.511}_{0.511}$ | $0.328^{0.032}_{0.034}$ | $(72.637, 39.432, 50.568)$ | $-526.129$ |
| CC+SN$_*$+BAO+$\hat{H}_0^{P}$ | $67.587^{0.494}_{0.501}$ | $0.287^{0.012}_{0.013}$ | $-19.436^{0.019}_{0.020}$ | $67.634^{0.497}_{0.497}$ | $0.285^{0.012}_{0.013}$ | $(73.193, 41.440, 22.771)$ | $-533.234$ |
| CC+$\hat{H}_0^{DES}$ | | | | $67.455^{1.076}_{1.070}$ | $0.332^{0.036}_{0.038}$ | | |
| CC+SN$_*$+$\hat{H}_0^{DES}$ | $67.248^{1.066}_{1.072}$ | $0.345^{0.036}_{0.038}$ | $-19.415^{0.033}_{0.033}$ | $67.527^{1.074}_{1.076}$ | $0.326^{0.035}_{0.037}$ | $(61.032, 39.231, 33.404)$ | $-520.635$ |
| CC+SN$_*$+BAO+$\hat{H}_0^{DES}$ | $68.117^{0.972}_{0.957}$ | $0.282^{0.015}_{0.016}$ | $-19.423^{0.028}_{0.028}$ | $68.278^{0.963}_{0.964}$ | $0.277^{0.015}_{0.016}$ | $(62.649, 34.270, 29.615)$ | $-528.668$ |

Table 5.3: Results of MCMC analysis with full Pantheon dataset

# Chapter 6

# Comparison of Results and Concluding Remarks

Within the previous chapters, we have obtained a large number of estimates for the Hubble constant $H_0$. These estimates are based on the different datasets, kernel functions, and 'prior' $H_0$ values as described in Sections 1.4, 2.2, and 2.4. This chapter is subdivided as follows. Firstly, in Section 6.1, we provide a summary of the work carried out throughout this dissertation. Then, in Section 6.2, we present the estimates obtained using the different statistical methods, prior values, and covariance kernels in the form of whisker plots and carry out a comparison of these estimates. Lastly, in Section 6.3, we state our concluding remarks and limitations encountered throughout the research, and we also mention some avenues for further research on this topic.

## 6.1 Summary of Dissertation

In this dissertation, we have concerned ourselves with one of the fundamental constants within the field of cosmology that explains the expansion of the Universe – this parameter is known as the Hubble constant and is denoted $H_0$. In particular, we set out to investigate an open question within cosmology known as the '$H_0$ tension', namely that different observations and various methodologies for estimating $H_0$ produce very different estimates which in some cases can be as high as 4 standard deviations from each other. This tension has been exacerbated in recent times owing to more accurate estimates, that is, estimates with lower standard errors.

Throughout most of the dissertation, we considered three different data sources, namely CC, CC+SN, and CC+SN+BAO. Then, we extended this by replacing the 'compressed' SN dataset with the full Pantheon dataset SN$_*$ in Section 5.2. We also considered a number of estimated values for $H_0$ from the established literature as 'priors', such as those due to Riess et al. (2019) and Aghanim et al. (2020b). The main differentiating factor between the different priors is that some give a lower value of around 67 km s$^{-1}$ Mpc$^{-1}$, while others give a higher estimate of around 74 km s$^{-1}$ Mpc$^{-1}$. Moreover, for the kernel methods considered, a variety of different kernel functions were trialled, such as the square exponential, rational quadratic, Matérn, and Cauchy kernels.

The different statistical methods employed within this dissertation are GPR, MCMC inference on the parameters of the $\Lambda CDM$ cosmological model, MCMC inference on the hyperparameters of the kernel within a GPR context, TPR, and heteroscedastic GPR and TPR. Out of these, the first two methods have already been carried out within the established literature for the purposes of estimating $H_0$ (see e.g. Briffa et al., 2022; Seikel et al., 2012). On the other hand, to our knowledge, the remaining methods have never been applied within the context of estimating cosmological parameters. Moreover, all the methods except for MCMC $\Lambda CDM$ are independent of the specification of any cosmological model in that they depend only on the observed data. However, each of the 'model-independent' methods still contains assumptions on the distribution of the observed data. In particular, the methods involving GPR assume that the data follows a multivariate normal distribution, and we generalise this assumption to the multivariate Student's $t$-distribution when considering TPR.

The results indicate that the method used may affect the estimates obtained for the Hubble constant. In most cases, the discrepancy between these methods is still within the margin of error and should be investigated further. Since both GPR and MCMC GPR gave similar results to MCMC $\Lambda CDM$, then this suggests that $\Lambda CDM$ seems to adequately model the large-scale structure of the universe, as can be found extensively within the literature. In this case one does not need to apply MCMC GPR and can instead use 'vanilla' GPR, since MCMC GPR is computationally intensive. The heteroscedastic variants of GPR and TPR gave similar estimates when compared with each other as well as with TPR, but the heteroscedastic methods performed better in this case since the standard errors are lower. Additionally, TPR and the heteroscedastic methods gave lower values for $\hat{H}_0$ when compared to GPR and both the parametric and non-parametric MCMC approaches considered.

The choice of prior significantly affects the estimate obtained in each case, with the 'lower' priors understandably leading to lower estimates for $H_0$ than the 'higher' priors. Out of the methods considered, GPR and MCMC GPR showed the greatest sensitivity to the prior specification, with TPR and the heteroscedastic regression methods being more robust to this. In other words, the choice of prior greatly influences the estimates obtained for GPR and MCMC GPR, but this effect is much less pronounced for TPR. One main drawback of TPR, however, is that the standard errors obtained are much larger than those for the other methods, and larger datasets may be needed to reduce this error. The use of heteroscedastic GPR and TPR produced similar estimates but with a significantly reduced standard error. Lastly, the choice of kernel has a negligible effect for GPR, MCMC GPR, TPR, as well as heteroscedastic GPR and TPR.

## 6.2 Comparison of Results

This section shall be concerned with a comparison of the different estimates obtained in order to determine the effect of different methods, datasets, and priors on the estimates obtained via the use of whisker plots. Figure 6.1, for example, contains all the $H_0$ estimates obtained using the CC+SN+BAO dataset, irrespective of the method and prior used. This allows us to determine the effect of different methods and priors given a fixed dataset. Similarly, we present the estimates obtained using the GPR method across all datasets and priors in Figure 6.2. We also show the estimates obtained using the priorless case as well as the Riess and Planck priors in Figures 6.3 through 6.5. The Riess and Planck priors specifically were picked here since they represent a 'higher' and a 'lower' estimate of $H_0$ respectively, and since the tension between these two priors is of more than $4\sigma$. The rest of the plots – i.e. those corresponding to the other methods, datasets, and priors – are shown in Appendix K.

As noted in Section 2.4, the effect of the different kernel functions has already been observed to be almost negligible. Therefore, for the purpose of this comparison, each value of $\hat{H}_0$ presented is the median of estimates obtained using a given prior, dataset, and method. The same applies for the standard error of this estimate. For each estimate, the actual estimated value is shown using the dot, while the errors are shown using the whiskers. More uncertain estimates, i.e. those with a higher standard error, consequently have longer whiskers than estimates with low standard error. The pre-established Riess and Planck priors $\hat{H}_0^R$ and $\hat{H}_0^P$ are superimposed on the plot as the

blue and red bars respectively. The $\sigma$-distance between each $H_0$ estimate and each of these two priors, as calculated in equation (2.18), is also shown in the whisker plot. Lastly, at the bottom of each plot, we show the median of all the values shown in that plot. The median $H_0$ estimate for each plot is also presented within Tables 6.1, 6.2, and 6.3. In order to obtain a suitable standard error for each median value, the median standard error of the relevant estimates is taken.

From these figures, one can immediately notice some patterns. In particular, the estimates obtained using GPR, MCMC $\Lambda CDM$, and MCMC GPR are highly dependent on the prior specification. For these methods, the tensions between each estimate and the pre-established Riess and Planck priors are often more than $2\sigma$, and even exceed $4\sigma$ in some cases. On the other hand, TPR as well as heteroscedastic GPR and TPR are less sensitive to the specification of the 'prior' $\hat{H}_0$ value, so that the estimates obtained for the Hubble constant using each of the different 'priors' are closer to each other. However, for TPR, only a few of the estimated values have more than $2\sigma$ tension with either the Riess or Planck priors, owing to the larger standard error associated with the TPR estimates. Therefore, further investigation of TPR with larger datasets is required so that these standard errors are reduced. Regardless of the method used, the standard errors for the priorless case are understandably much larger than for the cases where a prior $\hat{H}_0$ value was specified. This can be seen, for example, by comparing Figure 6.3 to Figures 6.4 and 6.5 and to similar figures within Appendix K. Additionally, a striking observation that can be made from Figure 6.5 is that the estimates for all methods when considering the Planck prior are very close to each other. In this case, the evidence suggests a lower $H_0$ value and one that is very close to $\hat{H}_0^P$.

Additionally, in Figure 6.6 we further summarise the results by presenting the median Hubble constant estimates obtained by prior, dataset, and method. In other words, given a fixed 'prior', we present the median of all the estimates involving that prior irrespective of the kernel, dataset, and method used. Similarly, given a dataset, we calculate the median over all priors, kernels, and methods, and given a method we get the median over all priors/kernels/datasets. At the bottom of this plot, we present the median estimate obtained across all methods, datasets, kernels, and priors. This value is $\hat{H}_0^{med} = 68.85 \pm 1.67$ km s$^{-1}$ Mpc$^{-1}$. As can be seen from the bottom part of Figure 6.6, this value is less than 1 standard deviation from the Planck prior $\hat{H}_0^P$ but is at a tension of more than $2\sigma$ from the Riess prior $\hat{H}_0^R$, which is further evidence in favour of a lower value for the Hubble constant.

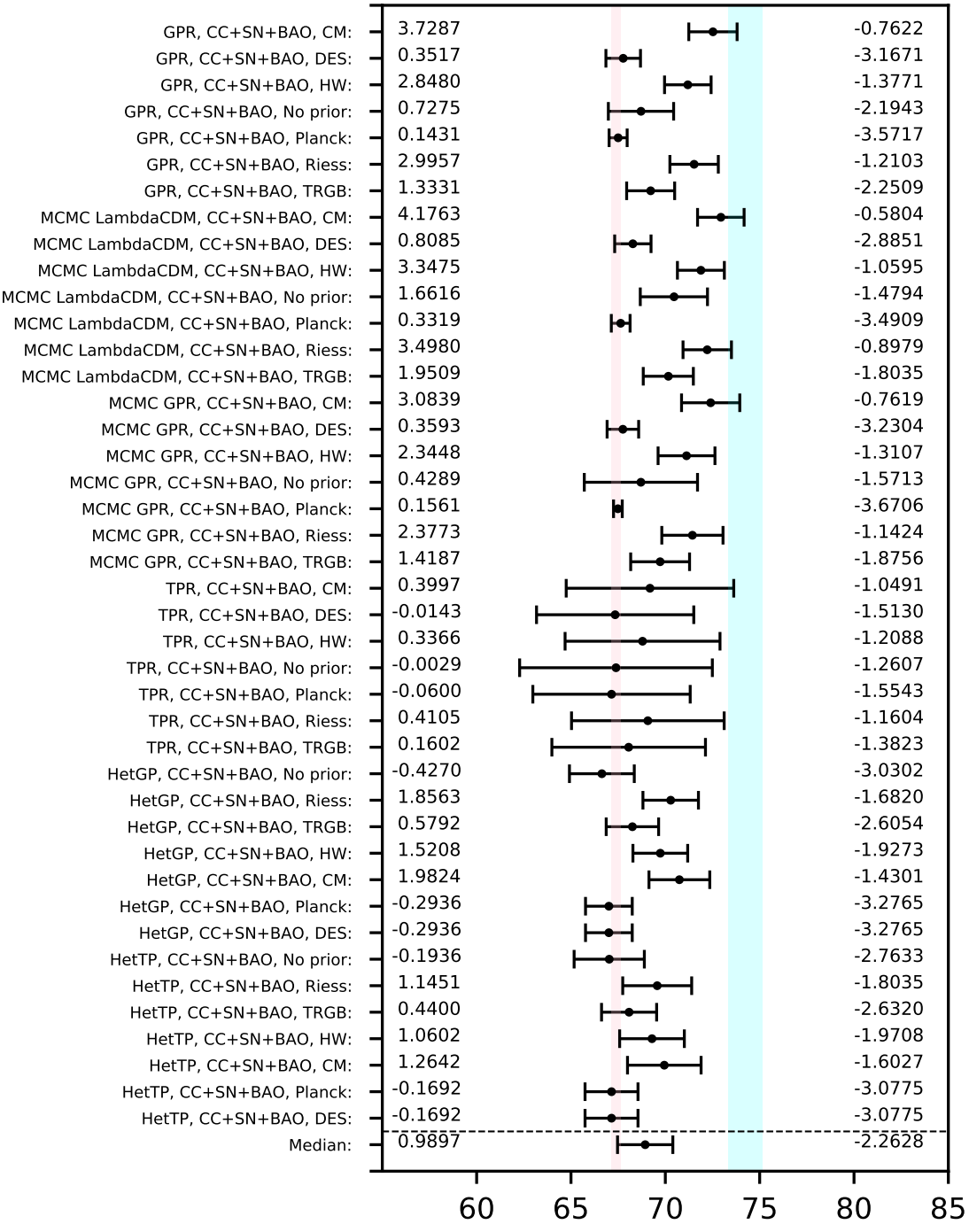



Figure 6.1: Comparison of estimates obtained using CC+SN+BAO dataset

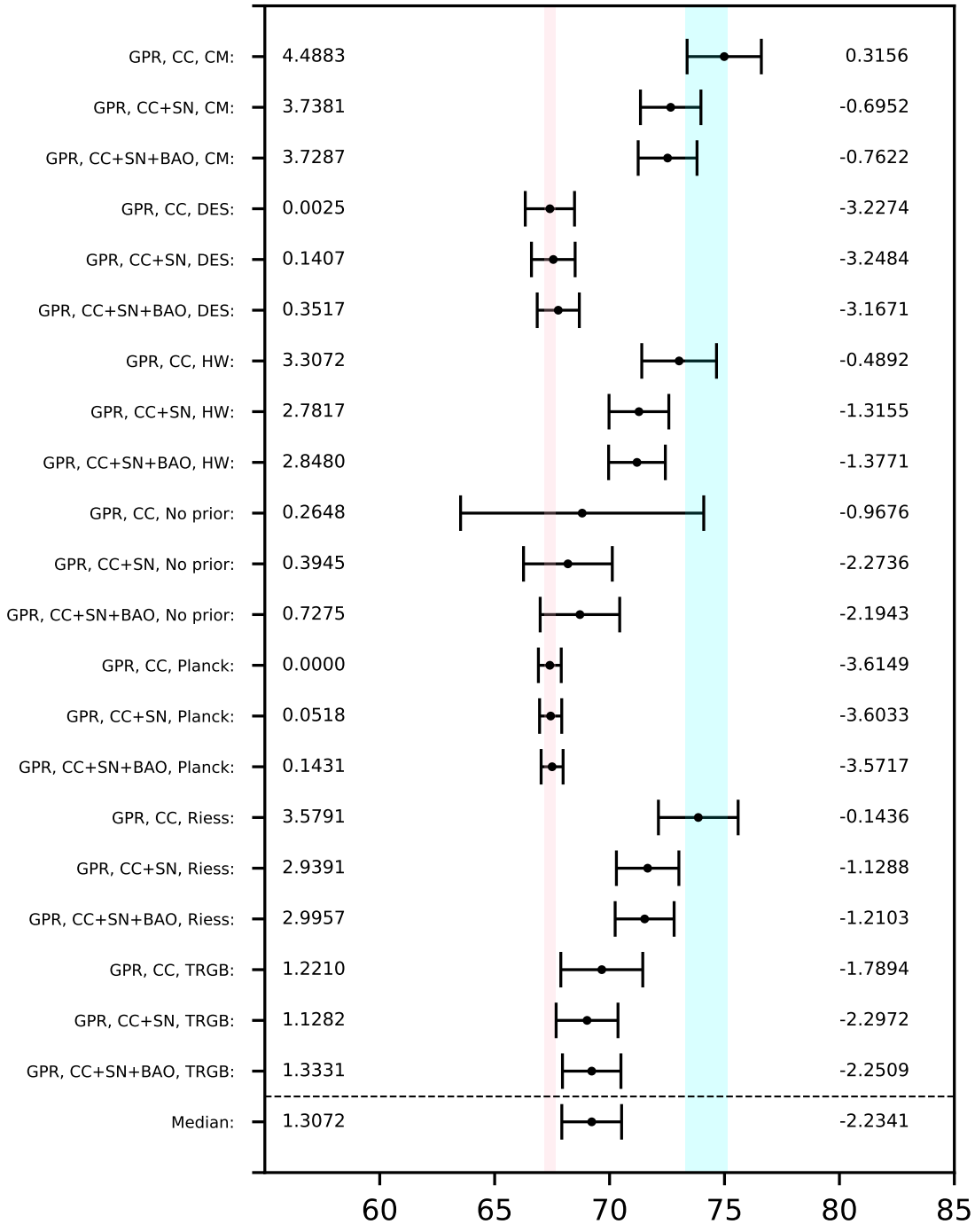



Figure 6.2: Comparison of estimates obtained using GPR method

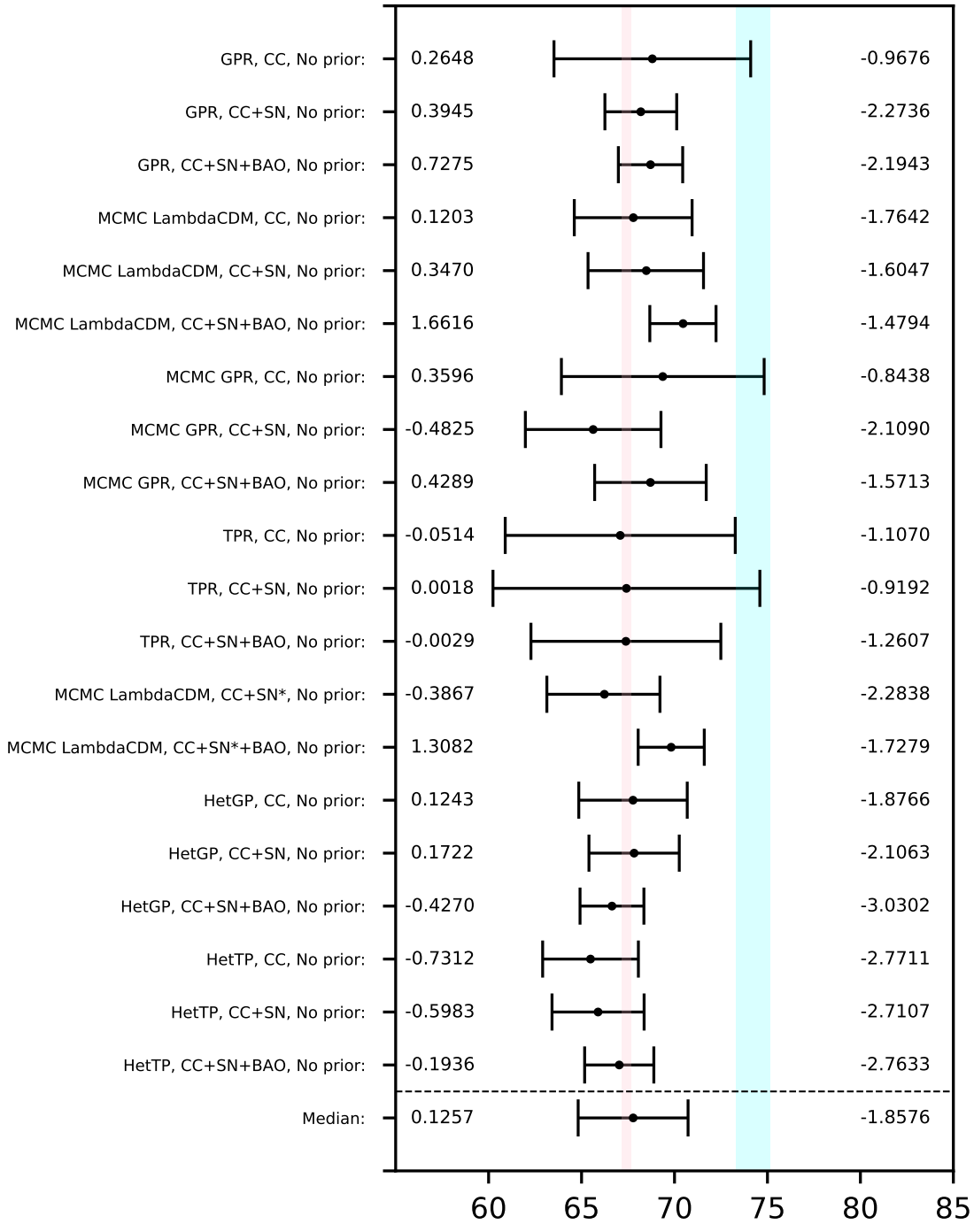



Figure 6.3: Comparison of estimates obtained in priorless case

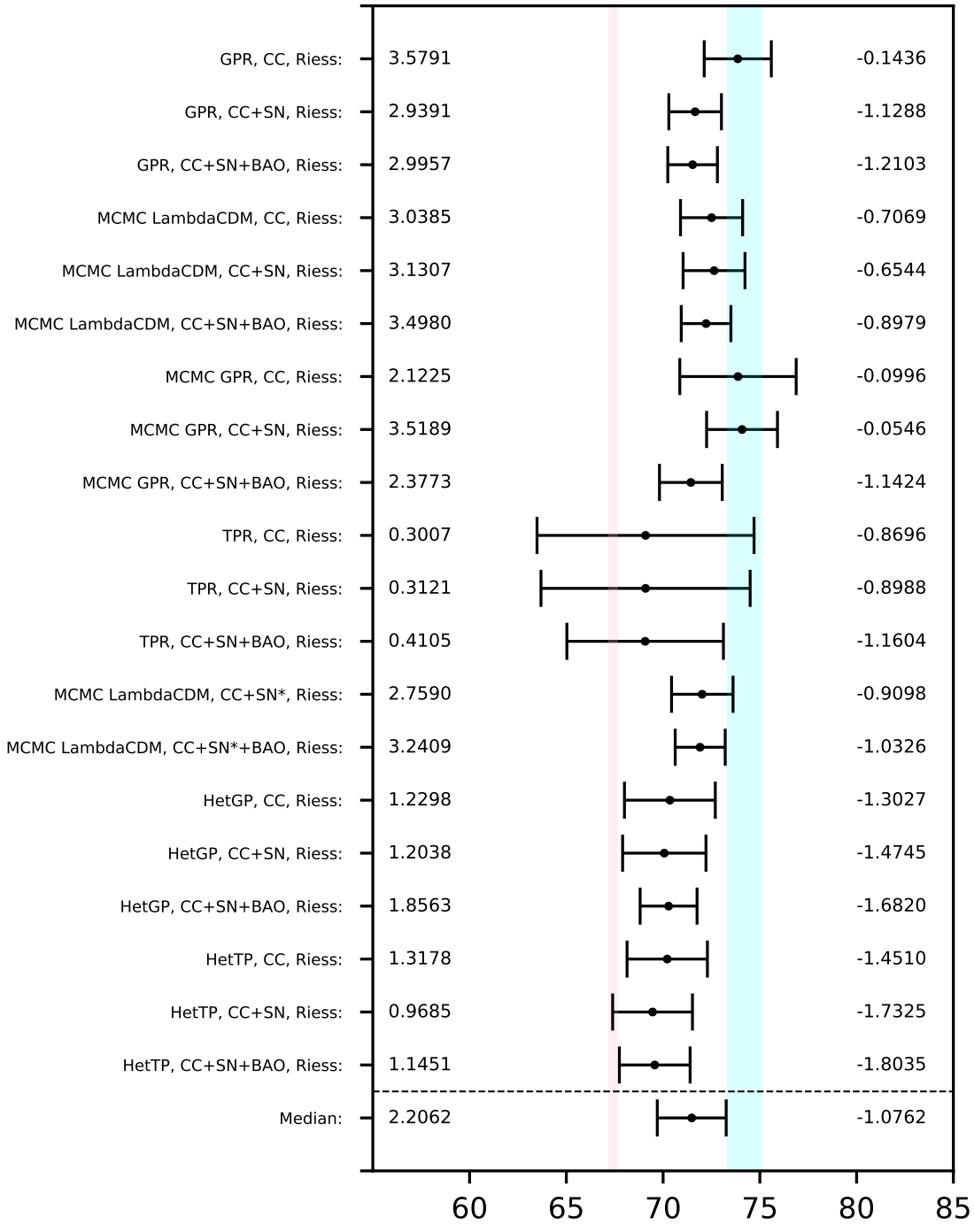



Figure 6.4: Comparison of estimates obtained using Riess prior

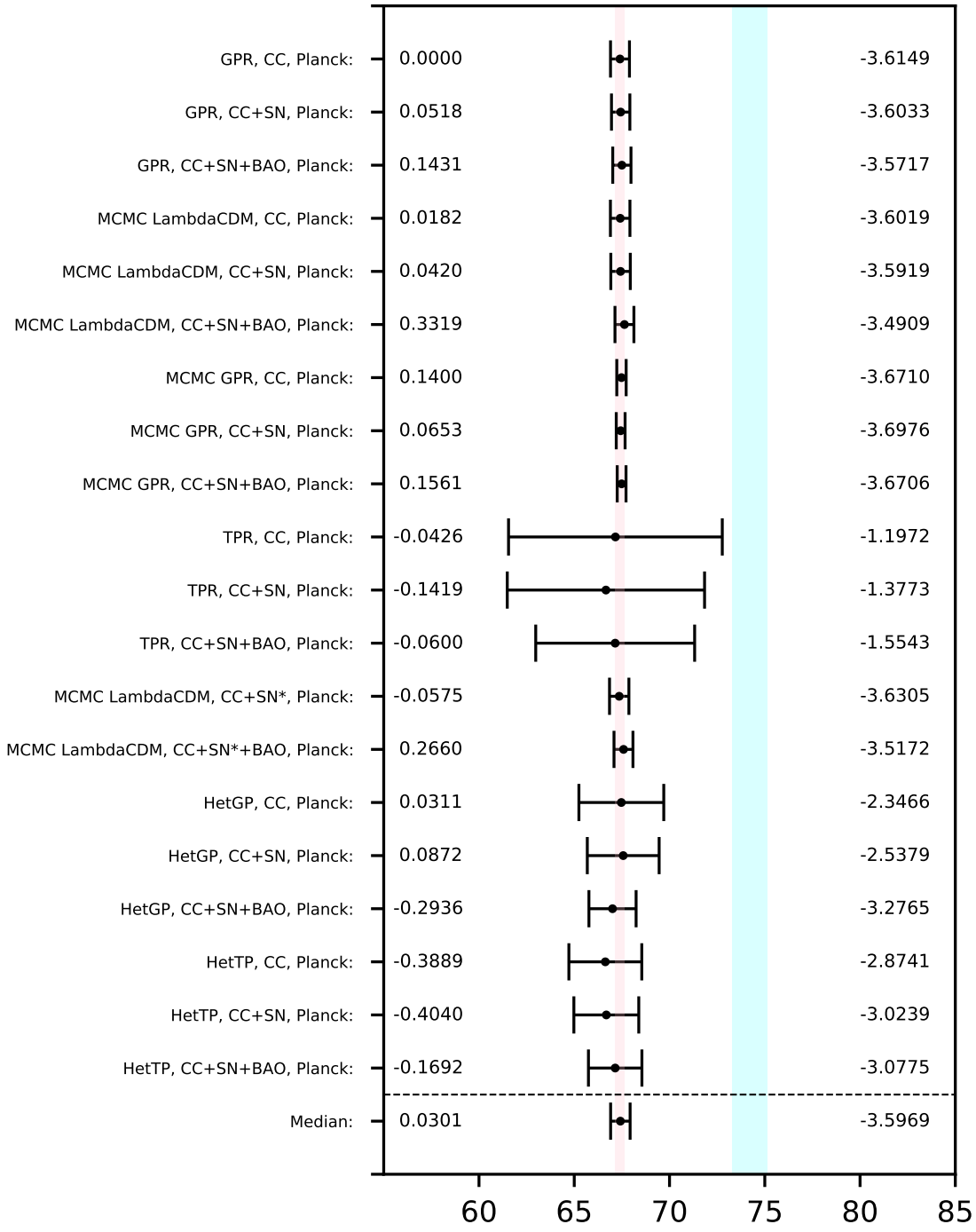



Figure 6.5: Comparison of estimates obtained using Planck prior

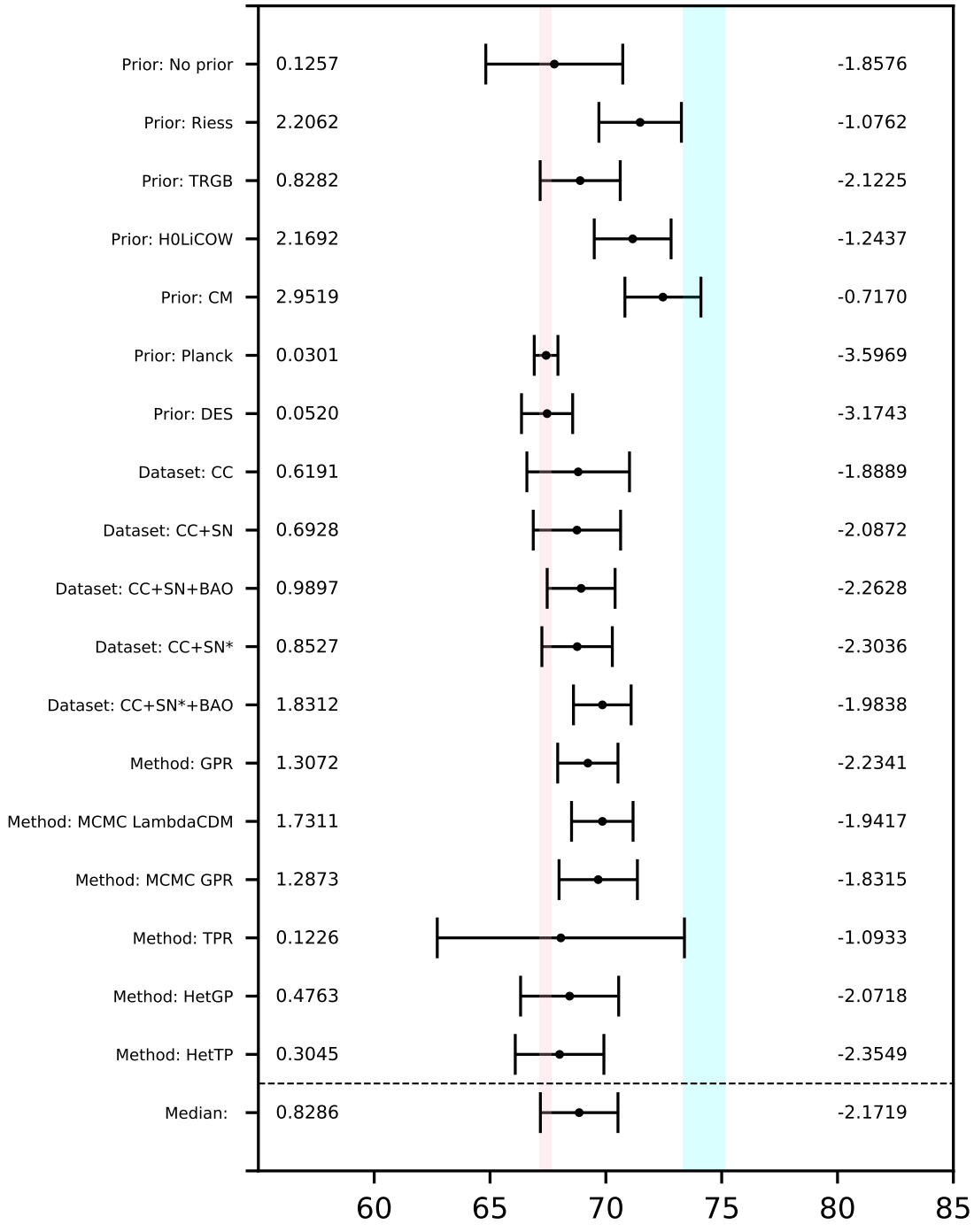



Figure 6.6: Comparison of estimates obtained: median values

We can also compare the median $\hat{H}_0$ value obtained for each method, prior, and dataset in order to determine the effect of the particular method, prior, or dataset on the estimates obtained. The table comparing the different datasets is presented as Table 6.1, and similarly Tables 6.2 and 6.3 respectively compare the methods and priors.

From these tables, we can see that larger datasets generally lead to more confident predictions, i.e. estimates with lower uncertainties. However, it should be noted that only the MCMC $\Lambda CDM$ method was applied to the datasets CC+SN$*$ and CC+SN$*$+BAO. When it comes to the method used, we see that GPR, MCMC $\Lambda CDM$, and MCMC GPR give very similar results, indicating that the model-independent GPR is in agreement with the widely-accepted $\Lambda CDM$ cosmological model. From the bottom part of Figure 6.2, we can see that the median value obtained for GPR is at a distance of 1.3072 standard deviations from $\hat{H}_0^P$, while there is a tension of $2.2341\sigma$ with $\hat{H}_0^R$. This provides some evidence in favour of a lower value for the Hubble constant. The rest of the methods considered, namely TPR and heteroscedastic GPR and TPR, produced even lower estimates for $H_0$. Therefore, these methods are providing further evidence in favour of a lower value of $H_0$ that is closer to the Planck and DES priors.

| **Dataset** | **Median $\hat{H}_0$ Value** |
|---|---|
| CC | $68.806 \pm 2.215$ |
| CC+SN | $68.751 \pm 1.885$ |
| CC+SN+BAO | $68.933 \pm 1.465$ |
| CC+SN$*$ | $68.762^{1.517}_{1.524}$ |
| CC+SN$*$+BAO | $69.850^{1.241}_{1.248}$ |

Table 6.1: Median $H_0$ estimate obtained by dataset

| **Method** | **Median $\hat{H}_0$ Value** |
|---|---|
| GPR | $69.222 \pm 1.301$ |
| MCMC $\Lambda CDM$ | $69.850^{1.324}_{1.331}$ |
| MCMC GPR | $69.670 \pm 1.691$ |
| TPR | $68.057 \pm 5.335$ |
| Heteroscedastic GPR | $68.436 \pm 2.117$ |
| Heteroscedastic TPR | $68.002 \pm 1.913$ |

Table 6.2: Median $H_0$ estimate obtained by method

As for the prior used, this understandably has a great effect on the $H_0$ estimates obtained, as the effect of adding a prior is to consider an additional 'artificial' data point at $z = 0$. As expected, using the Planck and DES priors produced lower values than the Riess, H0LiCOW, and CM priors. However, it is to be noted that the estimates obtained when no prior was included are closer to the Planck and DES priors than the Riess and H0LiCOW priors – this also suggests a lower value for $H_0$. This is arguably the main finding of our research.

| **Prior** | **Median $\hat{H}_0$ Value** |
|---|---|
| No prior | $67.777^{2.953}_{2.959}$ |
| $\hat{H}_0^R$ | $71.480^{1.781}_{1.780}$ |
| $\hat{H}_0^{TRGB}$ | $68.891^{1.727}_{1.730}$ |
| $\hat{H}_0^{HW}$ | $71.158^{1.658}_{1.657}$ |
| $\hat{H}_0^{CM}$ | $72.463^{1.640}_{1.638}$ |
| $\hat{H}_0^{P}$ | $67.421^{0.513}_{0.510}$ |
| $\hat{H}_0^{DES}$ | $67.463^{1.104}_{1.106}$ |

Table 6.3: Median $H_0$ estimate obtained by prior

## 6.3 Limitations and Possible Further Research Topics

A number of limitations and opportunities for further research within this area have been identified over the course of this research project. One such limitation was computational issues, especially for the MCMC approaches. For both the parametric and non-parametric MCMC approaches, specialised external computational resources were required. Another one of the more immediate limitations encountered is the use of very small datasets leading to a large standard error, especially for the TPR method. In most cases, we opted to use smaller datasets, with the exception being the use of the full Pantheon dataset in Section 5.2. Nevertheless, the results obtained with the smaller datasets largely agree with those using the full Pantheon dataset, but the use of the full Pantheon dataset requires one to assume some cosmological model such as $\Lambda CDM$. This is opposed to methods such as GPR and TPR, which are independent of any cosmological model and therefore depend only on the observational Hubble data.

Further work can include a meta-analysis considering both estimates obtained in this work as well as other estimates for $H_0$ already found in the literature. Additionally, a more thorough investigation the different datasets and priors available for the purposes of estimating $H_0$ can be carried out. Other possible avenues for further research include considering cosmological models apart from $\Lambda CDM$, for example modified gravity theories such as $f(T)$ models (Briffa et al., 2022) and $f(T, B)$ models (Briffa et al., 2023), as well as modified Newton dynamics (see e.g. Tian et al., 2013). Yet another possibility is the use of the statistical techniques explored within this dissertation for the estimation of different cosmological parameters such as $f\sigma_8$. To our knowledge, only GPR has been applied for such purposes (see e.g. Avila et al., 2022).

# Appendix A: Proofs of Some Theoretical Results on TPR

## Conditional Density of MVT Distribution

**Theorem A.1.** *Suppose that* $\mathbf{X} \sim MVT_p(\mathbf{m}, \mathbb{K}, \nu)$*, and let* $\mathbf{x}$ *be partitioned into two sub-vectors* $\mathbf{x}_1 = (x_1, x_2, ..., x_{p_1})^T$ *and* $\mathbf{x}_2 = (x_1^*, x_2^*, ..., x_{p_2}^*)^T$ *of lengths* $p_1$ *and* $p_2$ *respectively, with* $p = p_1 + p_2$*. Furthermore, define* $\mathbf{m}_1$ *and* $\mathbf{m}_2$ *as the mean functions corresponding to the entries of* $\mathbf{x}_1$ *and* $\mathbf{x}_2$ *respectively, and define* $\mathbb{K}_{11}$ *as the* $(p_1 \times p_1)$ *submatrix of* $\mathbb{K}$ *corresponding to the entries of* $\mathbf{x}_1$*, where the* $(i, j)^{th}$ *entry is given by* $k(x_i, x_j)$*. Also, define* $\mathbb{K}_{12}$*,* $\mathbb{K}_{21}$ *and* $\mathbb{K}_{22}$ *as matrices with dimensionalities* $(p_1 \times p_2)$*,* $(p_2 \times p_1)$*, and* $(p_2 \times p_2)$ *in a similar manner.*

*Then, the conditional density of* $\mathbf{X}_2$ *given* $\mathbf{X}_1$ *is the* $p_2$*-dimensional density given by*

$$\mathbf{X}_2|\mathbf{X}_1 \sim MVT_{p_2}\left[\widetilde{\mathbf{m}}_2,\ \frac{\beta_1 + \nu}{p_1 + \nu}\ \widetilde{\mathbb{K}}_{22},\ \nu + p_1\right],$$

*where*

$$\begin{aligned}
\widetilde{\mathbf{m}}_2 &= \mathbb{K}_{21}\mathbb{K}_{11}^{-1}(\mathbf{x}_1 - \mathbf{m}_1) - \mathbf{m}_2, \\
\beta_1 &= (\mathbf{x}_1 - \mathbf{m}_1)^T\mathbb{K}_{11}^{-1}(\mathbf{x}_1 - \mathbf{m}_1), \\
\widetilde{\mathbb{K}}_{22} &= \mathbb{K}_{22} - \mathbb{K}_{21}\mathbb{K}_{11}^{-1}\mathbb{K}_{12},
\end{aligned}$$

*and where* $\mathbb{K}_{12}^T = \mathbb{K}_{21}$ *by the symmetry property of the kernel* $k$*.*

*Proof.* Let $\beta_2 = (\mathbf{x}_2 - \widetilde{\mathbf{m}}_2)^T\widetilde{\mathbb{K}}_{22}^{-1}(\mathbf{x}_2 - \widetilde{\mathbf{m}}_2)$. Then,

$$\beta_1 + \beta_2 = (\mathbf{x} - \mathbf{m})^T\mathbb{K}^{-1}(\mathbf{x} - \mathbf{m}).$$

Therefore,

$$\begin{aligned}\mathbb{P}(\mathbf{x}_2|\mathbf{x}_1) &= \frac{\mathbb{P}(\mathbf{x}_1,\mathbf{x}_2)}{\mathbb{P}(\mathbf{x}_1)} \\ &\propto \left[1+\frac{\beta_1+\beta_2}{\nu}\right]^{-\frac{\nu+p}{2}} \left[1+\frac{\beta_1}{\nu}\right]^{\frac{\nu+p_1}{2}} \\ &\propto \left[1+\frac{\beta_2}{\beta_1+\nu}\right]^{-\frac{\nu+p_1}{2}},\end{aligned}$$

and the result may be obtained by comparing this expression to equation (4.1). □

## Equivalence of Two Generative Processes for TPs

**Theorem A.2.** *The two generative processes*

$$\begin{aligned}\sigma &\sim IWP(k,\nu) \\ x|\sigma &\sim GP(m,\nu\mathbb{\Sigma}).\end{aligned}$$

*and*

$$\begin{aligned}R^{-1} &\sim Gamma\left(\frac{\nu}{2},\frac{\rho}{2}\right) \\ \mathbf{X}|R &\sim MVN_p\left(\mathbf{m},\frac{\nu r}{\rho}\mathbb{K}\right)\end{aligned}$$

*mentioned in equations (4.3) and (4.4) are equivalent in the sense that they result in the same marginal distribution for* $\mathbf{X}$.

*Proof.* For the former, and assuming $\mathbf{m}=\mathbf{0}$ without loss of generality, the marginal distribution of $\mathbf{X}$ is given by

$$\begin{aligned}\mathbb{P}(\mathbf{x}) &= \int \mathbb{P}(\mathbf{x}|r)\ \mathbb{P}(r)\ \mathrm{d}r \\ &= \int (2\pi\nu r)^{-\frac{p}{2}}|\mathbb{K}|^{-\frac{1}{2}}\exp\left[-\frac{\mathbf{x}^T\mathbb{K}^{-1}\mathbf{x}}{2\nu r}\right] r^{-\frac{\nu}{2}-1}\ \frac{\exp(-\frac{1}{2r})}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}\ \mathrm{d}r \\ &= \frac{(2\pi\nu)^{-\frac{p}{2}}|\mathbb{K}|^{-\frac{1}{2}}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}\int r^{\left[-\frac{\nu+p}{2}-1\right]}\ \exp\left[-\left(1+\frac{\mathbf{x}^T\mathbb{K}\mathbf{x}}{\nu}\right)/2r\right]\ \mathrm{d}r \\ &= \frac{(2\pi\nu)^{-\frac{p}{2}}|\mathbb{K}|^{-\frac{1}{2}}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}\left[\left(1+\frac{\mathbf{x}^T\mathbb{K}^{-1}\mathbf{x}}{\nu}\right)/2\right]^{-\frac{\nu+p}{2}}\Gamma\left[\frac{\nu+p}{2}\right] \\ &= (\pi\nu)^{-\frac{p}{2}}|\mathbb{K}|^{-\frac{1}{2}}\left[1+\frac{\mathbf{x}^T\mathbb{K}^{-1}\mathbf{x}}{\nu}\right]^{-\frac{\nu+p}{2}}\frac{\Gamma\left(\frac{\nu+p}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}. \qquad \text{(A.1)}\end{aligned}$$

For the latter, and again assuming $\mathbf{m} = \mathbf{0}$ and $\rho = 1$ without loss of generality, let

$$c_p(\nu, \mathbb{K}) := \frac{|\mathbb{K}|^{\frac{\nu+p-1}{2}}}{2^{\frac{(\nu+p-1)p}{2}} \Gamma_p\left(\frac{\nu+p-1}{2}\right)}$$

where, for $\mathbb{S} \in \Pi(p)$,

$$\Gamma_p(x) = \int_{|\mathbb{S}|>0} |\mathbb{S}|^{x-\frac{p+1}{2}} \exp(\mathrm{Tr}(\mathbb{S}))\, \mathrm{d}\mathbb{S}.$$

Then, the marginal distribution of $\mathbf{X}$ is

$$\begin{aligned}
\mathbb{P}(\mathbf{x}) &= \int \mathbb{P}(\mathbf{x}|\sigma)\ \mathbb{P}(\sigma)\ \mathrm{d}\sigma \\
&= \int (2\pi\nu)^{-\frac{p}{2}} |\sigma|^{\frac{1}{2}} \exp\left[-\frac{\mathbf{x}^T\sigma\mathbf{x}}{2\nu}\right] \\
&\qquad \times c_p(\nu+p-1, \mathbb{K}^{-1}) |\sigma|^{\frac{\nu}{2}} \exp\left[-\frac{\mathrm{Tr}(\mathbb{K}\sigma)}{2}\right]\ \mathrm{d}\sigma \\
&= \int (2\pi\nu)^{-\frac{p}{2}}\ c_p(\nu+p-1, \mathbb{K}^{-1})\ c_p\left[\nu+p, \left(\mathbb{K} + \frac{\mathbf{x}\mathbf{x}^T}{\nu}\right)^{-1}\right]^{-1} \\
&= \int (2\pi\nu)^{-\frac{p}{2}} \left[|\mathbb{K}|^{-\frac{\nu+p-1}{2}} 2^{\frac{(\nu+p-1)p}{2}} \Gamma_p\left(\frac{\nu+p-1}{2}\right)\right]^{-1} \\
&\qquad \times |\mathbb{K}|^{-\frac{\nu+p}{2}} \left[1 + \frac{\mathbf{x}^T\mathbb{K}^{-1}\mathbf{x}}{\nu}\right]^{-\frac{\nu+p}{2}} 2^{\frac{(\nu+p)p}{2}} \Gamma_p\left[\frac{\nu+p}{2}\right] \\
&= (\pi\nu)^{-\frac{p}{2}} |\mathbb{K}|^{-\frac{1}{2}} \left[1 + \frac{\mathbf{x}^T\mathbb{K}^{-1}\mathbf{x}}{\nu}\right]^{-\frac{\nu+p}{2}} \frac{\Gamma_p\left(\frac{\nu+p}{2}\right)}{\Gamma_p\left(\frac{\nu+p-1}{2}\right)}.
\end{aligned} \tag{A.2}$$

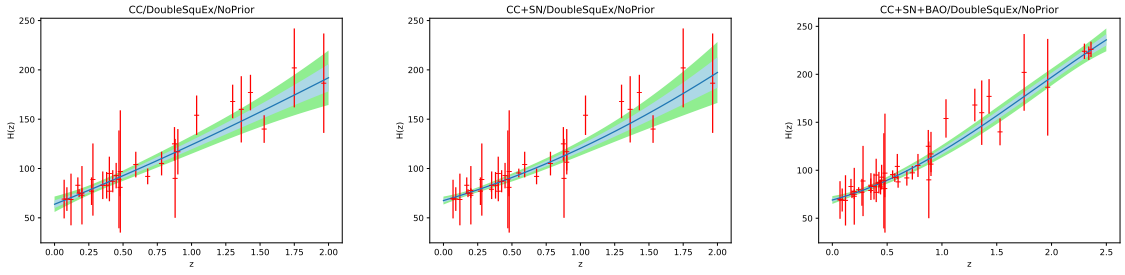
The equivalence of equations (A.1) and (A.2) then follows the identity

$$\frac{\Gamma_p(x)}{\Gamma_p\left(x-\frac{1}{2}\right)} = \frac{\Gamma(x)}{\Gamma\left(x-\frac{n}{2}\right)}.$$

Moreover, the marginal distribution obtained in each case is in fact that of the multivariate Student's $t$-distribution. □

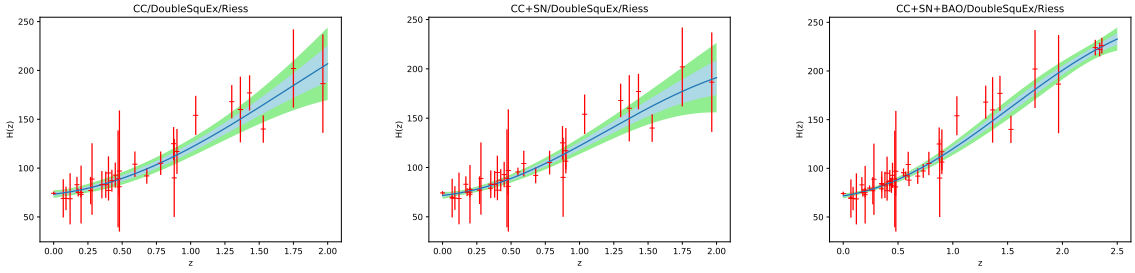
# Appendix B: GPR Plots

The figures for the square exponential kernel were presented in Section 2.4. Similar figures for all the other kernel functions considered are shown in this Appendix.



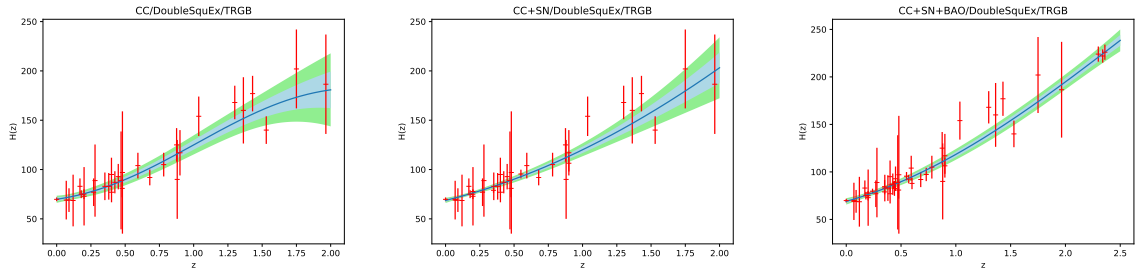

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.1: GPR reconstructions of $\hat{H}(z)$ against $z$ for double square exponential kernel function and without adding any $H_0$ estimate to the data



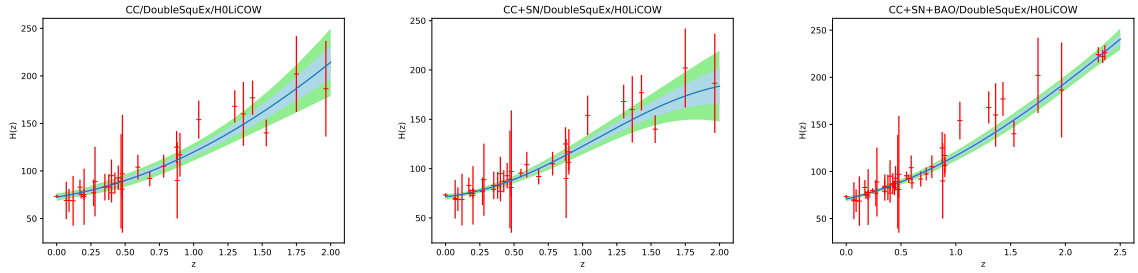

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.2: GPR reconstructions of $\hat{H}(z)$ against $z$ for double square exponential kernel function and using $\hat{H}_0^R$



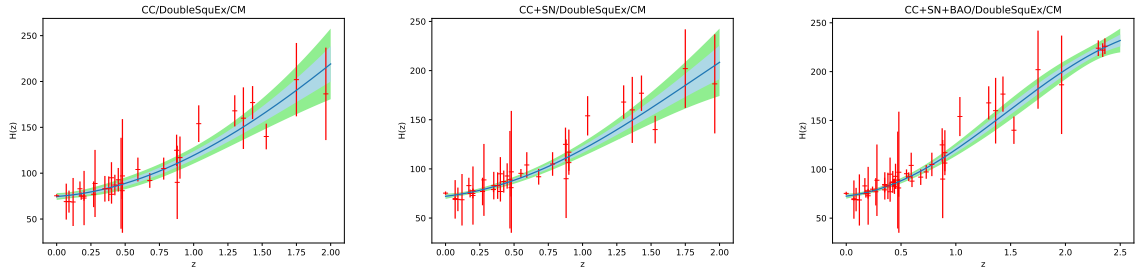

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.3: GPR reconstructions of $\hat{H}(z)$ against $z$ for double square exponential kernel function and using $\hat{H}_0^{TRGB}$



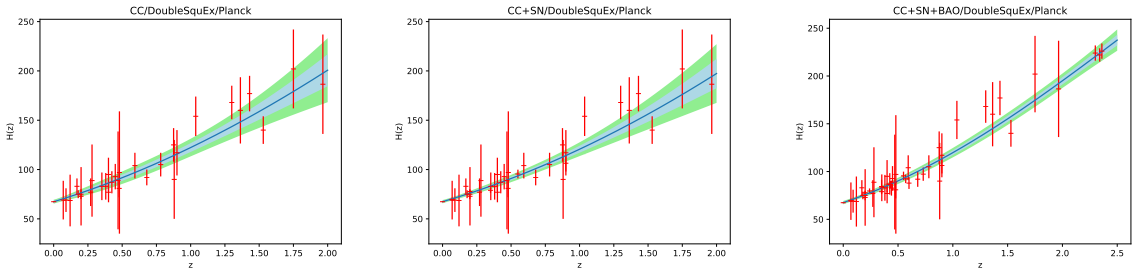

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.4: GPR reconstructions of $\hat{H}(z)$ against $z$ for double square exponential kernel function and using $\hat{H}_0^{HW}$



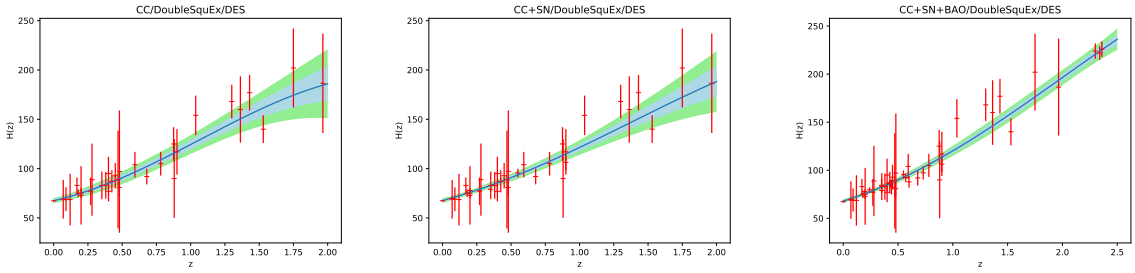

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.5: GPR reconstructions of $\hat{H}(z)$ against $z$ for double square exponential kernel function and using $\hat{H}_0^{CM}$



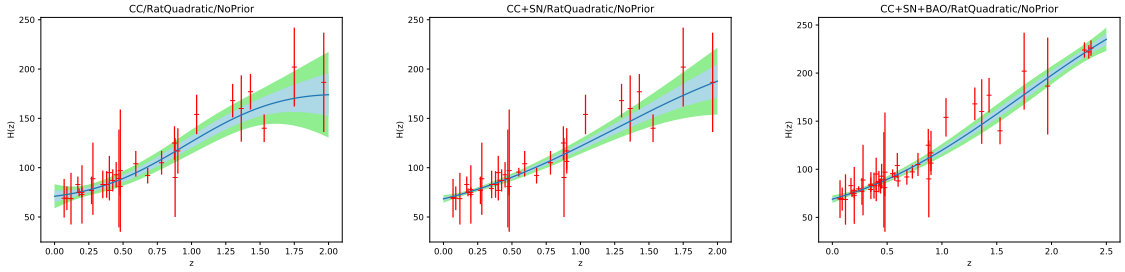

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.6: GPR reconstructions of $\hat{H}(z)$ against $z$ for double square exponential kernel function and using $\hat{H}_0^P$



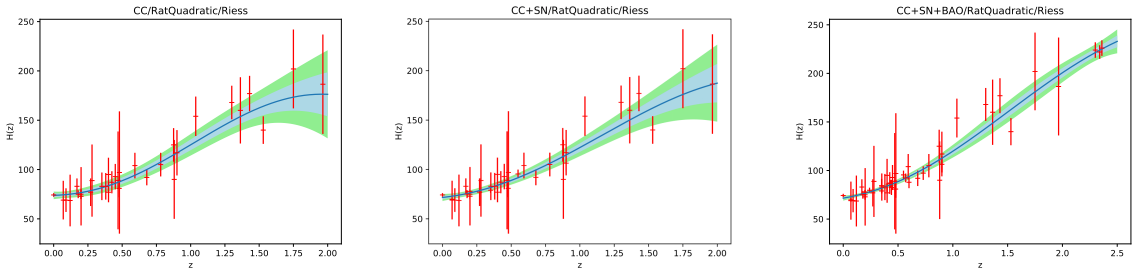

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.7: GPR reconstructions of $\hat{H}(z)$ against $z$ for double square exponential kernel function and using $\hat{H}_0^{DES}$



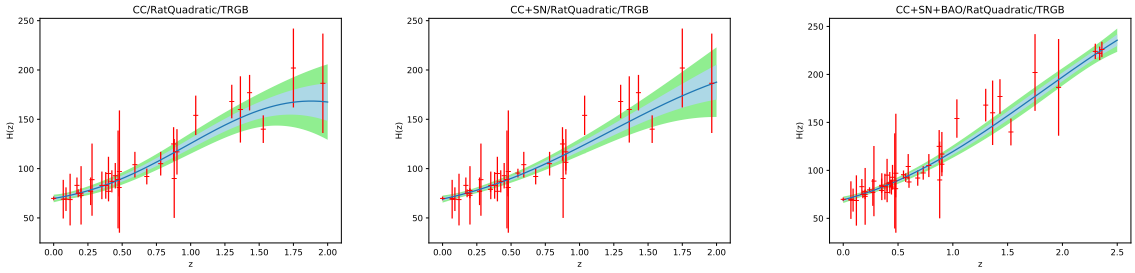

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.8: GPR reconstructions of $\hat{H}(z)$ against $z$ for rational quadratic kernel function and without adding any $H_0$ estimate to the data



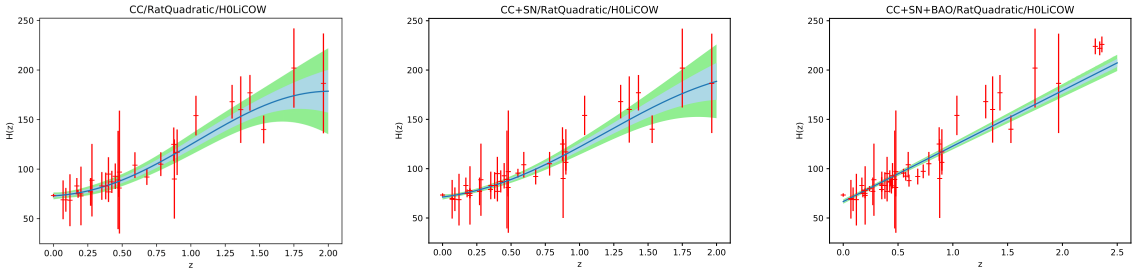

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.9: GPR reconstructions of $\hat{H}(z)$ against $z$ for rational quadratic kernel function and using $\hat{H}_0^R$



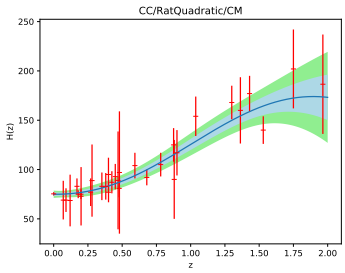

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.10: GPR reconstructions of $\hat{H}(z)$ against $z$ for rational quadratic kernel function and using $\hat{H}_0^{TRGB}$

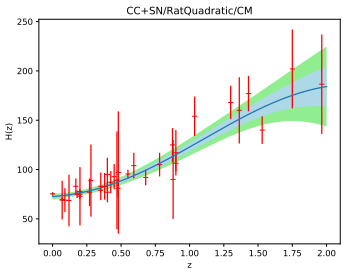



(a) CC (b) CC+SN (c) CC+SN+BAO

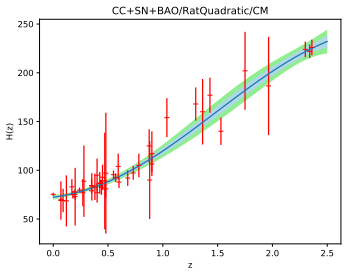
Figure B.11: GPR reconstructions of $\hat{H}(z)$ against $z$ for rational quadratic kernel function and using $\hat{H}_0^{HW}$




(a) CC

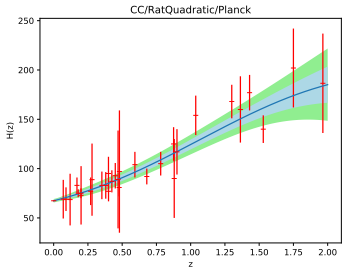



(b) CC+SN

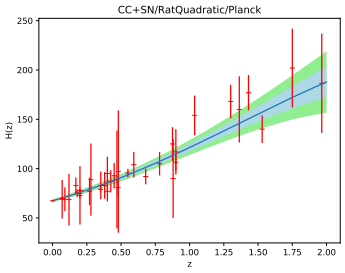



(c) CC+SN+BAO

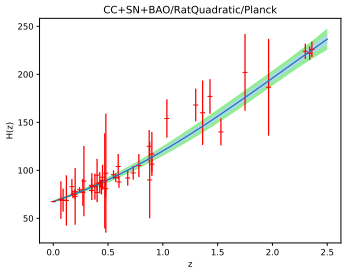
Figure B.12: GPR reconstructions of $\hat{H}(z)$ against $z$ for rational quadratic kernel function and using $\hat{H}_0^{CM}$




(a) CC

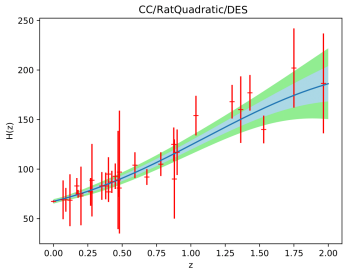



(b) CC+SN

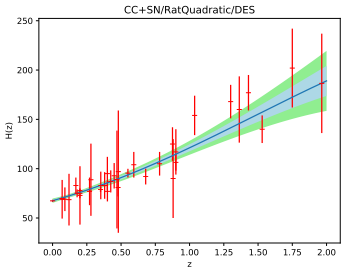



(c) CC+SN+BAO

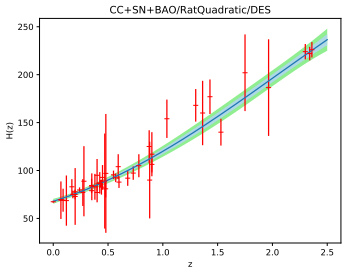
Figure B.13: GPR reconstructions of $\hat{H}(z)$ against $z$ for rational quadratic kernel function and using $\hat{H}_0^{P}$




(a) CC

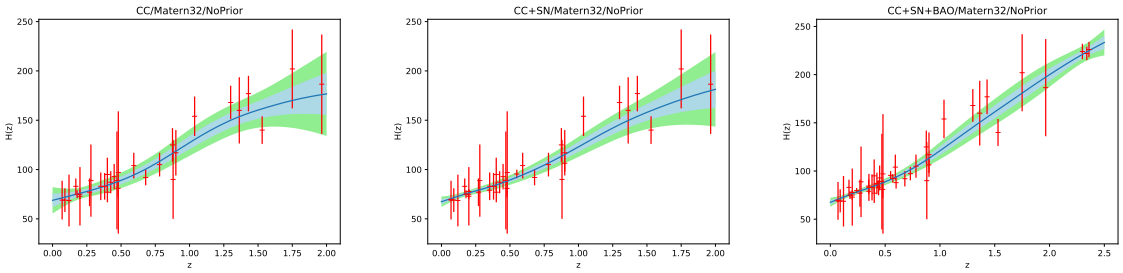



(b) CC+SN



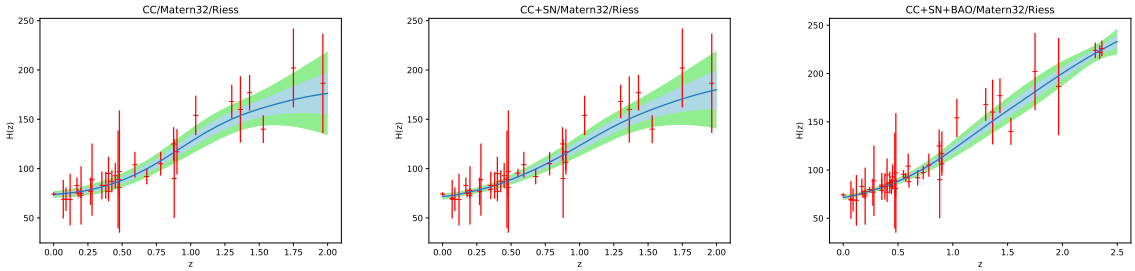

(c) CC+SN+BAO

Figure B.14: GPR reconstructions of $\hat{H}(z)$ against $z$ for rational quadratic kernel function and using $\hat{H}_0^{DES}$



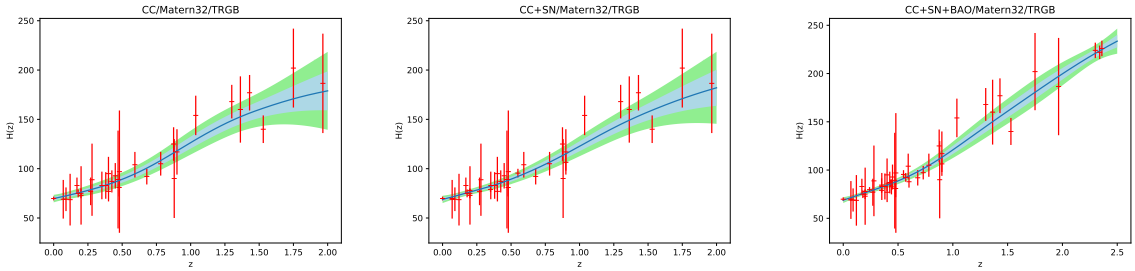

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.15: GPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{3}{2}\right)$ kernel function and without adding any $H_0$ estimate to the data



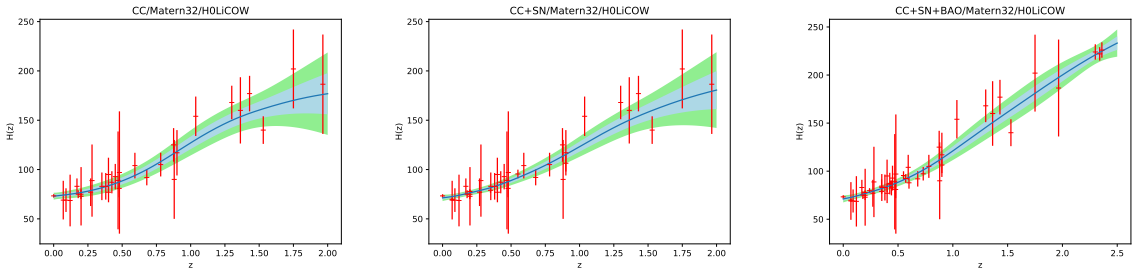

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.16: GPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{3}{2}\right)$ kernel function and using $\hat{H}_0^R$



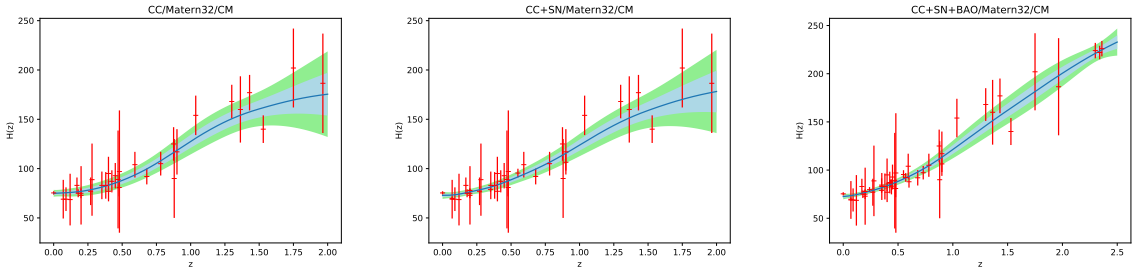

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.17: GPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{3}{2}\right)$ kernel function and using $\hat{H}_0^{TRGB}$



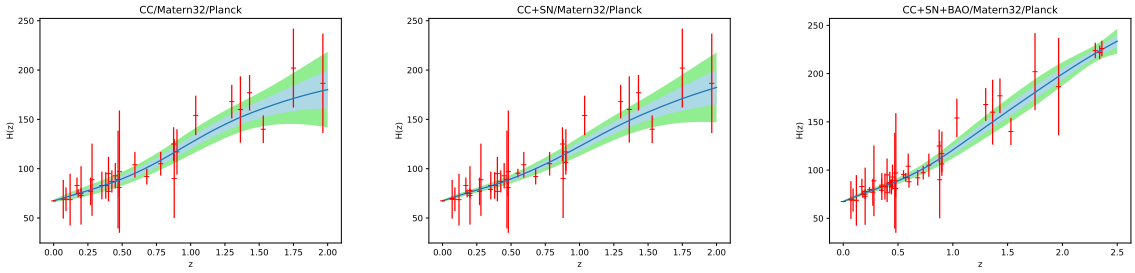

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.18: GPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{3}{2}\right)$ kernel function and using $\hat{H}_0^{HW}$



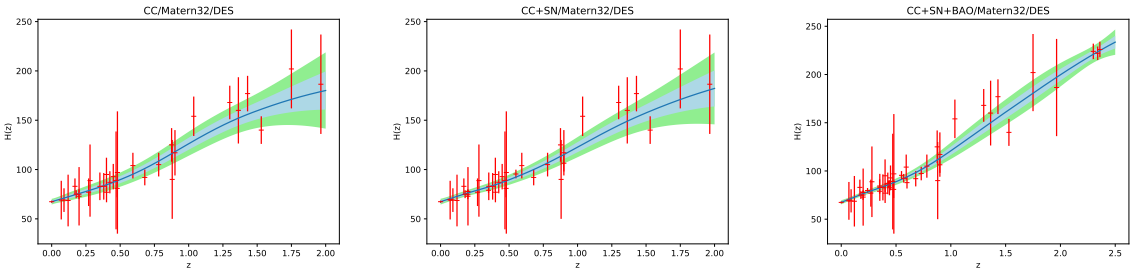

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.19: GPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{3}{2}\right)$ kernel function and using $\hat{H}_0^{CM}$



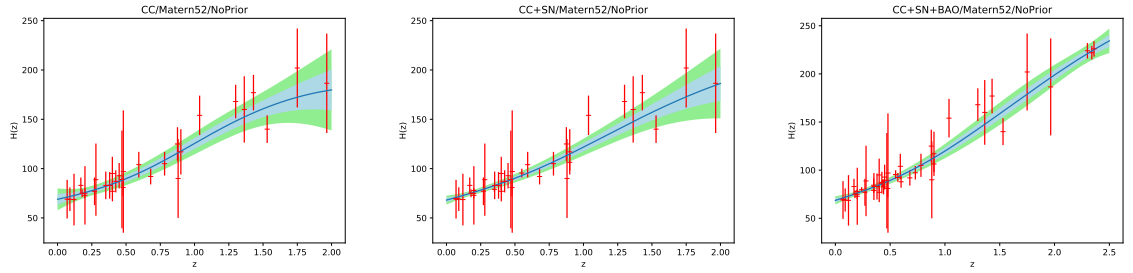

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.20: GPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{3}{2}\right)$ kernel function and using $\hat{H}_0^{P}$



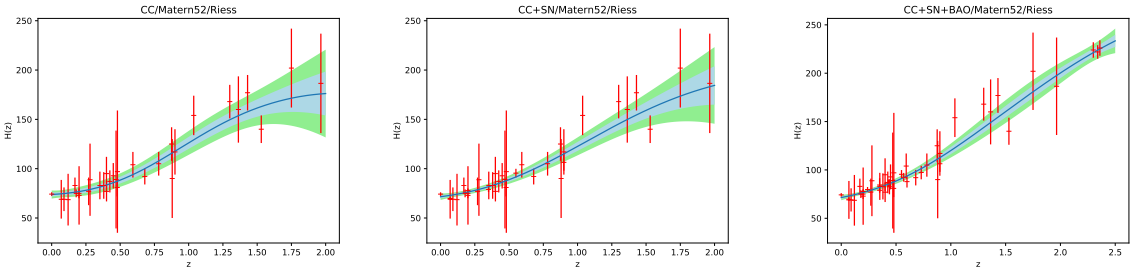

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.21: GPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{3}{2}\right)$ kernel function and using $\hat{H}_0^{DES}$



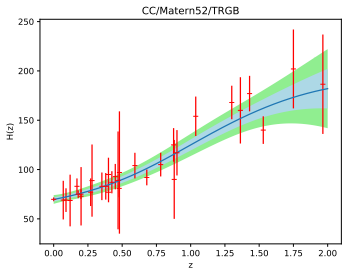

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.22: GPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{5}{2}\right)$ kernel function and without adding any $H_0$ estimate to the data

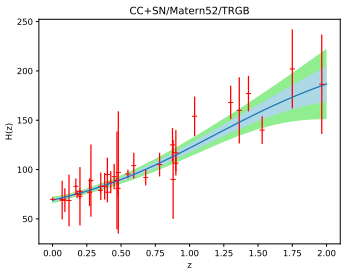



(a) CC (b) CC+SN (c) CC+SN+BAO

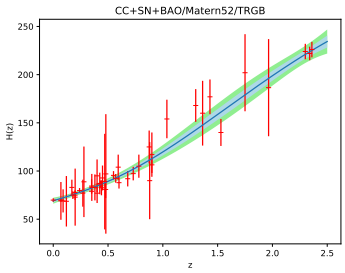
Figure B.23: GPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{5}{2}\right)$ kernel function and using $\hat{H}_0^{R}$




(a) CC

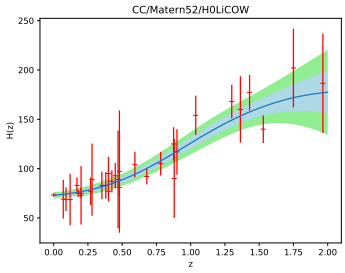



(b) CC+SN

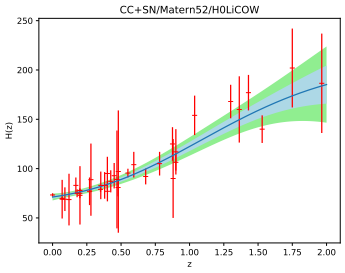



(c) CC+SN+BAO

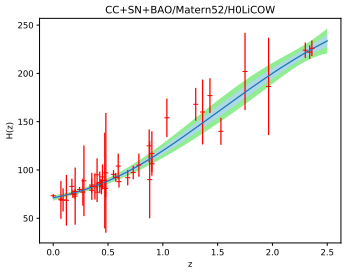
Figure B.24: GPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{5}{2}\right)$ kernel function and using $\hat{H}_0^{TRGB}$




(a) CC

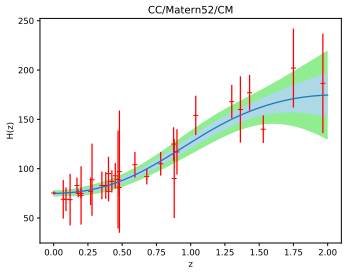



(b) CC+SN

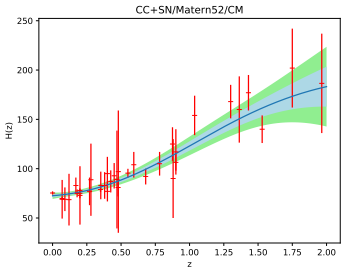



(c) CC+SN+BAO

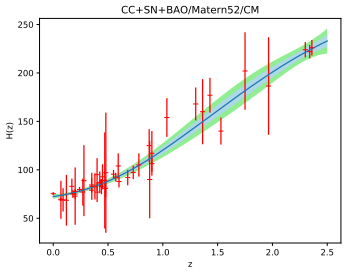
Figure B.25: GPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{5}{2}\right)$ kernel function and using $\hat{H}_0^{HW}$




(a) CC

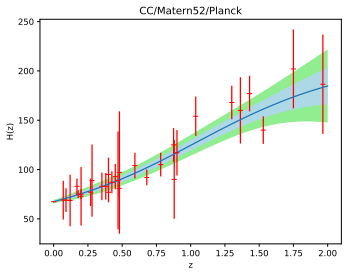



(b) CC+SN

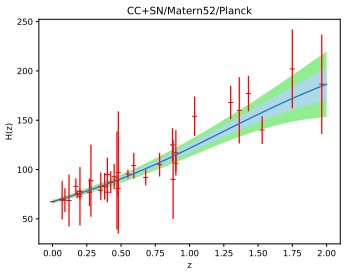



(c) CC+SN+BAO

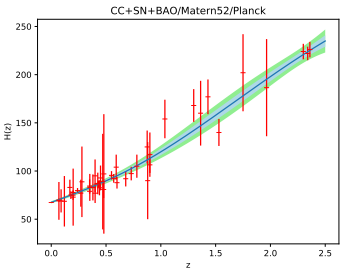
Figure B.26: GPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{5}{2}\right)$ kernel function and using $\hat{H}_0^{CM}$




(a) CC

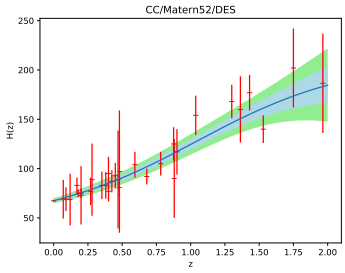



(b) CC+SN

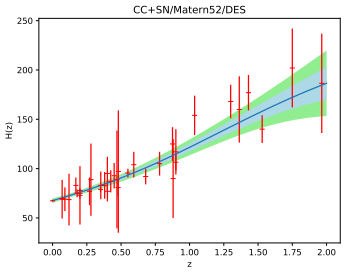



(c) CC+SN+BAO

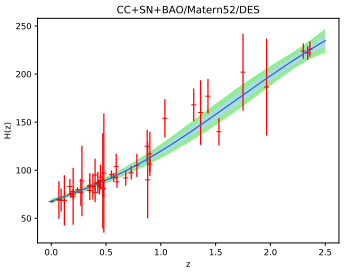
Figure B.27: GPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{5}{2}\right)$ kernel function and using $\hat{H}_0^P$




(a) CC

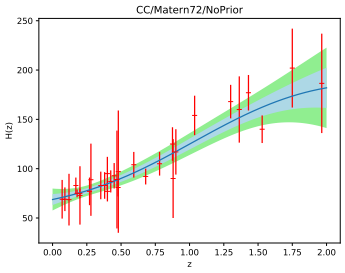



(b) CC+SN

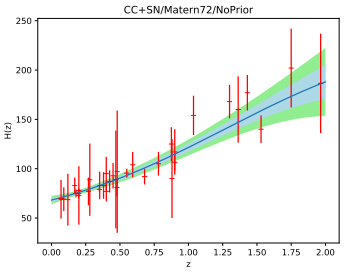



(c) CC+SN+BAO

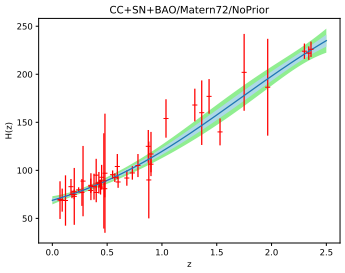
Figure B.28: GPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{5}{2}\right)$ kernel function and using $\hat{H}_0^{DES}$




(a) CC

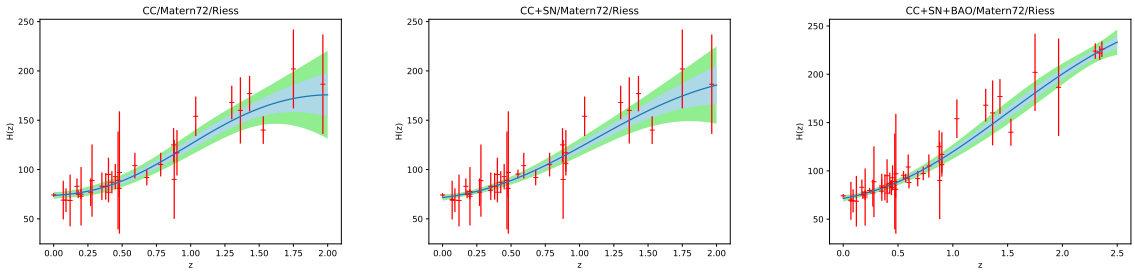



(b) CC+SN



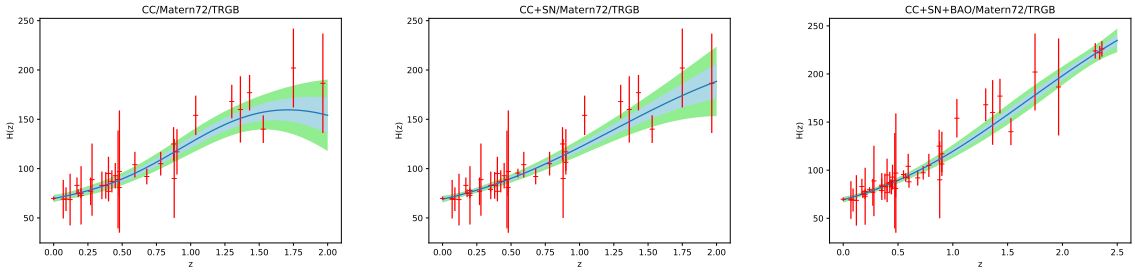

(c) CC+SN+BAO

Figure B.29: GPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{7}{2}\right)$ kernel function and without adding any $H_0$ estimate to the data



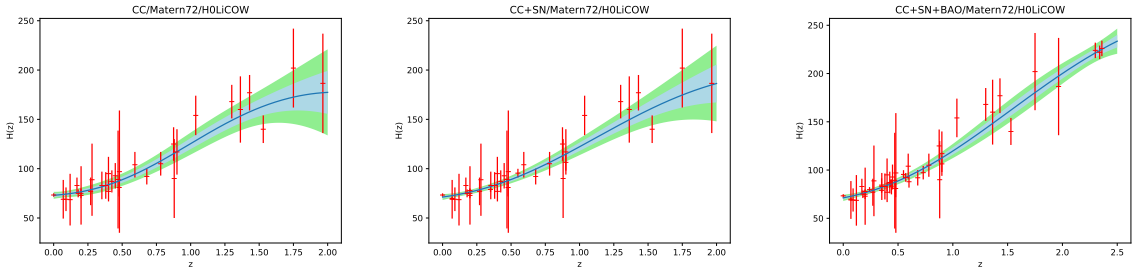

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.30: GPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{7}{2}\right)$ kernel function and using $\hat{H}_0^R$



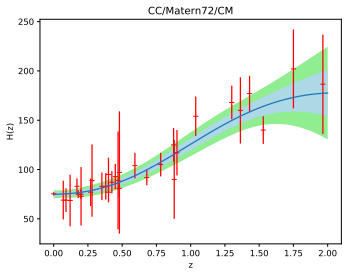

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.31: GPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{7}{2}\right)$ kernel function and using $\hat{H}_0^{TRGB}$

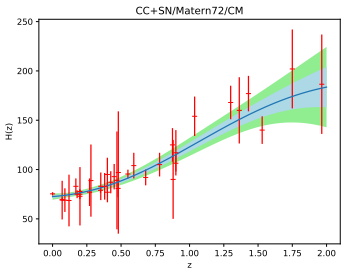



(a) CC (b) CC+SN (c) CC+SN+BAO

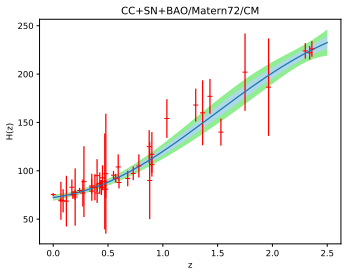
Figure B.32: GPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{7}{2}\right)$ kernel function and using $\hat{H}_0^{HW}$




(a) CC

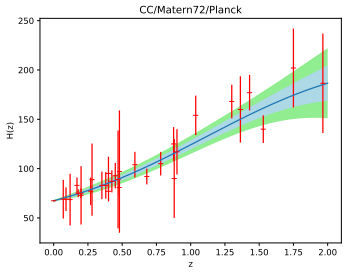



(b) CC+SN

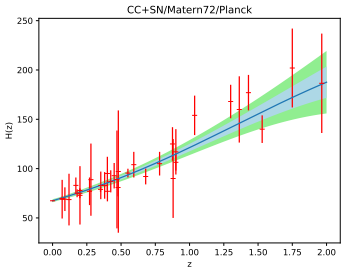



(c) CC+SN+BAO

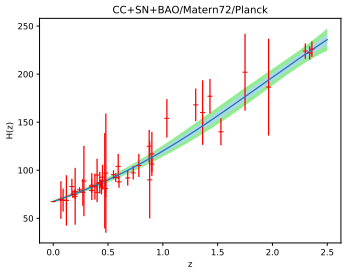
Figure B.33: GPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{7}{2}\right)$ kernel function and using $\hat{H}_0^{CM}$




(a) CC

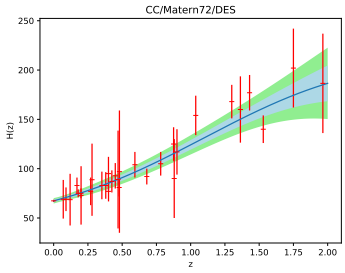



(b) CC+SN

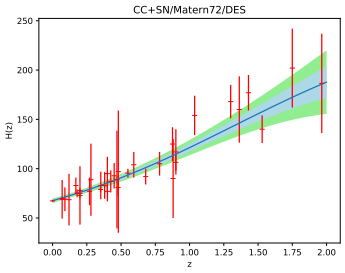



(c) CC+SN+BAO

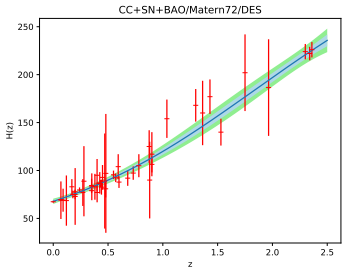
Figure B.34: GPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{7}{2}\right)$ kernel function and using $\hat{H}_0^{P}$




(a) CC

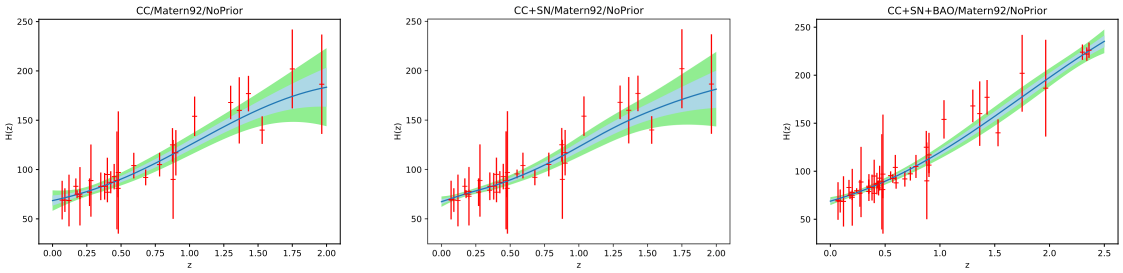



(b) CC+SN



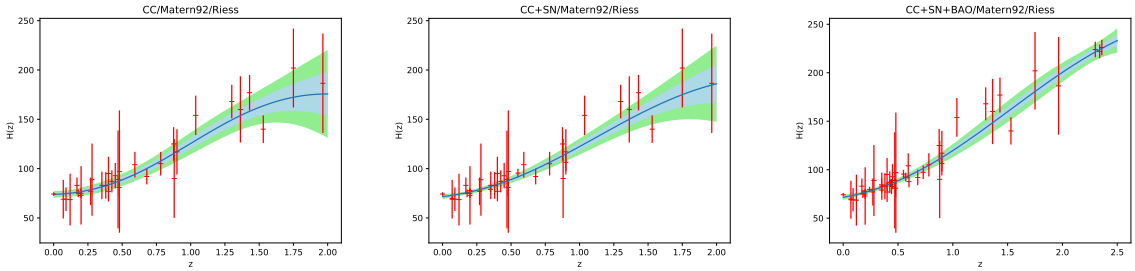

(c) CC+SN+BAO

Figure B.35: GPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{7}{2}\right)$ kernel function and using $\hat{H}_0^{DES}$



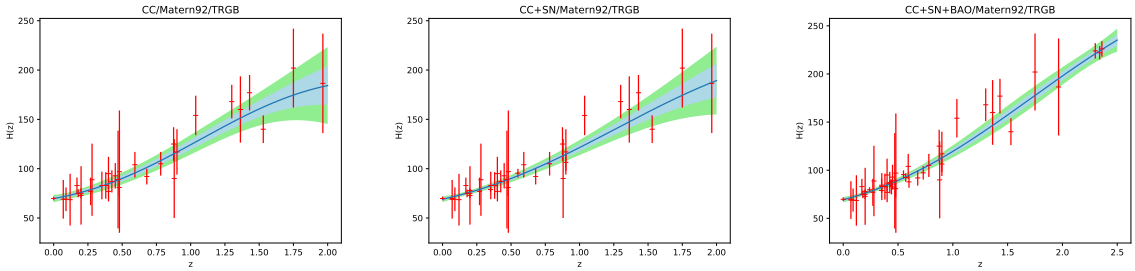

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.36: GPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{9}{2}\right)$ kernel function and without adding any $H_0$ estimate to the data



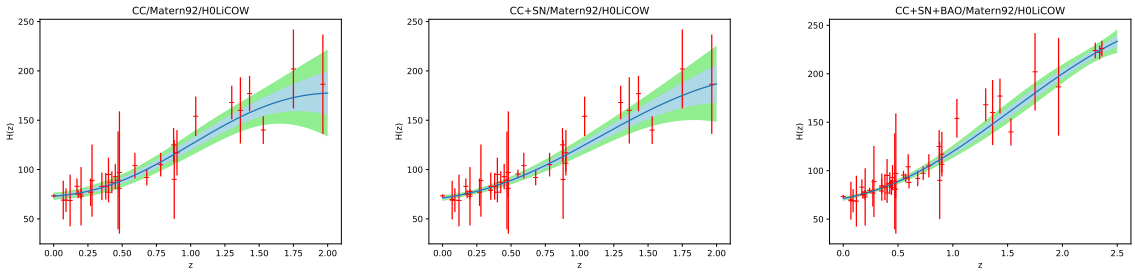

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.37: GPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{9}{2}\right)$ kernel function and using $\hat{H}_0^R$



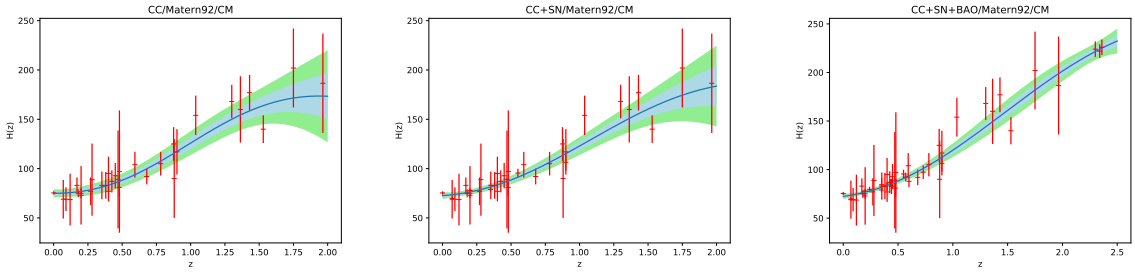

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.38: GPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{9}{2}\right)$ kernel function and using $\hat{H}_0^{TRGB}$



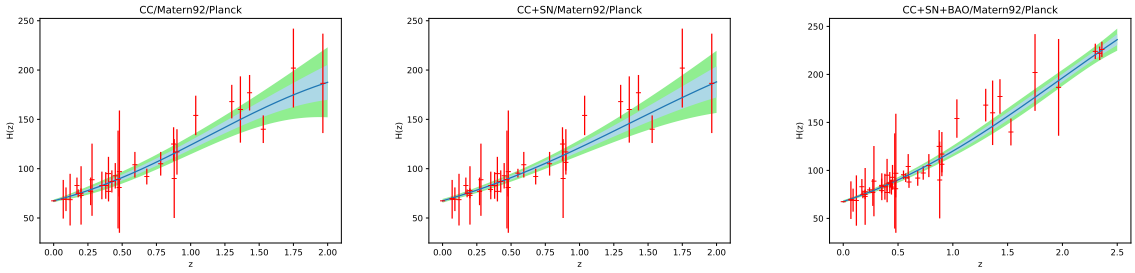

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.39: GPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{9}{2}\right)$ kernel function and using $\hat{H}_0^{HW}$



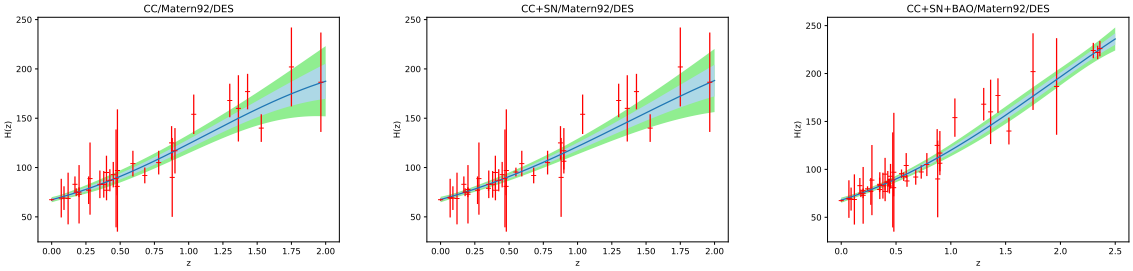

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.40: GPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{9}{2}\right)$ kernel function and using $\hat{H}_0^{CM}$



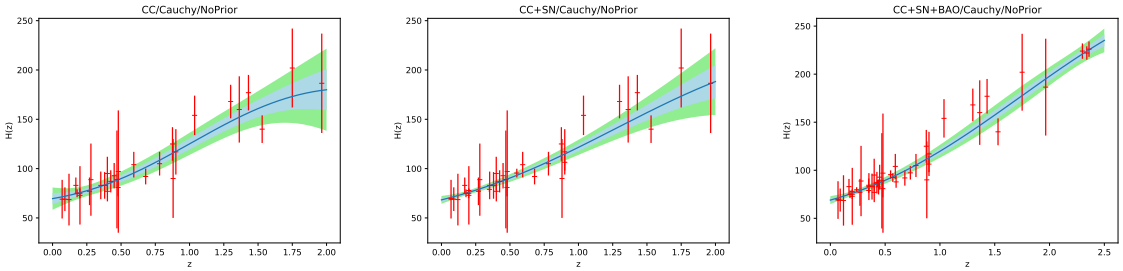

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.41: GPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{9}{2}\right)$ kernel function and using $\hat{H}_0^{P}$



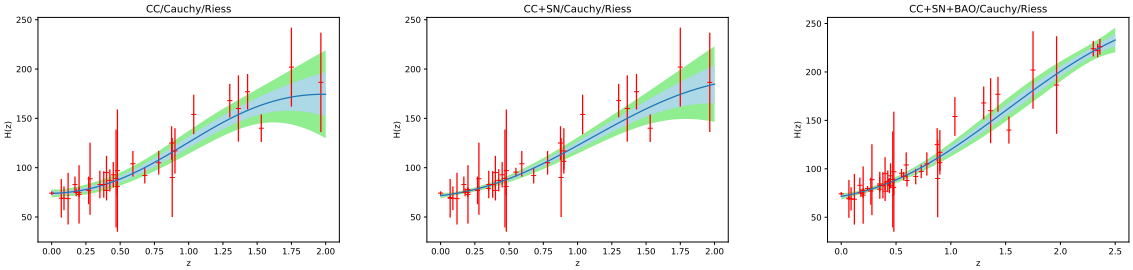

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.42: GPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{9}{2}\right)$ kernel function and using $\hat{H}_0^{DES}$



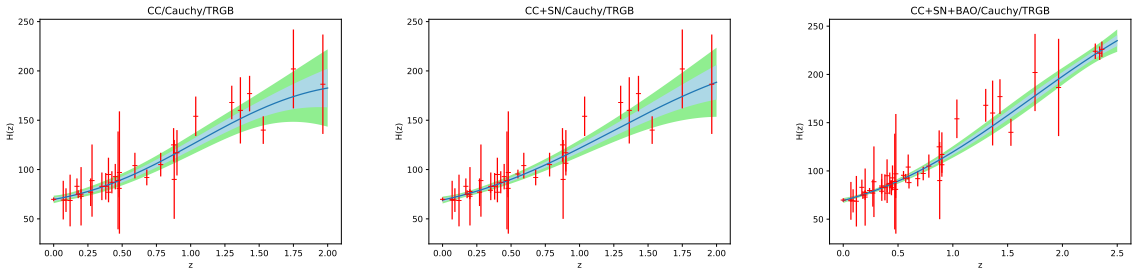

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.43: GPR reconstructions of $\hat{H}(z)$ against $z$ for Cauchy kernel function and without adding any $H_0$ estimate to the data



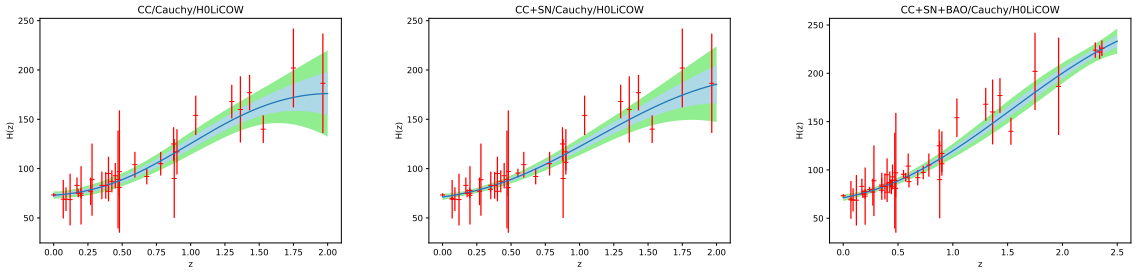

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.44: GPR reconstructions of $\hat{H}(z)$ against $z$ for Cauchy kernel function and using $\hat{H}_0^R$



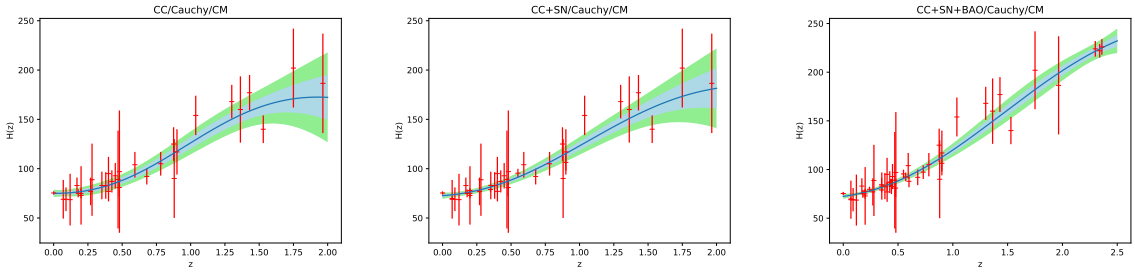

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.45: GPR reconstructions of $\hat{H}(z)$ against $z$ for Cauchy kernel function and using $\hat{H}_0^{TRGB}$



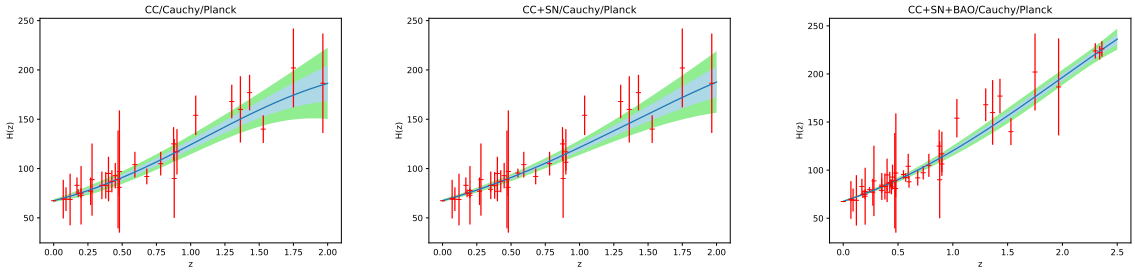

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.46: GPR reconstructions of $\hat{H}(z)$ against $z$ for Cauchy kernel function and using $\hat{H}_0^{HW}$



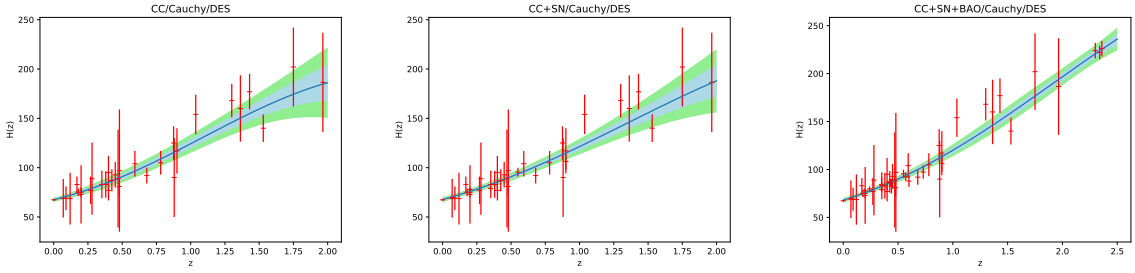

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.47: GPR reconstructions of $\hat{H}(z)$ against $z$ for Cauchy kernel function and using $\hat{H}_0^{CM}$




(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.48: GPR reconstructions of $\hat{H}(z)$ against $z$ for Cauchy kernel function and using $\hat{H}_0^P$




(a) CC (b) CC+SN (c) CC+SN+BAO

Figure B.49: GPR reconstructions of $\hat{H}(z)$ against $z$ for Cauchy kernel function and using $\hat{H}_0^{DES}$

# Appendix C: GPR Tables

The table for the square exponential kernel was presented in Section 2.4. Similar tables for all the other kernel functions considered are shown in this Appendix.

| **Dataset** | $\hat{H}_0$ (km s$^{-1}$ Mpc$^{-1}$) | $d(\hat{H}_0, \hat{H}_0^R)$ | $d(\hat{H}_0, \hat{H}_0^{TRGB})$ | $d(\hat{H}_0, \hat{H}_0^{HW})$ | $d(\hat{H}_0, \hat{H}_0^{CM})$ | $d(\hat{H}_0, \hat{H}_0^P)$ | $d(\hat{H}_0, \hat{H}_0^{DES})$ |
|---|---|---|---|---|---|---|---|
| CC | $67.448 \pm 4.753$ | $-1.3306$ | $-0.4595$ | $-1.1593$ | $-1.5675$ | 0.0100 | 0.0098 |
| CC+SN | $68.952 \pm 2.035$ | $-1.9296$ | $-0.3045$ | $-1.6397$ | $\mathbf{-2.4245}$ | 0.7408 | 0.6711 |
| CC+SN+BAO | $69.100 \pm 1.593$ | $\mathbf{-2.1170}$ | $-0.2823$ | $-1.8029$ | $\mathbf{-2.6997}$ | 1.0183 | 0.8783 |
| CC+$\hat{H}_0^R$ | $73.950 \pm 1.719$ | $-0.1078$ | 1.6198 | 0.2689 | $-0.5824$ | **3.6590** | **3.2097** |
| CC+SN+$\hat{H}_0^R$ | $71.704 \pm 1.315$ | $-1.1205$ | 0.8240 | $-0.7425$ | $-1.7089$ | **3.0590** | **2.5103** |
| CC+SN+BAO+$\hat{H}_0^R$ | $71.545 \pm 1.252$ | $-1.2110$ | 0.7666 | $-0.8314$ | $-1.8161$ | **3.0736** | **2.4865** |
| CC+$\hat{H}_0^{TRGB}$ | $69.602 \pm 1.773$ | $-1.8174$ | $-0.0762$ | $-1.5054$ | $\mathbf{-2.3532}$ | 1.1953 | 1.0554 |
| CC+SN+$\hat{H}_0^{TRGB}$ | $69.155 \pm 1.316$ | $\mathbf{-2.2552}$ | $-0.2790$ | $-1.9281$ | $\mathbf{-2.9030}$ | 1.2471 | 1.0235 |
| CC+SN+BAO+$\hat{H}_0^{TRGB}$ | $69.337 \pm 1.219$ | $\mathbf{-2.2293}$ | $-0.2051$ | $-1.8946$ | $\mathbf{-2.8972}$ | 1.4705 | 1.1799 |
| CC+$\hat{H}_0^{HW}$ | $73.106 \pm 1.614$ | $-0.4580$ | 1.3260 | $-0.0828$ | $-0.9632$ | **3.3765** | **2.9211** |
| CC+SN+$\hat{H}_0^{HW}$ | $71.353 \pm 1.267$ | $-1.2930$ | 0.6798 | $-0.9185$ | $-1.8997$ | **2.9017** | **2.3556** |
| CC+SN+BAO+$\hat{H}_0^{HW}$ | $71.189 \pm 1.201$ | $-1.3902$ | 0.6180 | $-1.0144$ | $\mathbf{-2.0152}$ | **2.9133** | **2.3269** |
| CC+$\hat{H}_0^{CM}$ | $75.063 \pm 1.602$ | 0.3477 | **2.1176** | 0.7547 | $-0.1236$ | **4.5659** | **3.9431** |
| CC+SN+$\hat{H}_0^{CM}$ | $72.686 \pm 1.289$ | $-0.6878$ | 1.2571 | $-0.2877$ | $-1.2580$ | **3.8240** | **3.1198** |
| CC+SN+BAO+$\hat{H}_0^{CM}$ | $73.235 \pm 1.497$ | $-0.4182$ | 1.4199 | $-0.0289$ | $-0.9401$ | **3.6970** | **3.1409** |
| CC+$\hat{H}_0^P$ | $67.393 \pm 0.497$ | $\mathbf{-3.6188}$ | $-1.2258$ | $\mathbf{-3.3353}$ | $\mathbf{-4.5419}$ | $-0.0105$ | $-0.0061$ |
| CC+SN+$\hat{H}_0^P$ | $67.448 \pm 0.481$ | $\mathbf{-3.5976}$ | $-1.2001$ | $\mathbf{-3.3125}$ | $\mathbf{-4.5222}$ | 0.0692 | 0.0400 |
| CC+SN+BAO+$\hat{H}_0^P$ | $67.513 \pm 0.476$ | $\mathbf{-3.5649}$ | $-1.1673$ | $\mathbf{-3.2777}$ | $\mathbf{-4.4878}$ | 0.1644 | 0.0947 |
| CC+$\hat{H}_0^{DES}$ | $67.366 \pm 1.070$ | $\mathbf{-3.2464}$ | $-1.1162$ | $\mathbf{-2.9541}$ | $\mathbf{-4.0084}$ | $-0.0287$ | $-0.0221$ |
| CC+SN+$\hat{H}_0^{DES}$ | $67.631 \pm 0.993$ | $\mathbf{-3.1779}$ | $-1.0117$ | $\mathbf{-2.8793}$ | $\mathbf{-3.9551}$ | 0.2077 | 0.1559 |
| CC+SN+BAO+$\hat{H}_0^{DES}$ | $67.827 \pm 0.895$ | $\mathbf{-3.1525}$ | $-0.9397$ | $\mathbf{-2.8491}$ | $\mathbf{-3.9525}$ | 0.4162 | 0.3008 |

Table C.1: GPR results with double square exponential kernel function

| **Dataset** | $\hat{H}_0$ (km s$^{-1}$ Mpc$^{-1}$) | $d(\hat{H}_0, \hat{H}_0^R)$ | $d(\hat{H}_0, \hat{H}_0^{TRGB})$ | $d(\hat{H}_0, \hat{H}_0^{HW})$ | $d(\hat{H}_0, \hat{H}_0^{CM})$ | $d(\hat{H}_0, \hat{H}_0^P)$ | $d(\hat{H}_0, \hat{H}_0^{DES})$ |
|---|---|---|---|---|---|---|---|
| CC | $71.023 \pm 5.647$ | $-0.5388$ | $0.2053$ | $-0.3861$ | $-0.7344$ | $0.6392$ | $0.6298$ |
| CC+SN | $68.353 \pm 1.846$ | $\mathbf{-2.2630}$ | $-0.5461$ | $-1.9711$ | $\mathbf{-2.8030}$ | $0.4984$ | $0.4436$ |
| CC+SN+BAO | $68.939 \pm 1.656$ | $\mathbf{-2.1462}$ | $-0.3416$ | $-1.8376$ | $\mathbf{-2.7177}$ | $0.8896$ | $0.7741$ |
| CC+$\hat{H}_0^R$ | $73.949 \pm 1.734$ | $-0.1077$ | $1.6128$ | $0.2673$ | $-0.5801$ | $\mathbf{3.6282}$ | $\mathbf{3.1887}$ |
| CC+SN+$\hat{H}_0^R$ | $71.706 \pm 1.351$ | $-1.1092$ | $0.8175$ | $-0.7341$ | $-1.6904$ | $\mathbf{2.9892}$ | $\mathbf{2.4716}$ |
| CC+SN+BAO+$\hat{H}_0^R$ | $71.560 \pm 1.265$ | $-1.1998$ | $0.7711$ | $-0.8209$ | $-1.8018$ | $\mathbf{3.0576}$ | $\mathbf{2.4812}$ |
| CC+$\hat{H}_0^{TRGB}$ | $69.796 \pm 1.783$ | $-1.7366$ | $-0.0016$ | $-1.4225$ | $\mathbf{-2.2674}$ | $1.2940$ | $1.1438$ |
| CC+SN+$\hat{H}_0^{TRGB}$ | $69.107 \pm 1.334$ | $\mathbf{-2.2656}$ | $-0.2983$ | $-1.9401$ | $\mathbf{-2.9099}$ | $1.1984$ | $0.9874$ |
| CC+SN+BAO+$\hat{H}_0^{TRGB}$ | $69.314 \pm 1.249$ | $\mathbf{-2.2225}$ | $-0.2137$ | $-1.8895$ | $\mathbf{-2.8833}$ | $1.4227$ | $1.1500$ |
| CC+$\hat{H}_0^{HW}$ | $73.111 \pm 1.627$ | $-0.4545$ | $1.3235$ | $-0.0805$ | $-0.9576$ | $\mathbf{3.3550}$ | $\mathbf{2.9077}$ |
| CC+SN+$\hat{H}_0^{HW}$ | $71.328 \pm 1.294$ | $-1.2949$ | $0.6649$ | $-0.9228$ | $-1.8965$ | $\mathbf{2.8320}$ | $\mathbf{2.3132}$ |
| CC+SN+BAO+$\hat{H}_0^{HW}$ | $71.238 \pm 1.218$ | $-1.3616$ | $0.6372$ | $-0.9859$ | $-1.9815$ | $\mathbf{2.9151}$ | $\mathbf{2.3386}$ |
| CC+$\hat{H}_0^{CM}$ | $75.058 \pm 1.615$ | $0.3444$ | $\mathbf{2.1085}$ | $0.7497$ | $-0.1253$ | $\mathbf{4.5297}$ | $\mathbf{3.9191}$ |
| CC+SN+$\hat{H}_0^{CM}$ | $72.668 \pm 1.303$ | $-0.6934$ | $1.2447$ | $-0.2951$ | $-1.2615$ | $\mathbf{3.7740}$ | $\mathbf{3.0890}$ |
| CC+SN+BAO+$\hat{H}_0^{CM}$ | $72.424 \pm 1.227$ | $-0.8183$ | $1.1600$ | $-0.4179$ | $-1.4065$ | $\mathbf{3.7912}$ | $\mathbf{3.0484}$ |
| CC+$\hat{H}_0^P$ | $67.409 \pm 0.497$ | $\mathbf{-3.6102}$ | $-1.2176$ | $\mathbf{-3.3262}$ | $\mathbf{-4.5326}$ | $0.0122$ | $0.0071$ |
| CC+SN+$\hat{H}_0^P$ | $67.445 \pm 0.481$ | $\mathbf{-3.5989}$ | $-1.2015$ | $\mathbf{-3.3139}$ | $\mathbf{-4.5235}$ | $0.0651$ | $0.0376$ |
| CC+SN+BAO+$\hat{H}_0^P$ | $67.518 \pm 0.476$ | $\mathbf{-3.5623}$ | $-1.1649$ | $\mathbf{-3.2749}$ | $\mathbf{-4.4850}$ | $0.1712$ | $0.0987$ |
| CC+$\hat{H}_0^{DES}$ | $67.443 \pm 1.072$ | $\mathbf{-3.2087}$ | $-1.0807$ | $\mathbf{-2.9145}$ | $\mathbf{-3.9679}$ | $0.0360$ | $0.0277$ |
| CC+SN+$\hat{H}_0^{DES}$ | $67.587 \pm 0.935$ | $\mathbf{-3.2414}$ | $-1.0448$ | $\mathbf{-2.9442}$ | $\mathbf{-4.0372}$ | $0.1768$ | $0.1298$ |
| CC+SN+BAO+$\hat{H}_0^{DES}$ | $67.837 \pm 0.904$ | $\mathbf{-3.1411}$ | $-0.9329$ | $\mathbf{-2.8374}$ | $\mathbf{-3.9383}$ | $0.4235$ | $0.3072$ |

Table C.2: GPR results with rational quadratic kernel function

| **Dataset** | $\hat{H}_0$ (km s$^{-1}$ Mpc$^{-1}$) | $d(\hat{H}_0, \hat{H}_0^R)$ | $d(\hat{H}_0, \hat{H}_0^{TRGB})$ | $d(\hat{H}_0, \hat{H}_0^{HW})$ | $d(\hat{H}_0, \hat{H}_0^{CM})$ | $d(\hat{H}_0, \hat{H}_0^P)$ | $d(\hat{H}_0, \hat{H}_0^{DES})$ |
|---|---|---|---|---|---|---|---|
| CC | $68.798 \pm 6.340$ | $-0.8220$ | $-0.1514$ | $-0.6858$ | $-0.9989$ | 0.2198 | 0.2173 |
| CC+SN | $67.394 \pm 2.344$ | $\mathbf{-2.3002}$ | $-0.7973$ | $\mathbf{-2.0397}$ | $\mathbf{-2.7589}$ | $-0.0023$ | $-0.0021$ |
| CC+SN+BAO | $67.519 \pm 2.200$ | $\mathbf{-2.3467}$ | $-0.7846$ | $\mathbf{-2.0791}$ | $\mathbf{-2.8287}$ | 0.0528 | 0.0484 |
| CC+$\hat{H}_0^R$ | $73.820 \pm 1.751$ | $-0.1585$ | 1.5558 | 0.2129 | $-0.6307$ | **3.5255** | **3.1046** |
| CC+SN+$\hat{H}_0^R$ | $71.741 \pm 1.458$ | $-1.0631$ | 0.8104 | $-0.6962$ | $-1.6225$ | **2.8163** | **2.3768** |
| CC+SN+BAO+$\hat{H}_0^R$ | $71.519 \pm 1.408$ | $-1.1739$ | 0.7269 | $-0.8069$ | $-1.7479$ | **2.7571** | **2.3055** |
| CC+$\hat{H}_0^{TRGB}$ | $69.673 \pm 1.808$ | $-1.7726$ | $-0.0484$ | $-1.4616$ | $\mathbf{-2.3004}$ | 1.2119 | 1.0742 |
| CC+SN+$\hat{H}_0^{TRGB}$ | $68.834 \pm 1.454$ | $\mathbf{-2.3119}$ | $-0.4038$ | $-1.9962$ | $\mathbf{-2.9324}$ | 0.9322 | 0.7863 |
| CC+SN+BAO+$\hat{H}_0^{TRGB}$ | $68.828 \pm 1.419$ | $\mathbf{-2.3364}$ | $-0.4100$ | $\mathbf{-2.0195}$ | $\mathbf{-2.9657}$ | 0.9487 | 0.7950 |
| CC+$\hat{H}_0^{HW}$ | $72.992 \pm 1.641$ | $-0.5013$ | 1.2712 | $-0.1306$ | $-1.0042$ | **3.2593** | **2.8302** |
| CC+SN+$\hat{H}_0^{HW}$ | $71.301 \pm 1.385$ | $-1.2760$ | 0.6385 | $-0.9114$ | $-1.8593$ | **2.6489** | **2.2054** |
| CC+SN+BAO+$\hat{H}_0^{HW}$ | $71.138 \pm 1.345$ | $-1.3620$ | 0.5749 | $-0.9975$ | $-1.9574$ | **2.6059** | **2.1518** |
| CC+$\hat{H}_0^{CM}$ | $74.958 \pm 1.628$ | 0.3024 | **2.0617** | 0.7046 | $-0.1674$ | **4.4383** | **3.8470** |
| CC+SN+$\hat{H}_0^{CM}$ | $72.912 \pm 1.410$ | $-0.5680$ | 1.3156 | $-0.1755$ | $-1.1115$ | **3.6854** | **3.0829** |
| CC+SN+BAO+$\hat{H}_0^{CM}$ | $72.579 \pm 1.357$ | $-0.7230$ | 1.1903 | $-0.3316$ | $-1.2833$ | **3.5820** | **2.9652** |
| CC+$\hat{H}_0^P$ | $67.406 \pm 0.498$ | $\mathbf{-3.6110}$ | $-1.2186$ | $\mathbf{-3.3270}$ | $\mathbf{-4.5333}$ | 0.0091 | 0.0053 |
| CC+SN+$\hat{H}_0^P$ | $67.404 \pm 0.488$ | $\mathbf{-3.6173}$ | $-1.2214$ | $\mathbf{-3.3336}$ | $\mathbf{-4.5421}$ | 0.0058 | 0.0034 |
| CC+SN+BAO+$\hat{H}_0^P$ | $67.410 \pm 0.487$ | $\mathbf{-3.6146}$ | $-1.2185$ | $\mathbf{-3.3307}$ | $\mathbf{-4.5393}$ | 0.0145 | 0.0084 |
| CC+$\hat{H}_0^{DES}$ | $67.430 \pm 1.080$ | $\mathbf{-3.2085}$ | $-1.0843$ | $\mathbf{-2.9146}$ | $\mathbf{-3.9657}$ | 0.0256 | 0.0198 |
| CC+SN+$\hat{H}_0^{DES}$ | $67.414 \pm 0.988$ | $\mathbf{-3.2864}$ | $-1.1140$ | $\mathbf{-2.9934}$ | $\mathbf{-4.0718}$ | 0.0131 | 0.0098 |
| CC+SN+BAO+$\hat{H}_0^{DES}$ | $67.438 \pm 0.979$ | $\mathbf{-3.2816}$ | $-1.1049$ | $\mathbf{-2.9880}$ | $\mathbf{-4.0689}$ | 0.0349 | 0.0260 |

Table C.3: GPR results with Matérn $\left(\nu = \frac{3}{2}\right)$ kernel function

| **Dataset** | $\hat{H}_0$ (km s$^{-1}$ Mpc$^{-1}$) | $d(\hat{H}_0, \hat{H}_0^R)$ | $d(\hat{H}_0, \hat{H}_0^{TRGB})$ | $d(\hat{H}_0, \hat{H}_0^{HW})$ | $d(\hat{H}_0, \hat{H}_0^{CM})$ | $d(\hat{H}_0, \hat{H}_0^P)$ | $d(\hat{H}_0, \hat{H}_0^{DES})$ |
|---|---|---|---|---|---|---|---|
| CC | $68.901 \pm 5.437$ | $-0.9277$ | $-0.1560$ | $-0.7722$ | $-1.1332$ | 0.2750 | 0.2706 |
| CC+SN | $68.076 \pm 1.929$ | $\mathbf{-2.3165}$ | $-0.6367$ | $\mathbf{-2.0315}$ | $\mathbf{-2.8433}$ | 0.3392 | 0.3044 |
| CC+SN+BAO | $68.617 \pm 1.788$ | $\mathbf{-2.1959}$ | $-0.4534$ | $-1.8979$ | $\mathbf{-2.7440}$ | 0.6554 | 0.5796 |
| CC+$\hat{H}_0^R$ | $73.806 \pm 1.738$ | $-0.1644$ | 1.5560 | 0.2083 | $-0.6387$ | **3.5431** | **3.1151** |
| CC+SN+$\hat{H}_0^R$ | $71.574 \pm 1.368$ | $-1.1622$ | 0.7575 | $-0.7910$ | $-1.7428$ | **2.8646** | **2.3771** |
| CC+SN+BAO+$\hat{H}_0^R$ | $71.464 \pm 1.299$ | $-1.2327$ | 0.7231 | $-0.8582$ | $-1.8301$ | **2.9207** | **2.3881** |
| CC+$\hat{H}_0^{TRGB}$ | $69.646 \pm 1.786$ | $-1.7938$ | $-0.0589$ | $-1.4819$ | $\mathbf{-2.3263}$ | 1.2114 | 1.0711 |
| CC+SN+$\hat{H}_0^{TRGB}$ | $68.962 \pm 1.358$ | $\mathbf{-2.3158}$ | $-0.3589$ | $-1.9940$ | $\mathbf{-2.9575}$ | 1.0797 | 0.8940 |
| CC+SN+BAO+$\hat{H}_0^{TRGB}$ | $69.166 \pm 1.297$ | $\mathbf{-2.2613}$ | $-0.2755$ | $-1.9332$ | $\mathbf{-2.9135}$ | 1.2707 | 1.0386 |
| CC+$\hat{H}_0^{HW}$ | $72.981 \pm 1.629$ | $-0.5074$ | 1.2707 | $-0.1356$ | $-1.0124$ | **3.2741** | **2.8385** |
| CC+SN+$\hat{H}_0^{HW}$ | $71.200 \pm 1.309$ | $-1.3469$ | 0.6068 | $-0.9786$ | $-1.9483$ | **2.7111** | **2.2220** |
| CC+SN+BAO+$\hat{H}_0^{HW}$ | $71.140 \pm 1.248$ | $-1.3957$ | 0.5894 | $-1.0243$ | $\mathbf{-2.0116}$ | **2.7813** | **2.2478** |
| CC+$\hat{H}_0^{CM}$ | $74.943 \pm 1.618$ | 0.2971 | **2.0611** | 0.7003 | $-0.1743$ | **4.4546** | **3.8557** |
| CC+SN+$\hat{H}_0^{CM}$ | $72.574 \pm 1.320$ | $-0.7320$ | 1.1992 | $-0.3371$ | $-1.2992$ | **3.6661** | **3.0116** |
| CC+SN+BAO+$\hat{H}_0^{CM}$ | $72.367 \pm 1.257$ | $-0.8381$ | 1.1267 | $-0.4416$ | $-1.4221$ | **3.6724** | **2.9740** |
| CC+$\hat{H}_0^P$ | $67.402 \pm 0.497$ | $\mathbf{-3.6136}$ | $-1.2210$ | $\mathbf{-3.3298}$ | $\mathbf{-4.5363}$ | 0.0028 | 0.0017 |
| CC+SN+$\hat{H}_0^P$ | $67.436 \pm 0.483$ | $\mathbf{-3.6027}$ | $-1.2057$ | $\mathbf{-3.3180}$ | $\mathbf{-4.5274}$ | 0.0525 | 0.0304 |
| CC+SN+BAO+$\hat{H}_0^P$ | $67.491 \pm 0.481$ | $\mathbf{-3.5747}$ | $-1.1781$ | $\mathbf{-3.2882}$ | $\mathbf{-4.4976}$ | 0.1313 | 0.0759 |
| CC+$\hat{H}_0^{DES}$ | $67.411 \pm 1.073$ | $\mathbf{-3.2229}$ | $-1.0949$ | $\mathbf{-2.9295}$ | $\mathbf{-3.9827}$ | 0.0092 | 0.0071 |
| CC+SN+$\hat{H}_0^{DES}$ | $67.547 \pm 0.947$ | $\mathbf{-3.2527}$ | $-1.0614$ | $\mathbf{-2.9565}$ | $\mathbf{-4.0463}$ | 0.1372 | 0.1012 |
| CC+SN+BAO+$\hat{H}_0^{DES}$ | $67.740 \pm 0.931$ | $\mathbf{-3.1698}$ | $-0.9736$ | $\mathbf{-2.8686}$ | $\mathbf{-3.9621}$ | 0.3218 | 0.2360 |

Table C.4: GPR results with Matérn $\left(\nu = \frac{5}{2}\right)$ kernel function

| **Dataset** | $\hat{H}_0$ (km s$^{-1}$ Mpc$^{-1}$) | $d(\hat{H}_0, \hat{H}_0^R)$ | $d(\hat{H}_0, \hat{H}_0^{TRGB})$ | $d(\hat{H}_0, \hat{H}_0^{HW})$ | $d(\hat{H}_0, \hat{H}_0^{CM})$ | $d(\hat{H}_0, \hat{H}_0^P)$ | $d(\hat{H}_0, \hat{H}_0^{DES})$ |
|---|---|---|---|---|---|---|---|
| CC | $68.734 \pm 5.181$ | $-0.9990$ | $-0.1931$ | $-0.8373$ | $-1.2147$ | 0.2564 | 0.2519 |
| CC+SN | $68.156 \pm 1.851$ | $\mathbf{-2.3360}$ | $-0.6199$ | $\mathbf{-2.0468}$ | $\mathbf{-2.8779}$ | 0.3940 | 0.3509 |
| CC+SN+BAO | $68.809 \pm 1.690$ | $\mathbf{-2.1788}$ | $-0.3897$ | $-1.8736$ | $\mathbf{-2.7451}$ | 0.7996 | 0.6988 |
| CC+$\hat{H}_0^R$ | $73.820 \pm 1.734$ | $-0.1592$ | 1.5628 | 0.2140 | $-0.6339$ | **3.5577** | **3.1265** |
| CC+SN+$\hat{H}_0^R$ | $71.600 \pm 1.352$ | $-1.1555$ | 0.7721 | $-0.7825$ | $-1.7389$ | **2.9141** | **2.4100** |
| CC+SN+BAO+$\hat{H}_0^R$ | $71.500 \pm 1.274$ | $-1.2244$ | 0.7432 | $-0.8473$ | $-1.8261$ | **2.9963** | **2.4362** |
| CC+$\hat{H}_0^{TRGB}$ | $69.638 \pm 1.780$ | $-1.8000$ | $-0.0623$ | $-1.4879$ | $\mathbf{-2.3339}$ | 1.2104 | 1.0695 |
| CC+SN+$\hat{H}_0^{TRGB}$ | $69.002 \pm 1.336$ | $\mathbf{-2.3110}$ | $-0.3434$ | $-1.9876$ | $\mathbf{-2.9572}$ | 1.1232 | 0.9258 |
| CC+SN+BAO+$\hat{H}_0^{TRGB}$ | $69.248 \pm 1.263$ | $\mathbf{-2.2445}$ | $-0.2421$ | $-1.9135$ | $\mathbf{-2.9034}$ | 1.3603 | 1.1032 |
| CC+$\hat{H}_0^{HW}$ | $72.991 \pm 1.626$ | $-0.5037$ | 1.2758 | $-0.1315$ | $-1.0090$ | **3.2862** | **2.8477** |
| CC+SN+$\hat{H}_0^{HW}$ | $71.229 \pm 1.295$ | $-1.3393$ | 0.6214 | $-0.9693$ | $-1.9431$ | **2.7588** | **2.2537** |
| CC+SN+BAO+$\hat{H}_0^{HW}$ | $71.180 \pm 1.226$ | $-1.3856$ | 0.6102 | $-1.0117$ | $\mathbf{-2.0054}$ | **2.8554** | **2.2951** |
| CC+$\hat{H}_0^{CM}$ | $74.956 \pm 1.615$ | 0.3023 | **2.0675** | 0.7061 | $-0.1692$ | **4.4692** | **3.8668** |
| CC+SN+$\hat{H}_0^{CM}$ | $72.583 \pm 1.305$ | $-0.7311$ | 1.2072 | $-0.3347$ | $-1.3008$ | **3.7084** | **3.0365** |
| CC+SN+BAO+$\hat{H}_0^{CM}$ | $72.377 \pm 1.235$ | $-0.8380$ | 1.1374 | $-0.4393$ | $-1.4259$ | **3.7365** | **3.0099** |
| CC+$\hat{H}_0^P$ | $67.399 \pm 0.497$ | $\mathbf{-3.6153}$ | $-1.2225$ | $\mathbf{-3.3315}$ | $\mathbf{-4.5381}$ | $-0.0013$ | $-0.0008$ |
| CC+SN+$\hat{H}_0^P$ | $67.438 \pm 0.481$ | $\mathbf{-3.6026}$ | $-1.2052$ | $\mathbf{-3.3179}$ | $\mathbf{-4.5275}$ | 0.0545 | 0.0315 |
| CC+SN+BAO+$\hat{H}_0^P$ | $67.509 \pm 0.478$ | $\mathbf{-3.5664}$ | $-1.1693$ | $\mathbf{-3.2793}$ | $\mathbf{-4.4892}$ | 0.1580 | 0.0911 |
| CC+$\hat{H}_0^{DES}$ | $67.397 \pm 1.071$ | $\mathbf{-3.2306}$ | $-1.1016$ | $\mathbf{-2.9375}$ | $\mathbf{-3.9912}$ | $-0.0024$ | $-0.0018$ |
| CC+SN+$\hat{H}_0^{DES}$ | $67.555 \pm 0.937$ | $\mathbf{-3.2555}$ | $-1.0595$ | $\mathbf{-2.9592}$ | $\mathbf{-4.0517}$ | 0.1461 | 0.1074 |
| CC+SN+BAO+$\hat{H}_0^{DES}$ | $67.804 \pm 0.912$ | $\mathbf{-3.1517}$ | $-0.9471$ | $\mathbf{-2.8489}$ | $\mathbf{-3.9475}$ | 0.3882 | 0.2826 |

Table C.5: GPR results with Matérn $\left(\nu = \frac{7}{2}\right)$ kernel function

| **Dataset** | $\hat{H}_0$ (km s$^{-1}$ Mpc$^{-1}$) | $d(\hat{H}_0, \hat{H}_0^R)$ | $d(\hat{H}_0, \hat{H}_0^{TRGB})$ | $d(\hat{H}_0, \hat{H}_0^{HW})$ | $d(\hat{H}_0, \hat{H}_0^{CM})$ | $d(\hat{H}_0, \hat{H}_0^P)$ | $d(\hat{H}_0, \hat{H}_0^{DES})$ |
|---|---|---|---|---|---|---|---|
| CC | $68.528 \pm 5.060$ | $-1.0585$ | $-0.2353$ | $-0.8940$ | $-1.2795$ | 0.2218 | 0.2178 |
| CC+SN | $68.160 \pm 1.825$ | $\mathbf{-2.3510}$ | $-0.6224$ | $\mathbf{-2.0607}$ | $\mathbf{-2.8984}$ | 0.4017 | 0.3567 |
| CC+SN+BAO | $68.854 \pm 1.656$ | $\mathbf{-2.1809}$ | $-0.3753$ | $-1.8735$ | $\mathbf{-2.7539}$ | 0.8408 | 0.7316 |
| CC+$\hat{H}_0^R$ | $73.826 \pm 1.732$ | $-0.1568$ | 1.5659 | 0.2168 | $-0.6315$ | **3.5644** | **3.1318** |
| CC+SN+$\hat{H}_0^R$ | $71.617 \pm 1.346$ | $-1.1501$ | 0.7803 | $-0.7763$ | $-1.7343$ | **2.9372** | **2.4261** |
| CC+SN+BAO+$\hat{H}_0^R$ | $71.516 \pm 1.265$ | $-1.2201$ | 0.7517 | $-0.8420$ | $-1.8233$ | **3.0262** | **2.4554** |
| CC+$\hat{H}_0^{TRGB}$ | $69.631 \pm 1.778$ | $-1.8040$ | $-0.0651$ | $-1.4919$ | $\mathbf{-2.3384}$ | 1.2080 | 1.0671 |
| CC+SN+$\hat{H}_0^{TRGB}$ | $69.013 \pm 1.328$ | $\mathbf{-2.3111}$ | $-0.3396$ | $-1.9873$ | $\mathbf{-2.9591}$ | 1.1363 | 0.9351 |
| CC+SN+BAO+$\hat{H}_0^{TRGB}$ | $69.269 \pm 1.250$ | $\mathbf{-2.2423}$ | $-0.2334$ | $-1.9102$ | $\mathbf{-2.9038}$ | 1.3883 | 1.1225 |
| CC+$\hat{H}_0^{HW}$ | $72.995 \pm 1.625$ | $-0.5021$ | 1.2781 | $-0.1296$ | $-1.0076$ | **3.2914** | **2.8516** |
| CC+SN+$\hat{H}_0^{HW}$ | $71.243 \pm 1.289$ | $-1.3348$ | 0.6286 | $-0.9641$ | $-1.9395$ | **2.7800** | **2.2681** |
| CC+SN+BAO+$\hat{H}_0^{HW}$ | $71.196 \pm 1.217$ | $-1.3812$ | 0.6185 | $-1.0064$ | $\mathbf{-2.0023}$ | **2.8841** | **2.3134** |
| CC+$\hat{H}_0^{CM}$ | $74.962 \pm 1.614$ | 0.3051 | **2.0708** | 0.7091 | $-0.1665$ | **4.4763** | **3.8722** |
| CC+SN+$\hat{H}_0^{CM}$ | $72.595 \pm 1.300$ | $-0.7267$ | 1.2140 | $-0.3296$ | $-1.2972$ | **3.7300** | **3.0507** |
| CC+SN+BAO+$\hat{H}_0^{CM}$ | $72.385 \pm 1.227$ | $-0.8359$ | 1.1432 | $-0.4363$ | $-1.4251$ | **3.7633** | **3.0257** |
| CC+$\hat{H}_0^P$ | $67.397 \pm 0.497$ | $\mathbf{-3.6162}$ | $-1.2233$ | $\mathbf{-3.3325}$ | $\mathbf{-4.5391}$ | $-0.0036$ | $-0.0021$ |
| CC+SN+$\hat{H}_0^P$ | $67.437 \pm 0.481$ | $\mathbf{-3.6033}$ | $-1.2058$ | $\mathbf{-3.3187}$ | $\mathbf{-4.5283}$ | 0.0530 | 0.0306 |
| CC+SN+BAO+$\hat{H}_0^P$ | $67.513 \pm 0.477$ | $\mathbf{-3.5649}$ | $-1.1675$ | $\mathbf{-3.2777}$ | $\mathbf{-4.4877}$ | 0.1636 | 0.0943 |
| CC+$\hat{H}_0^{DES}$ | $67.389 \pm 1.071$ | $\mathbf{-3.2347}$ | $-1.1053$ | $\mathbf{-2.9418}$ | $\mathbf{-3.9957}$ | $-0.0090$ | $-0.0069$ |
| CC+SN+$\hat{H}_0^{DES}$ | $67.552 \pm 0.935$ | $\mathbf{-3.2591}$ | $-1.0616$ | $\mathbf{-2.9629}$ | $\mathbf{-4.0562}$ | 0.1435 | 0.1054 |
| CC+SN+BAO+$\hat{H}_0^{DES}$ | $67.817 \pm 0.906$ | $\mathbf{-3.1496}$ | $-0.9420$ | $\mathbf{-2.8465}$ | $\mathbf{-3.9468}$ | 0.4032 | 0.2928 |

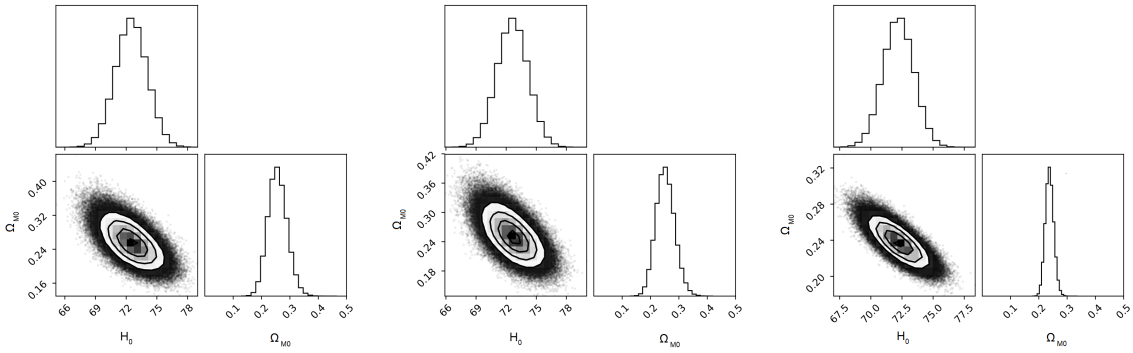
Table C.6: GPR results with Matérn $\left(\nu = \frac{9}{2}\right)$ kernel function

| **Dataset** | $\hat{H}_0$ (km s$^{-1}$ Mpc$^{-1}$) | $d(\hat{H}_0, \hat{H}_0^R)$ | $d(\hat{H}_0, \hat{H}_0^{TRGB})$ | $d(\hat{H}_0, \hat{H}_0^{HW})$ | $d(\hat{H}_0, \hat{H}_0^{CM})$ | $d(\hat{H}_0, \hat{H}_0^P)$ | $d(\hat{H}_0, \hat{H}_0^{DES})$ |
|---|---|---|---|---|---|---|---|
| CC | $69.578 \pm 5.252$ | $-0.8351$ | $-0.0397$ | $-0.6742$ | $-1.0468$ | 0.4130 | 0.4060 |
| CC+SN | $68.304 \pm 1.835$ | $\mathbf{-2.2892}$ | $-0.5664$ | $-1.9974$ | $\mathbf{-2.8324}$ | 0.4753 | 0.4225 |
| CC+SN+BAO | $68.941 \pm 1.658$ | $\mathbf{-2.1444}$ | $-0.3406$ | $-1.8358$ | $\mathbf{-2.7155}$ | 0.8903 | 0.7748 |
| CC+$\hat{H}_0^R$ | $73.867 \pm 1.732$ | $-0.1404$ | 1.5821 | 0.2337 | $-0.6145$ | **3.5878** | **3.1522** |
| CC+SN+$\hat{H}_0^R$ | $71.663 \pm 1.345$ | $-1.1296$ | 0.8004 | $-0.7549$ | $-1.7129$ | **2.9704** | **2.4533** |
| CC+SN+BAO+$\hat{H}_0^R$ | $71.552 \pm 1.263$ | $-1.2042$ | 0.7678 | $-0.8253$ | $-1.8068$ | **3.0557** | **2.4785** |
| CC+$\hat{H}_0^{TRGB}$ | $69.702 \pm 1.779$ | $-1.7752$ | $-0.0375$ | $-1.4621$ | $\mathbf{-2.3082}$ | 1.2462 | 1.1009 |
| CC+SN+$\hat{H}_0^{TRGB}$ | $69.077 \pm 1.330$ | $\mathbf{-2.2816}$ | $-0.3117$ | $-1.9566$ | $\mathbf{-2.9277}$ | 1.1806 | 0.9719 |
| CC+SN+BAO+$\hat{H}_0^{TRGB}$ | $69.317 \pm 1.250$ | $\mathbf{-2.2208}$ | $-0.2126$ | $-1.8878$ | $\mathbf{-2.8812}$ | 1.4235 | 1.1510 |
| CC+$\hat{H}_0^{HW}$ | $73.035 \pm 1.625$ | $-0.4858$ | 1.2940 | $-0.1128$ | $-0.9907$ | **3.3150** | **2.8720** |
| CC+SN+$\hat{H}_0^{HW}$ | $71.290 \pm 1.289$ | $-1.3139$ | 0.6489 | $-0.9422$ | $-1.9174$ | **2.8135** | **2.2955** |
| CC+SN+BAO+$\hat{H}_0^{HW}$ | $71.231 \pm 1.216$ | $-1.3652$ | 0.6345 | $-0.9895$ | $-1.9857$ | **2.9134** | **2.3363** |
| CC+$\hat{H}_0^{CM}$ | $74.993 \pm 1.613$ | 0.3179 | **2.0835** | 0.7225 | $-0.1532$ | **4.4958** | **3.8888** |
| CC+SN+$\hat{H}_0^{CM}$ | $72.629 \pm 1.299$ | $-0.7116$ | 1.2292 | $-0.3137$ | $-1.2814$ | **3.7574** | **3.0723** |
| CC+SN+BAO+$\hat{H}_0^{CM}$ | $72.412 \pm 1.225$ | $-0.8241$ | 1.1555 | $-0.4238$ | $-1.4131$ | **3.7885** | **3.0445** |
| CC+$\hat{H}_0^P$ | $67.404 \pm 0.497$ | $\mathbf{-3.6126}$ | $-1.2199$ | $\mathbf{-3.3288}$ | $\mathbf{-4.5353}$ | 0.0057 | 0.0034 |
| CC+SN+$\hat{H}_0^P$ | $67.446 \pm 0.481$ | $\mathbf{-3.5985}$ | $-1.2011$ | $\mathbf{-3.3135}$ | $\mathbf{-4.5231}$ | 0.0661 | 0.0382 |
| CC+SN+BAO+$\hat{H}_0^P$ | $67.520 \pm 0.477$ | $\mathbf{-3.5614}$ | $-1.1641$ | $\mathbf{-3.2740}$ | $\mathbf{-4.4840}$ | 0.1732 | 0.0998 |
| CC+$\hat{H}_0^{DES}$ | $67.421 \pm 1.071$ | $\mathbf{-3.2195}$ | $-1.0909$ | $\mathbf{-2.9259}$ | $\mathbf{-3.9795}$ | 0.0174 | 0.0134 |
| CC+SN+$\hat{H}_0^{DES}$ | $67.587 \pm 0.935$ | $\mathbf{-3.2414}$ | $-1.0449$ | $\mathbf{-2.9443}$ | $\mathbf{-4.0373}$ | 0.1766 | 0.1297 |
| CC+SN+BAO+$\hat{H}_0^{DES}$ | $67.842 \pm 0.905$ | $\mathbf{-3.1377}$ | $-0.9304$ | $\mathbf{-2.8339}$ | $\mathbf{-3.9343}$ | 0.4272 | 0.3101 |

Table C.7: GPR results with Cauchy kernel function

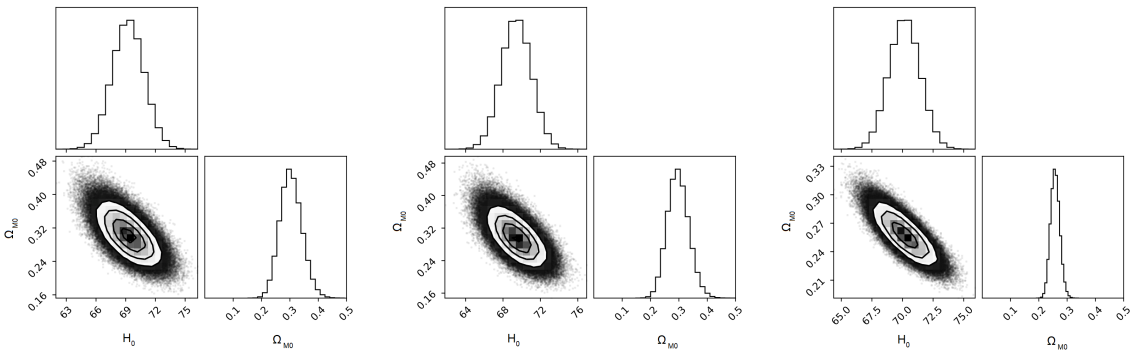
# Appendix D: MCMC $\Lambda CDM$ Plots

The figures for the priorless case were presented in Section 3.3. Similar figures for all the other prior $H_0$ estimates considered are shown in this Appendix.



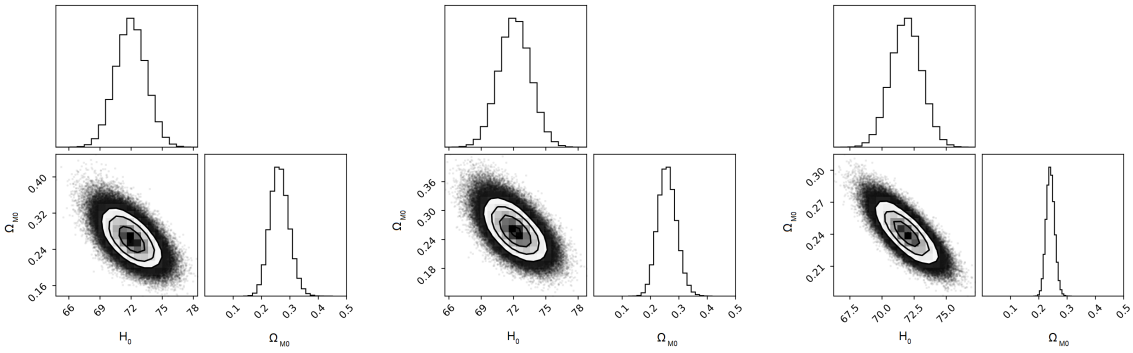

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure D.1: Corner plots for MCMC on parameters of $\Lambda CDM$ model using $\hat{H}_0^R$



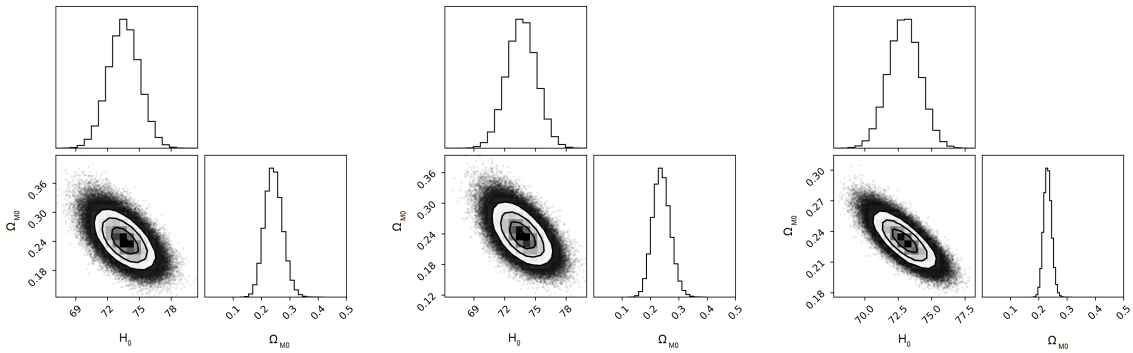

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure D.2: Corner plots for MCMC on parameters of $\Lambda CDM$ model using $\hat{H}_0^{TRGB}$



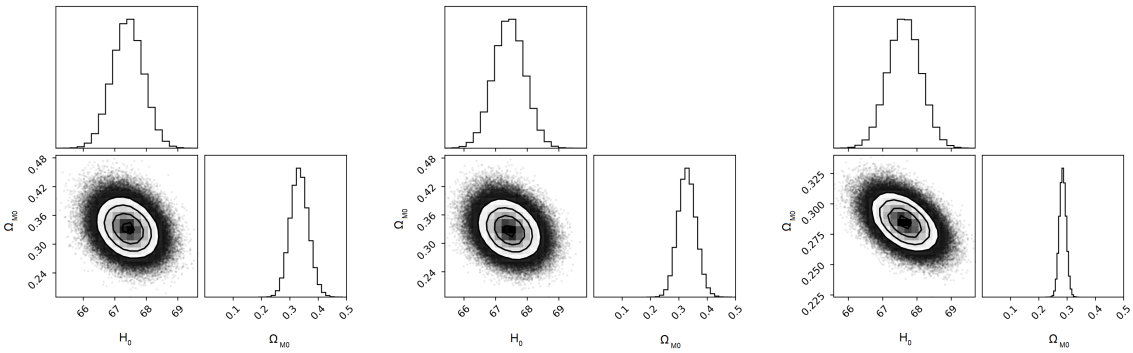

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure D.3: Corner plots for MCMC on parameters of $\Lambda CDM$ model using $\hat{H}_0^{HW}$



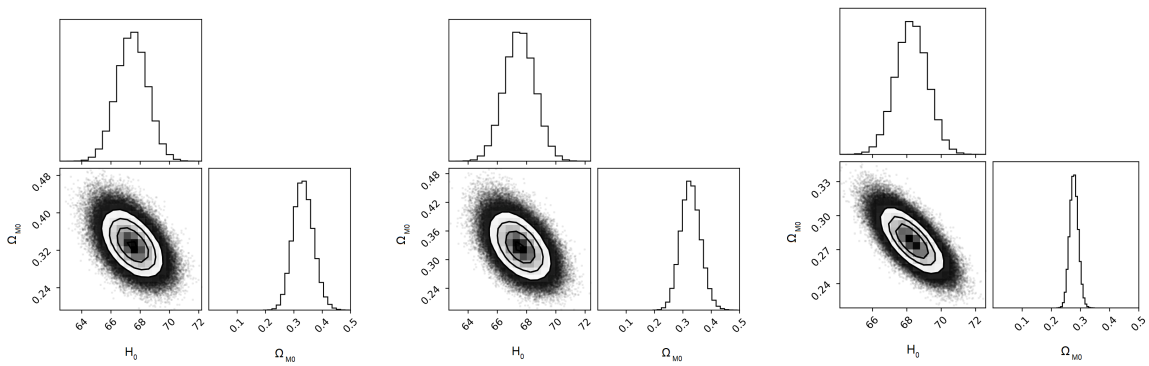

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure D.4: Corner plots for MCMC on parameters of $\Lambda CDM$ model using $\hat{H}_0^{CM}$




(a) CC (b) CC+SN (c) CC+SN+BAO

Figure D.5: Corner plots for MCMC on parameters of $\Lambda CDM$ model using $\hat{H}_0^{P}$

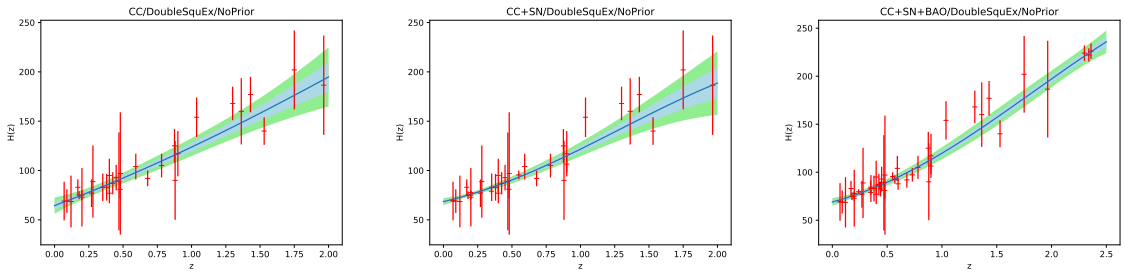



(a) CC (b) CC+SN (c) CC+SN+BAO

Figure D.6: Corner plots for MCMC on parameters of $\Lambda CDM$ model using $\hat{H}_0^{DES}$

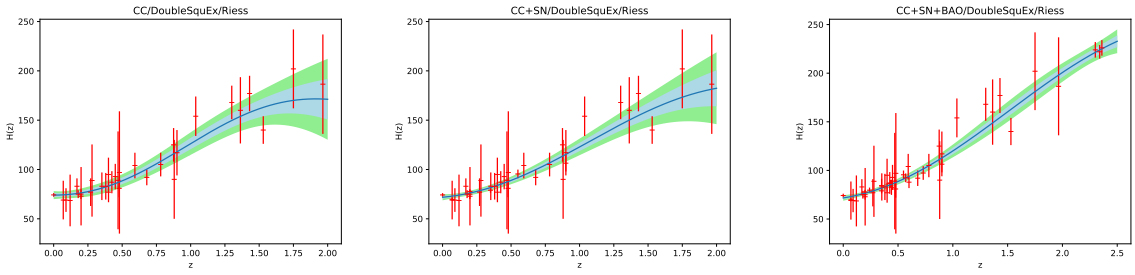
# Appendix E: MCMC GPR Plots

The figures for the square exponential kernel were presented in Section 3.4. Similar figures for all the other kernel functions considered are shown in this Appendix.



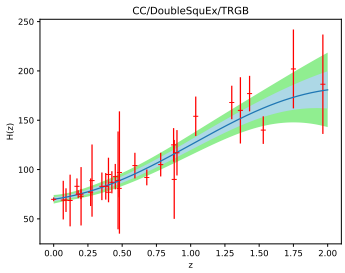

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure E.1: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for double square exponential kernel function and without adding any $H_0$ estimate to the data

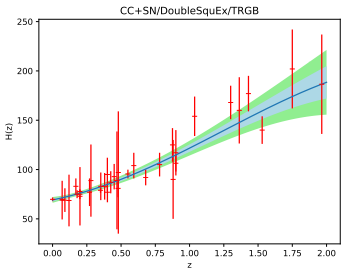



(a) CC (b) CC+SN (c) CC+SN+BAO

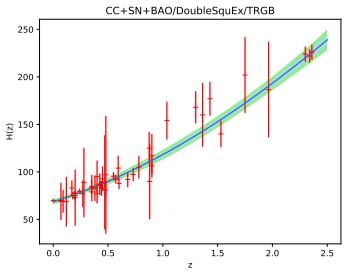
Figure E.2: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for double square exponential kernel function and using $\hat{H}_0^R$




(a) CC

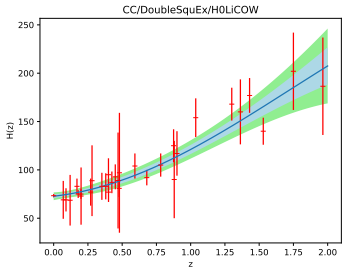



(b) CC+SN

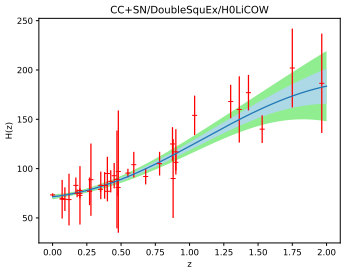



(c) CC+SN+BAO

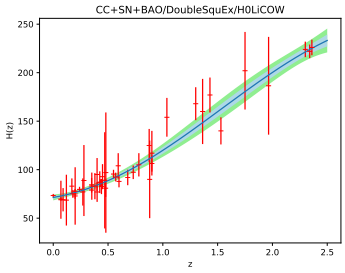
Figure E.3: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for double square exponential kernel function and using $\hat{H}_0^{TRGB}$




(a) CC

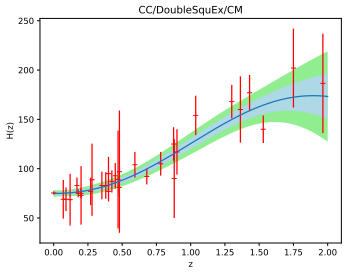



(b) CC+SN

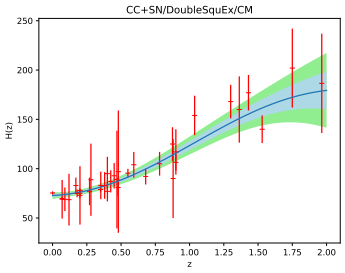



(c) CC+SN+BAO

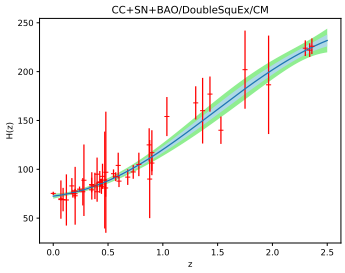
Figure E.4: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for double square exponential kernel function and using $\hat{H}_0^{HW}$




(a) CC

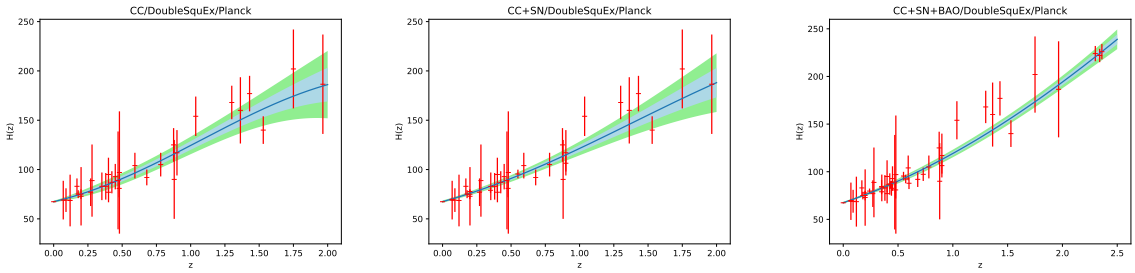



(b) CC+SN



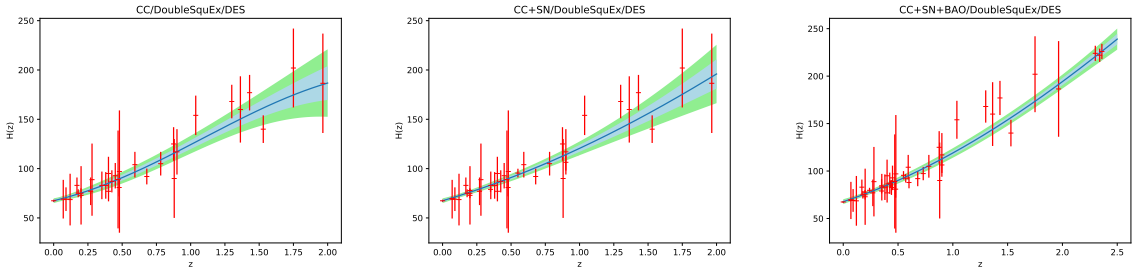

(c) CC+SN+BAO

Figure E.5: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for double square exponential kernel function and using $\hat{H}_0^{CM}$



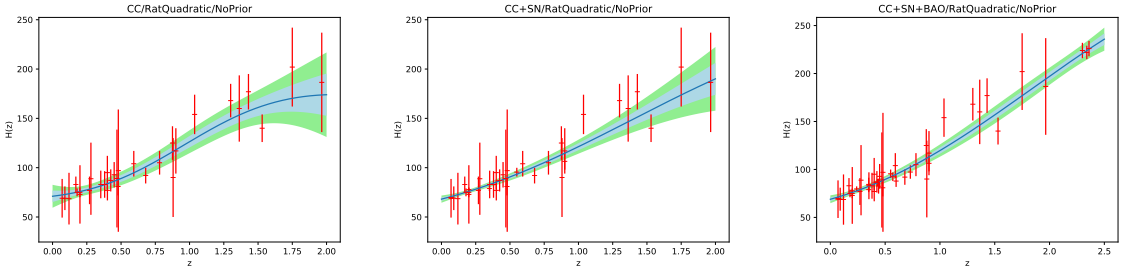

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure E.6: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for double square exponential kernel function and using $\hat{H}_0^P$



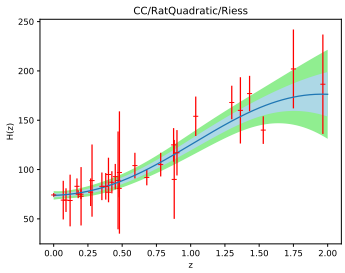

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure E.7: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for double square exponential kernel function and using $\hat{H}_0^{DES}$

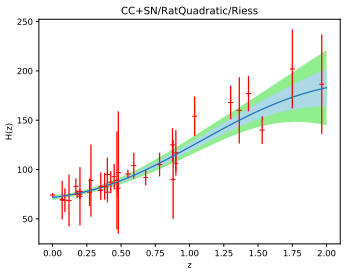



(a) CC (b) CC+SN (c) CC+SN+BAO

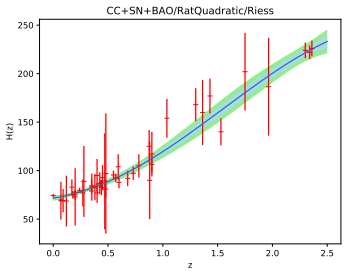
Figure E.8: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for rational quadratic kernel function and without adding any $H_0$ estimate to the data




(a) CC

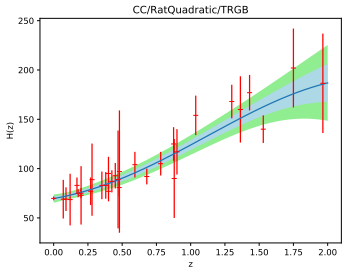



(b) CC+SN

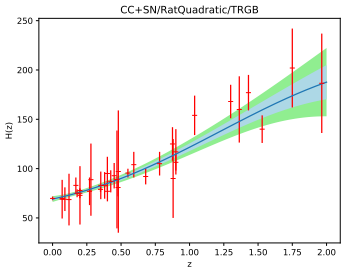



(c) CC+SN+BAO

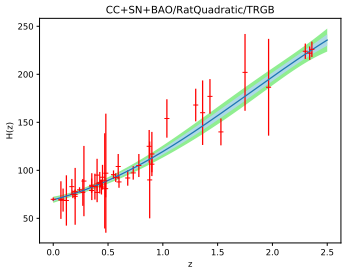
Figure E.9: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for rational quadratic kernel function and using $\hat{H}_0^R$




(a) CC

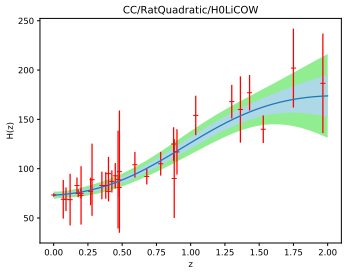



(b) CC+SN

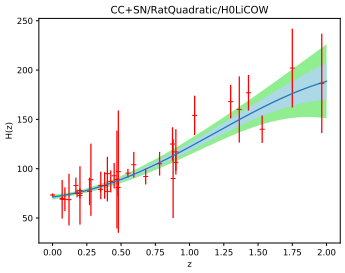



(c) CC+SN+BAO

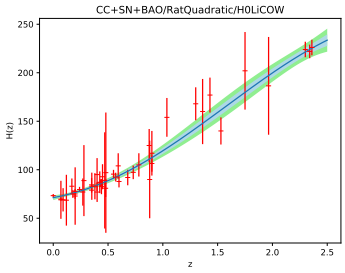
Figure E.10: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for rational quadratic kernel function and using $\hat{H}_0^{TRGB}$




(a) CC

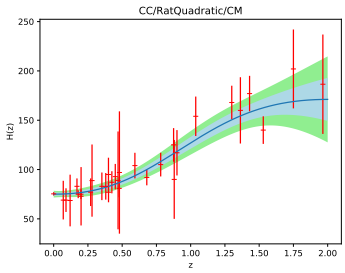



(b) CC+SN

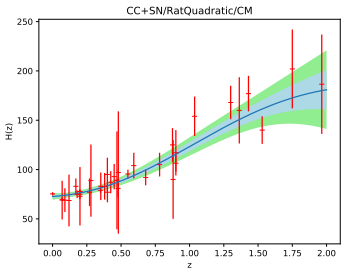



(c) CC+SN+BAO

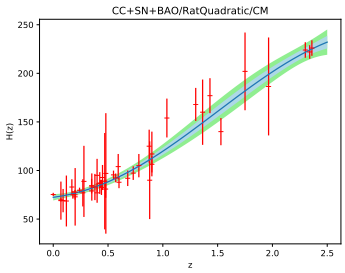
Figure E.11: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for rational quadratic kernel function and using $\hat{H}_0^{HW}$




(a) CC

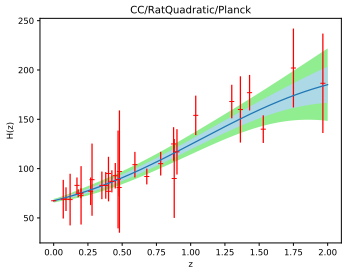



(b) CC+SN

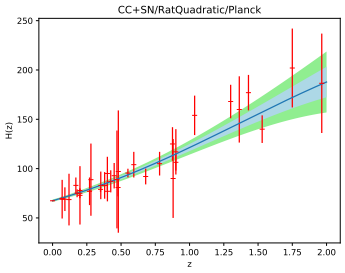



(c) CC+SN+BAO

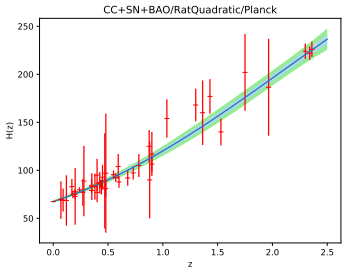
Figure E.12: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for rational quadratic kernel function and using $\hat{H}_0^{CM}$




(a) CC

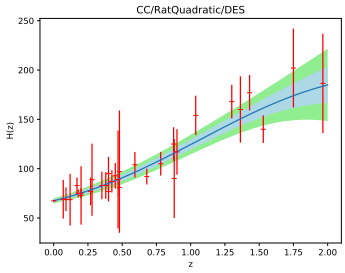



(b) CC+SN

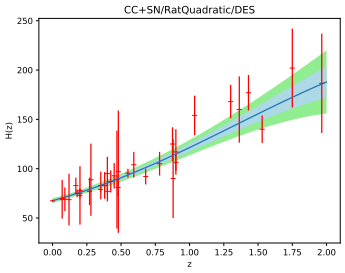



(c) CC+SN+BAO

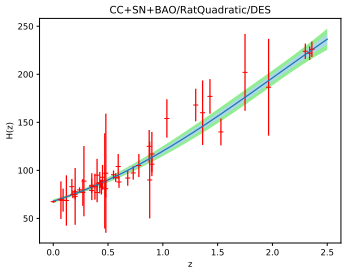
Figure E.13: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for rational quadratic kernel function and using $\hat{H}_0^{P}$




(a) CC

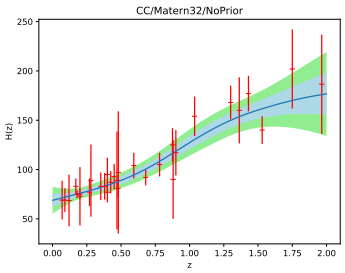



(b) CC+SN

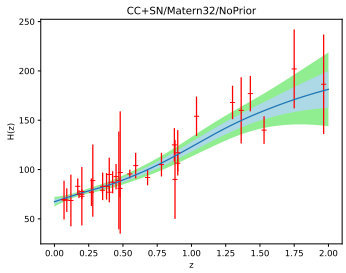



(c) CC+SN+BAO

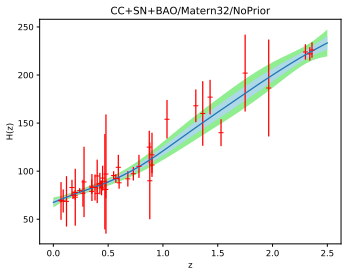
Figure E.14: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for rational quadratic kernel function and using $\hat{H}_0^{DES}$




(a) CC

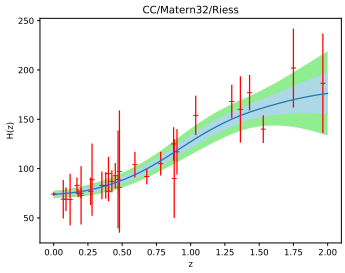



(b) CC+SN

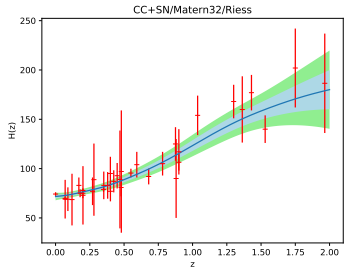



(c) CC+SN+BAO

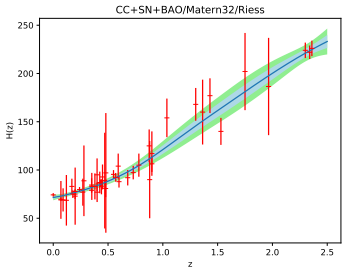
Figure E.15: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{3}{2}\right)$ kernel function and without adding any $H_0$ estimate to the data




(a) CC

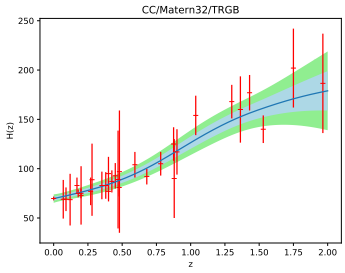



(b) CC+SN

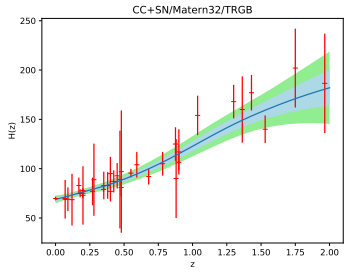



(c) CC+SN+BAO

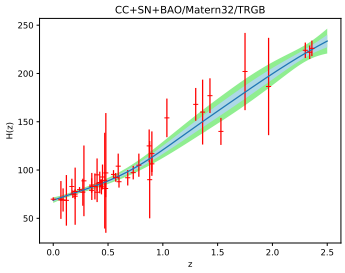
Figure E.16: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{3}{2}\right)$ kernel function and using $\hat{H}_0^R$




(a) CC

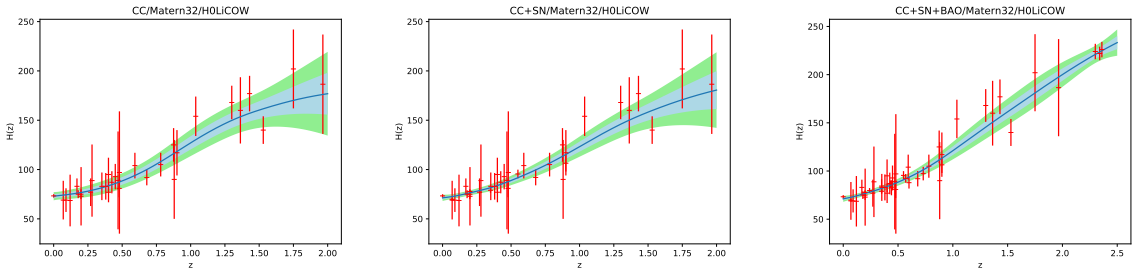



(b) CC+SN



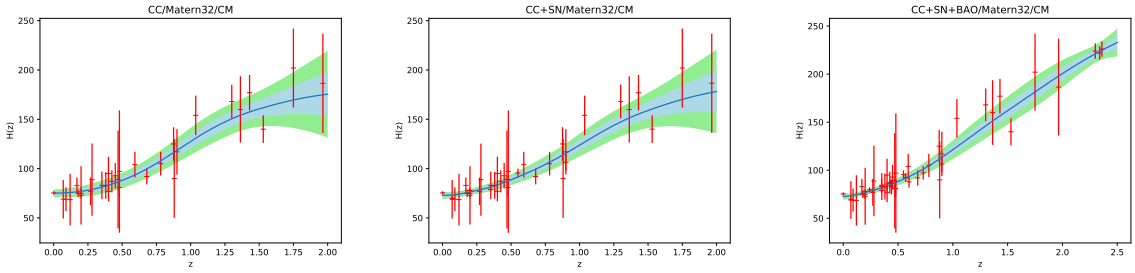

(c) CC+SN+BAO

Figure E.17: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{3}{2}\right)$ kernel function and using $\hat{H}_0^{TRGB}$



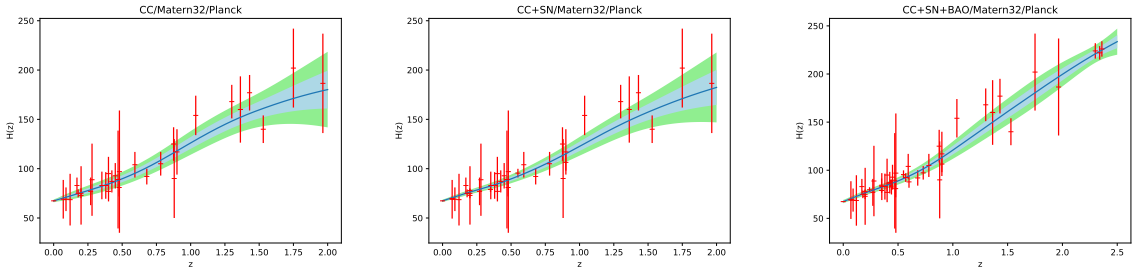

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure E.18: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{3}{2}\right)$ kernel function and using $\hat{H}_0^{HW}$



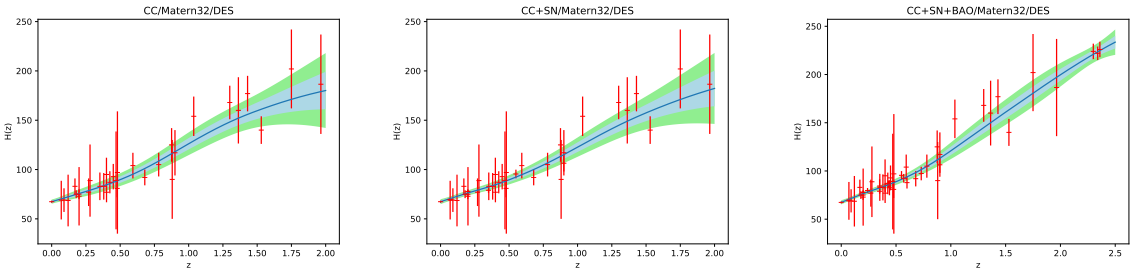

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure E.19: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{3}{2}\right)$ kernel function and using $\hat{H}_0^{CM}$



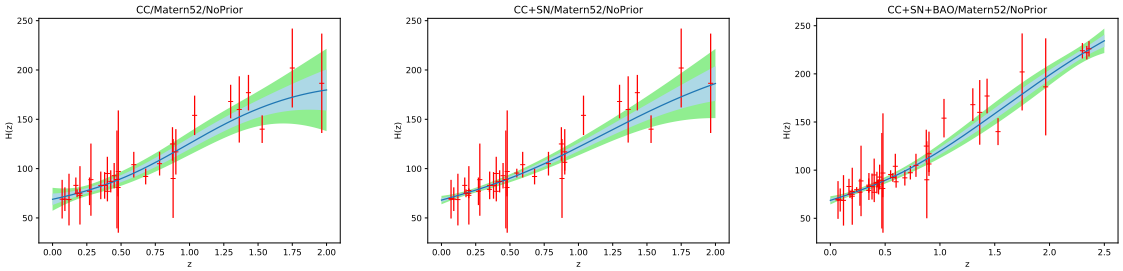

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure E.20: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{3}{2}\right)$ kernel function and using $\hat{H}_0^{P}$



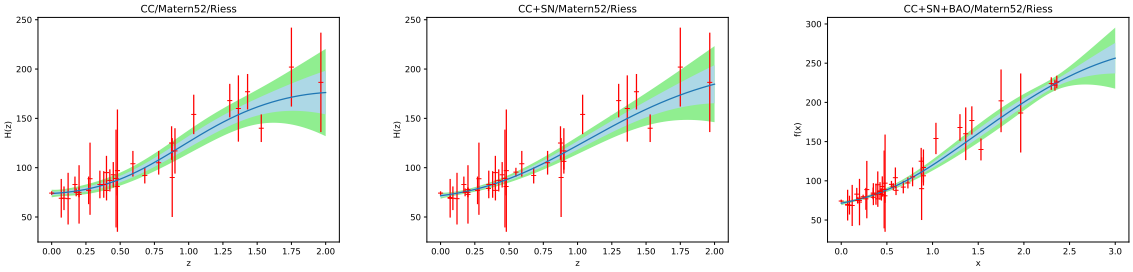

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure E.21: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{3}{2}\right)$ kernel function and using $\hat{H}_0^{DES}$



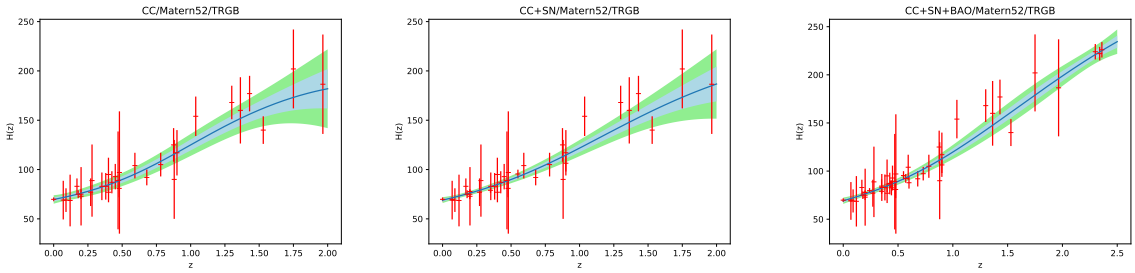

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure E.22: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{5}{2}\right)$ kernel function and without adding any $H_0$ estimate to the data



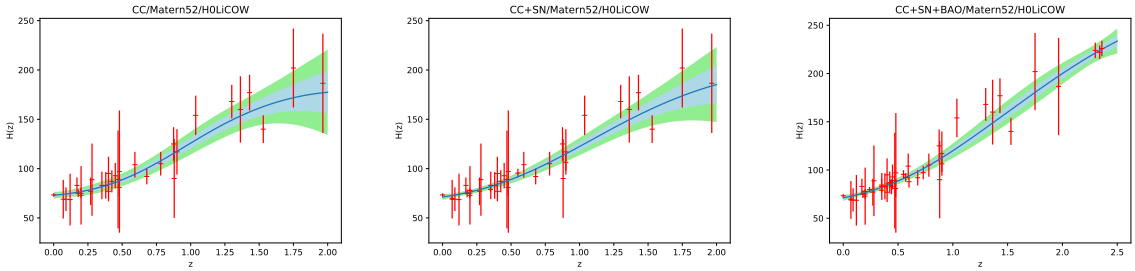

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure E.23: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{5}{2}\right)$ kernel function and using $\hat{H}_0^{R}$



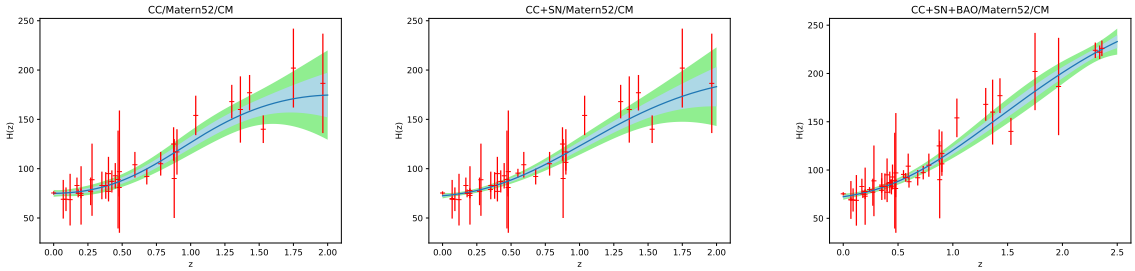

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure E.24: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{5}{2}\right)$ kernel function and using $\hat{H}_0^{TRGB}$



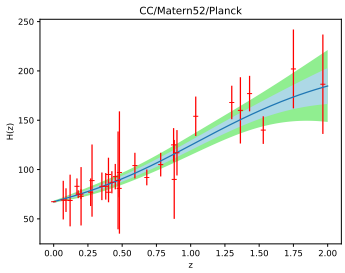

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure E.25: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{5}{2}\right)$ kernel function and using $\hat{H}_0^{HW}$

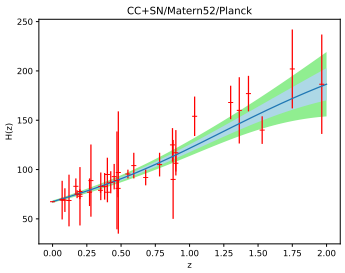



(a) CC (b) CC+SN (c) CC+SN+BAO

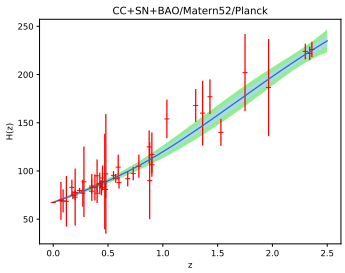
Figure E.26: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{5}{2}\right)$ kernel function and using $\hat{H}_0^{CM}$




(a) CC

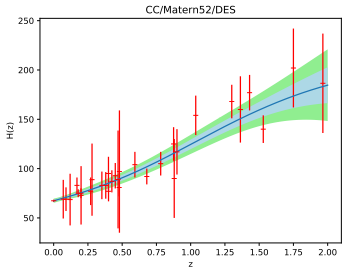



(b) CC+SN

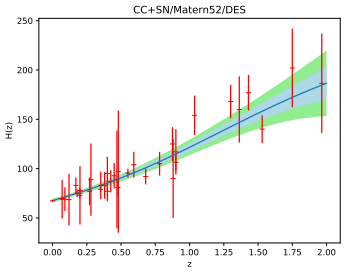



(c) CC+SN+BAO

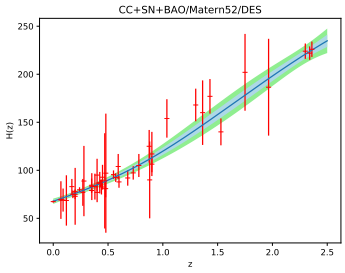
Figure E.27: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{5}{2}\right)$ kernel function and using $\hat{H}_0^P$




(a) CC

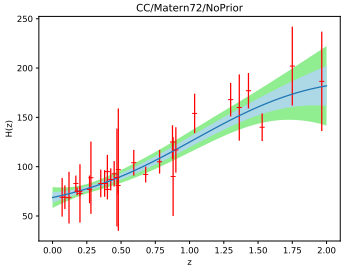



(b) CC+SN

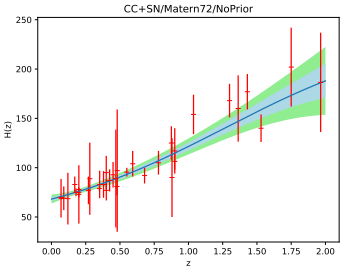



(c) CC+SN+BAO

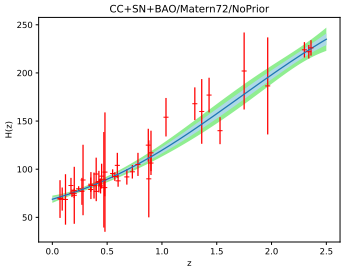
Figure E.28: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{5}{2}\right)$ kernel function and using $\hat{H}_0^{DES}$




(a) CC

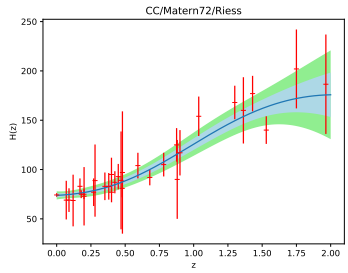



(b) CC+SN

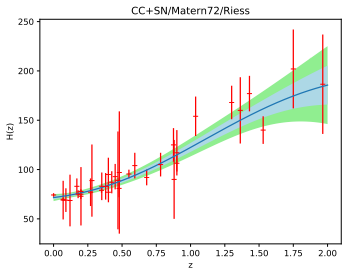



(c) CC+SN+BAO

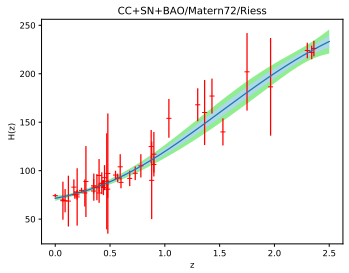
Figure E.29: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{7}{2}\right)$ kernel function and without adding any $H_0$ estimate to the data




(a) CC

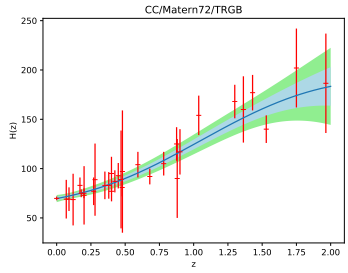



(b) CC+SN

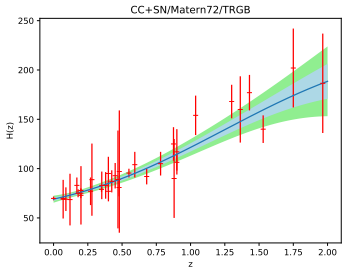



(c) CC+SN+BAO

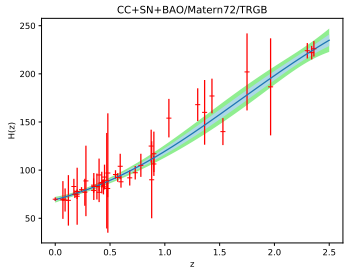
Figure E.30: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{7}{2}\right)$ kernel function and using $\hat{H}_0^R$




(a) CC

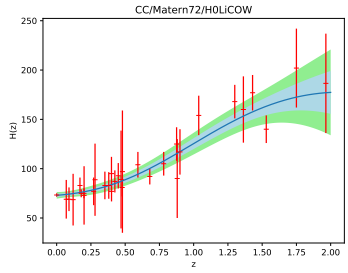



(b) CC+SN

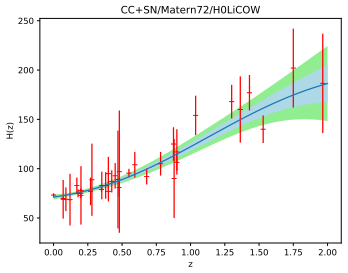



(c) CC+SN+BAO

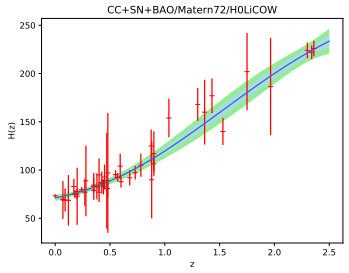
Figure E.31: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{7}{2}\right)$ kernel function and using $\hat{H}_0^{TRGB}$




(a) CC

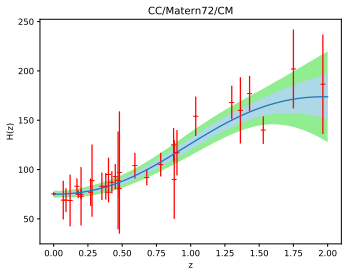



(b) CC+SN

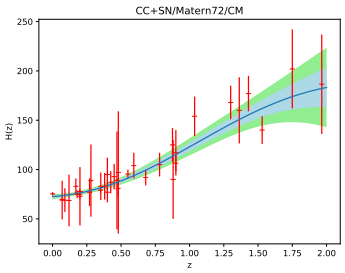



(c) CC+SN+BAO

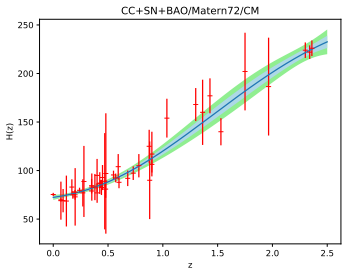
Figure E.32: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{7}{2}\right)$ kernel function and using $\hat{H}_0^{HW}$




(a) CC

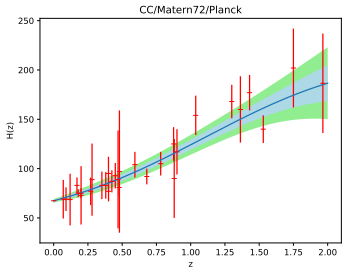



(b) CC+SN

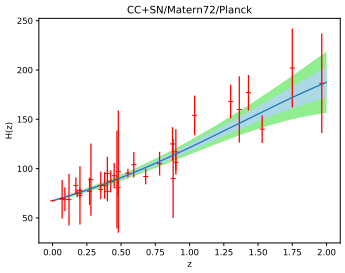



(c) CC+SN+BAO

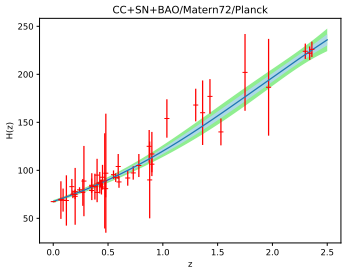
Figure E.33: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{7}{2}\right)$ kernel function and using $\hat{H}_0^{CM}$




(a) CC

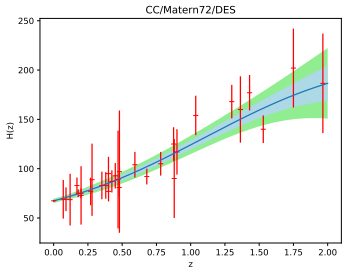



(b) CC+SN

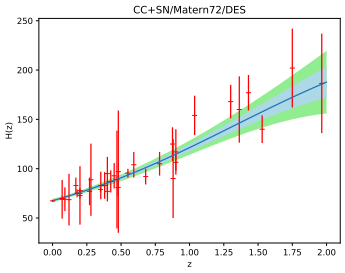



(c) CC+SN+BAO

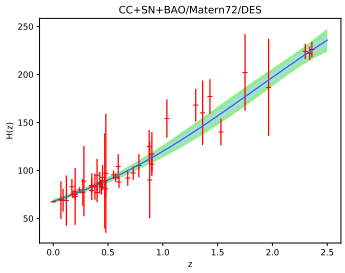
Figure E.34: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{7}{2}\right)$ kernel function and using $\hat{H}_0^{P}$




(a) CC

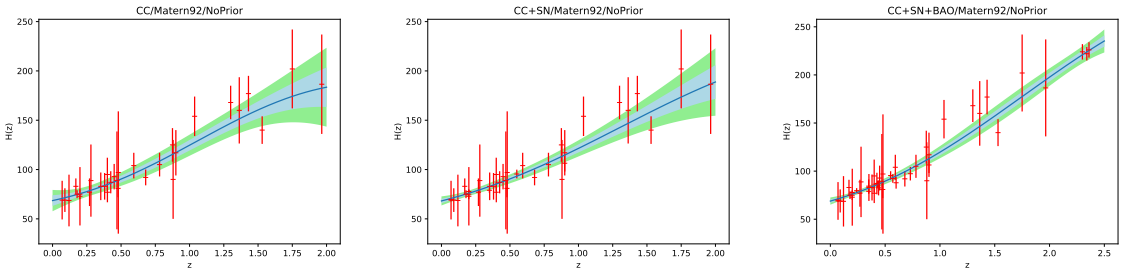



(b) CC+SN



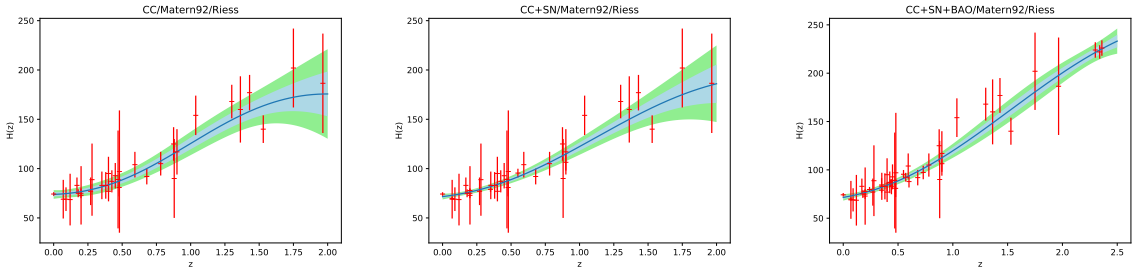

(c) CC+SN+BAO

Figure E.35: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{7}{2}\right)$ kernel function and using $\hat{H}_0^{DES}$



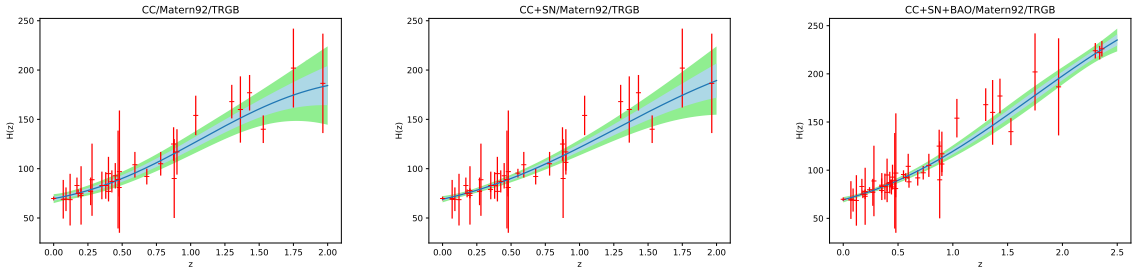

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure E.36: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{9}{2}\right)$ kernel function and without adding any $H_0$ estimate to the data



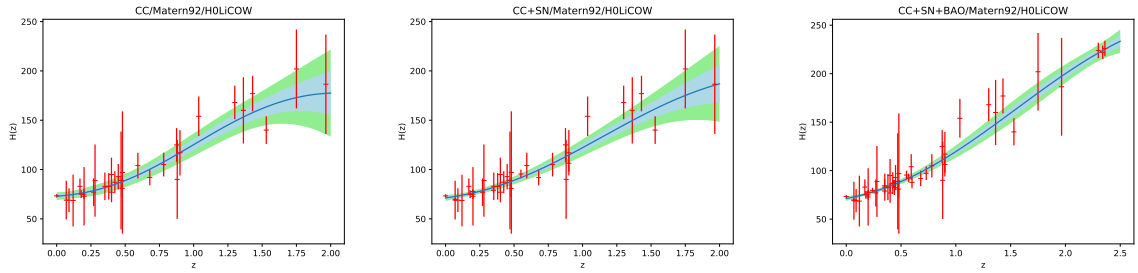

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure E.37: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{9}{2}\right)$ kernel function and using $\hat{H}_0^R$



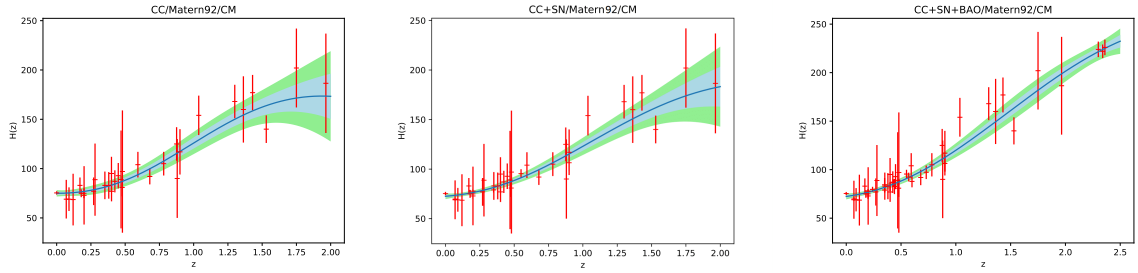

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure E.38: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{9}{2}\right)$ kernel function and using $\hat{H}_0^{TRGB}$



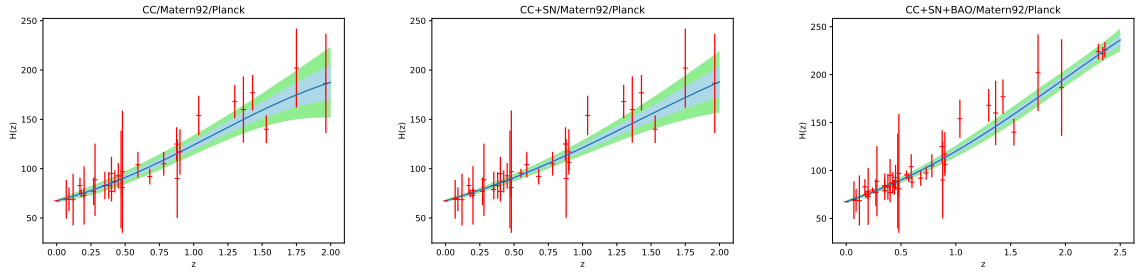

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure E.39: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{9}{2}\right)$ kernel function and using $\hat{H}_0^{HW}$



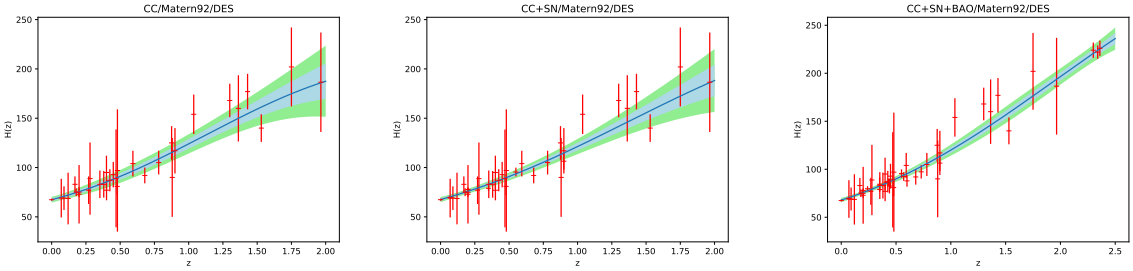

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure E.40: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{9}{2}\right)$ kernel function and using $\hat{H}_0^{CM}$



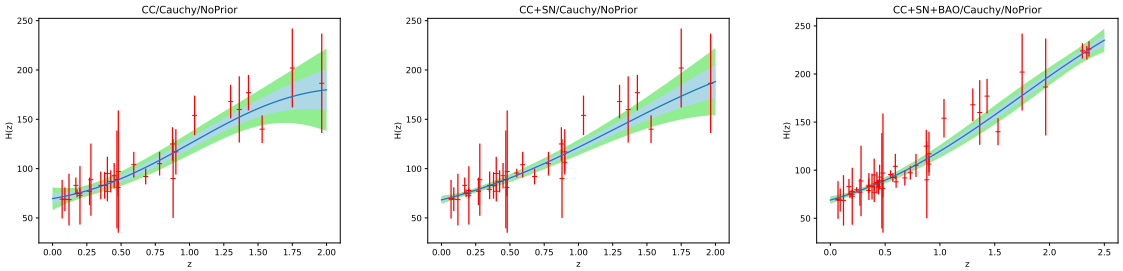

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure E.41: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{9}{2}\right)$ kernel function and using $\hat{H}_0^{P}$



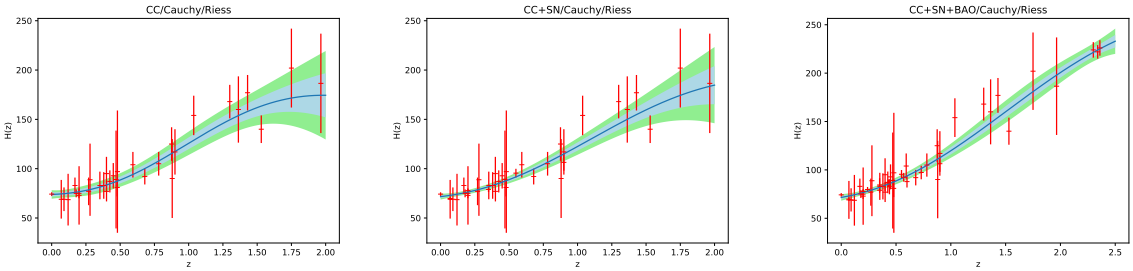

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure E.42: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{9}{2}\right)$ kernel function and using $\hat{H}_0^{DES}$



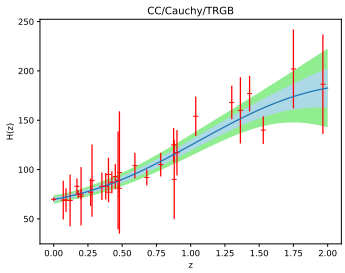

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure E.43: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Cauchy kernel function and without adding any $H_0$ estimate to the data

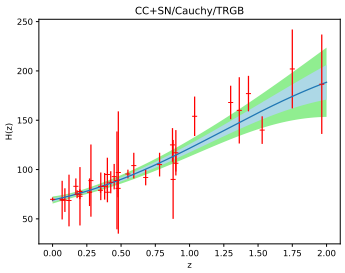



(a) CC (b) CC+SN (c) CC+SN+BAO

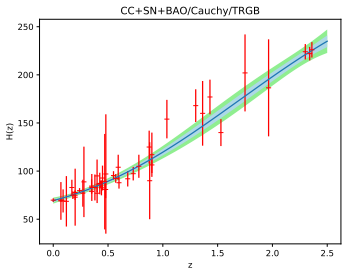
Figure E.44: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Cauchy kernel function and using $\hat{H}_0^R$




(a) CC

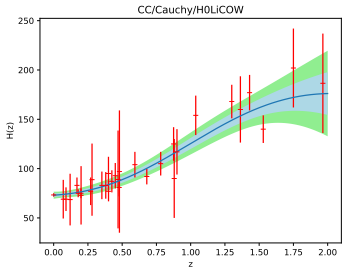



(b) CC+SN

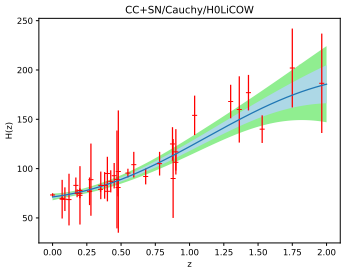



(c) CC+SN+BAO

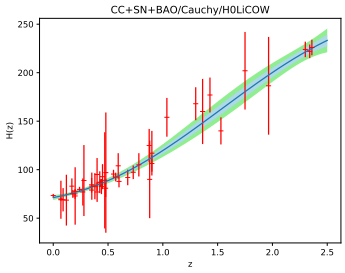
Figure E.45: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Cauchy kernel function and using $\hat{H}_0^{TRGB}$




(a) CC

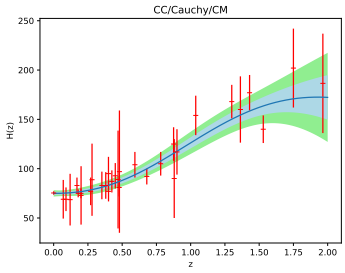



(b) CC+SN

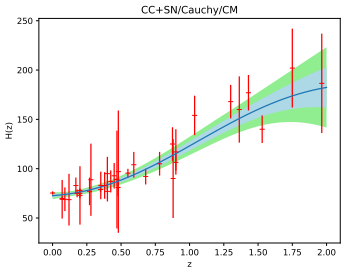



(c) CC+SN+BAO

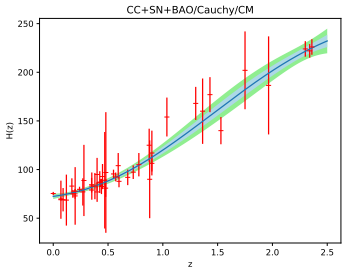
Figure E.46: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Cauchy kernel function and using $\hat{H}_0^{HW}$




(a) CC

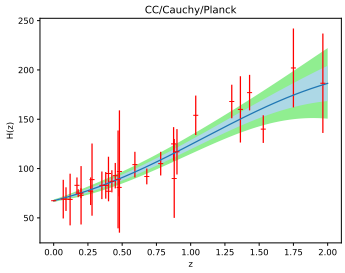



(b) CC+SN

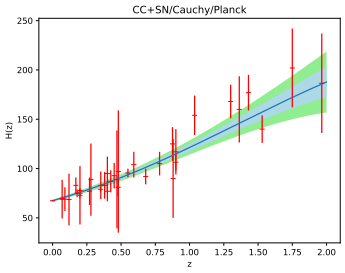



(c) CC+SN+BAO

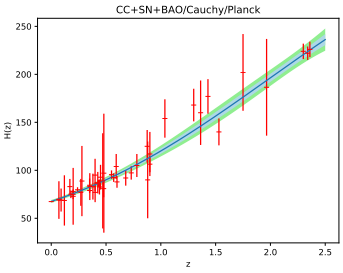
Figure E.47: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Cauchy kernel function and using $\hat{H}_0^{CM}$




(a) CC

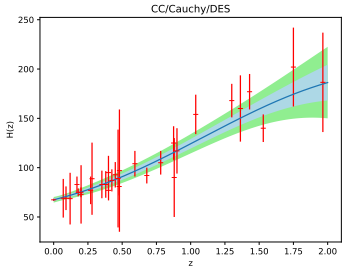



(b) CC+SN

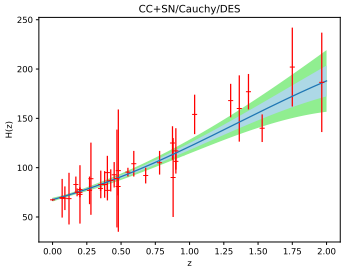



(c) CC+SN+BAO

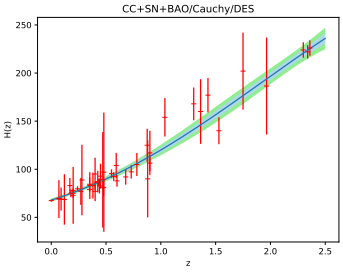
Figure E.48: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Cauchy kernel function and using $\hat{H}_0^P$




(a) CC




(b) CC+SN




(c) CC+SN+BAO

Figure E.49: MCMC GP reconstructions of $\hat{H}(z)$ against $z$ for Cauchy kernel function and using $\hat{H}_0^{DES}$

# Appendix F: MCMC GPR Tables

The table for the square exponential kernel was presented in Section 3.4. Similar tables for all the other kernel functions considered are shown in this Appendix.

| **Dataset** | $\hat{H}_0$ (km s$^{-1}$ Mpc$^{-1}$) | $d(\hat{H}_0, \hat{H}_0^R)$ | $d(\hat{H}_0, \hat{H}_0^{TRGB})$ | $d(\hat{H}_0, \hat{H}_0^{HW})$ | $d(\hat{H}_0, \hat{H}_0^{CM})$ | $d(\hat{H}_0, \hat{H}_0^P)$ | $d(\hat{H}_0, \hat{H}_0^{DES})$ |
|---|---|---|---|---|---|---|---|
| CC | $71.934 \pm 5.605$ | $-0.3880$ | $0.3606$ | $-0.2333$ | $-0.5838$ | $0.8057$ | $0.7938$ |
| CC+SN | $68.448 \pm 3.242$ | $-1.5526$ | $-0.3598$ | $-1.3256$ | $-1.8904$ | $0.3195$ | $0.3061$ |
| CC+SN+BAO | $69.100 \pm 2.537$ | $-1.6398$ | $-0.2208$ | $-1.3753$ | $\mathbf{-2.0540}$ | $0.6575$ | $0.6148$ |
| CC+$\hat{H}_0^R$ | $74.039 \pm 2.970$ | $-0.0521$ | $1.2024$ | $0.2159$ | $-0.3843$ | $\mathbf{2.2046}$ | $\mathbf{2.0964}$ |
| CC+SN+$\hat{H}_0^R$ | $71.146 \pm 1.640$ | $-1.2546$ | $0.5363$ | $-0.9118$ | $-1.7905$ | $\mathbf{2.1846}$ | $1.8968$ |
| CC+SN+BAO+$\hat{H}_0^R$ | $70.833 \pm 1.369$ | $-1.4873$ | $0.4409$ | $-1.1303$ | $\mathbf{-2.0843}$ | $\mathbf{2.3547}$ | $1.9543$ |
| CC+$\hat{H}_0^{TRGB}$ | $69.195 \pm 3.056$ | $-1.4128$ | $-0.1681$ | $-1.1739$ | $-1.7650$ | $0.5796$ | $0.5526$ |
| CC+SN+$\hat{H}_0^{TRGB}$ | $68.681 \pm 1.642$ | $\mathbf{-2.2599}$ | $-0.4458$ | $-1.9546$ | $\mathbf{-2.8392}$ | $0.7461$ | $0.6480$ |
| CC+SN+BAO+$\hat{H}_0^{TRGB}$ | $69.412 \pm 1.496$ | $\mathbf{-2.0411}$ | $-0.1606$ | $-1.7172$ | $\mathbf{-2.6400}$ | $1.2756$ | $1.0835$ |
| CC+$\hat{H}_0^{HW}$ | $71.982 \pm 2.431$ | $-0.7370$ | $0.7071$ | $-0.4444$ | $-1.1398$ | $1.8461$ | $1.7171$ |
| CC+SN+$\hat{H}_0^{HW}$ | $71.456 \pm 1.626$ | $-1.1327$ | $0.6620$ | $-0.7840$ | $-1.6657$ | $\mathbf{2.3841}$ | $\mathbf{2.0659}$ |
| CC+SN+BAO+$\hat{H}_0^{HW}$ | $70.632 \pm 1.298$ | $-1.6048$ | $0.3617$ | $-1.2472$ | $\mathbf{-2.2221}$ | $\mathbf{2.3236}$ | $1.8997$ |
| CC+$\hat{H}_0^{CM}$ | $74.423 \pm 2.513$ | $0.0653$ | $1.4673$ | $0.3700$ | $-0.3068$ | $\mathbf{2.7408}$ | $\mathbf{2.5600}$ |
| CC+SN+$\hat{H}_0^{CM}$ | $72.366 \pm 1.600$ | $-0.7652$ | $1.0331$ | $-0.4001$ | $-1.2864$ | $\mathbf{2.9631}$ | $\mathbf{2.5580}$ |
| CC+SN+BAO+$\hat{H}_0^{CM}$ | $72.281 \pm 1.454$ | $-0.8325$ | $1.0367$ | $-0.4556$ | $-1.3813$ | $\mathbf{3.1736}$ | $\mathbf{2.6765}$ |
| CC+$\hat{H}_0^P$ | $67.378 \pm 0.247$ | $\mathbf{-3.7250}$ | $-1.2639$ | $\mathbf{-3.4471}$ | $\mathbf{-4.6946}$ | $-0.0387$ | $-0.0191$ |
| CC+SN+$\hat{H}_0^P$ | $67.448 \pm 0.231$ | $\mathbf{-3.6913}$ | $-1.2288$ | $\mathbf{-3.4110}$ | $\mathbf{-4.6597}$ | $0.0872$ | $0.0427$ |
| CC+SN+BAO+$\hat{H}_0^P$ | $67.526 \pm 0.226$ | $\mathbf{-3.6498}$ | $-1.1883$ | $\mathbf{-3.3667}$ | $\mathbf{-4.6154}$ | $0.2305$ | $0.1126$ |
| CC+$\hat{H}_0^{DES}$ | $67.455 \pm 1.147$ | $\mathbf{-3.1449}$ | $-1.0567$ | $\mathbf{-2.8504}$ | $\mathbf{-3.8814}$ | $0.0439$ | $0.0345$ |
| CC+SN+$\hat{H}_0^{DES}$ | $67.413 \pm 0.848$ | $\mathbf{-3.3900}$ | $-1.1473$ | $\mathbf{-3.0987}$ | $\mathbf{-4.2173}$ | $0.0130$ | $0.0092$ |
| CC+SN+BAO+$\hat{H}_0^{DES}$ | $67.694 \pm 0.782$ | $\mathbf{-3.2946}$ | $-1.0251$ | $\mathbf{-2.9960}$ | $\mathbf{-4.1316}$ | $0.3167$ | $0.2178$ |

Table F.1: MCMC GPR Results with double square exponential kernel function

| **Dataset** | $\hat{H}_0$ (km s$^{-1}$ Mpc$^{-1}$) | $d(\hat{H}_0, \hat{H}_0^R)$ | $d(\hat{H}_0, \hat{H}_0^{TRGB})$ | $d(\hat{H}_0, \hat{H}_0^{HW})$ | $d(\hat{H}_0, \hat{H}_0^{CM})$ | $d(\hat{H}_0, \hat{H}_0^P)$ | $d(\hat{H}_0, \hat{H}_0^{DES})$ |
|---|---|---|---|---|---|---|---|
| CC | $71.023 \pm 5.289$ | $-0.5715$ | $0.2177$ | $-0.4098$ | $-0.7797$ | $0.6821$ | $0.6708$ |
| CC+SN | $68.353 \pm 3.408$ | $-1.5185$ | $-0.3708$ | $-1.2989$ | $-1.8414$ | $0.2767$ | $0.2661$ |
| CC+SN+BAO | $68.903 \pm 2.546$ | $-1.6991$ | $-0.2824$ | $-1.4364$ | $\mathbf{-2.1138}$ | $0.5794$ | $0.5420$ |
| CC+$\hat{H}_0^R$ | $73.949 \pm 3.008$ | $-0.0770$ | $1.1661$ | $0.1879$ | $-0.4066$ | $\mathbf{2.1475}$ | $\mathbf{2.0446}$ |
| CC+SN+$\hat{H}_0^R$ | $71.646 \pm 1.779$ | $-1.0112$ | $0.7093$ | $-0.6720$ | $-1.5135$ | $\mathbf{2.2975}$ | $\mathbf{2.0299}$ |
| CC+SN+BAO+$\hat{H}_0^R$ | $71.546 \pm 1.559$ | $-1.1157$ | $0.7105$ | $-0.7603$ | $-1.6596$ | $\mathbf{2.5324}$ | $\mathbf{2.1730}$ |
| CC+$\hat{H}_0^{TRGB}$ | $69.778 \pm 3.609$ | $-1.0990$ | $-0.0054$ | $-0.8829$ | $-1.3997$ | $0.6527$ | $0.6303$ |
| CC+SN+$\hat{H}_0^{TRGB}$ | $69.778 \pm 3.608$ | $-1.0991$ | $-0.0055$ | $-0.8830$ | $-1.3999$ | $0.6526$ | $0.6302$ |
| CC+SN+BAO+$\hat{H}_0^{TRGB}$ | $73.259 \pm 1.062$ | $-0.4560$ | $1.5890$ | $-0.0204$ | $-1.0520$ | $\mathbf{4.9902}$ | $\mathbf{3.8314}$ |
| CC+$\hat{H}_0^{HW}$ | $73.111 \pm 2.647$ | $-0.3454$ | $1.0160$ | $-0.0602$ | $-0.7143$ | $\mathbf{2.1198}$ | $1.9921$ |
| CC+SN+$\hat{H}_0^{HW}$ | $71.269 \pm 1.634$ | $-1.2066$ | $0.5862$ | $-0.8615$ | $-1.7415$ | $\mathbf{2.2644}$ | $1.9644$ |
| CC+SN+BAO+$\hat{H}_0^{HW}$ | $71.238 \pm 1.484$ | $-1.2699$ | $0.5966$ | $-0.9138$ | $-1.8346$ | $\mathbf{2.4517}$ | $\mathbf{2.0782}$ |
| CC+$\hat{H}_0^{CM}$ | $74.970 \pm 2.595$ | $0.2368$ | $1.6077$ | $0.5385$ | $-0.1228$ | $\mathbf{2.8649}$ | $\mathbf{2.6862}$ |
| CC+SN+$\hat{H}_0^{CM}$ | $72.624 \pm 1.665$ | $-0.6472$ | $1.1178$ | $-0.2842$ | $-1.1527$ | $\mathbf{3.0054}$ | $\mathbf{2.6181}$ |
| CC+SN+BAO+$\hat{H}_0^{CM}$ | $72.408 \pm 1.472$ | $-0.7740$ | $1.0850$ | $-0.3966$ | $-1.3170$ | $\mathbf{3.2207}$ | $\mathbf{2.7249}$ |
| CC+$\hat{H}_0^P$ | $68.046 \pm 0.239$ | $\mathbf{-3.3632}$ | $-0.9157$ | $\mathbf{-3.0603}$ | $\mathbf{-4.3041}$ | $1.1665$ | $0.5743$ |
| CC+SN+$\hat{H}_0^P$ | $67.445 \pm 0.231$ | $\mathbf{-3.6927}$ | $-1.2303$ | $\mathbf{-3.4125}$ | $\mathbf{-4.6612}$ | $0.0820$ | $0.0402$ |
| CC+SN+BAO+$\hat{H}_0^P$ | $67.517 \pm 0.226$ | $\mathbf{-3.6550}$ | $-1.1932$ | $\mathbf{-3.3722}$ | $\mathbf{-4.6210}$ | $0.2131$ | $0.1041$ |
| CC+$\hat{H}_0^{DES}$ | $71.625 \pm 0.943$ | $-1.2661$ | $0.8601$ | $-0.8618$ | $-1.9336$ | $\mathbf{3.9572}$ | $\mathbf{2.9155}$ |
| CC+SN+$\hat{H}_0^{DES}$ | $67.533 \pm 0.866$ | $\mathbf{-3.3174}$ | $-1.0856$ | $\mathbf{-3.0224}$ | $\mathbf{-4.1354}$ | $0.1331$ | $0.0951$ |
| CC+SN+BAO+$\hat{H}_0^{DES}$ | $67.831 \pm 0.802$ | $\mathbf{-3.2122}$ | $-0.9547$ | $\mathbf{-2.9093}$ | $\mathbf{-4.0387}$ | $0.4560$ | $0.3166$ |

Table F.2: MCMC GPR Results with rational quadratic kernel function

| **Dataset** | $\hat{H}_0$ (km s$^{-1}$ Mpc$^{-1}$) | $d(\hat{H}_0, \hat{H}_0^R)$ | $d(\hat{H}_0, \hat{H}_0^{TRGB})$ | $d(\hat{H}_0, \hat{H}_0^{HW})$ | $d(\hat{H}_0, \hat{H}_0^{CM})$ | $d(\hat{H}_0, \hat{H}_0^P)$ | $d(\hat{H}_0, \hat{H}_0^{DES})$ |
|---|---|---|---|---|---|---|---|
| CC | $68.798 \pm 6.040$ | $-0.8595$ | $-0.1582$ | $-0.7175$ | $-1.0451$ | 0.2307 | 0.2277 |
| CC+SN | $67.394 \pm 5.493$ | $-1.1796$ | $-0.4139$ | $-1.0271$ | $-1.3851$ | $-0.0010$ | $-0.0010$ |
| CC+SN+BAO | $67.519 \pm 4.842$ | $-1.2955$ | $-0.4386$ | $-1.1266$ | $-1.5281$ | 0.0245 | 0.0240 |
| CC+$\hat{H}_0^R$ | $73.820 \pm 3.066$ | $-0.1123$ | 1.1145 | 0.1482 | $-0.4378$ | **2.0667** | 1.9710 |
| CC+SN+$\hat{H}_0^R$ | $71.741 \pm 2.126$ | $-0.8859$ | 0.6808 | $-0.5728$ | $-1.3321$ | 1.9879 | 1.8137 |
| CC+SN+BAO+$\hat{H}_0^R$ | $71.519 \pm 1.982$ | $-1.0039$ | 0.6260 | $-0.6822$ | $-1.4747$ | **2.0152** | 1.8172 |
| CC+$\hat{H}_0^{TRGB}$ | $69.673 \pm 3.268$ | $-1.2156$ | $-0.0336$ | $-0.9846$ | $-1.5450$ | 0.6876 | 0.6592 |
| CC+SN+$\hat{H}_0^{TRGB}$ | $68.834 \pm 2.115$ | $-1.9301$ | $-0.3398$ | $-1.6457$ | **$-2.4121$** | 0.6596 | 0.6013 |
| CC+SN+BAO+$\hat{H}_0^{TRGB}$ | $68.828 \pm 2.014$ | $-1.9863$ | $-0.3512$ | $-1.6967$ | **$-2.4866$** | 0.6878 | 0.6220 |
| CC+$\hat{H}_0^{HW}$ | $72.992 \pm 2.693$ | $-0.3779$ | 0.9683 | $-0.0969$ | $-0.7430$ | **2.0413** | 1.9220 |
| CC+SN+$\hat{H}_0^{HW}$ | $71.301 \pm 1.919$ | $-1.1035$ | 0.5559 | $-0.7796$ | $-1.5873$ | 1.9671 | 1.7636 |
| CC+SN+BAO+$\hat{H}_0^{HW}$ | $71.138 \pm 1.808$ | $-1.2014$ | 0.5102 | $-0.8712$ | $-1.7067$ | 1.9930 | 1.7665 |
| CC+$\hat{H}_0^{CM}$ | $74.959 \pm 2.650$ | 0.2297 | 1.5819 | 0.5267 | $-0.1248$ | **2.8026** | **2.6341** |
| CC+SN+$\hat{H}_0^{CM}$ | $72.912 \pm 1.987$ | $-0.4853$ | 1.1320 | $-0.1482$ | $-0.9368$ | **2.6901** | **2.4269** |
| CC+SN+BAO+$\hat{H}_0^{CM}$ | $72.579 \pm 1.840$ | $-0.6341$ | 1.0505 | $-0.2878$ | $-1.1121$ | **2.7156** | **2.4155** |
| CC+$\hat{H}_0^P$ | $67.406 \pm 0.248$ | **$-3.7095$** | $-1.2492$ | **$-3.4305$** | **$-4.6776$** | 0.0114 | 0.0057 |
| CC+SN+$\hat{H}_0^P$ | $67.404 \pm 0.238$ | **$-3.7134$** | $-1.2512$ | **$-3.4347$** | **$-4.6829$** | 0.0073 | 0.0036 |
| CC+SN+BAO+$\hat{H}_0^P$ | $67.410 \pm 0.237$ | **$-3.7103$** | $-1.2481$ | **$-3.4314$** | **$-4.6798$** | 0.0183 | 0.0090 |
| CC+$\hat{H}_0^{DES}$ | $67.430 \pm 1.166$ | **$-3.1415$** | $-1.0630$ | **$-2.8476$** | **$-3.8731$** | 0.0240 | 0.0190 |
| CC+SN+$\hat{H}_0^{DES}$ | $67.414 \pm 0.976$ | **$-3.2954$** | $-1.1168$ | **$-3.0024$** | **$-4.0843$** | 0.0131 | 0.0098 |
| CC+SN+BAO+$\hat{H}_0^{DES}$ | $67.438 \pm 0.958$ | **$-3.2970$** | $-1.1098$ | **$-3.0036$** | **$-4.0905$** | 0.0355 | 0.0263 |

Table F.3: MCMC GPR Results with Matérn $\left(\nu = \frac{3}{2}\right)$ kernel function

| **Dataset** | $\hat{H}_0$ (km s$^{-1}$ Mpc$^{-1}$) | $d(\hat{H}_0, \hat{H}_0^R)$ | $d(\hat{H}_0, \hat{H}_0^{TRGB})$ | $d(\hat{H}_0, \hat{H}_0^{HW})$ | $d(\hat{H}_0, \hat{H}_0^{CM})$ | $d(\hat{H}_0, \hat{H}_0^P)$ | $d(\hat{H}_0, \hat{H}_0^{DES})$ |
|---|---|---|---|---|---|---|---|
| CC | $68.902 \pm 5.356$ | $-0.9402$ | $-0.1581$ | $-0.7827$ | $-1.1488$ | 0.2791 | 0.2746 |
| CC+SN | $68.076 \pm 3.722$ | $-1.4829$ | $-0.4125$ | $-1.2767$ | $-1.7813$ | 0.1800 | 0.1742 |
| CC+SN+BAO | $68.617 \pm 3.198$ | $-1.5227$ | $-0.3180$ | $-1.2930$ | $-1.8638$ | 0.3760 | 0.3598 |
| CC+$\hat{H}_0^R$ | $73.806 \pm 3.019$ | $-0.1173$ | 1.1230 | 0.1461 | $-0.4467$ | **2.0933** | 1.9936 |
| CC+SN+$\hat{H}_0^R$ | $71.574 \pm 1.873$ | $-1.0134$ | 0.6648 | $-0.6826$ | $-1.5010$ | **2.1532** | 1.9216 |
| CC+SN+BAO+$\hat{H}_0^R$ | $71.464 \pm 1.686$ | $-1.1108$ | 0.6551 | $-0.7668$ | $-1.6326$ | **2.3107** | **2.0186** |
| CC+$\hat{H}_0^{TRGB}$ | $69.646 \pm 3.189$ | $-1.2457$ | $-0.0414$ | $-1.0111$ | $-1.5825$ | 0.6959 | 0.6659 |
| CC+SN+$\hat{H}_0^{TRGB}$ | $68.962 \pm 1.843$ | **$-2.0300$** | $-0.3166$ | $-1.7302$ | **$-2.5616$** | 0.8180 | 0.7278 |
| CC+SN+BAO+$\hat{H}_0^{TRGB}$ | $69.166 \pm 1.682$ | **$-2.0391$** | $-0.2497$ | $-1.7284$ | **$-2.6010$** | 1.0064 | 0.8788 |
| CC+$\hat{H}_0^{HW}$ | $72.981 \pm 2.655$ | $-0.3850$ | 0.9741 | $-0.1013$ | $-0.7541$ | **2.0655** | 1.9417 |
| CC+SN+$\hat{H}_0^{HW}$ | $71.200 \pm 1.715$ | $-1.2077$ | 0.5471 | $-0.8697$ | $-1.7287$ | **2.1275** | 1.8653 |
| CC+SN+BAO+$\hat{H}_0^{HW}$ | $71.140 \pm 1.558$ | $-1.2856$ | 0.5453 | $-0.9368$ | $-1.8375$ | **2.2855** | 1.9608 |
| CC+$\hat{H}_0^{CM}$ | $74.943 \pm 2.618$ | 0.2269 | 1.5902 | 0.5266 | $-0.1307$ | **2.8306** | **2.6568** |
| CC+SN+$\hat{H}_0^{CM}$ | $72.574 \pm 1.742$ | $-0.6532$ | 1.0763 | $-0.2981$ | $-1.1469$ | **2.8550** | **2.5115** |
| CC+SN+BAO+$\hat{H}_0^{CM}$ | $72.366 \pm 1.579$ | $-0.7693$ | 1.0389 | $-0.4024$ | $-1.2941$ | **2.9988** | **2.5809** |
| CC+$\hat{H}_0^P$ | $67.402 \pm 0.247$ | **$-3.7120$** | $-1.2515$ | **$-3.4332$** | **$-4.6805$** | 0.0036 | 0.0018 |
| CC+SN+$\hat{H}_0^P$ | $67.436 \pm 0.233$ | **$-3.6971$** | $-1.2347$ | **$-3.4172$** | **$-4.6658$** | 0.0661 | 0.0324 |
| CC+SN+BAO+$\hat{H}_0^P$ | $67.491 \pm 0.231$ | **$-3.6678$** | $-1.2064$ | **$-3.3859$** | **$-4.6344$** | 0.1653 | 0.0810 |
| CC+$\hat{H}_0^{DES}$ | $67.411 \pm 1.151$ | **$-3.1618$** | $-1.0754$ | **$-2.8683$** | **$-3.8981$** | 0.0086 | 0.0068 |
| CC+SN+$\hat{H}_0^{DES}$ | $67.547 \pm 0.897$ | **$-3.2891$** | $-1.0724$ | **$-2.9934$** | **$-4.0977$** | 0.1431 | 0.1035 |
| CC+SN+BAO+$\hat{H}_0^{DES}$ | $67.740 \pm 0.867$ | **$-3.2146$** | $-0.9864$ | **$-2.9138$** | **$-4.0256$** | 0.3399 | 0.2428 |

Table F.4: MCMC GPR Results with Matérn $\left(\nu = \frac{5}{2}\right)$ kernel function

| **Dataset** | $\hat{H}_0$ (km s$^{-1}$ Mpc$^{-1}$) | $d(\hat{H}_0, \hat{H}_0^R)$ | $d(\hat{H}_0, \hat{H}_0^{TRGB})$ | $d(\hat{H}_0, \hat{H}_0^{HW})$ | $d(\hat{H}_0, \hat{H}_0^{CM})$ | $d(\hat{H}_0, \hat{H}_0^P)$ | $d(\hat{H}_0, \hat{H}_0^{DES})$ |
|---|---|---|---|---|---|---|---|
| CC | $68.735 \pm 5.284$ | $-0.9814$ | $-0.1897$ | $-0.8224$ | $-1.1930$ | $0.2515$ | $0.2473$ |
| CC+SN | $68.156 \pm 3.427$ | $-1.5629$ | $-0.4197$ | $-1.3448$ | $-1.8851$ | $0.2182$ | $0.2099$ |
| CC+SN+BAO | $68.809 \pm 2.855$ | $-1.5980$ | $-0.2889$ | $-1.3514$ | $-1.9744$ | $0.4860$ | $0.4605$ |
| CC+$\hat{H}_0^R$ | $73.820 \pm 3.006$ | $-0.1139$ | $1.1303$ | $0.1505$ | $-0.4444$ | $\mathbf{2.1066}$ | $\mathbf{2.0055}$ |
| CC+SN+$\hat{H}_0^R$ | $71.600 \pm 1.828$ | $-1.0156$ | $0.6829$ | $-0.6809$ | $-1.5104$ | $\mathbf{2.2169}$ | $1.9692$ |
| CC+SN+BAO+$\hat{H}_0^R$ | $71.500 \pm 1.622$ | $-1.1155$ | $0.6804$ | $-0.7659$ | $-1.6484$ | $\mathbf{2.4150}$ | $\mathbf{2.0917}$ |
| CC+$\hat{H}_0^{TRGB}$ | $69.638 \pm 3.168$ | $-1.2541$ | $-0.0438$ | $-1.0185$ | $-1.5928$ | $0.6978$ | $0.6674$ |
| CC+SN+$\hat{H}_0^{TRGB}$ | $69.002 \pm 1.785$ | $\mathbf{-2.0467}$ | $-0.3060$ | $-1.7435$ | $\mathbf{-2.5895}$ | $0.8643$ | $0.7641$ |
| CC+SN+BAO+$\hat{H}_0^{TRGB}$ | $69.248 \pm 1.595$ | $\mathbf{-2.0546}$ | $-0.2226$ | $-1.7383$ | $\mathbf{-2.6342}$ | $1.1053$ | $0.9536$ |
| CC+$\hat{H}_0^{HW}$ | $72.991 \pm 2.644$ | $-0.3829$ | $0.9799$ | $-0.0984$ | $-0.7530$ | $\mathbf{2.0774}$ | $1.9520$ |
| CC+SN+$\hat{H}_0^{HW}$ | $71.229 \pm 1.676$ | $-1.2090$ | $0.5639$ | $-0.8676$ | $-1.7366$ | $\mathbf{2.1890}$ | $1.9098$ |
| CC+SN+BAO+$\hat{H}_0^{HW}$ | $71.180 \pm 1.502$ | $-1.2883$ | $0.5696$ | $-0.9346$ | $-1.8504$ | $\mathbf{2.3873}$ | $\mathbf{2.0300}$ |
| CC+$\hat{H}_0^{CM}$ | $74.951 \pm 2.607$ | $0.2298$ | $1.5968$ | $0.5304$ | $-0.1288$ | $\mathbf{2.8448}$ | $\mathbf{2.6688}$ |
| CC+SN+$\hat{H}_0^{CM}$ | $72.583 \pm 1.703$ | $-0.6569$ | $1.0905$ | $-0.2981$ | $-1.1568$ | $\mathbf{2.9197}$ | $\mathbf{2.5562}$ |
| CC+SN+BAO+$\hat{H}_0^{CM}$ | $72.377 \pm 1.524$ | $-0.7763$ | $1.0580$ | $-0.4042$ | $-1.3105$ | $\mathbf{3.1026}$ | $\mathbf{2.6478}$ |
| CC+$\hat{H}_0^P$ | $67.399 \pm 0.247$ | $\mathbf{-3.7136}$ | $-1.2531$ | $\mathbf{-3.4350}$ | $\mathbf{-4.6823}$ | $-0.0016$ | $-0.0008$ |
| CC+SN+$\hat{H}_0^P$ | $67.438 \pm 0.232$ | $\mathbf{-3.6966}$ | $-1.2341$ | $\mathbf{-3.4168}$ | $\mathbf{-4.6655}$ | $0.0686$ | $0.0337$ |
| CC+SN+BAO+$\hat{H}_0^P$ | $67.509 \pm 0.228$ | $\mathbf{-3.6586}$ | $-1.1971$ | $\mathbf{-3.3761}$ | $\mathbf{-4.6247}$ | $0.1988$ | $0.0972$ |
| CC+$\hat{H}_0^{DES}$ | $67.393 \pm 1.210$ | $\mathbf{-3.1239}$ | $-1.0686$ | $\mathbf{-2.8310}$ | $\mathbf{-3.8434}$ | $-0.0054$ | $-0.0043$ |
| CC+SN+$\hat{H}_0^{DES}$ | $67.555 \pm 0.879$ | $\mathbf{-3.2977}$ | $-1.0723$ | $\mathbf{-3.0020}$ | $\mathbf{-4.1113}$ | $0.1536$ | $0.1103$ |
| CC+SN+BAO+$\hat{H}_0^{DES}$ | $67.804 \pm 0.832$ | $\mathbf{-3.2063}$ | $-0.9624$ | $\mathbf{-2.9040}$ | $\mathbf{-4.0253}$ | $0.4161$ | $0.2928$ |

Table F.5: MCMC GPR Results with Matérn $\left(\nu = \frac{7}{2}\right)$ kernel function

| **Dataset** | $\hat{H}_0$ (km s$^{-1}$ Mpc$^{-1}$) | $d(\hat{H}_0, \hat{H}_0^R)$ | $d(\hat{H}_0, \hat{H}_0^{TRGB})$ | $d(\hat{H}_0, \hat{H}_0^{HW})$ | $d(\hat{H}_0, \hat{H}_0^{CM})$ | $d(\hat{H}_0, \hat{H}_0^P)$ | $d(\hat{H}_0, \hat{H}_0^{DES})$ |
|---|---|---|---|---|---|---|---|
| CC | $68.528 \pm 5.360$ | $-1.0055$ | $-0.2237$ | $-0.8486$ | $-1.2144$ | $0.2095$ | $0.2061$ |
| CC+SN | $68.160 \pm 3.331$ | $-1.5964$ | $-0.4276$ | $-1.3744$ | $-1.9272$ | $0.2257$ | $0.2167$ |
| CC+SN+BAO | $68.854 \pm 2.741$ | $-1.6308$ | $-0.2836$ | $-1.3783$ | $\mathbf{-2.0204}$ | $0.5219$ | $0.4923$ |
| CC+$\hat{H}_0^R$ | $73.826 \pm 3.000$ | $-0.1122$ | $1.1337$ | $0.1526$ | $-0.4432$ | $\mathbf{2.1127}$ | $\mathbf{2.0110}$ |
| CC+SN+$\hat{H}_0^R$ | $71.617 \pm 1.811$ | $-1.0140$ | $0.6921$ | $-0.6777$ | $-1.5113$ | $\mathbf{2.2444}$ | $1.9901$ |
| CC+SN+BAO+$\hat{H}_0^R$ | $71.516 \pm 1.600$ | $-1.1160$ | $0.6908$ | $-0.7643$ | $-1.6528$ | $\mathbf{2.4557}$ | $\mathbf{2.1200}$ |
| CC+$\hat{H}_0^{TRGB}$ | $69.631 \pm 3.160$ | $-1.2586$ | $-0.0459$ | $-1.0227$ | $-1.5983$ | $0.6973$ | $0.6667$ |
| CC+SN+$\hat{H}_0^{TRGB}$ | $69.013 \pm 1.764$ | $\mathbf{-2.0543}$ | $-0.3036$ | $-1.7499$ | $\mathbf{-2.6013}$ | $0.8795$ | $0.7757$ |
| CC+SN+BAO+$\hat{H}_0^{TRGB}$ | $69.269 \pm 1.563$ | $\mathbf{-2.0638}$ | $-0.2158$ | $-1.7456$ | $\mathbf{-2.6502}$ | $1.1392$ | $0.9781$ |
| CC+$\hat{H}_0^{HW}$ | $72.995 \pm 2.640$ | $-0.3819$ | $0.9825$ | $-0.0970$ | $-0.7525$ | $\mathbf{2.0826}$ | $1.9566$ |
| CC+SN+$\hat{H}_0^{HW}$ | $71.244 \pm 1.662$ | $-1.2076$ | $0.5721$ | $-0.8649$ | $-1.7377$ | $\mathbf{2.2153}$ | $1.9290$ |
| CC+SN+BAO+$\hat{H}_0^{HW}$ | $71.196 \pm 1.482$ | $-1.2885$ | $0.5792$ | $-0.9330$ | $-1.8543$ | $\mathbf{2.4267}$ | $\mathbf{2.0566}$ |
| CC+$\hat{H}_0^{CM}$ | $74.962 \pm 2.604$ | $0.2336$ | $1.6014$ | $0.5345$ | $-0.1252$ | $\mathbf{2.8520}$ | $\mathbf{2.6752}$ |
| CC+SN+$\hat{H}_0^{CM}$ | $72.595 \pm 1.689$ | $-0.6545$ | $1.0992$ | $-0.2943$ | $-1.1565$ | $\mathbf{2.9483}$ | $\mathbf{2.5767}$ |
| CC+SN+BAO+$\hat{H}_0^{CM}$ | $72.385 \pm 1.505$ | $-0.7769$ | $1.0667$ | $-0.4028$ | $-1.3144$ | $\mathbf{3.1438}$ | $\mathbf{2.6745}$ |
| CC+$\hat{H}_0^P$ | $67.397 \pm 0.247$ | $\mathbf{-3.7146}$ | $-1.2539$ | $\mathbf{-3.4360}$ | $\mathbf{-4.6832}$ | $-0.0046$ | $-0.0023$ |
| CC+SN+$\hat{H}_0^P$ | $67.437 \pm 0.231$ | $\mathbf{-3.6973}$ | $-1.2347$ | $\mathbf{-3.4175}$ | $\mathbf{-4.6662}$ | $0.0667$ | $0.0327$ |
| CC+SN+BAO+$\hat{H}_0^P$ | $67.399 \pm 0.250$ | $\mathbf{-3.7131}$ | $-1.2530$ | $\mathbf{-3.4344}$ | $\mathbf{-4.6813}$ | $-0.0022$ | $-0.0011$ |
| CC+$\hat{H}_0^{DES}$ | $67.389 \pm 1.147$ | $\mathbf{-3.1753}$ | $-1.0862$ | $\mathbf{-2.8823}$ | $\mathbf{-3.9136}$ | $-0.0085$ | $-0.0067$ |
| CC+SN+$\hat{H}_0^{DES}$ | $67.552 \pm 0.873$ | $\mathbf{-3.3030}$ | $-1.0750$ | $\mathbf{-3.0074}$ | $\mathbf{-4.1183}$ | $0.1512$ | $0.1083$ |
| CC+SN+BAO+$\hat{H}_0^{DES}$ | $67.817 \pm 0.820$ | $\mathbf{-3.2073}$ | $-0.9581$ | $\mathbf{-2.9047}$ | $\mathbf{-4.0292}$ | $0.4343$ | $0.3040$ |

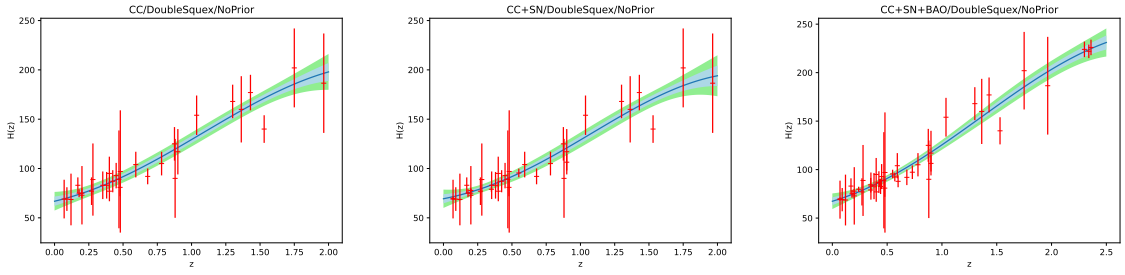
Table F.6: MCMC GPR Results with Matérn $\left(\nu = \frac{9}{2}\right)$ kernel function

| **Dataset** | $\hat{H}_0$ (km s$^{-1}$ Mpc$^{-1}$) | $d(\hat{H}_0, \hat{H}_0^R)$ | $d(\hat{H}_0, \hat{H}_0^{TRGB})$ | $d(\hat{H}_0, \hat{H}_0^{HW})$ | $d(\hat{H}_0, \hat{H}_0^{CM})$ | $d(\hat{H}_0, \hat{H}_0^P)$ | $d(\hat{H}_0, \hat{H}_0^{DES})$ |
|---|---|---|---|---|---|---|---|
| CC | $69.583 \pm 5.329$ | $-0.8234$ | $-0.0383$ | $-0.6644$ | $-1.0320$ | $0.4078$ | $0.4012$ |
| CC+SN | $68.303 \pm 3.365$ | $-1.5467$ | $-0.3875$ | $-1.3255$ | $-1.8738$ | $0.2653$ | $0.2549$ |
| CC+SN+BAO | $68.941 \pm 2.747$ | $-1.6019$ | $-0.2571$ | $-1.3492$ | $-1.9902$ | $0.5519$ | $0.5208$ |
| CC+$\hat{H}_0^R$ | $73.867 \pm 2.999$ | $-0.1006$ | $1.1456$ | $0.1645$ | $-0.4314$ | $\mathbf{2.1270}$ | $\mathbf{2.0245}$ |
| CC+SN+$\hat{H}_0^R$ | $71.663 \pm 1.810$ | $-0.9960$ | $0.7101$ | $-0.6591$ | $-1.4928$ | $\mathbf{2.2703}$ | $\mathbf{2.0128}$ |
| CC+SN+BAO+$\hat{H}_0^R$ | $71.555 \pm 1.598$ | $-1.1005$ | $0.7069$ | $-0.7480$ | $-1.6369$ | $\mathbf{2.4819}$ | $\mathbf{2.1420}$ |
| CC+$\hat{H}_0^{TRGB}$ | $69.700 \pm 3.163$ | $-1.2385$ | $-0.0271$ | $-1.0025$ | $-1.5775$ | $0.7182$ | $0.6868$ |
| CC+SN+$\hat{H}_0^{TRGB}$ | $69.077 \pm 1.768$ | $\mathbf{-2.0267}$ | $-0.2785$ | $-1.7216$ | $\mathbf{-2.5719}$ | $0.9128$ | $0.8054$ |
| CC+SN+BAO+$\hat{H}_0^{TRGB}$ | $69.317 \pm 1.563$ | $\mathbf{-2.0441}$ | $-0.1965$ | $-1.7251$ | $\mathbf{-2.6296}$ | $1.1681$ | $1.0029$ |
| CC+$\hat{H}_0^{HW}$ | $73.035 \pm 2.639$ | $-0.3697$ | $0.9947$ | $-0.0845$ | $-0.7400$ | $\mathbf{2.0977}$ | $1.9707$ |
| CC+SN+$\hat{H}_0^{HW}$ | $71.290 \pm 1.661$ | $-1.1890$ | $0.5903$ | $-0.8457$ | $-1.7184$ | $\mathbf{2.2419}$ | $1.9521$ |
| CC+SN+BAO+$\hat{H}_0^{HW}$ | $71.235 \pm 1.481$ | $-1.2722$ | $0.5956$ | $-0.9159$ | $-1.8374$ | $\mathbf{2.4532}$ | $\mathbf{2.0786}$ |
| CC+$\hat{H}_0^{CM}$ | $74.993 \pm 2.603$ | $0.2434$ | $1.6116$ | $0.5446$ | $-0.1152$ | $\mathbf{2.8651}$ | $\mathbf{2.6873}$ |
| CC+SN+$\hat{H}_0^{CM}$ | $72.629 \pm 1.687$ | $-0.6412$ | $1.1135$ | $-0.2802$ | $-1.1430$ | $\mathbf{2.9723}$ | $\mathbf{2.5967}$ |
| CC+SN+BAO+$\hat{H}_0^{CM}$ | $72.412 \pm 1.500$ | $-0.7666$ | $1.0790$ | $-0.3917$ | $-1.3044$ | $\mathbf{3.1695}$ | $\mathbf{2.6942}$ |
| CC+$\hat{H}_0^P$ | $67.404 \pm 0.247$ | $\mathbf{-3.7109}$ | $-1.2505$ | $\mathbf{-3.4321}$ | $\mathbf{-4.6793}$ | $0.0073$ | $0.0036$ |
| CC+SN+$\hat{H}_0^P$ | $67.446 \pm 0.231$ | $\mathbf{-3.6923}$ | $-1.2299$ | $\mathbf{-3.4121}$ | $\mathbf{-4.6608}$ | $0.0833$ | $0.0408$ |
| CC+SN+BAO+$\hat{H}_0^P$ | $67.520 \pm 0.227$ | $\mathbf{-3.6531}$ | $-1.1917$ | $\mathbf{-3.3702}$ | $\mathbf{-4.6188}$ | $0.2180$ | $0.1066$ |
| CC+$\hat{H}_0^{DES}$ | $67.421 \pm 1.148$ | $\mathbf{-3.1600}$ | $-1.0719$ | $\mathbf{-2.8663}$ | $\mathbf{-3.8971}$ | $0.0164$ | $0.0129$ |
| CC+SN+$\hat{H}_0^{DES}$ | $67.587 \pm 0.875$ | $\mathbf{-3.2847}$ | $-1.0579$ | $\mathbf{-2.9881}$ | $\mathbf{-4.0985}$ | $0.1860$ | $0.1333$ |
| CC+SN+BAO+$\hat{H}_0^{DES}$ | $67.845 \pm 0.821$ | $\mathbf{-3.1932}$ | $-0.9447$ | $\mathbf{-2.8898}$ | $\mathbf{-4.0140}$ | $0.4627$ | $0.3240$ |

Table F.7: MCMC GPR Results with Cauchy kernel function

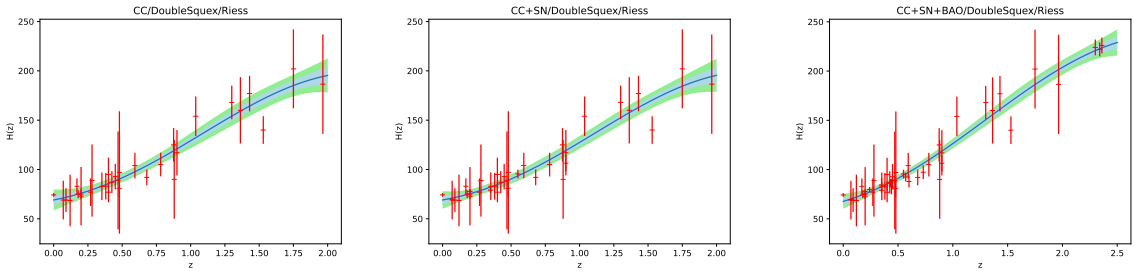
# Appendix G: TPR Plots

Figures for the square exponential kernel were presented in Section 4.4. Similar figures for all the other kernel functions considered are shown in this Appendix.



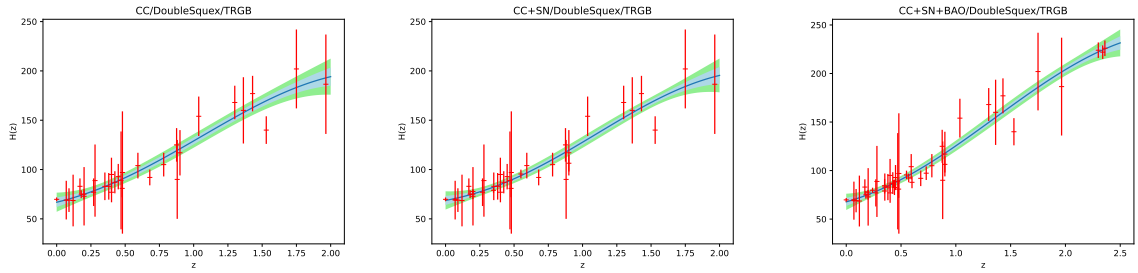

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.1: TPR reconstructions of $\hat{H}(z)$ against $z$ for double square exponential kernel function and without adding any $H_0$ estimate to the data



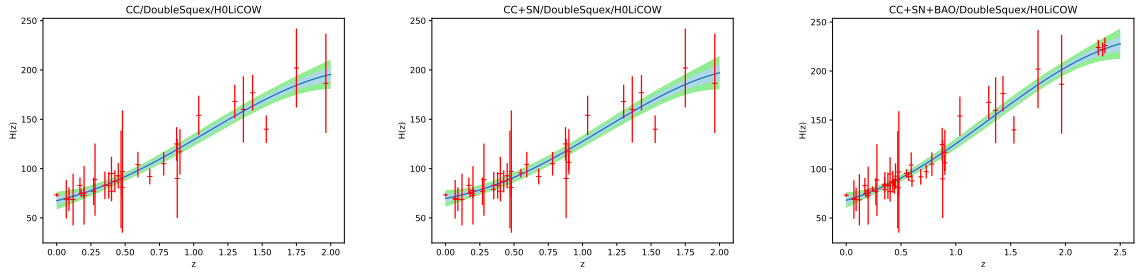

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.2: TPR reconstructions of $\hat{H}(z)$ against $z$ for double square exponential kernel function and using $\hat{H}_0^R$



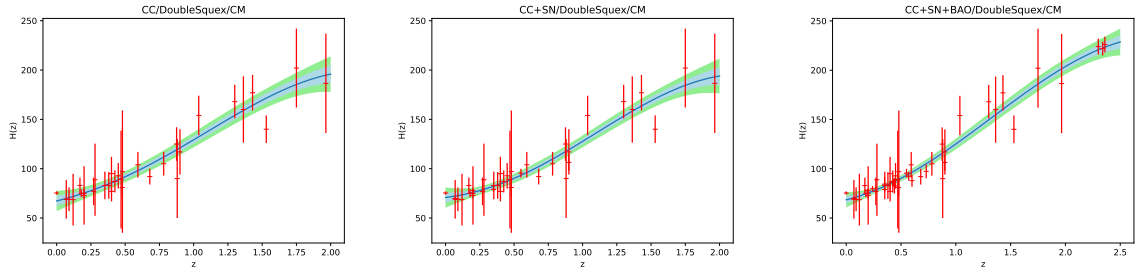

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.3: TPR reconstructions of $\hat{H}(z)$ against $z$ for double square exponential kernel function and using $\hat{H}_0^{TRGB}$



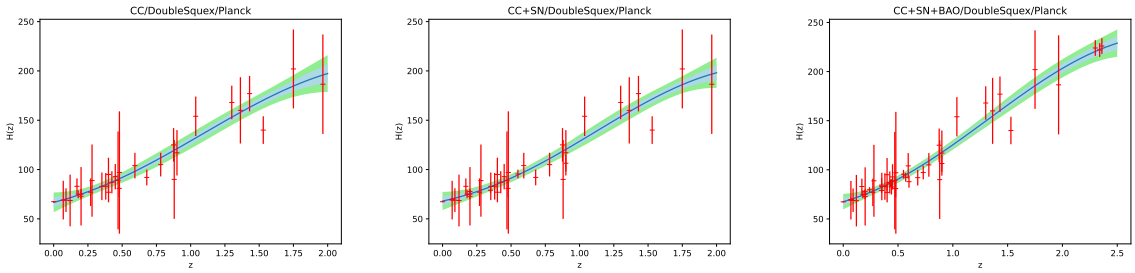

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.4: TPR reconstructions of $\hat{H}(z)$ against $z$ for double square exponential kernel function and using $\hat{H}_0^{HW}$



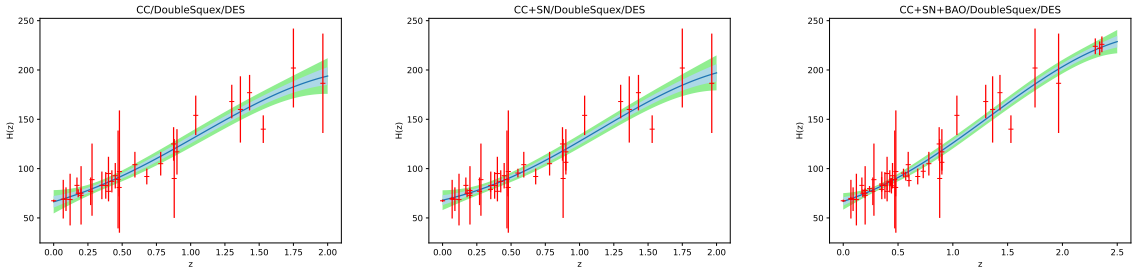

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.5: TPR reconstructions of $\hat{H}(z)$ against $z$ for double square exponential kernel function and using $\hat{H}_0^{CM}$



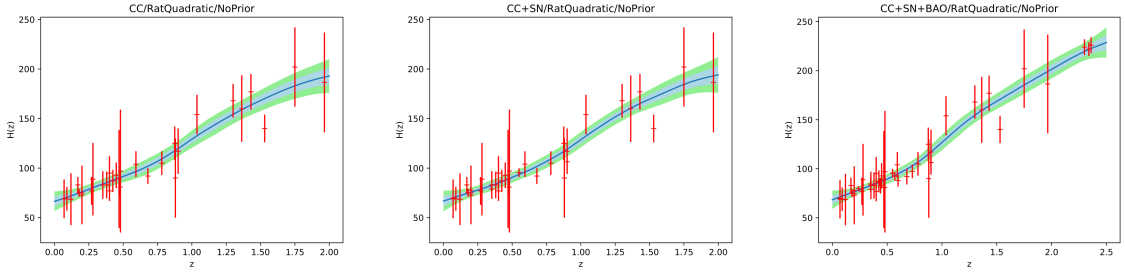

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.6: TPR reconstructions of $\hat{H}(z)$ against $z$ for double square exponential kernel function and using $\hat{H}_0^P$



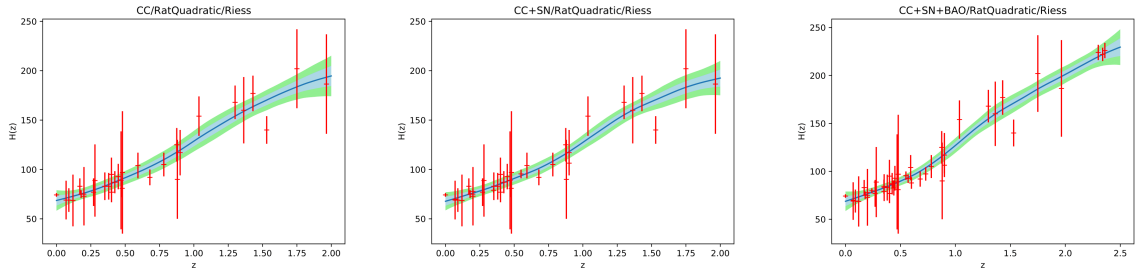

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.7: TPR reconstructions of $\hat{H}(z)$ against $z$ for double square exponential kernel function and using $\hat{H}_0^{DES}$



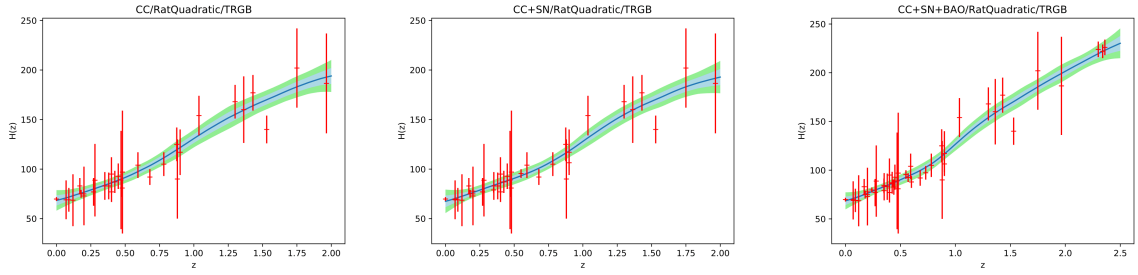

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.8: TPR reconstructions of $\hat{H}(z)$ against $z$ for rational quadratic kernel function and without adding any $H_0$ estimate to the data



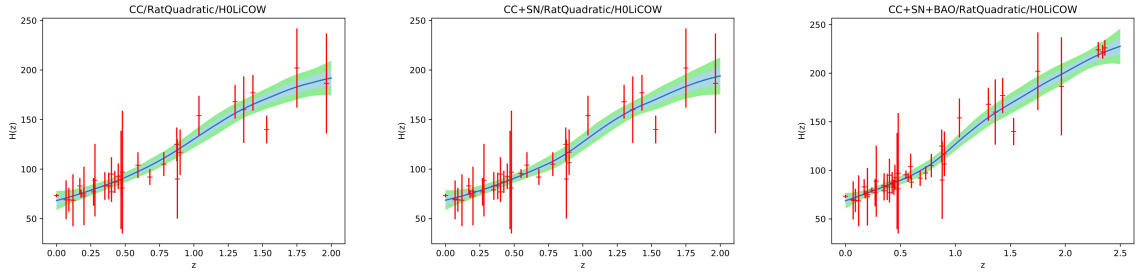

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.9: TPR reconstructions of $\hat{H}(z)$ against $z$ for rational quadratic kernel function and using $\hat{H}_0^R$



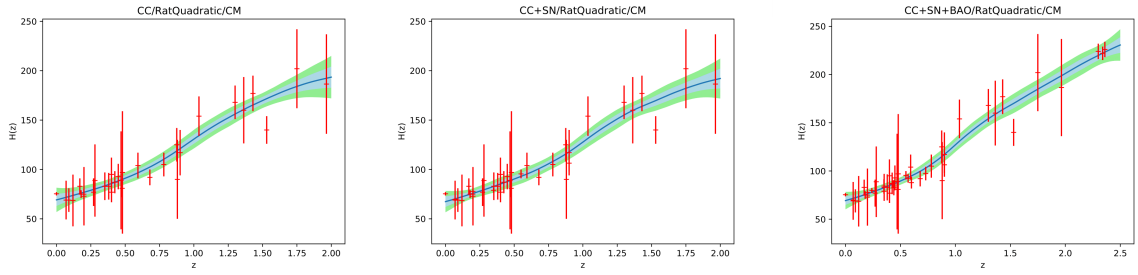

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.10: TPR reconstructions of $\hat{H}(z)$ against $z$ for rational quadratic kernel function and using $\hat{H}_0^{TRGB}$



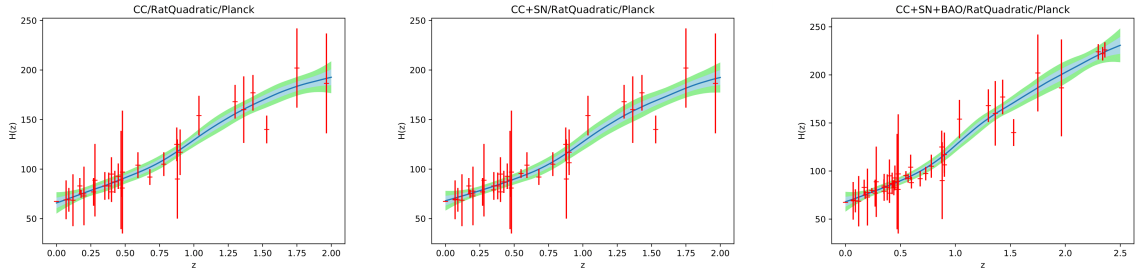

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.11: TPR reconstructions of $\hat{H}(z)$ against $z$ for rational quadratic kernel function and using $\hat{H}_0^{HW}$



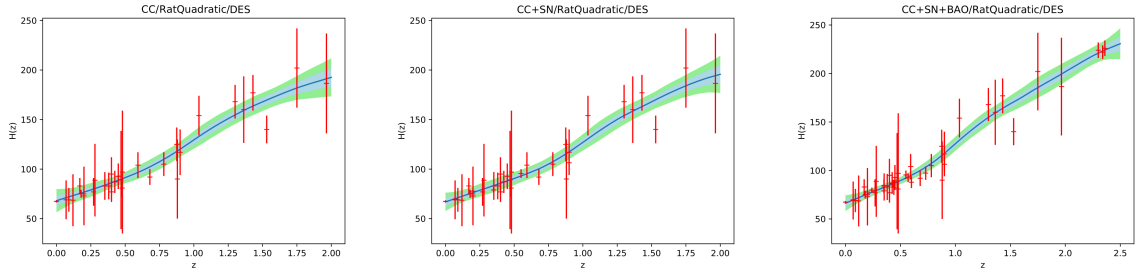

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.12: TPR reconstructions of $\hat{H}(z)$ against $z$ for rational quadratic kernel function and using $\hat{H}_0^{CM}$



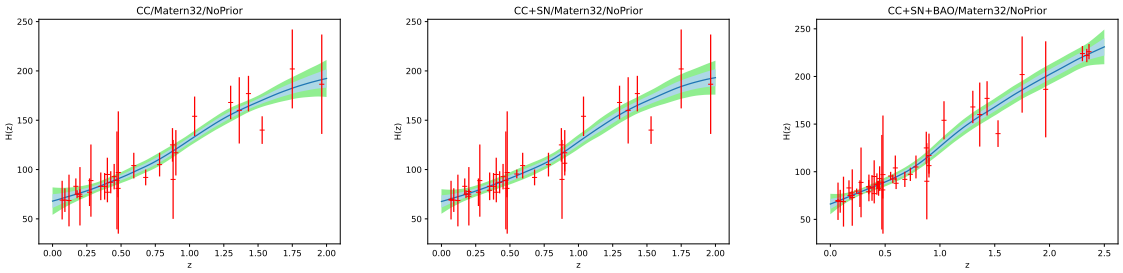

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.13: TPR reconstructions of $\hat{H}(z)$ against $z$ for rational quadratic kernel function and using $\hat{H}_0^{P}$



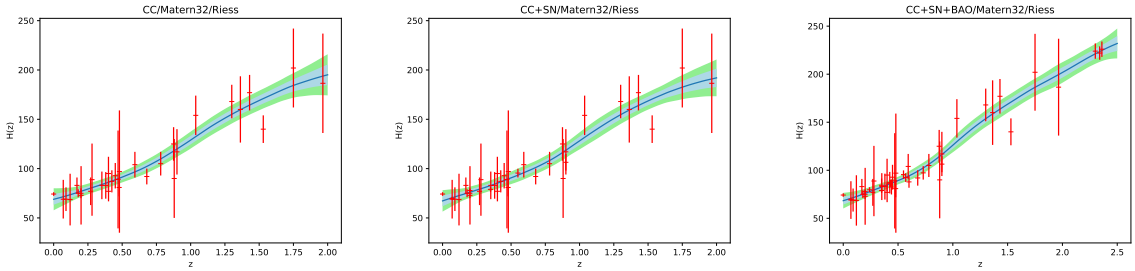

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.14: TPR reconstructions of $\hat{H}(z)$ against $z$ for rational quadratic kernel function and using $\hat{H}_0^{DES}$



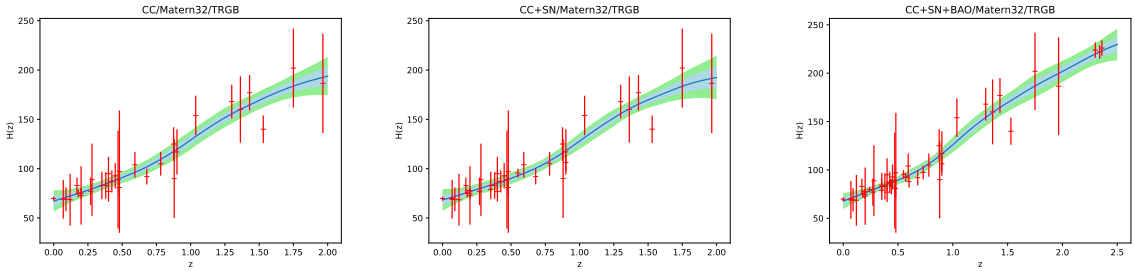

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.15: TPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{3}{2}\right)$ kernel function and without adding any $H_0$ estimate to the data



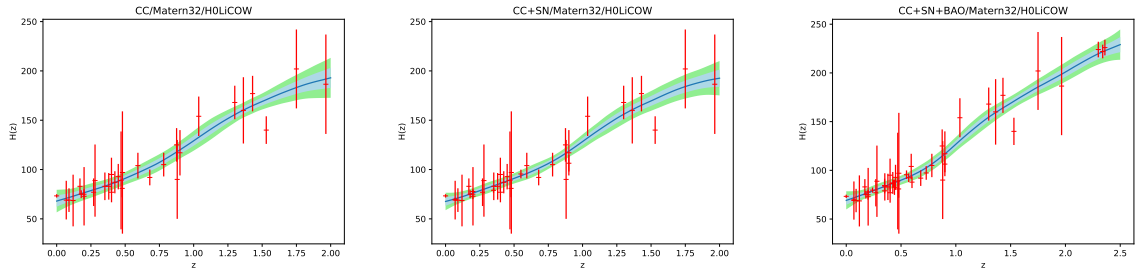

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.16: TPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{3}{2}\right)$ kernel function and using $\hat{H}_0^R$



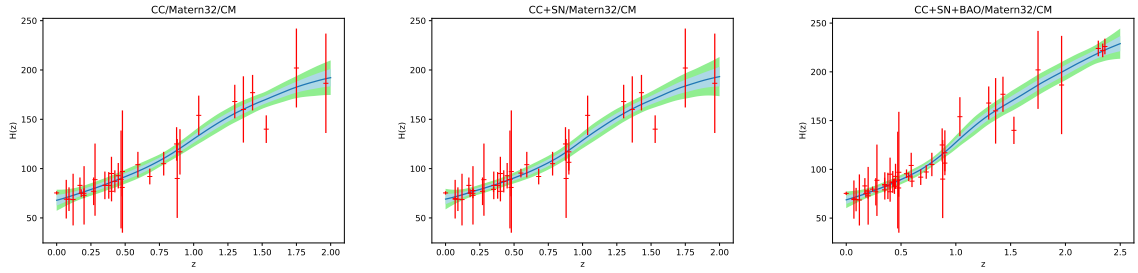

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.17: TPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{3}{2}\right)$ kernel function and using $\hat{H}_0^{TRGB}$



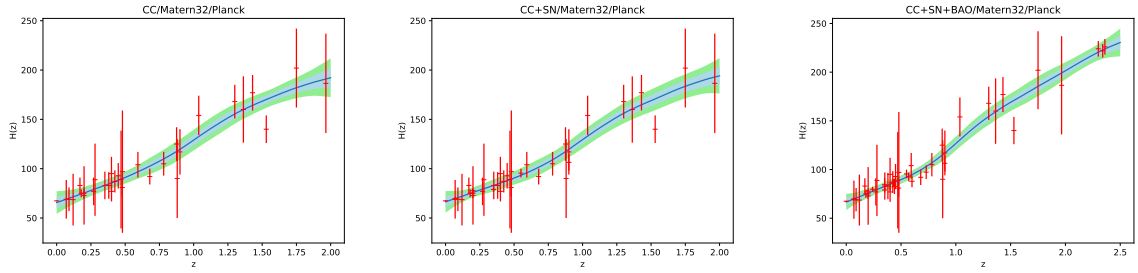

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.18: TPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{3}{2}\right)$ kernel function and using $\hat{H}_0^{HW}$



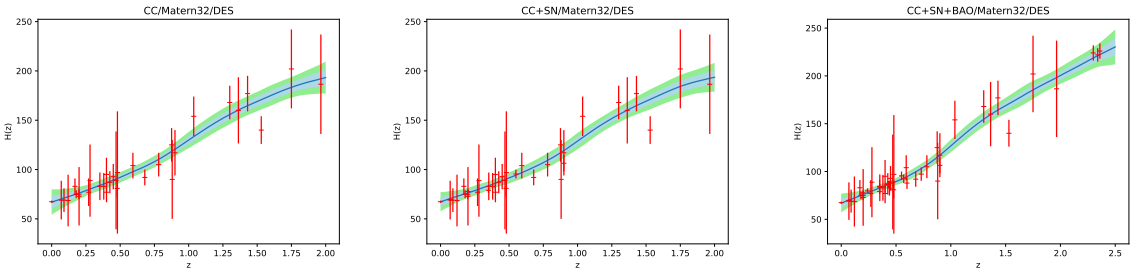

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.19: TPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{3}{2}\right)$ kernel function and using $\hat{H}_0^{CM}$



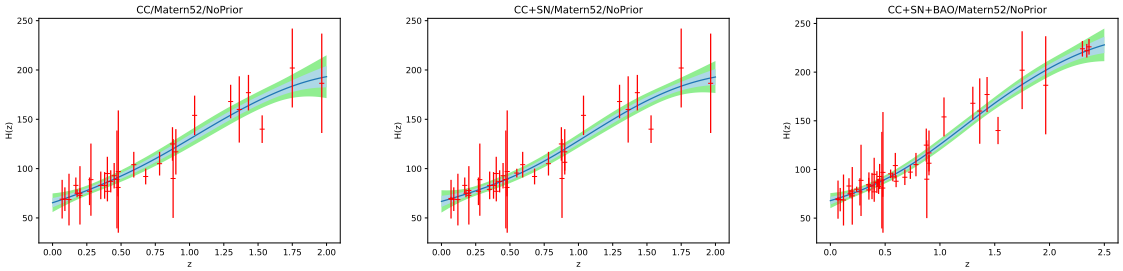

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.20: TPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{3}{2}\right)$ kernel function and using $\hat{H}_0^{P}$



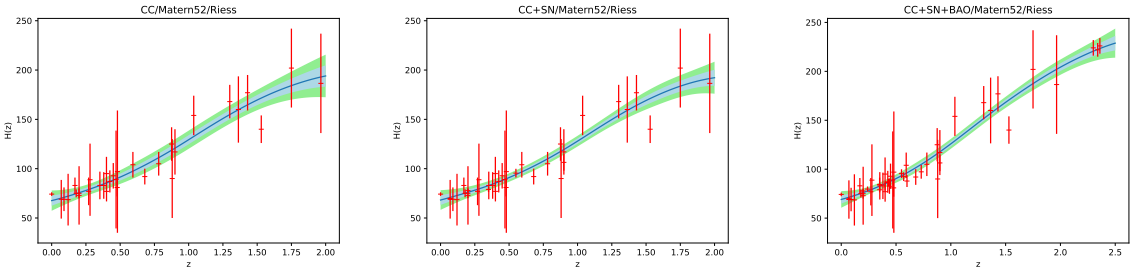

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.21: TPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{3}{2}\right)$ kernel function and using $\hat{H}_0^{DES}$



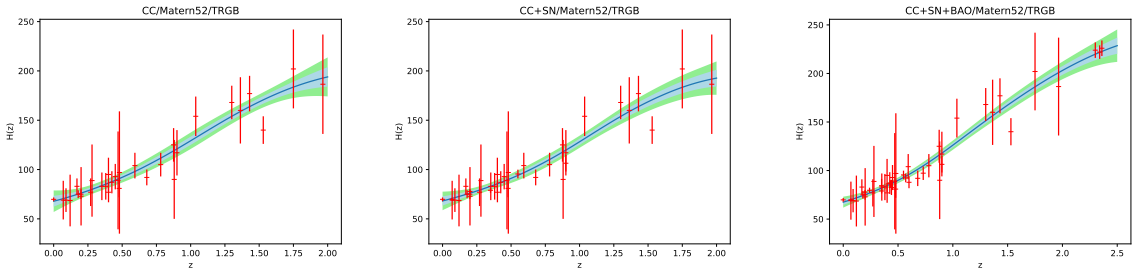

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.22: TPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{5}{2}\right)$ kernel function and without adding any $H_0$ estimate to the data



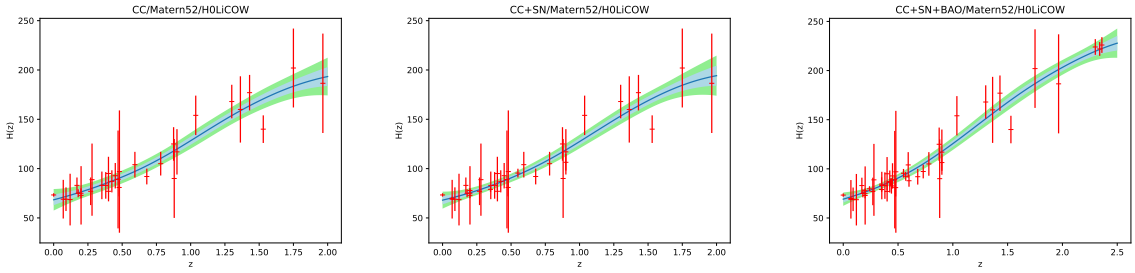

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.23: TPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{5}{2}\right)$ kernel function and using $\hat{H}_0^R$



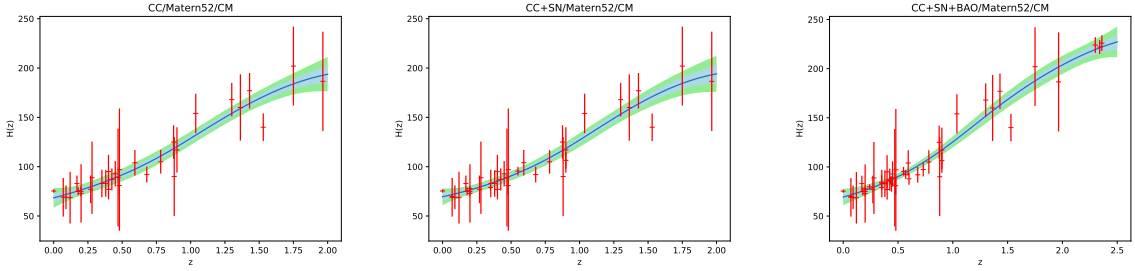

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.24: TPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{5}{2}\right)$ kernel function and using $\hat{H}_0^{TRGB}$



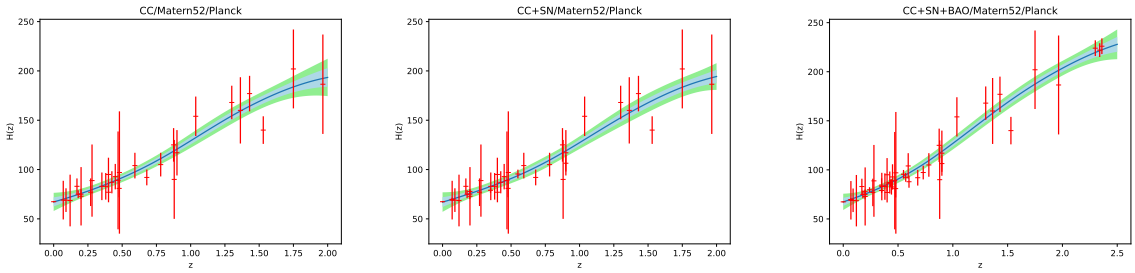

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.25: TPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{5}{2}\right)$ kernel function and using $\hat{H}_0^{HW}$



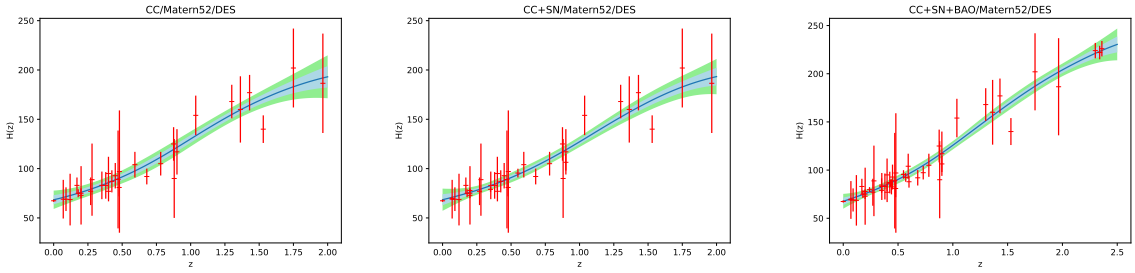

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.26: TPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{5}{2}\right)$ kernel function and using $\hat{H}_0^{CM}$



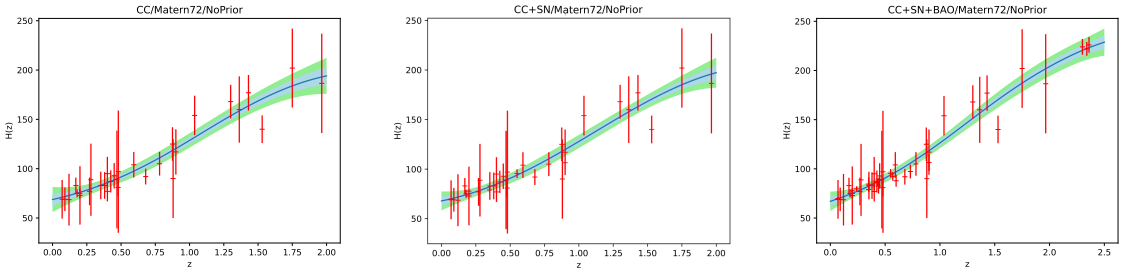

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.27: TPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{5}{2}\right)$ kernel function and using $\hat{H}_0^P$



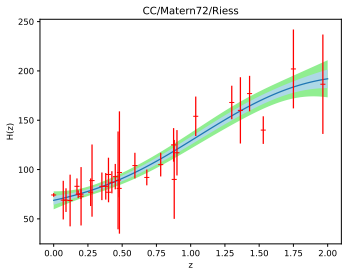

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.28: TPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{5}{2}\right)$ kernel function and using $\hat{H}_0^{DES}$

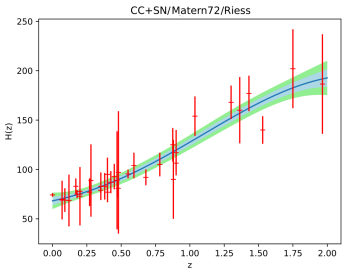



(a) CC (b) CC+SN (c) CC+SN+BAO

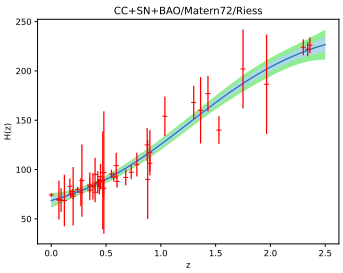
Figure G.29: TPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{7}{2}\right)$ kernel function and without adding any $H_0$ estimate to the data




(a) CC

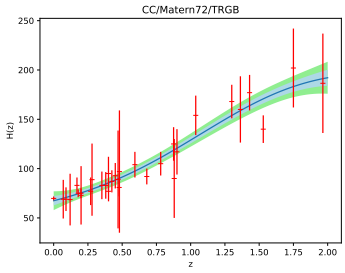



(b) CC+SN

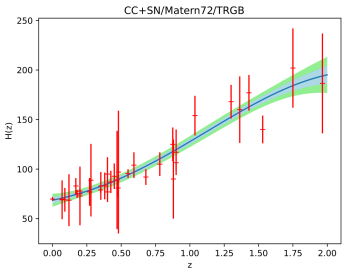



(c) CC+SN+BAO

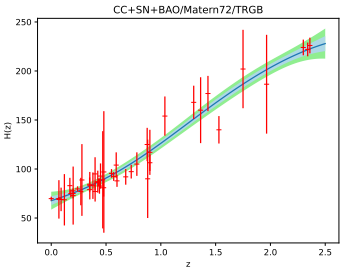
Figure G.30: TPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{7}{2}\right)$ kernel function and using $\hat{H}_0^R$




(a) CC

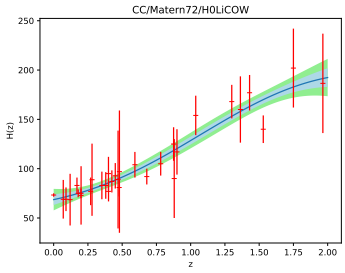



(b) CC+SN

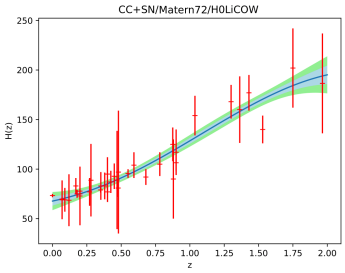



(c) CC+SN+BAO

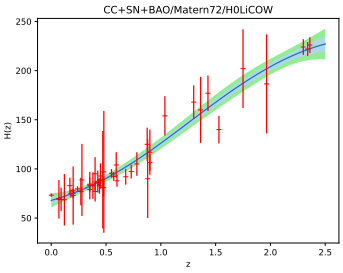
Figure G.31: TPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{7}{2}\right)$ kernel function and using $\hat{H}_0^{TRGB}$




(a) CC

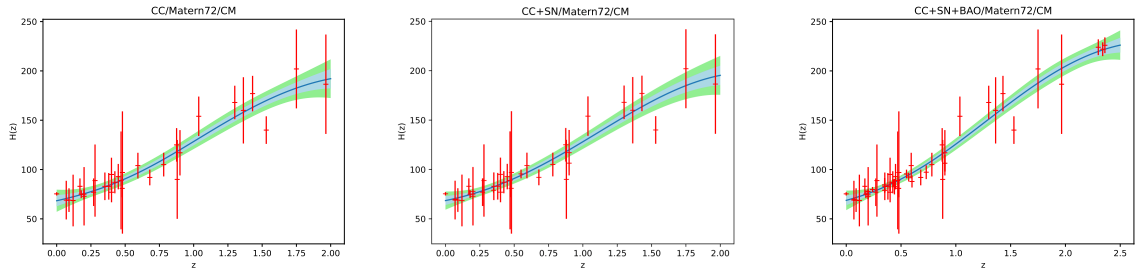



(b) CC+SN



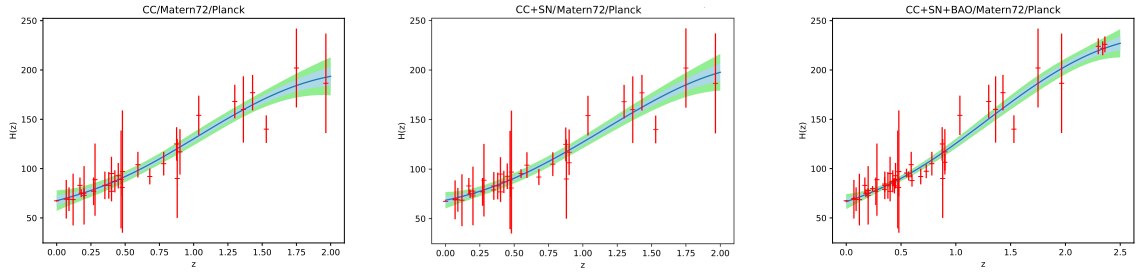

(c) CC+SN+BAO

Figure G.32: TPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{7}{2}\right)$ kernel function and using $\hat{H}_0^{HW}$



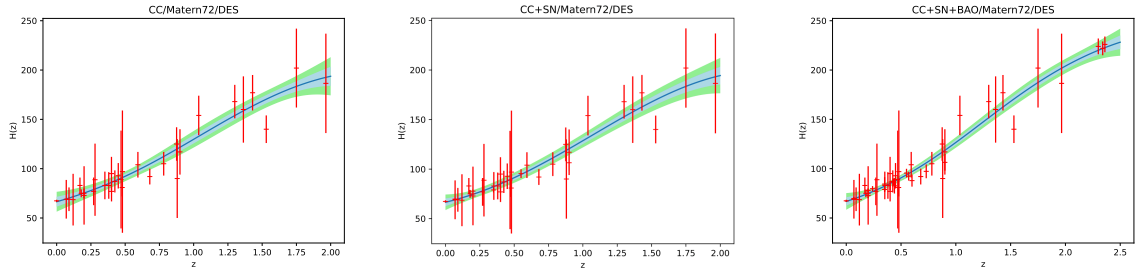

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.33: TPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{7}{2}\right)$ kernel function and using $\hat{H}_0^{CM}$



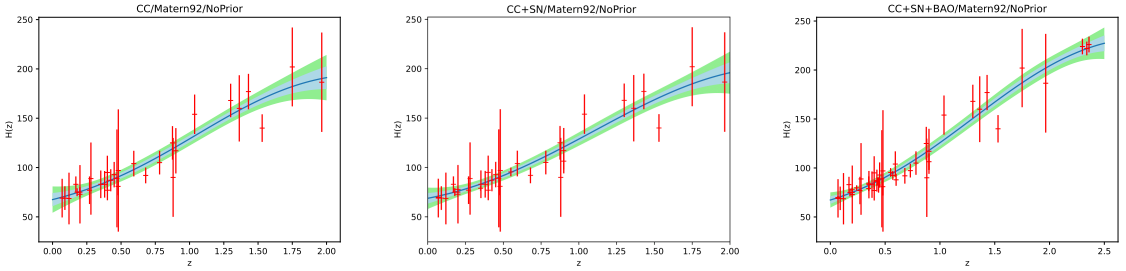

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.34: TPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{7}{2}\right)$ kernel function and using $\hat{H}_0^{P}$



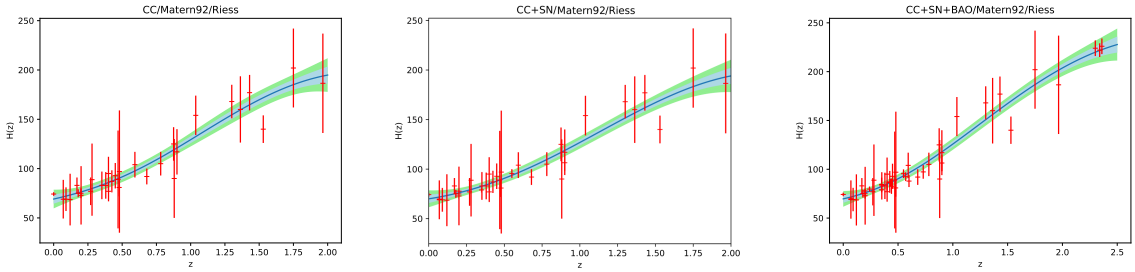

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.35: TPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{7}{2}\right)$ kernel function and using $\hat{H}_0^{DES}$



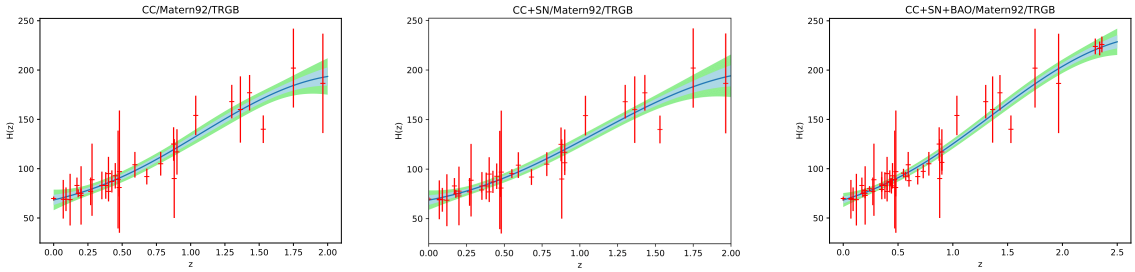

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.36: TPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{9}{2}\right)$ kernel function and without adding any $H_0$ estimate to the data



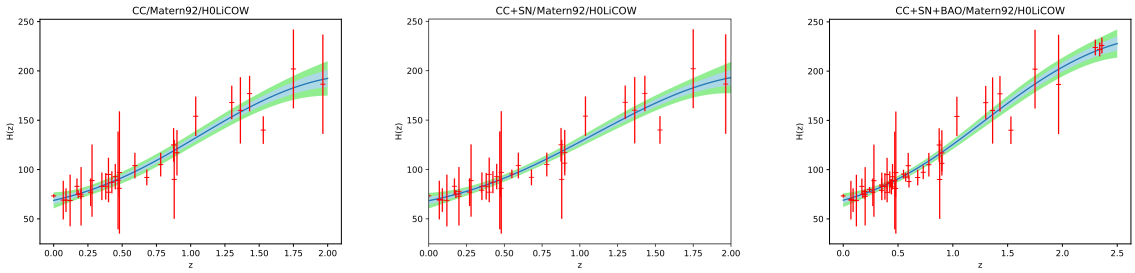

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.37: TPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{9}{2}\right)$ kernel function and using $\hat{H}_0^R$



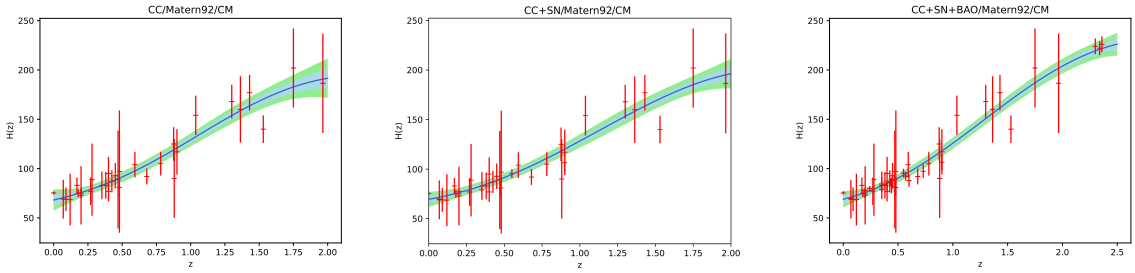

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.38: TPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{9}{2}\right)$ kernel function and using $\hat{H}_0^{TRGB}$



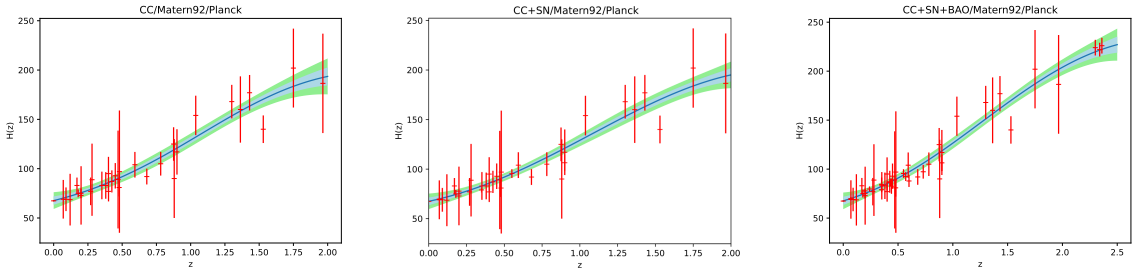

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.39: TPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{9}{2}\right)$ kernel function and using $\hat{H}_0^{HW}$



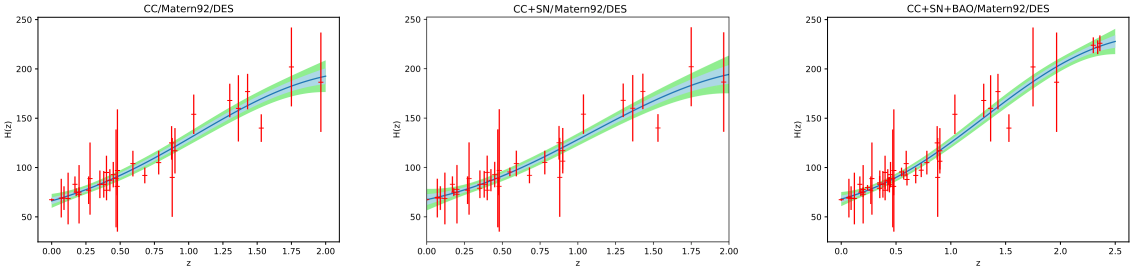

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.40: TPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{9}{2}\right)$ kernel function and using $\hat{H}_0^{CM}$



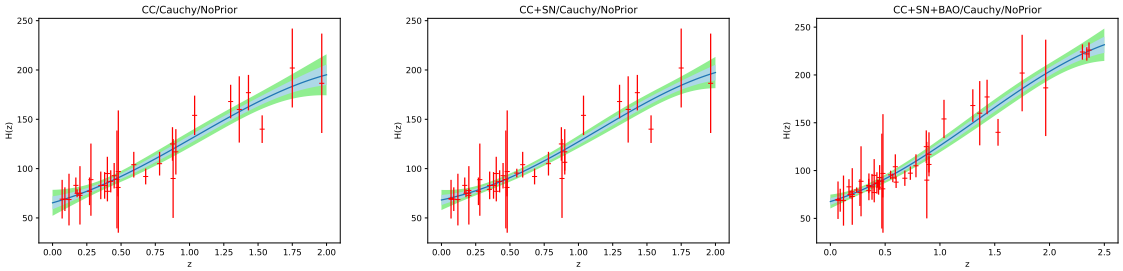

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.41: TPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{9}{2}\right)$ kernel function and using $\hat{H}_0^{P}$



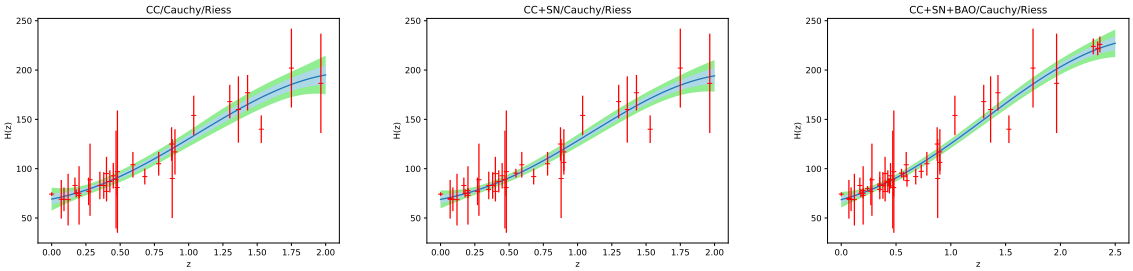

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.42: TPR reconstructions of $\hat{H}(z)$ against $z$ for Matérn $\left(\nu = \frac{9}{2}\right)$ kernel function and using $\hat{H}_0^{DES}$



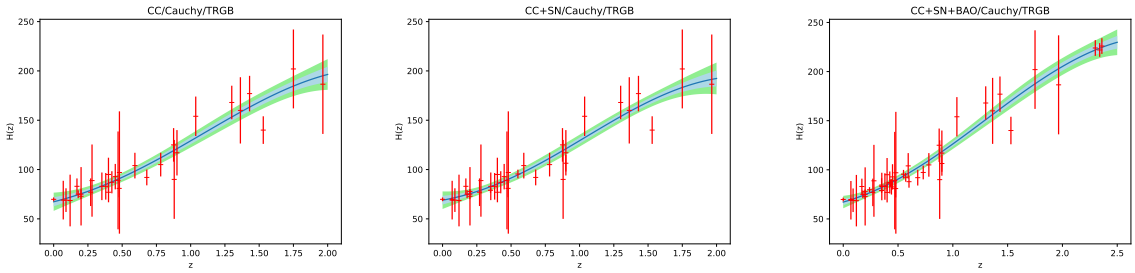

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.43: TPR reconstructions of $\hat{H}(z)$ against $z$ for Cauchy kernel function and without adding any $H_0$ estimate to the data



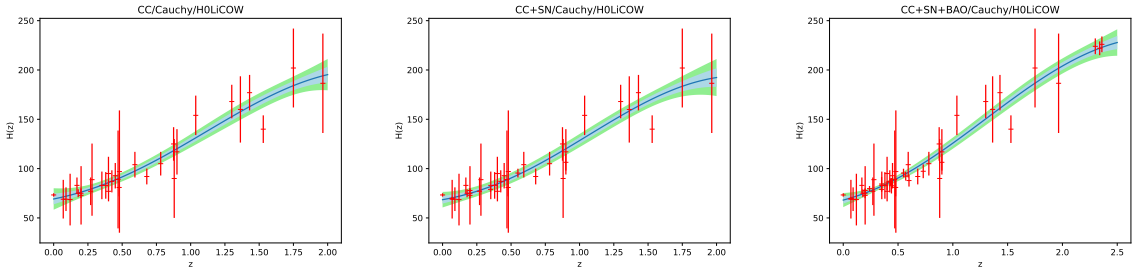

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.44: TPR reconstructions of $\hat{H}(z)$ against $z$ for Cauchy kernel function and using $\hat{H}_0^R$



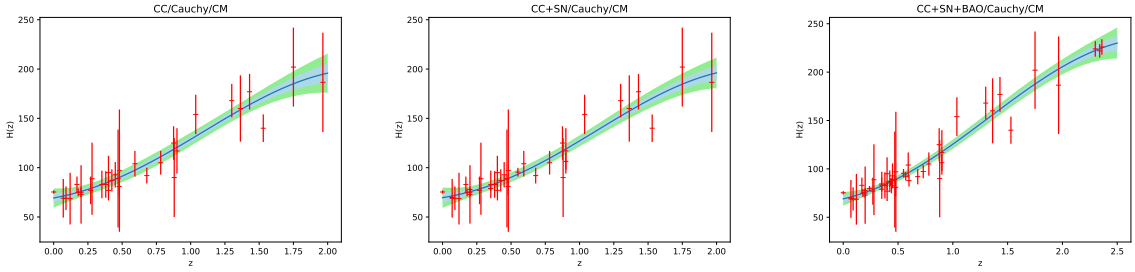

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.45: TPR reconstructions of $\hat{H}(z)$ against $z$ for Cauchy kernel function and using $\hat{H}_0^{TRGB}$



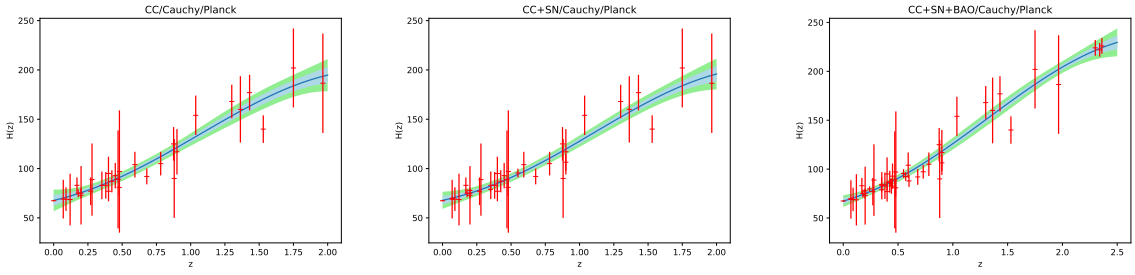

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.46: TPR reconstructions of $\hat{H}(z)$ against $z$ for Cauchy kernel function and using $\hat{H}_0^{HW}$



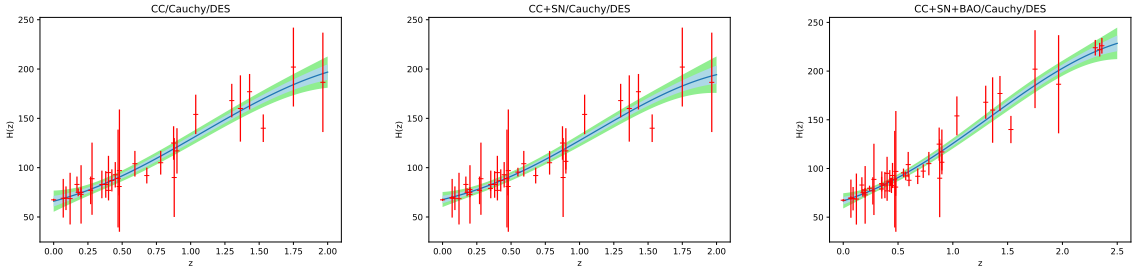

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.47: TPR reconstructions of $\hat{H}(z)$ against $z$ for Cauchy kernel function and using $\hat{H}_0^{CM}$




(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.48: TPR reconstructions of $\hat{H}(z)$ against $z$ for Cauchy kernel function and using $\hat{H}_0^P$




(a) CC (b) CC+SN (c) CC+SN+BAO

Figure G.49: TPR reconstructions of $\hat{H}(z)$ against $z$ for Cauchy kernel function and using $\hat{H}_0^{DES}$

# Appendix H: TPR Tables

The table for the square exponential kernel was presented in Section 4.4. Similar tables for all the other kernel functions considered are shown in this Appendix.

| **Dataset** | $\hat{H}_0$ | $d(\hat{H}_0, \hat{H}_0^R)$ | $d(\hat{H}_0, \hat{H}_0^{TRGB})$ | $d(\hat{H}_0, \hat{H}_0^{HW})$ | $d(\hat{H}_0, \hat{H}_0^{CM})$ | $d(\hat{H}_0, \hat{H}_0^P)$ | $d(\hat{H}_0, \hat{H}_0^{DES})$ |
|---|---|---|---|---|---|---|---|
| CC | $66.824 \pm 4.720$ | $-1.4620$ | $-0.5849$ | $-1.2908$ | $-1.7017$ | $-0.1214$ | $-0.1189$ |
| CC+SN | $69.335 \pm 4.644$ | $-0.9794$ | $-0.0927$ | $-0.8018$ | $-1.2180$ | $0.4142$ | $0.4054$ |
| CC+SN+BAO | $67.349 \pm 4.035$ | $-1.5523$ | $-0.5496$ | $-1.3592$ | $-1.8306$ | $-0.0126$ | $-0.0122$ |
| CC+$\hat{H}_0^R$ | $69.082 \pm 5.270$ | $-0.9216$ | $-0.1282$ | $-0.7617$ | $-1.1332$ | $0.3178$ | $0.3125$ |
| CC+SN+$\hat{H}_0^R$ | $69.080 \pm 4.455$ | $-1.0681$ | $-0.1486$ | $-0.8850$ | $-1.3169$ | $0.3749$ | $0.3662$ |
| CC+SN+BAO+$\hat{H}_0^R$ | $67.717 \pm 3.752$ | $-1.5594$ | $-0.4952$ | $-1.3553$ | $-1.8567$ | $0.0838$ | $0.0812$ |
| CC+$\hat{H}_0^{TRGB}$ | $66.868 \pm 4.850$ | $-1.4193$ | $-0.5629$ | $-1.2516$ | $-1.6526$ | $-0.1091$ | $-0.1069$ |
| CC+SN+$\hat{H}_0^{TRGB}$ | $68.689 \pm 4.623$ | $-1.1133$ | $-0.2223$ | $-0.9362$ | $-1.3543$ | $0.2772$ | $0.2712$ |
| CC+SN+BAO+$\hat{H}_0^{TRGB}$ | $67.751 \pm 4.187$ | $-1.4169$ | $-0.4456$ | $-1.2279$ | $-1.6843$ | $0.0832$ | $0.0811$ |
| CC+$\hat{H}_0^{HW}$ | $67.479 \pm 4.538$ | $-1.3788$ | $-0.4719$ | $-1.2013$ | $-1.6267$ | $0.0172$ | $0.0168$ |
| CC+SN+$\hat{H}_0^{HW}$ | $69.802 \pm 4.299$ | $-0.9463$ | $0.0005$ | $-0.7566$ | $-1.2019$ | $0.5551$ | $0.5414$ |
| CC+SN+BAO+$\hat{H}_0^{HW}$ | $68.163 \pm 3.989$ | $-1.3815$ | $-0.3705$ | $-1.1848$ | $-1.6605$ | $0.1898$ | $0.1844$ |
| CC+$\hat{H}_0^{CM}$ | $67.289 \pm 5.172$ | $-1.2641$ | $-0.4557$ | $-1.1040$ | $-1.4823$ | $-0.0213$ | $-0.0209$ |
| CC+SN+$\hat{H}_0^{CM}$ | $70.791 \pm 5.148$ | $-0.6279$ | $0.1806$ | $-0.4627$ | $-0.8418$ | $0.6556$ | $0.6441$ |
| CC+SN+BAO+$\hat{H}_0^{CM}$ | $68.441 \pm 3.809$ | $-1.3692$ | $-0.3194$ | $-1.1651$ | $-1.6598$ | $0.2709$ | $0.2625$ |
| CC+$\hat{H}_0^P$ | $66.786 \pm 4.997$ | $-1.3979$ | $-0.5637$ | $-1.2341$ | $-1.6245$ | $-0.1222$ | $-0.1199$ |
| CC+SN+$\hat{H}_0^P$ | $68.076 \pm 4.610$ | $-1.2398$ | $-0.3458$ | $-1.0633$ | $-1.4827$ | $0.1457$ | $0.1426$ |
| CC+SN+BAO+$\hat{H}_0^P$ | $67.591 \pm 3.948$ | $-1.5248$ | $-0.5041$ | $-1.3281$ | $-1.8083$ | $0.0480$ | $0.0466$ |
| CC+$\hat{H}_0^{DES}$ | $66.368 \pm 5.909$ | $-1.2700$ | $-0.5530$ | $-1.1275$ | $-1.4622$ | $-0.1741$ | $-0.1718$ |
| CC+SN+$\hat{H}_0^{DES}$ | $67.981 \pm 5.019$ | $-1.1687$ | $-0.3390$ | $-1.0039$ | $-1.3924$ | $0.1151$ | $0.1130$ |
| CC+SN+BAO+$\hat{H}_0^{DES}$ | $66.756 \pm 4.193$ | $-1.6330$ | $-0.6612$ | $-1.4464$ | $-1.9027$ | $-0.1524$ | $-0.1485$ |

Table H.1: TPR Results with double square exponential kernel function

| **Dataset** | $\hat{H}_0$ | $d(\hat{H}_0, \hat{H}_0^R)$ | $d(\hat{H}_0, \hat{H}_0^{TRGB})$ | $d(\hat{H}_0, \hat{H}_0^{HW})$ | $d(\hat{H}_0, \hat{H}_0^{CM})$ | $d(\hat{H}_0, \hat{H}_0^P)$ | $d(\hat{H}_0, \hat{H}_0^{DES})$ |
|---|---|---|---|---|---|---|---|
| CC | $69.142 \pm 8.041$ | $-0.6159$ | $-0.0796$ | $-0.5059$ | $-0.7557$ | 0.2163 | 0.2147 |
| CC+SN | $69.911 \pm 12.147$ | $-0.3508$ | 0.0091 | $-0.2763$ | $-0.4435$ | 0.2066 | 0.2059 |
| CC+SN+BAO | $68.333 \pm 12.075$ | $-0.4821$ | $-0.1200$ | $-0.4073$ | $-0.5755$ | 0.0772 | 0.0770 |
| CC+$\hat{H}_0^R$ | $70.353 \pm 8.270$ | $-0.4566$ | 0.0652 | $-0.3490$ | $-0.5921$ | 0.3565 | 0.3540 |
| CC+SN+$\hat{H}_0^R$ | $71.054 \pm 5.797$ | $-0.5210$ | 0.2056 | $-0.3717$ | $-0.7117$ | 0.6281 | 0.6193 |
| CC+SN+BAO+$\hat{H}_0^R$ | $71.847 \pm 5.286$ | $-0.4244$ | 0.3645 | $-0.2616$ | $-0.6315$ | 0.8375 | 0.8236 |
| CC+$\hat{H}_0^{TRGB}$ | $67.905 \pm 5.908$ | $-1.0215$ | $-0.3053$ | $-0.8775$ | $-1.2121$ | 0.0852 | 0.0841 |
| CC+SN+$\hat{H}_0^{TRGB}$ | $69.568 \pm 6.277$ | $-0.7119$ | $-0.0354$ | $-0.5739$ | $-0.8899$ | 0.3443 | 0.3402 |
| CC+SN+BAO+$\hat{H}_0^{TRGB}$ | $70.454 \pm 5.908$ | $-0.6091$ | 0.1054 | $-0.4629$ | $-0.7970$ | 0.5151 | 0.5082 |
| CC+$\hat{H}_0^{HW}$ | $70.910 \pm 7.264$ | $-0.4420$ | 0.1478 | $-0.3204$ | $-0.5955$ | 0.4820 | 0.4777 |
| CC+SN+$\hat{H}_0^{HW}$ | $71.847 \pm 5.838$ | $-0.3880$ | 0.3335 | $-0.2389$ | $-0.5766$ | 0.7590 | 0.7487 |
| CC+SN+BAO+$\hat{H}_0^{HW}$ | $70.336 \pm 6.590$ | $-0.5681$ | 0.0782 | $-0.4355$ | $-0.7373$ | 0.4443 | 0.4395 |
| CC+$\hat{H}_0^{CM}$ | $72.809 \pm 6.590$ | $-0.2063$ | 0.4388 | $-0.0721$ | $-0.3736$ | 0.8185 | 0.8097 |
| CC+SN+$\hat{H}_0^{CM}$ | $72.332 \pm 6.479$ | $-0.2805$ | 0.3750 | $-0.1445$ | $-0.4509$ | 0.7590 | 0.7505 |
| CC+SN+BAO+$\hat{H}_0^{CM}$ | $71.639 \pm 5.787$ | $-0.4254$ | 0.3019 | $-0.2754$ | $-0.6158$ | 0.7298 | 0.7196 |
| CC+$\hat{H}_0^P$ | $67.203 \pm 8.744$ | $-0.7857$ | $-0.2902$ | $-0.6845$ | $-0.9150$ | $-0.0225$ | $-0.0224$ |
| CC+SN+$\hat{H}_0^P$ | $67.557 \pm 6.157$ | $-1.0377$ | $-0.3481$ | $-0.8991$ | $-1.2210$ | 0.0254 | 0.0251 |
| CC+SN+BAO+$\hat{H}_0^P$ | $66.938 \pm 6.368$ | $-1.0996$ | $-0.4307$ | $-0.9654$ | $-1.2774$ | $-0.0724$ | $-0.0715$ |
| CC+$\hat{H}_0^{DES}$ | $67.963 \pm 7.137$ | $-0.8495$ | $-0.2487$ | $-0.7274$ | $-1.0074$ | 0.0788 | 0.0780 |
| CC+SN+$\hat{H}_0^{DES}$ | $67.134 \pm 5.516$ | $-1.2200$ | $-0.4570$ | $-1.0683$ | $-1.4249$ | $-0.0481$ | $-0.0473$ |
| CC+SN+BAO+$\hat{H}_0^{DES}$ | $67.669 \pm 5.451$ | $-1.1399$ | $-0.3691$ | $-0.9861$ | $-1.3466$ | 0.0492 | 0.0485 |

Table H.2: TPR Results with rational quadratic kernel function

| **Dataset** | $\hat{H}_0$ | $d(\hat{H}_0, \hat{H}_0^R)$ | $d(\hat{H}_0, \hat{H}_0^{TRGB})$ | $d(\hat{H}_0, \hat{H}_0^{HW})$ | $d(\hat{H}_0, \hat{H}_0^{CM})$ | $d(\hat{H}_0, \hat{H}_0^P)$ | $d(\hat{H}_0, \hat{H}_0^{DES})$ |
|---|---|---|---|---|---|---|---|
| CC | $67.996 \pm 6.956$ | $-0.8657$ | $-0.2502$ | $-0.7408$ | $-1.0277$ | $0.0854$ | $0.0846$ |
| CC+SN | $67.625 \pm 6.220$ | $-1.0177$ | $-0.3344$ | $-0.8801$ | $-1.1990$ | $0.0360$ | $0.0356$ |
| CC+SN+BAO | $66.102 \pm 5.363$ | $-1.4335$ | $-0.6500$ | $-1.2795$ | $-1.6456$ | $-0.2410$ | $-0.2371$ |
| CC+$\hat{H}_0^R$ | $68.899 \pm 5.583$ | $-0.9061$ | $-0.1527$ | $-0.7541$ | $-1.1064$ | $0.2675$ | $0.2635$ |
| CC+SN+$\hat{H}_0^R$ | $67.254 \pm 5.427$ | $-1.2170$ | $-0.4428$ | $-1.0632$ | $-1.4251$ | $-0.0268$ | $-0.0263$ |
| CC+SN+BAO+$\hat{H}_0^R$ | $68.356 \pm 4.075$ | $-1.3139$ | $-0.3212$ | $-1.1197$ | $-1.5868$ | $0.2328$ | $0.2264$ |
| CC+$\hat{H}_0^{TRGB}$ | $67.167 \pm 5.438$ | $-1.2300$ | $-0.4571$ | $-1.0765$ | $-1.4378$ | $-0.0427$ | $-0.0420$ |
| CC+SN+$\hat{H}_0^{TRGB}$ | $68.332 \pm 5.520$ | $-1.0129$ | $-0.2514$ | $-0.8600$ | $-1.2162$ | $0.1682$ | $0.1656$ |
| CC+SN+BAO+$\hat{H}_0^{TRGB}$ | $67.762 \pm 4.136$ | $-1.4293$ | $-0.4479$ | $-1.2386$ | $-1.6999$ | $0.0868$ | $0.0845$ |
| CC+$\hat{H}_0^{HW}$ | $68.017 \pm 5.721$ | $-1.0333$ | $-0.2958$ | $-0.8853$ | $-1.2299$ | $0.1074$ | $0.1059$ |
| CC+SN+$\hat{H}_0^{HW}$ | $67.657 \pm 4.419$ | $-1.3733$ | $-0.4455$ | $-1.1918$ | $-1.6273$ | $0.0579$ | $0.0565$ |
| CC+SN+BAO+$\hat{H}_0^{HW}$ | $69.170 \pm 4.667$ | $-1.0081$ | $-0.1250$ | $-0.8315$ | $-1.2460$ | $0.3772$ | $0.3692$ |
| CC+$\hat{H}_0^{CM}$ | $67.926 \pm 5.343$ | $-1.1150$ | $-0.3304$ | $-0.9584$ | $-1.3254$ | $0.0981$ | $0.0965$ |
| CC+SN+$\hat{H}_0^{CM}$ | $69.088 \pm 5.207$ | $-0.9304$ | $-0.1285$ | $-0.7690$ | $-1.1445$ | $0.3227$ | $0.3172$ |
| CC+SN+BAO+$\hat{H}_0^{CM}$ | $68.706 \pm 4.267$ | $-1.1886$ | $-0.2342$ | $-1.0002$ | $-1.4488$ | $0.3040$ | $0.2964$ |
| CC+$\hat{H}_0^P$ | $65.694 \pm 5.744$ | $-1.4150$ | $-0.6786$ | $-1.2697$ | $-1.6135$ | $-0.2958$ | $-0.2916$ |
| CC+SN+$\hat{H}_0^P$ | $66.439 \pm 5.509$ | $-1.3410$ | $-0.5767$ | $-1.1899$ | $-1.5471$ | $-0.1737$ | $-0.1710$ |
| CC+SN+BAO+$\hat{H}_0^P$ | $66.639 \pm 4.203$ | $-1.6553$ | $-0.6854$ | $-1.4693$ | $-1.9246$ | $-0.1799$ | $-0.1753$ |
| CC+$\hat{H}_0^{DES}$ | $67.044 \pm 6.435$ | $-1.0730$ | $-0.4108$ | $-0.9399$ | $-1.2489$ | $-0.0552$ | $-0.0546$ |
| CC+SN+$\hat{H}_0^{DES}$ | $67.316 \pm 4.873$ | $-1.3273$ | $-0.4749$ | $-1.1595$ | $-1.5587$ | $-0.0171$ | $-0.0168$ |
| CC+SN+BAO+$\hat{H}_0^{DES}$ | $67.030 \pm 4.908$ | $-1.3734$ | $-0.5263$ | $-1.2070$ | $-1.6037$ | $-0.0750$ | $-0.0735$ |

Table H.3: TPR Results with Matérn $\left(\nu = \frac{3}{2}\right)$ kernel function

| **Dataset** | $\hat{H}_0$ | $d(\hat{H}_0, \hat{H}_0^R)$ | $d(\hat{H}_0, \hat{H}_0^{TRGB})$ | $d(\hat{H}_0, \hat{H}_0^{HW})$ | $d(\hat{H}_0, \hat{H}_0^{CM})$ | $d(\hat{H}_0, \hat{H}_0^P)$ | $d(\hat{H}_0, \hat{H}_0^{DES})$ |
|---|---|---|---|---|---|---|---|
| CC | $65.406 \pm 4.731$ | $-1.7388$ | $-0.8619$ | $-1.5702$ | $-1.9807$ | $-0.4191$ | $-0.4105$ |
| CC+SN | $66.750 \pm 5.662$ | $-1.2560$ | $-0.5107$ | $-1.1079$ | $-1.4561$ | $-0.1144$ | $-0.1127$ |
| CC+SN+BAO | $67.928 \pm 3.888$ | $-1.4656$ | $-0.4325$ | $-1.2659$ | $-1.7523$ | $0.1348$ | $0.1308$ |
| CC+$\hat{H}_0^R$ | $67.617 \pm 5.144$ | $-1.2102$ | $-0.3981$ | $-1.0490$ | $-1.4291$ | $0.0420$ | $0.0413$ |
| CC+SN+$\hat{H}_0^R$ | $68.211 \pm 4.985$ | $-1.1323$ | $-0.2979$ | $-0.9662$ | $-1.3571$ | $0.1618$ | $0.1588$ |
| CC+SN+BAO+$\hat{H}_0^R$ | $69.082 \pm 4.249$ | $-1.1116$ | $-0.1542$ | $-0.9217$ | $-1.3719$ | $0.3932$ | $0.3833$ |
| CC+$\hat{H}_0^{TRGB}$ | $67.870 \pm 5.458$ | $-1.1037$ | $-0.3340$ | $-0.9499$ | $-1.3098$ | $0.0857$ | $0.0844$ |
| CC+SN+$\hat{H}_0^{TRGB}$ | $68.169 \pm 4.698$ | $-1.2010$ | $-0.3219$ | $-1.0270$ | $-1.4392$ | $0.1627$ | $0.1593$ |
| CC+SN+BAO+$\hat{H}_0^{TRGB}$ | $67.444 \pm 2.821$ | $\mathbf{-2.0184}$ | $-0.6927$ | $-1.7780$ | $\mathbf{-2.4080}$ | $0.0154$ | $0.0146$ |
| CC+$\hat{H}_0^{HW}$ | $68.387 \pm 5.450$ | $-1.0150$ | $-0.2447$ | $-0.8605$ | $-1.2208$ | $0.1804$ | $0.1776$ |
| CC+SN+$\hat{H}_0^{HW}$ | $68.024 \pm 4.306$ | $-1.3254$ | $-0.3774$ | $-1.1396$ | $-1.5849$ | $0.1439$ | $0.1403$ |
| CC+SN+BAO+$\hat{H}_0^{HW}$ | $69.151 \pm 3.399$ | $-1.3149$ | $-0.1668$ | $-1.0919$ | $-1.6351$ | $0.5096$ | $0.4900$ |
| CC+$\hat{H}_0^{CM}$ | $68.435 \pm 5.006$ | $-1.0860$ | $-0.2548$ | $-0.9201$ | $-1.3095$ | $0.2058$ | $0.2020$ |
| CC+SN+$\hat{H}_0^{CM}$ | $69.584 \pm 4.251$ | $-1.0026$ | $-0.0465$ | $-0.8117$ | $-1.2615$ | $0.5101$ | $0.4973$ |
| CC+SN+BAO+$\hat{H}_0^{CM}$ | $69.339 \pm 4.186$ | $-1.0693$ | $-0.1003$ | $-0.8767$ | $-1.3326$ | $0.4599$ | $0.4480$ |
| CC+$\hat{H}_0^P$ | $67.173 \pm 4.608$ | $-1.4223$ | $-0.5270$ | $-1.2474$ | $-1.6671$ | $-0.0489$ | $-0.0478$ |
| CC+SN+$\hat{H}_0^P$ | $67.011 \pm 4.946$ | $-1.3679$ | $-0.5265$ | $-1.2026$ | $-1.5965$ | $-0.0783$ | $-0.0768$ |
| CC+SN+BAO+$\hat{H}_0^P$ | $67.379 \pm 4.156$ | $-1.5076$ | $-0.5297$ | $-1.3184$ | $-1.7779$ | $-0.0049$ | $-0.0048$ |
| CC+$\hat{H}_0^{DES}$ | $68.394 \pm 4.644$ | $-1.1679$ | $-0.2802$ | $-0.9919$ | $-1.4084$ | $0.2128$ | $0.2083$ |
| CC+SN+$\hat{H}_0^{DES}$ | $68.463 \pm 5.628$ | $-0.9733$ | $-0.2251$ | $-0.8227$ | $-1.1726$ | $0.1881$ | $0.1854$ |
| CC+SN+BAO+$\hat{H}_0^{DES}$ | $67.631 \pm 3.841$ | $-1.5502$ | $-0.5061$ | $-1.3496$ | $-1.8413$ | $0.0598$ | $0.0579$ |

Table H.4: TPR Results with Matérn $\left(\nu = \frac{5}{2}\right)$ kernel function

| **Dataset** | $\hat{H}_0$ | $d(\hat{H}_0, \hat{H}_0^R)$ | $d(\hat{H}_0, \hat{H}_0^{TRGB})$ | $d(\hat{H}_0, \hat{H}_0^{HW})$ | $d(\hat{H}_0, \hat{H}_0^{CM})$ | $d(\hat{H}_0, \hat{H}_0^P)$ | $d(\hat{H}_0, \hat{H}_0^{DES})$ |
|---|---|---|---|---|---|---|---|
| CC | $68.737 \pm 6.363$ | $-0.8284$ | $-0.1600$ | $-0.6927$ | $-1.0048$ | $0.2095$ | $0.2071$ |
| CC+SN | $67.162 \pm 8.831$ | $-0.7827$ | $-0.2920$ | $-0.6824$ | $-0.9108$ | $-0.0269$ | $-0.0267$ |
| CC+SN+BAO | $67.020 \pm 4.933$ | $-1.3693$ | $-0.5259$ | $-1.2036$ | $-1.5985$ | $-0.0766$ | $-0.0751$ |
| CC+$\hat{H}_0^R$ | $68.792 \pm 4.559$ | $-1.1057$ | $-0.2040$ | $-0.9265$ | $-1.3497$ | $0.3036$ | $0.2969$ |
| CC+SN+$\hat{H}_0^R$ | $68.429 \pm 7.592$ | $-0.7418$ | $-0.1752$ | $-0.6261$ | $-0.8901$ | $0.1352$ | $0.1341$ |
| CC+SN+BAO+$\hat{H}_0^R$ | $68.629 \pm 3.698$ | $-1.3566$ | $-0.2817$ | $-1.1478$ | $-1.6549$ | $0.3294$ | $0.3186$ |
| CC+$\hat{H}_0^{TRGB}$ | $67.443 \pm 4.771$ | $-1.3271$ | $-0.4589$ | $-1.1564$ | $-1.5632$ | $0.0090$ | $0.0088$ |
| CC+SN+$\hat{H}_0^{TRGB}$ | $66.392 \pm 6.499$ | $-1.1598$ | $-0.5033$ | $-1.0283$ | $-1.3345$ | $-0.1547$ | $-0.1529$ |
| CC+SN+BAO+$\hat{H}_0^{TRGB}$ | $67.688 \pm 4.561$ | $-1.3301$ | $-0.4274$ | $-1.1529$ | $-1.5763$ | $0.0629$ | $0.0615$ |
| CC+$\hat{H}_0^{HW}$ | $68.631 \pm 5.455$ | $-0.9719$ | $-0.2024$ | $-0.8171$ | $-1.1772$ | $0.2247$ | $0.2212$ |
| CC+SN+$\hat{H}_0^{HW}$ | $69.760 \pm 6.391$ | $-0.6712$ | $-0.0060$ | $-0.5353$ | $-0.8460$ | $0.3681$ | $0.3639$ |
| CC+SN+BAO+$\hat{H}_0^{HW}$ | $67.929 \pm 3.577$ | $-1.5673$ | $-0.4618$ | $-1.3560$ | $-1.8776$ | $0.1465$ | $0.1414$ |
| CC+$\hat{H}_0^{CM}$ | $68.371 \pm 5.615$ | $-0.9910$ | $-0.2411$ | $-0.8402$ | $-1.1908$ | $0.1722$ | $0.1697$ |
| CC+SN+$\hat{H}_0^{CM}$ | $69.573 \pm 6.499$ | $-0.6885$ | $-0.0335$ | $-0.5548$ | $-0.8606$ | $0.3334$ | $0.3296$ |
| CC+SN+BAO+$\hat{H}_0^{CM}$ | $68.764 \pm 4.961$ | $-1.0324$ | $-0.1949$ | $-0.8649$ | $-1.2573$ | $0.2736$ | $0.2685$ |
| CC+$\hat{H}_0^P$ | $67.540 \pm 5.194$ | $-1.2137$ | $-0.4086$ | $-1.0539$ | $-1.4306$ | $0.0269$ | $0.0264$ |
| CC+SN+$\hat{H}_0^P$ | $63.681 \pm 5.524$ | $-1.8121$ | $-1.0475$ | $-1.6643$ | $\mathbf{-2.0211}$ | $-0.6705$ | $-0.6603$ |
| CC+SN+BAO+$\hat{H}_0^P$ | $66.553 \pm 3.829$ | $-1.8084$ | $-0.7597$ | $-1.6105$ | $\mathbf{-2.1038}$ | $-0.2194$ | $-0.2127$ |
| CC+$\hat{H}_0^{DES}$ | $66.487 \pm 5.044$ | $-1.4421$ | $-0.6147$ | $-1.2799$ | $-1.6670$ | $-0.1802$ | $-0.1769$ |
| CC+SN+$\hat{H}_0^{DES}$ | $64.497 \pm 8.127$ | $-1.1676$ | $-0.6355$ | $-1.0603$ | $-1.3079$ | $-0.3566$ | $-0.3540$ |
| CC+SN+BAO+$\hat{H}_0^{DES}$ | $66.786 \pm 4.227$ | $-1.6152$ | $-0.6503$ | $-1.4297$ | $-1.8827$ | $-0.1442$ | $-0.1405$ |

Table H.5: TPR Results with Matérn $\left(\nu = \frac{7}{2}\right)$ kernel function

| **Dataset** | $\hat{H}_0$ | $d(\hat{H}_0, \hat{H}_0^R)$ | $d(\hat{H}_0, \hat{H}_0^{TRGB})$ | $d(\hat{H}_0, \hat{H}_0^{HW})$ | $d(\hat{H}_0, \hat{H}_0^{CM})$ | $d(\hat{H}_0, \hat{H}_0^P)$ | $d(\hat{H}_0, \hat{H}_0^{DES})$ |
|---|---|---|---|---|---|---|---|
| CC | $67.512 \pm 6.674$ | $-0.9697$ | $-0.3297$ | $-0.8404$ | $-1.1388$ | $0.0168$ | $0.0166$ |
| CC+SN | $61.833 \pm 9.943$ | $-1.1873$ | $-0.7626$ | $-1.1012$ | $-1.2985$ | $-0.5412$ | $-0.5388$ |
| CC+SN+BAO | $67.293 \pm 3.900$ | $-1.6094$ | $-0.5778$ | $-1.4119$ | $-1.8973$ | $-0.0272$ | $-0.0264$ |
| CC+$\hat{H}_0^R$ | $69.141 \pm 4.739$ | $-1.0004$ | $-0.1290$ | $-0.8260$ | $-1.2348$ | $0.3654$ | $0.3580$ |
| CC+SN+$\hat{H}_0^R$ | $69.856 \pm 6.321$ | $-0.6635$ | $0.0084$ | $-0.5262$ | $-0.8401$ | $0.3873$ | $0.3827$ |
| CC+SN+BAO+$\hat{H}_0^R$ | $69.694 \pm 4.045$ | $-1.0204$ | $-0.0237$ | $-0.8218$ | $-1.2913$ | $0.5629$ | $0.5473$ |
| CC+$\hat{H}_0^{TRGB}$ | $68.315 \pm 5.221$ | $-1.0679$ | $-0.2672$ | $-0.9078$ | $-1.2827$ | $0.1746$ | $0.1716$ |
| CC+SN+$\hat{H}_0^{TRGB}$ | $68.741 \pm 7.498$ | $-0.7100$ | $-0.1369$ | $-0.5929$ | $-0.8600$ | $0.1785$ | $0.1770$ |
| CC+SN+BAO+$\hat{H}_0^{TRGB}$ | $68.217 \pm 3.554$ | $-1.5033$ | $-0.3927$ | $-1.2901$ | $-1.8144$ | $0.2277$ | $0.2197$ |
| CC+$\hat{H}_0^{HW}$ | $68.751 \pm 4.093$ | $-1.2208$ | $-0.2324$ | $-1.0263$ | $-1.4914$ | $0.3278$ | $0.3189$ |
| CC+SN+$\hat{H}_0^{HW}$ | $69.219 \pm 6.975$ | $-0.6938$ | $-0.0804$ | $-0.5685$ | $-0.8546$ | $0.2601$ | $0.2575$ |
| CC+SN+BAO+$\hat{H}_0^{HW}$ | $68.976 \pm 3.514$ | $-1.3249$ | $-0.2062$ | $-1.1075$ | $-1.6362$ | $0.4441$ | $0.4281$ |
| CC+$\hat{H}_0^{CM}$ | $68.213 \pm 5.478$ | $-1.0406$ | $-0.2737$ | $-0.8869$ | $-1.2456$ | $0.1479$ | $0.1456$ |
| CC+SN+$\hat{H}_0^{CM}$ | $68.178 \pm 7.801$ | $-0.7542$ | $-0.2020$ | $-0.6415$ | $-0.8987$ | $0.0996$ | $0.0988$ |
| CC+SN+BAO+$\hat{H}_0^{CM}$ | $68.956 \pm 4.173$ | $-1.1564$ | $-0.1842$ | $-0.9642$ | $-1.4215$ | $0.3702$ | $0.3605$ |
| CC+$\hat{H}_0^P$ | $67.657 \pm 4.244$ | $-1.4211$ | $-0.4608$ | $-1.2342$ | $-1.6853$ | $0.0602$ | $0.0587$ |
| CC+SN+$\hat{H}_0^P$ | $65.664 \pm 6.669$ | $-1.2377$ | $-0.5965$ | $-1.1096$ | $-1.4084$ | $-0.2596$ | $-0.2569$ |
| CC+SN+BAO+$\hat{H}_0^P$ | $67.603 \pm 4.247$ | $-1.4322$ | $-0.4723$ | $-1.2454$ | $-1.6963$ | $0.0474$ | $0.0462$ |
| CC+$\hat{H}_0^{DES}$ | $66.204 \pm 3.533$ | $\mathbf{-2.0171}$ | $-0.8965$ | $-1.8099$ | $\mathbf{-2.3380}$ | $-0.3352$ | $-0.3232$ |
| CC+SN+$\hat{H}_0^{DES}$ | $65.413 \pm 5.997$ | $-1.4053$ | $-0.6974$ | $-1.2653$ | $-1.5956$ | $-0.3302$ | $-0.3259$ |
| CC+SN+BAO+$\hat{H}_0^{DES}$ | $67.999 \pm 3.594$ | $-1.5442$ | $-0.4431$ | $-1.3333$ | $-1.8529$ | $0.1650$ | $0.1593$ |

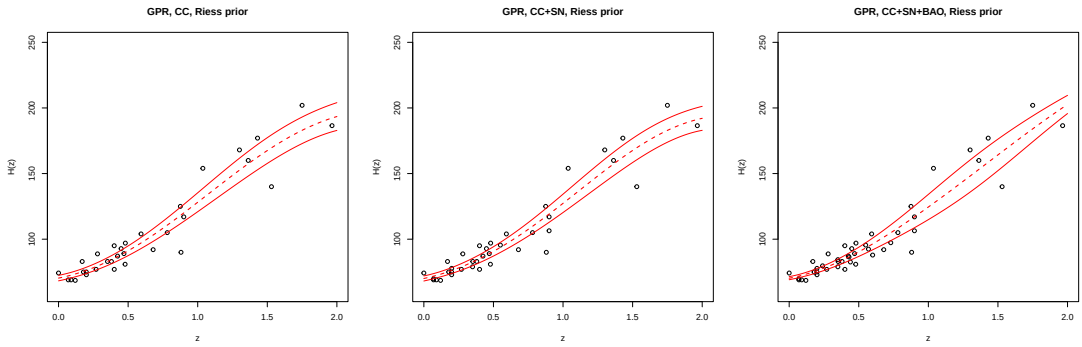
Table H.6: TPR Results with Matérn $\left(\nu = \frac{9}{2}\right)$ kernel function

| **Dataset** | $\hat{H}_0$ | $d(\hat{H}_0, \hat{H}_0^R)$ | $d(\hat{H}_0, \hat{H}_0^{TRGB})$ | $d(\hat{H}_0, \hat{H}_0^{HW})$ | $d(\hat{H}_0, \hat{H}_0^{CM})$ | $d(\hat{H}_0, \hat{H}_0^P)$ | $d(\hat{H}_0, \hat{H}_0^{DES})$ |
|---|---|---|---|---|---|---|---|
| CC | $65.318 \pm 6.581$ | $-1.3038$ | $-0.6544$ | $-1.1744$ | $-1.4771$ | $-0.3155$ | $-0.3121$ |
| CC+SN | $68.197 \pm 5.118$ | $-1.1089$ | $-0.2937$ | $-0.9463$ | $-1.3280$ | $0.1550$ | $0.1522$ |
| CC+SN+BAO | $67.607 \pm 3.627$ | $-1.6296$ | $-0.5356$ | $-1.4212$ | $-1.9371$ | $0.0566$ | $0.0546$ |
| CC+$\hat{H}_0^R$ | $69.160 \pm 5.812$ | $-0.8308$ | $-0.1046$ | $-0.6837$ | $-1.0231$ | $0.3017$ | $0.2976$ |
| CC+SN+$\hat{H}_0^R$ | $68.897 \pm 4.559$ | $-1.0844$ | $-0.1828$ | $-0.9049$ | $-1.3282$ | $0.3265$ | $0.3193$ |
| CC+SN+BAO+$\hat{H}_0^R$ | $68.603 \pm 3.924$ | $-1.2986$ | $-0.2747$ | $-1.0984$ | $-1.5807$ | $0.3040$ | $0.2951$ |
| CC+$\hat{H}_0^{TRGB}$ | $67.324 \pm 4.621$ | $-1.3884$ | $-0.4955$ | $-1.2136$ | $-1.6322$ | $-0.0163$ | $-0.0159$ |
| CC+SN+$\hat{H}_0^{TRGB}$ | $68.925 \pm 4.449$ | $-1.1015$ | $-0.1808$ | $-0.9185$ | $-1.3509$ | $0.3407$ | $0.3328$ |
| CC+SN+BAO+$\hat{H}_0^{TRGB}$ | $67.140 \pm 3.162$ | $-1.9404$ | $-0.7209$ | $-1.7157$ | $\mathbf{-2.2927}$ | $-0.0811$ | $-0.0776$ |
| CC+$\hat{H}_0^{HW}$ | $69.206 \pm 5.444$ | $-0.8734$ | $-0.1030$ | $-0.7178$ | $-1.0783$ | $0.3304$ | $0.3252$ |
| CC+SN+$\hat{H}_0^{HW}$ | $68.498 \pm 3.946$ | $-1.3168$ | $-0.2973$ | $-1.1176$ | $-1.5977$ | $0.2761$ | $0.2681$ |
| CC+SN+BAO+$\hat{H}_0^{HW}$ | $68.089 \pm 3.576$ | $-1.5279$ | $-0.4225$ | $-1.3160$ | $-1.8377$ | $0.1909$ | $0.1842$ |
| CC+$\hat{H}_0^{CM}$ | $69.299 \pm 4.834$ | $-0.9526$ | $-0.0964$ | $-0.7807$ | $-1.1823$ | $0.3908$ | $0.3831$ |
| CC+SN+$\hat{H}_0^{CM}$ | $69.600 \pm 5.167$ | $-0.8433$ | $-0.0363$ | $-0.6802$ | $-1.0583$ | $0.4239$ | $0.4165$ |
| CC+SN+BAO+$\hat{H}_0^{CM}$ | $69.045 \pm 3.514$ | $-1.3077$ | $-0.1891$ | $-1.0901$ | $-1.6188$ | $0.4634$ | $0.4467$ |
| CC+$\hat{H}_0^P$ | $67.709 \pm 5.509$ | $-1.1222$ | $-0.3589$ | $-0.9698$ | $-1.3266$ | $0.0558$ | $0.0549$ |
| CC+SN+$\hat{H}_0^P$ | $67.804 \pm 4.343$ | $-1.3625$ | $-0.4211$ | $-1.1784$ | $-1.6205$ | $0.0923$ | $0.0901$ |
| CC+SN+BAO+$\hat{H}_0^P$ | $67.302 \pm 3.009$ | $-1.9671$ | $-0.7018$ | $-1.7354$ | $\mathbf{-2.3351}$ | $-0.0320$ | $-0.0305$ |
| CC+$\hat{H}_0^{DES}$ | $66.002 \pm 5.358$ | $-1.4523$ | $-0.6681$ | $-1.2983$ | $-1.6648$ | $-0.2598$ | $-0.2556$ |
| CC+SN+$\hat{H}_0^{DES}$ | $67.699 \pm 3.802$ | $-1.5472$ | $-0.4944$ | $-1.3450$ | $-1.8408$ | $0.0779$ | $0.0754$ |
| CC+SN+BAO+$\hat{H}_0^{DES}$ | $66.748 \pm 3.916$ | $-1.7302$ | $-0.7011$ | $-1.5346$ | $\mathbf{-2.0185}$ | $-0.1650$ | $-0.1602$ |

Table H.7: TPR Results with Cauchy kernel function

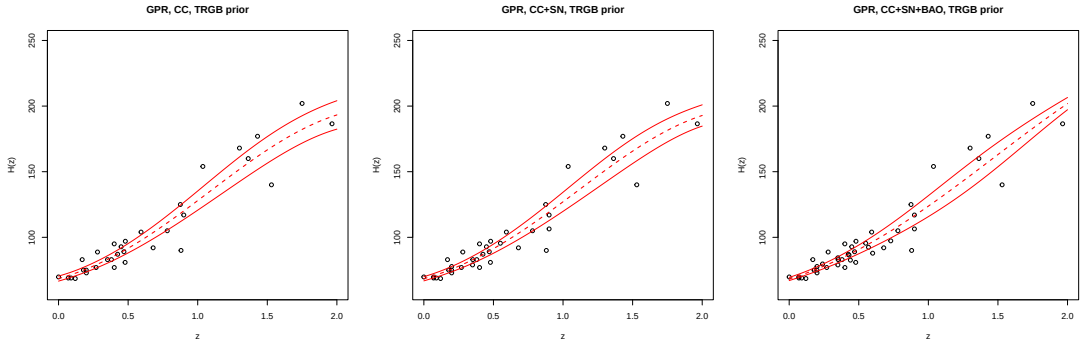
# Appendix I: Heteroscedastic GPR and TPR Plots

The figures for the priorless case were presented in Section 5.1. Similar figures for all the other prior $H_0$ estimates considered are shown in this Appendix.



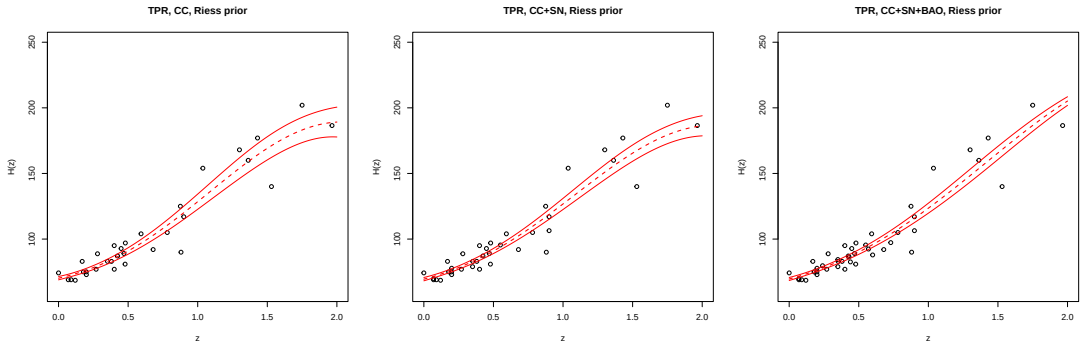

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure I.1: Heteroscedastic GPR reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^R$



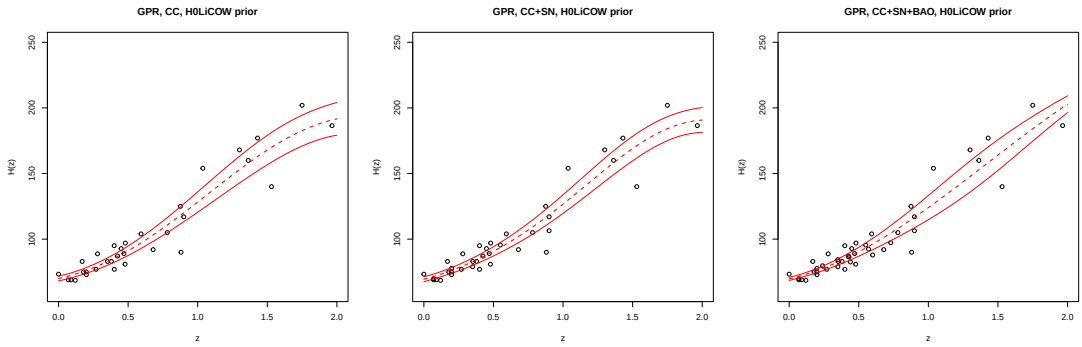

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure I.2: Heteroscedastic GPR reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^{TRGB}$



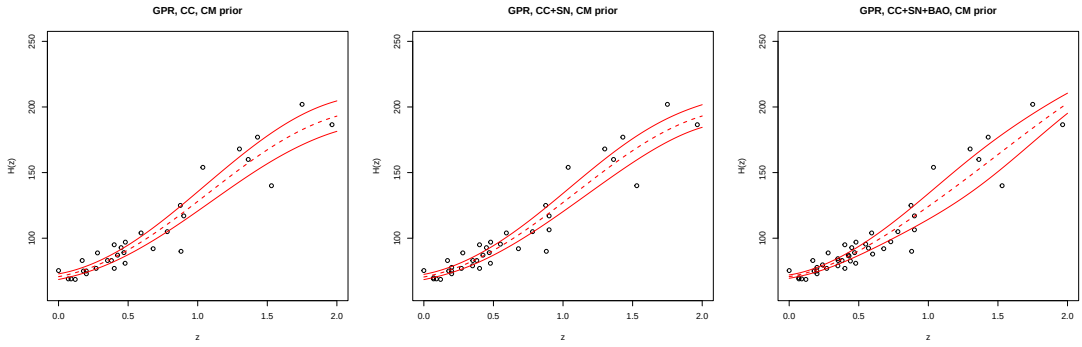

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure I.3: Heteroscedastic TPR reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^R$



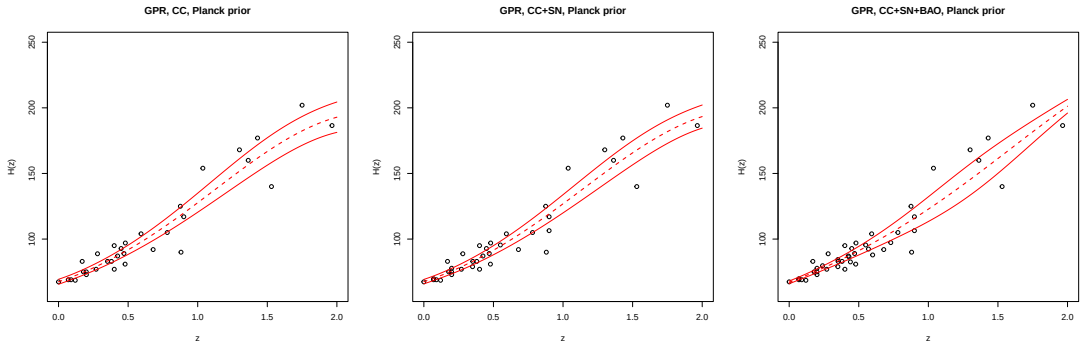

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure I.4: Heteroscedastic GPR reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^{HW}$



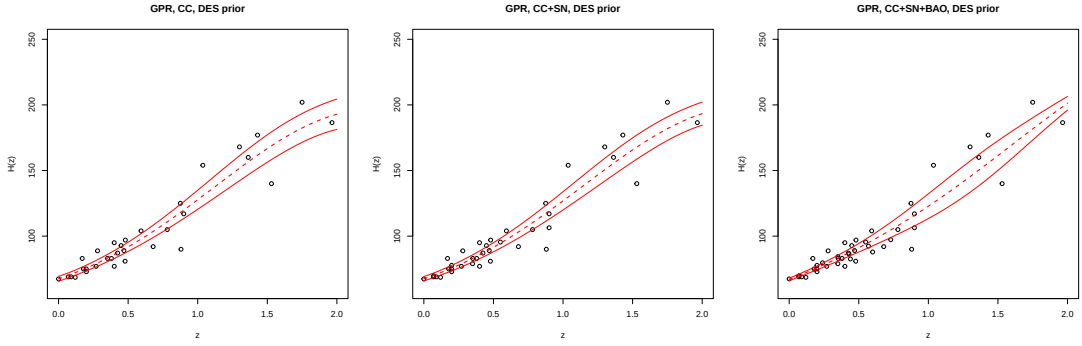

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure I.5: Heteroscedastic GPR reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^{CM}$



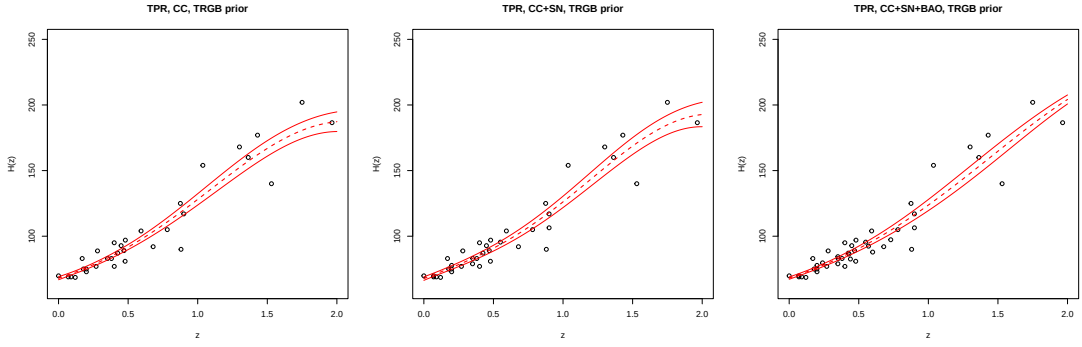

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure I.6: Heteroscedastic GPR reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^P$



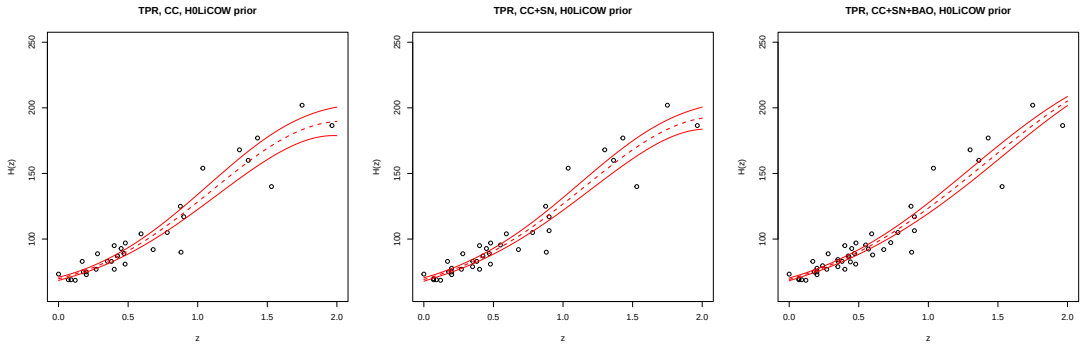

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure I.7: Heteroscedastic GPR reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^{DES}$



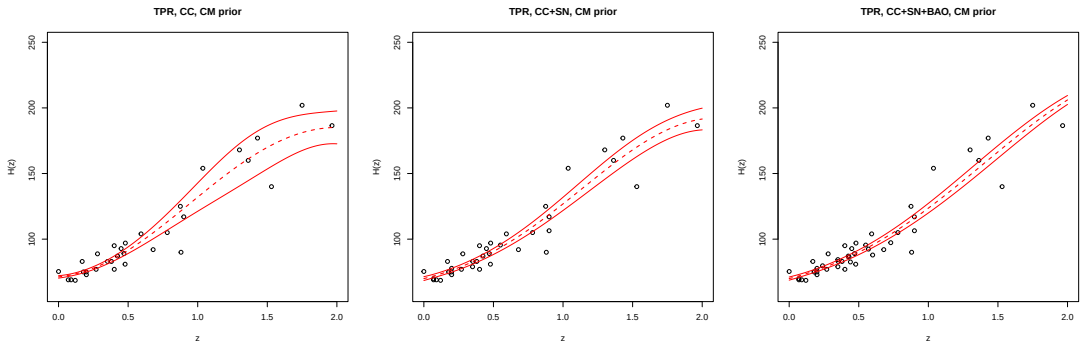

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure I.8: Heteroscedastic TPR reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^{TRGB}$



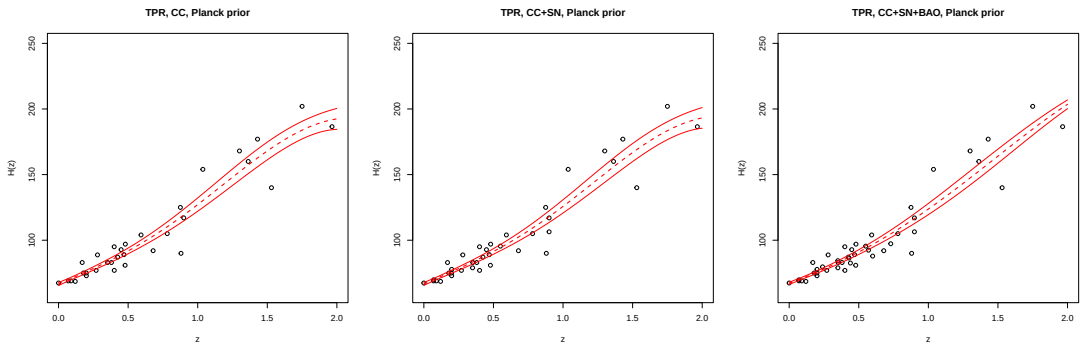

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure I.9: Heteroscedastic TPR reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^{HW}$



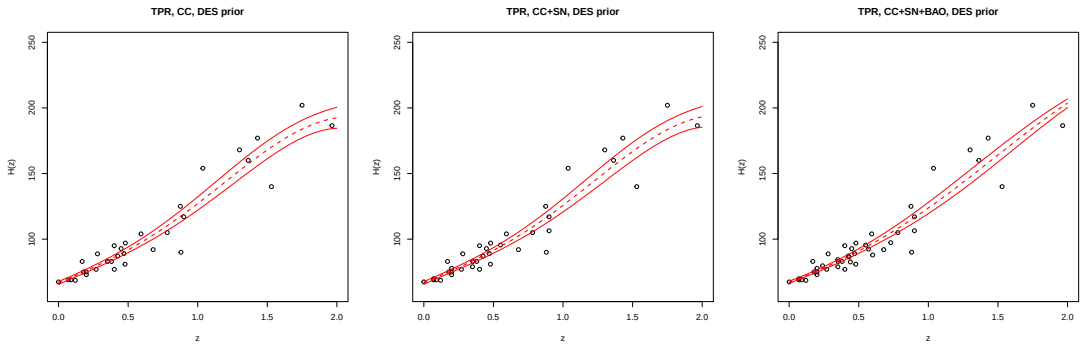

(a) CC (b) CC+SN (c) CC+SN+BAO

Figure I.10: Heteroscedastic TPR reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^{CM}$




(a) CC (b) CC+SN (c) CC+SN+BAO

Figure I.11: Heteroscedastic TPR reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^{P}$

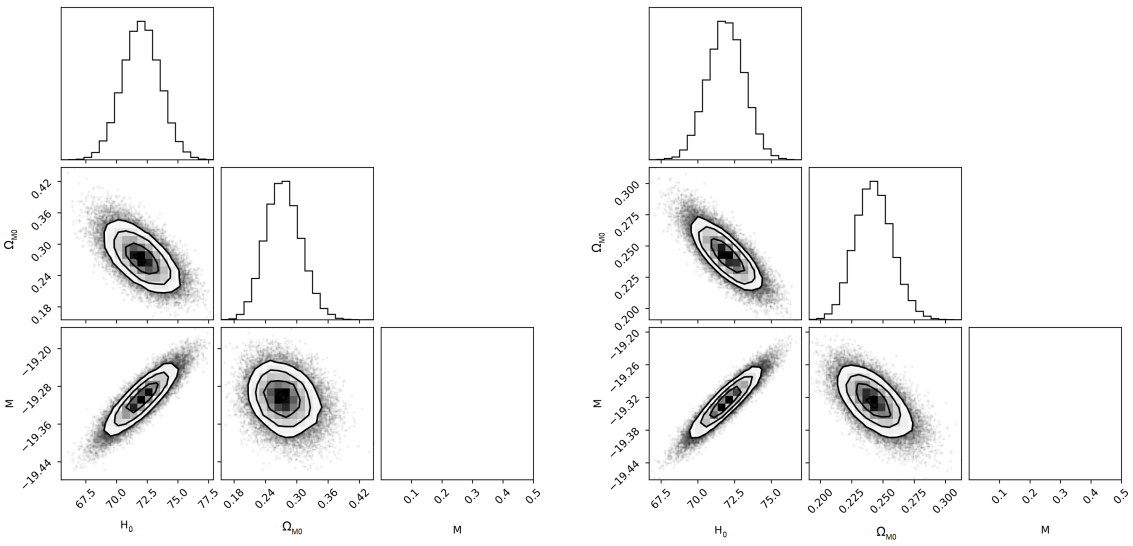



(a) CC (b) CC+SN (c) CC+SN+BAO

Figure I.12: Heteroscedastic TPR reconstructions of $\hat{H}(z)$ against $z$ for square exponential kernel function and using $\hat{H}_0^{DES}$

# Appendix J: MCMC $\Lambda CDM$ Plots - Full Pantheon Data

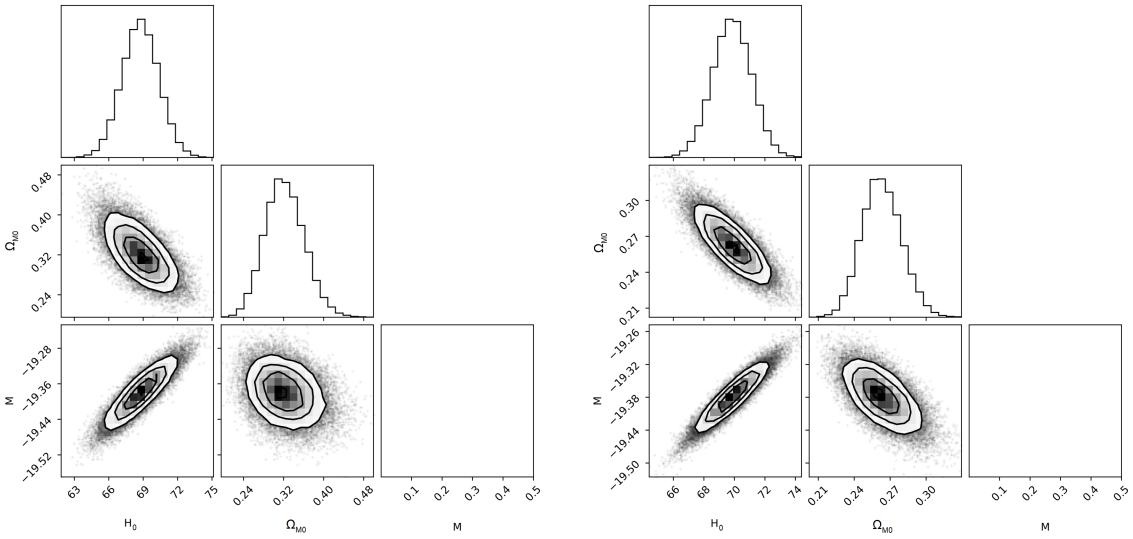
The figures for the priorless case were presented in Section 5.2. Similar figures for all the other prior $H_0$ estimates considered are shown in this Appendix.




(a) CC+SN*

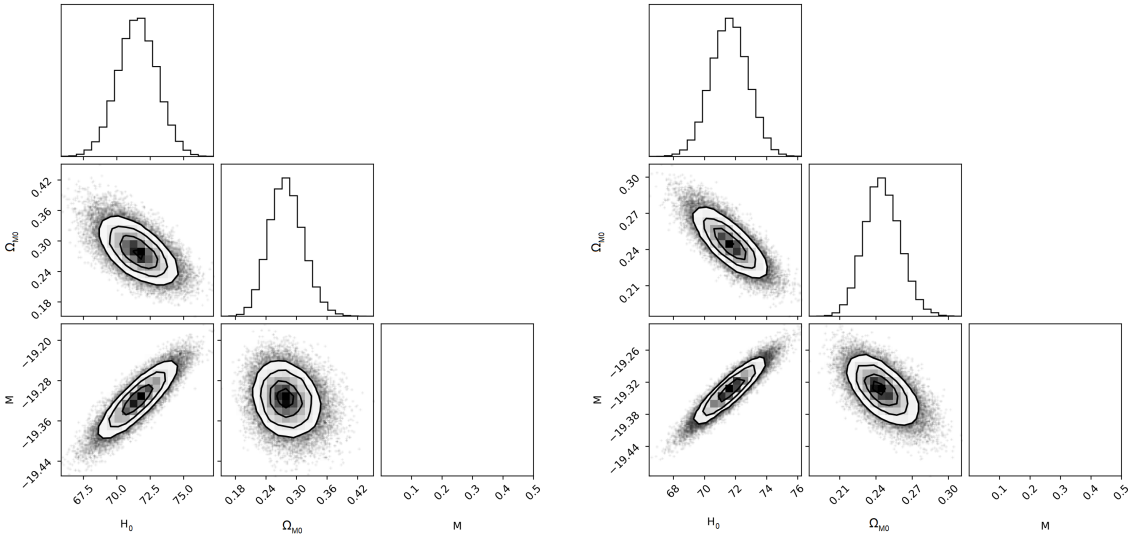
(b) CC+SN*+BAO

Figure J.1: Corner plots for MCMC on parameters of ΛCDM model using $\hat{H}_0^R$ case using the full Pantheon dataset



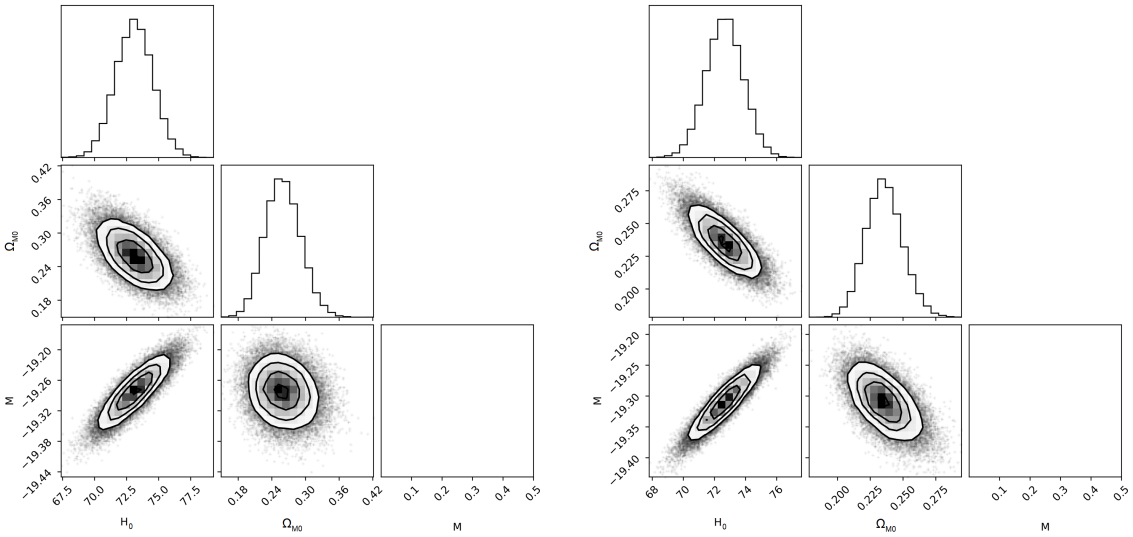

(a) CC+SN* (b) CC+SN*+BAO

Figure J.2: Corner plots for MCMC on parameters of ΛCDM model using $\hat{H}_0^{TRGB}$ case using the full Pantheon dataset



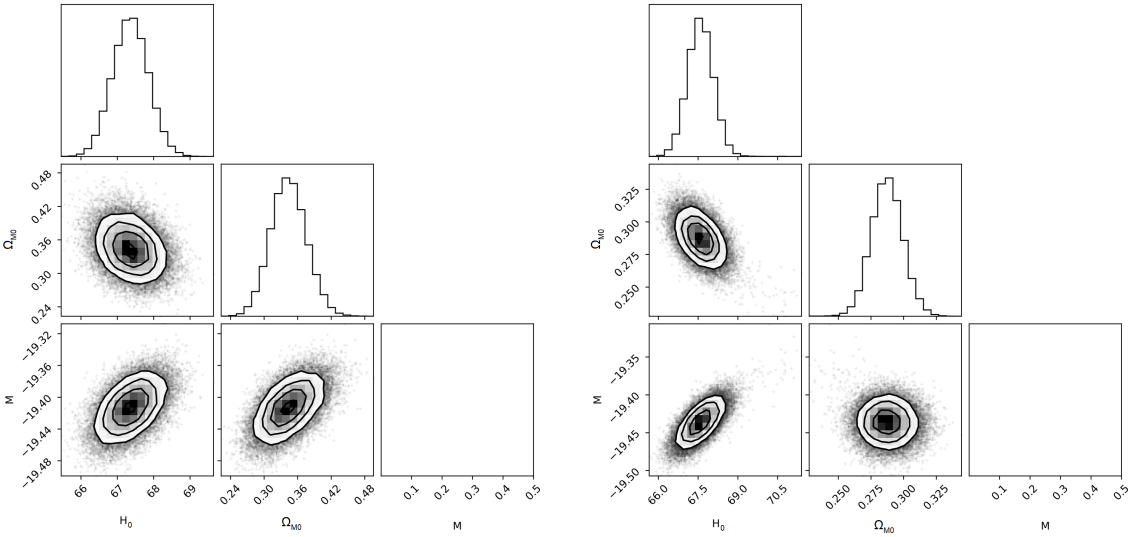

(a) CC+SN* (b) CC+SN*+BAO

Figure J.3: Corner plots for MCMC on parameters of ΛCDM model using $\hat{H}_0^{HW}$ case using the full Pantheon dataset



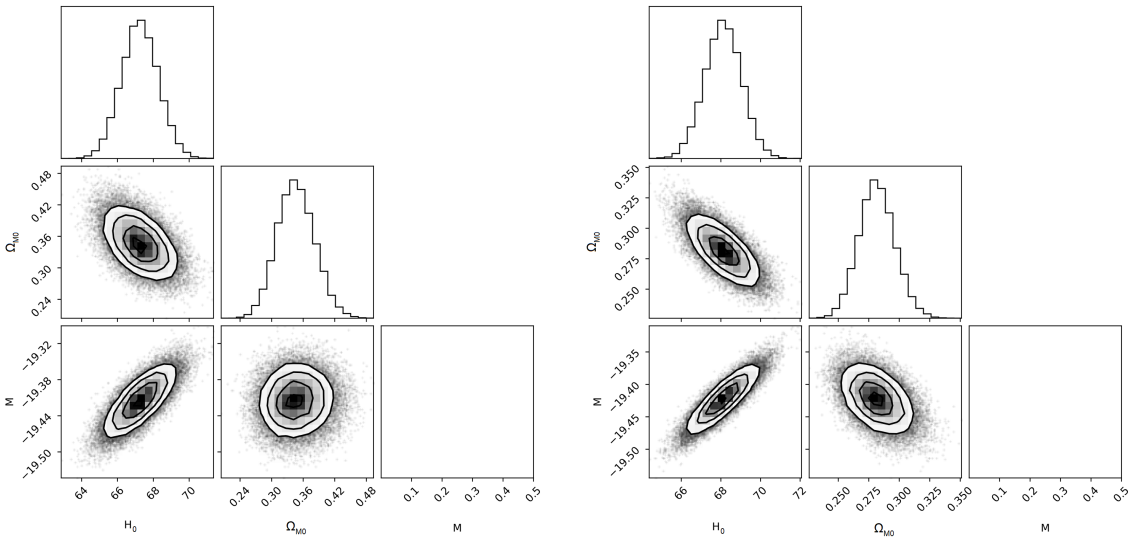

(a) CC+SN* (b) CC+SN*+BAO

Figure J.4: Corner plots for MCMC on parameters of ΛCDM model using $\hat{H}_0^{CM}$ case using the full Pantheon dataset




(a) CC+SN* (b) CC+SN*+BAO

Figure J.5: Corner plots for MCMC on parameters of ΛCDM model using $\hat{H}_0^{P}$ case using the full Pantheon dataset




(a) CC+SN*

(b) CC+SN*+BAO

Figure J.6: Corner plots for MCMC on parameters of ΛCDM model using $\hat{H}_0^{DES}$ case using the full Pantheon dataset

# Appendix K: Comparison Plots

The figures for the priorless case and the case where the Planck prior was added to the data, as well as the figure for the CC+SN+BAO dataset and that for GPR, were presented in Section 5.1. Similar figures for all the other priors, datasets, and methods considered are shown in this Appendix.

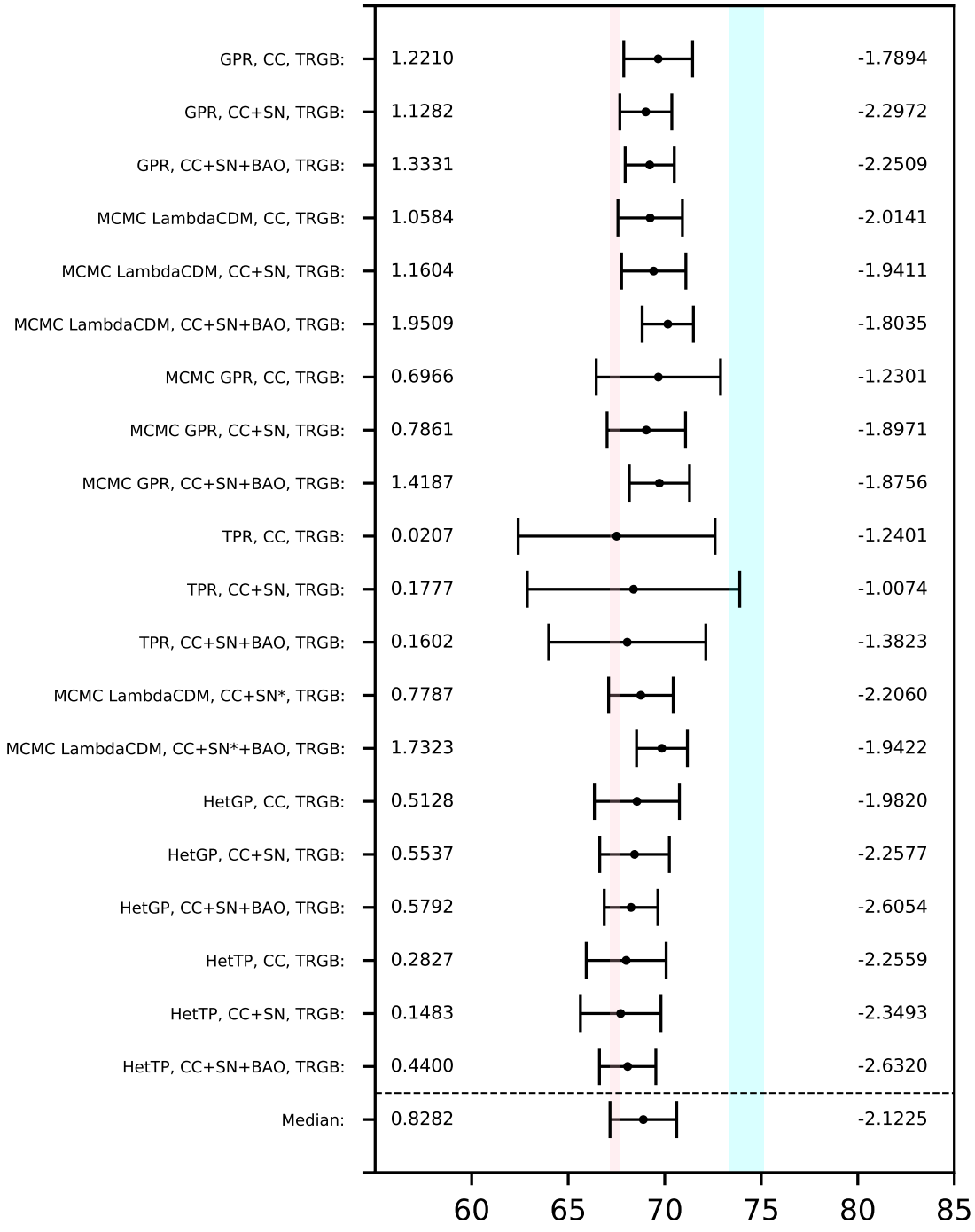



Figure K.1: Comparison of estimates obtained using TRGB prior

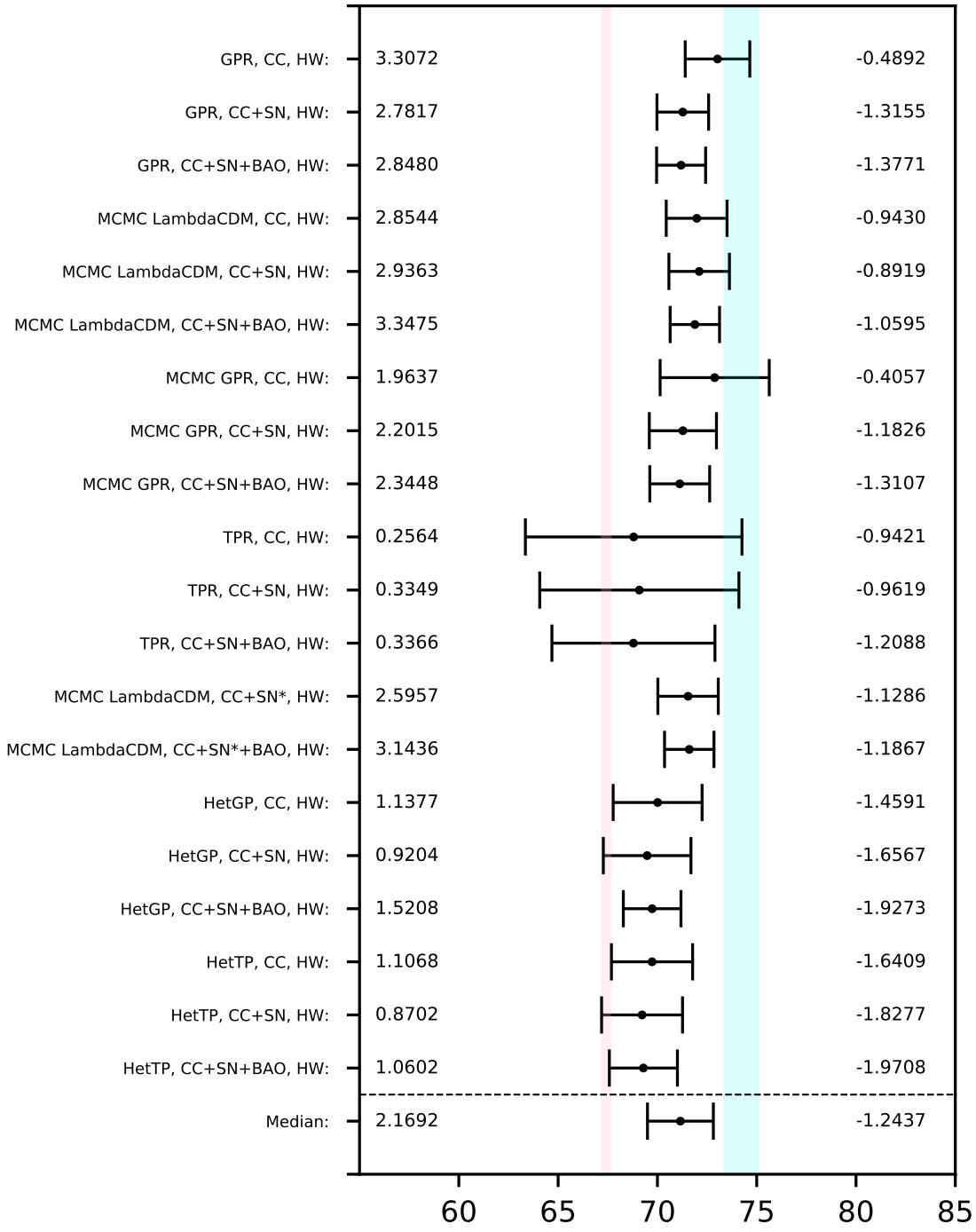



Figure K.2: Comparison of estimates obtained using H0LiCOW prior

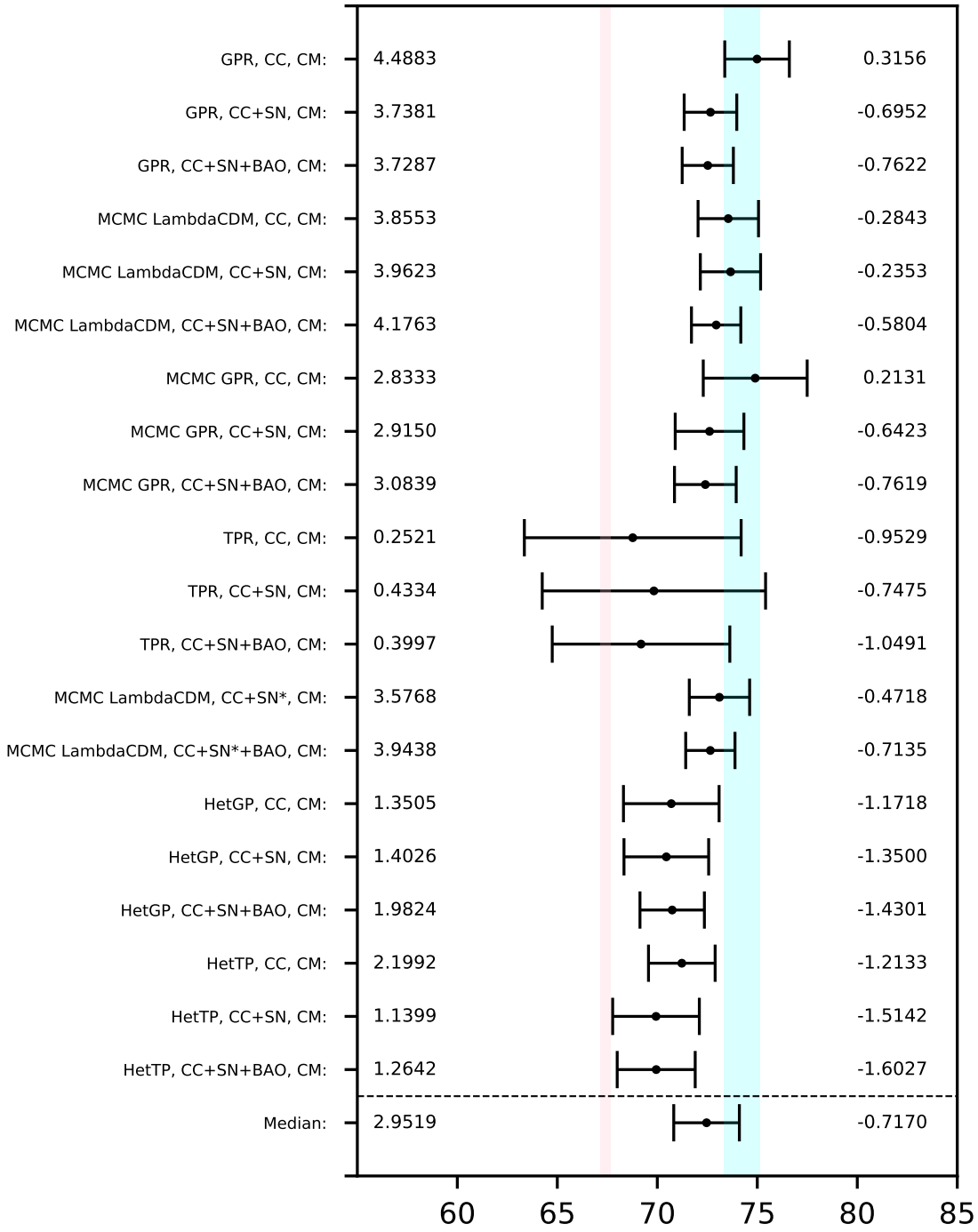



Figure K.3: Comparison of estimates obtained using CM prior

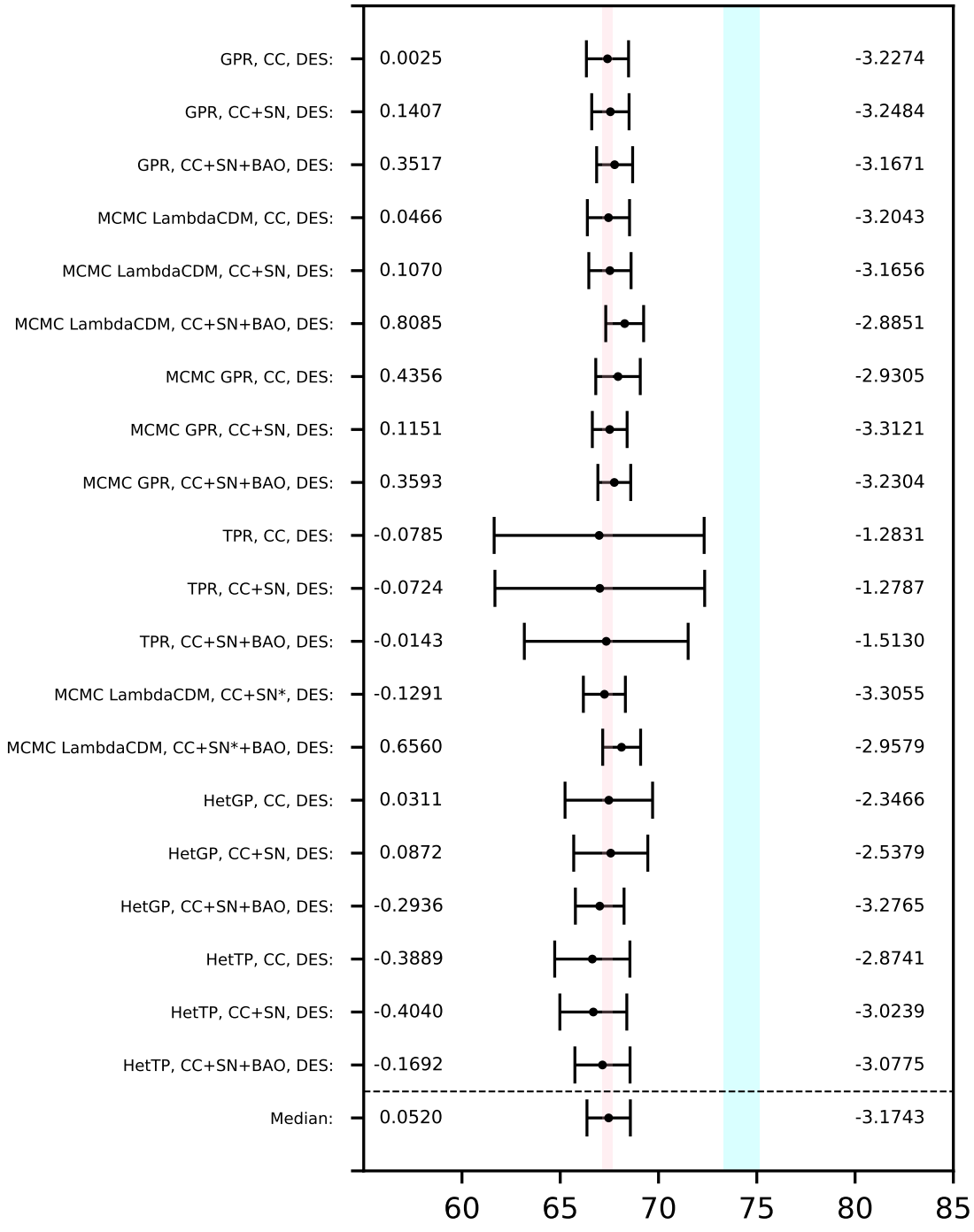



Figure K.4: Comparison of estimates obtained using DES prior

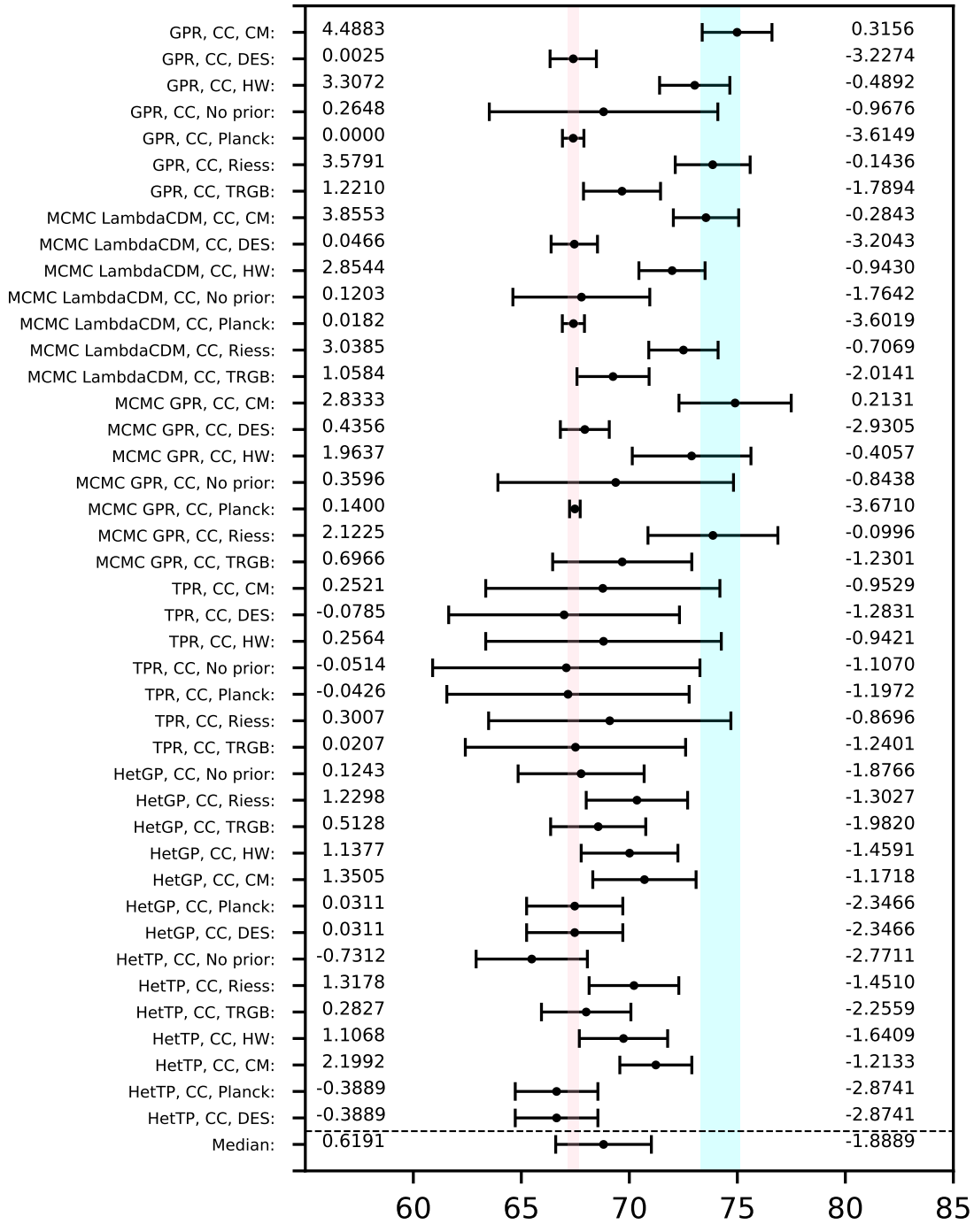



Figure K.5: Comparison of estimates obtained using CC dataset

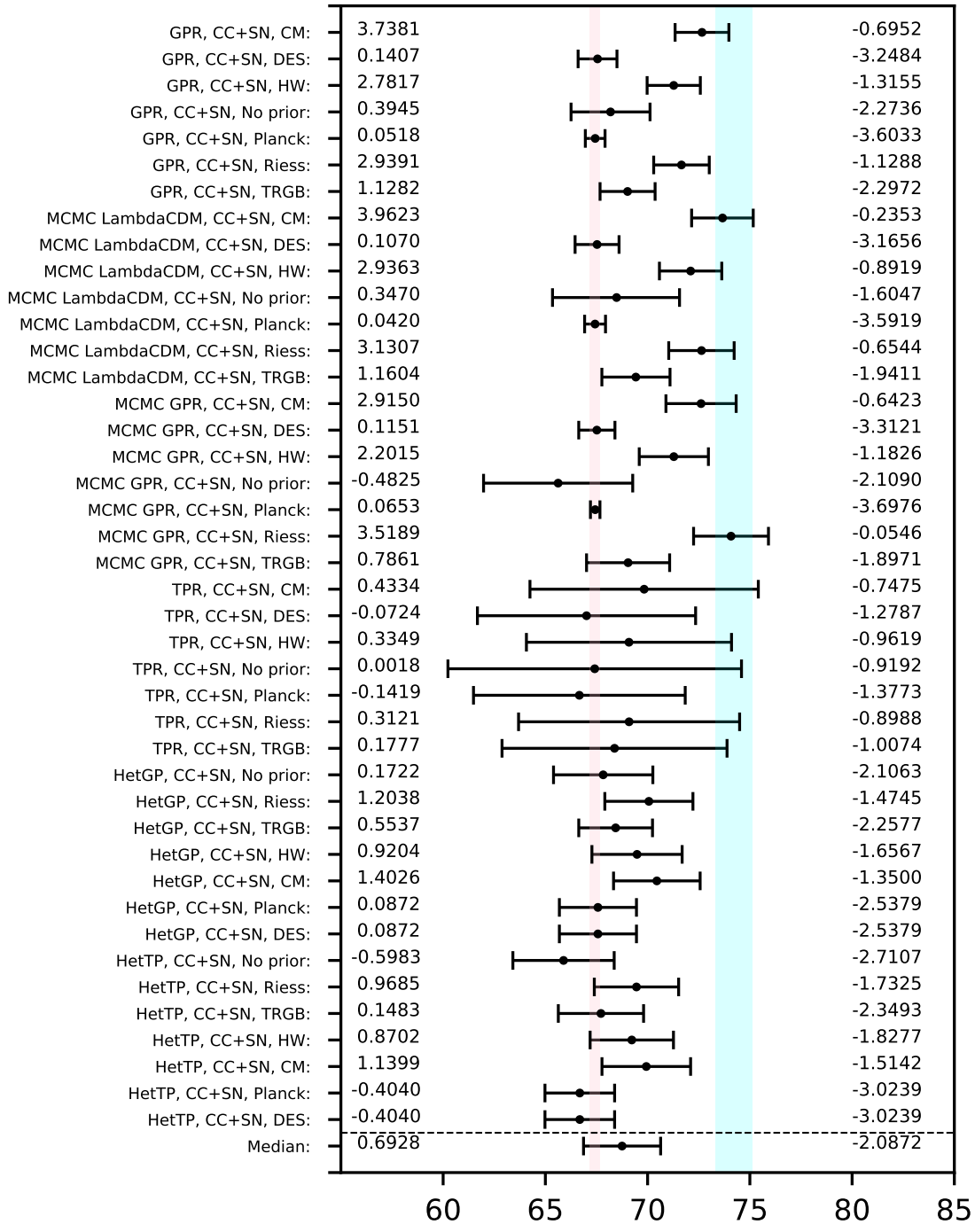



Figure K.6: Comparison of estimates obtained using CC+SN dataset

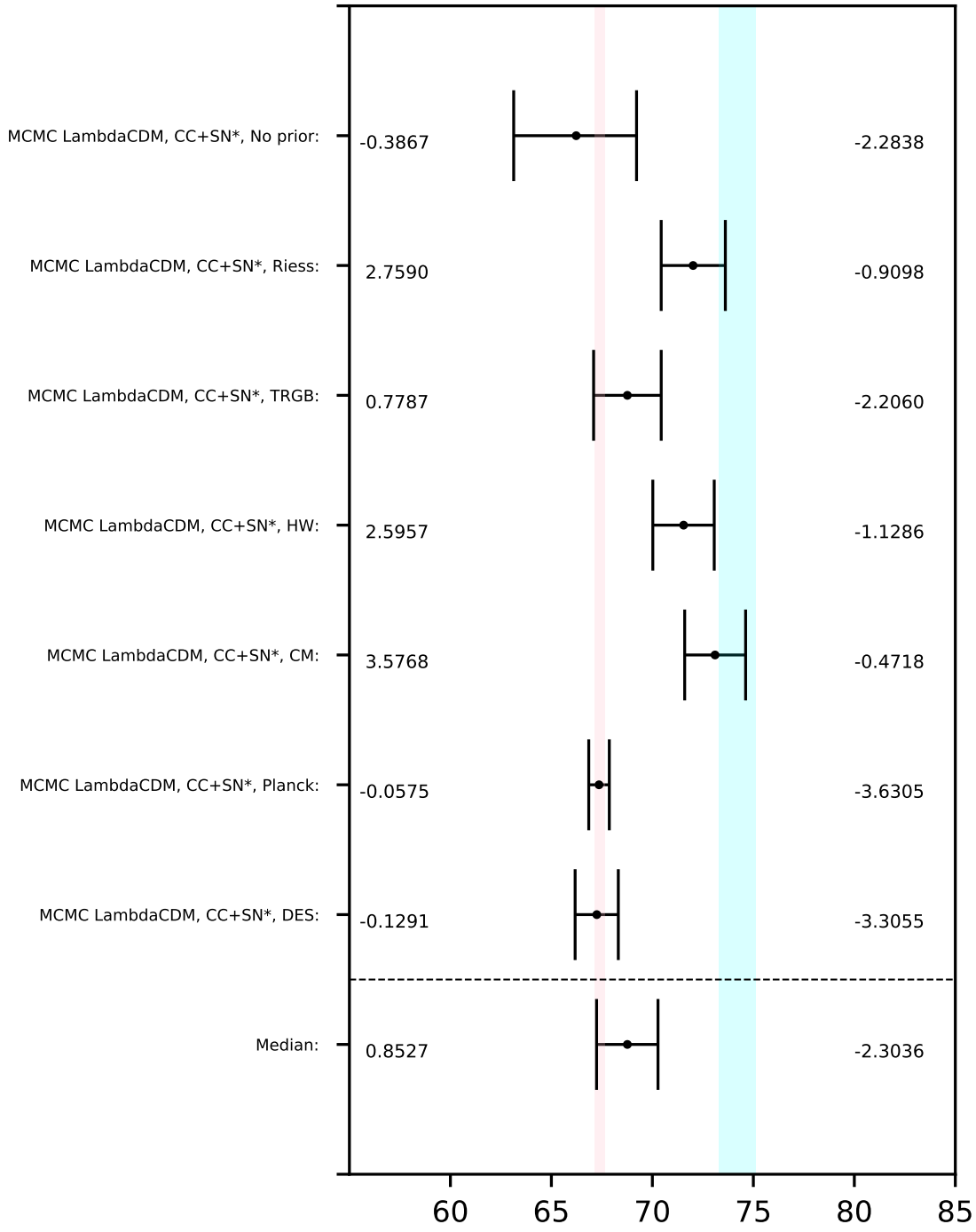



Figure K.7: Comparison of estimates obtained using CC+SN$_*$ dataset

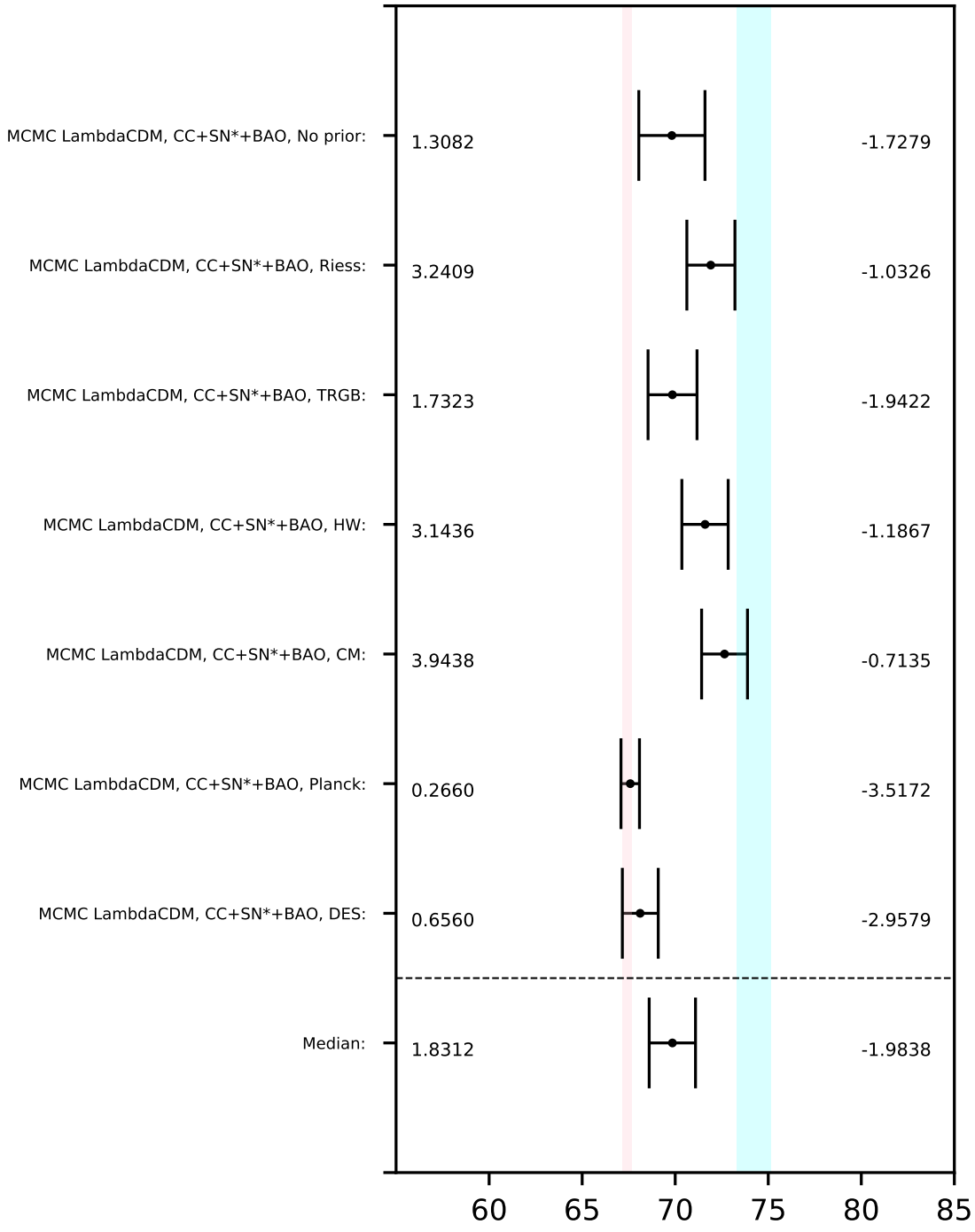



Figure K.8: Comparison of estimates obtained using CC+SN$_*$+BAO dataset

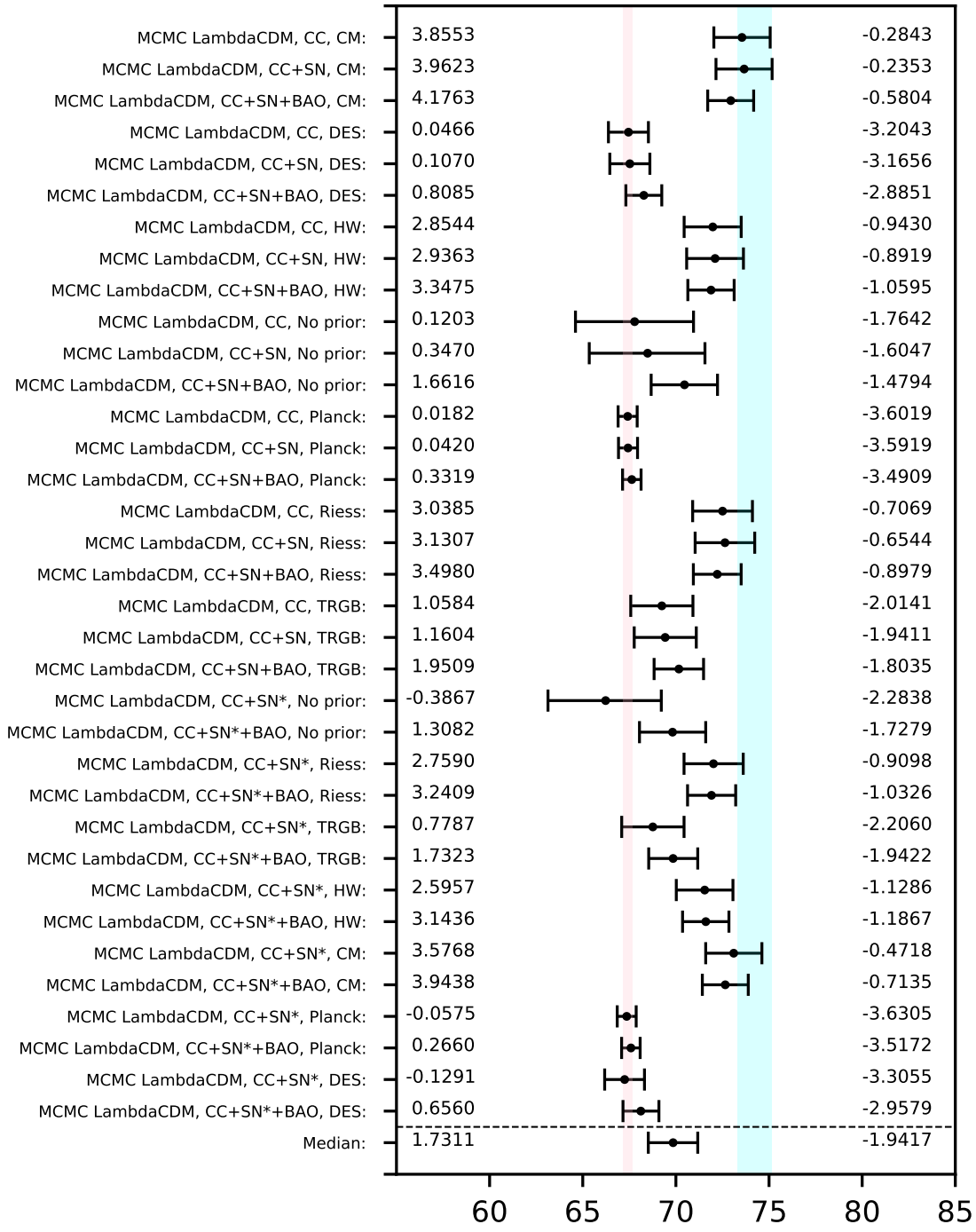



Figure K.9: Comparison of estimates obtained using MCMC $\Lambda CDM$ method

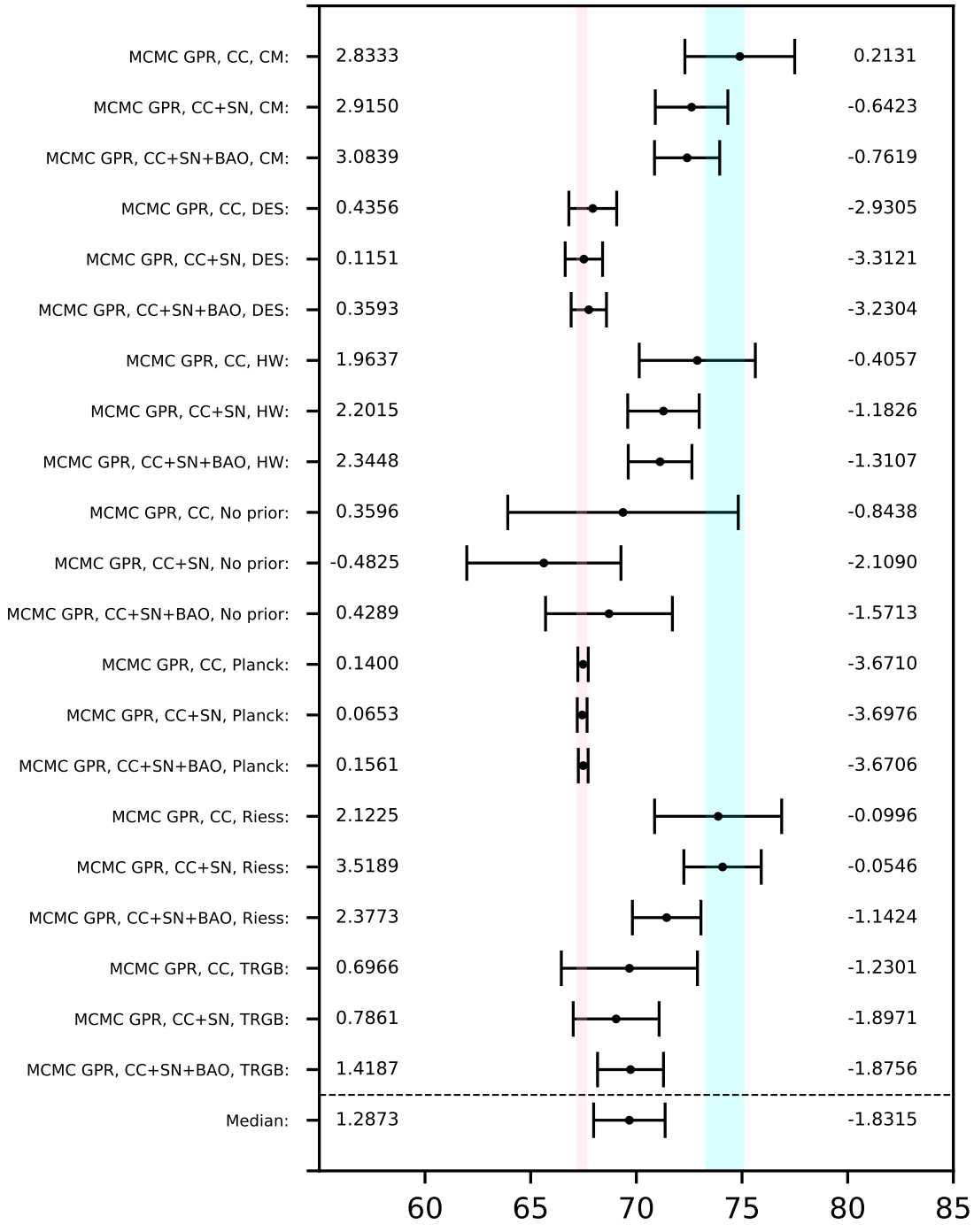



Figure K.10: Comparison of estimates obtained using MCMC GPR method

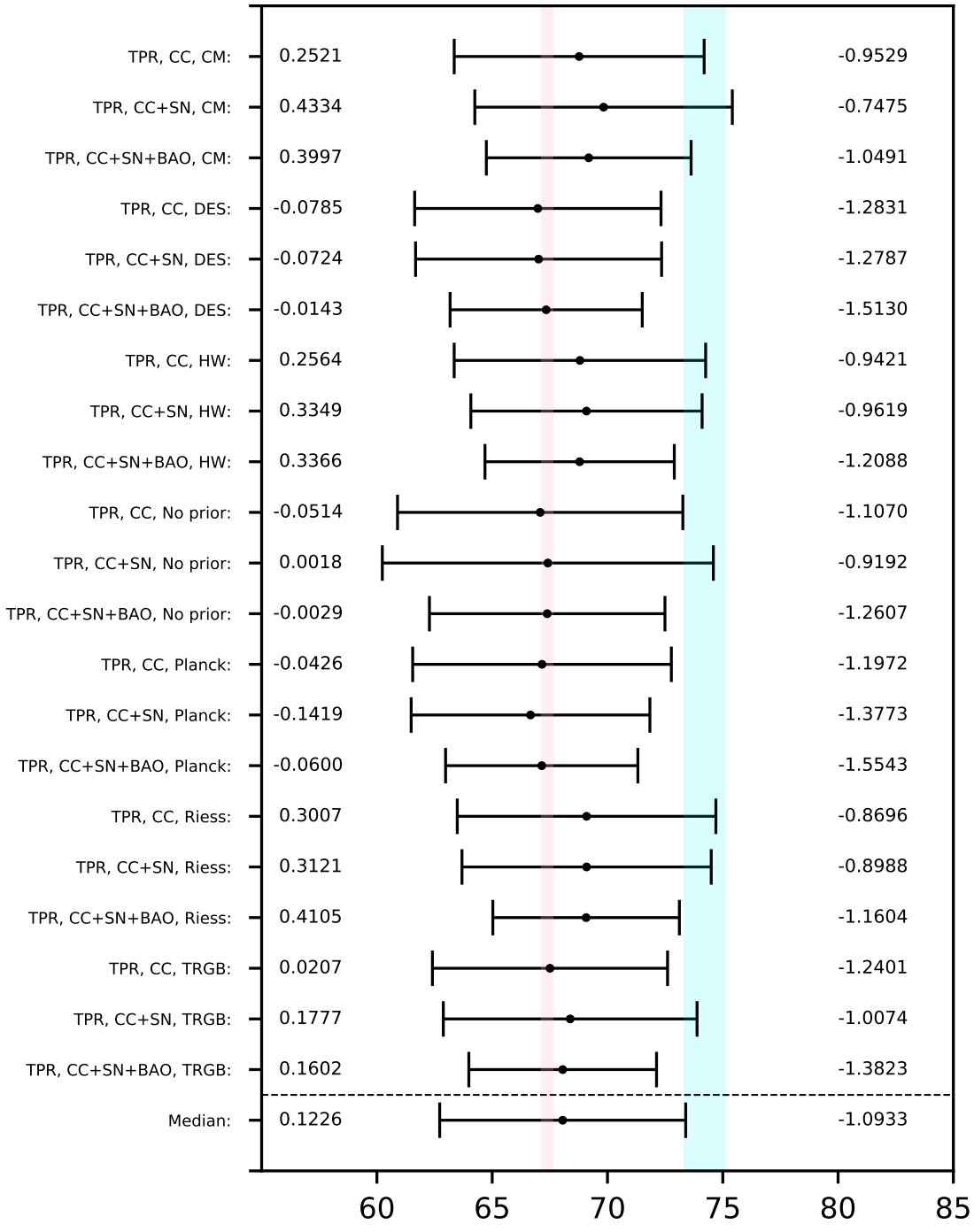



Figure K.11: Comparison of estimates obtained using TPR method

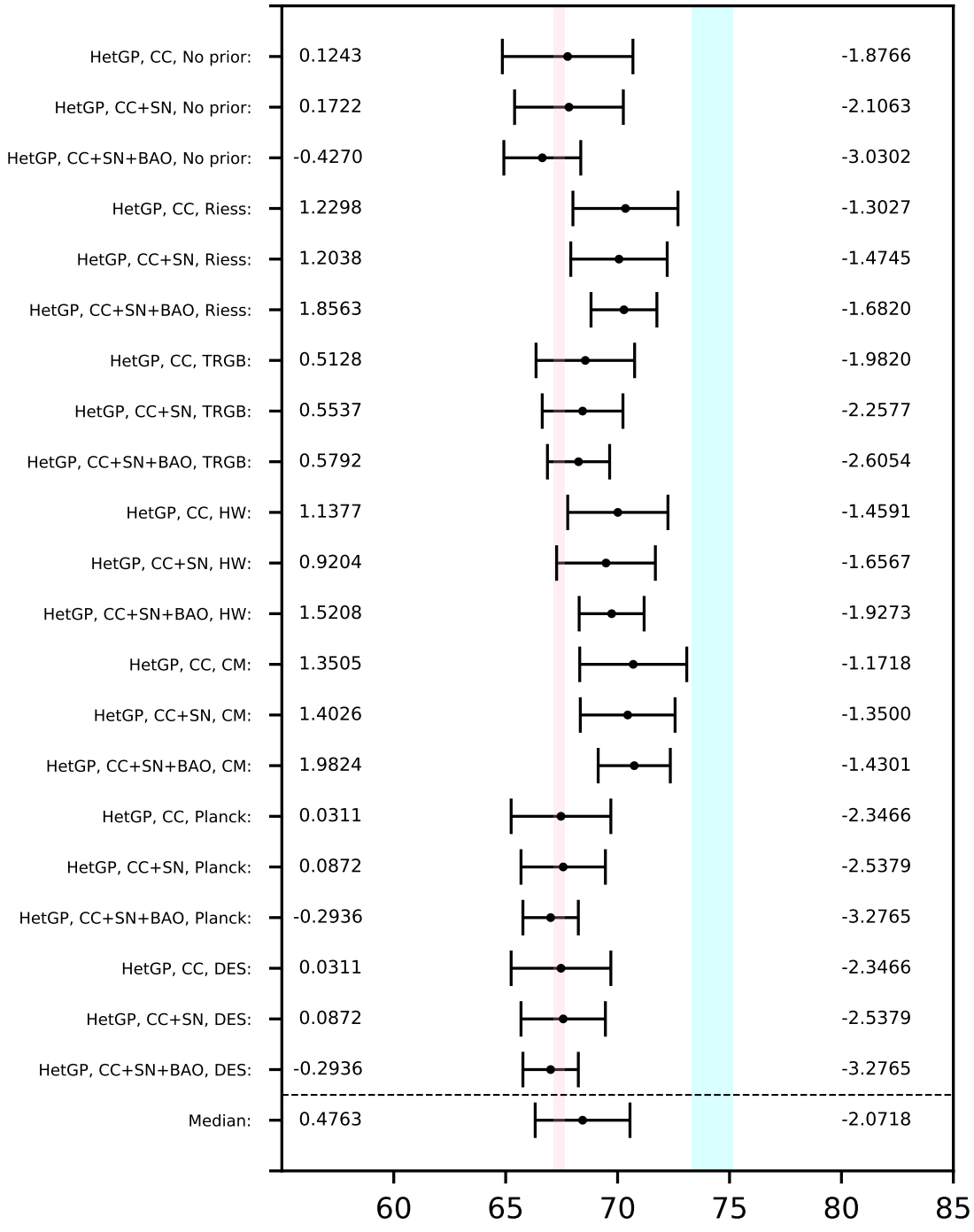



Figure K.12: Comparison of estimates obtained using 'heteroscedastic' GPR method

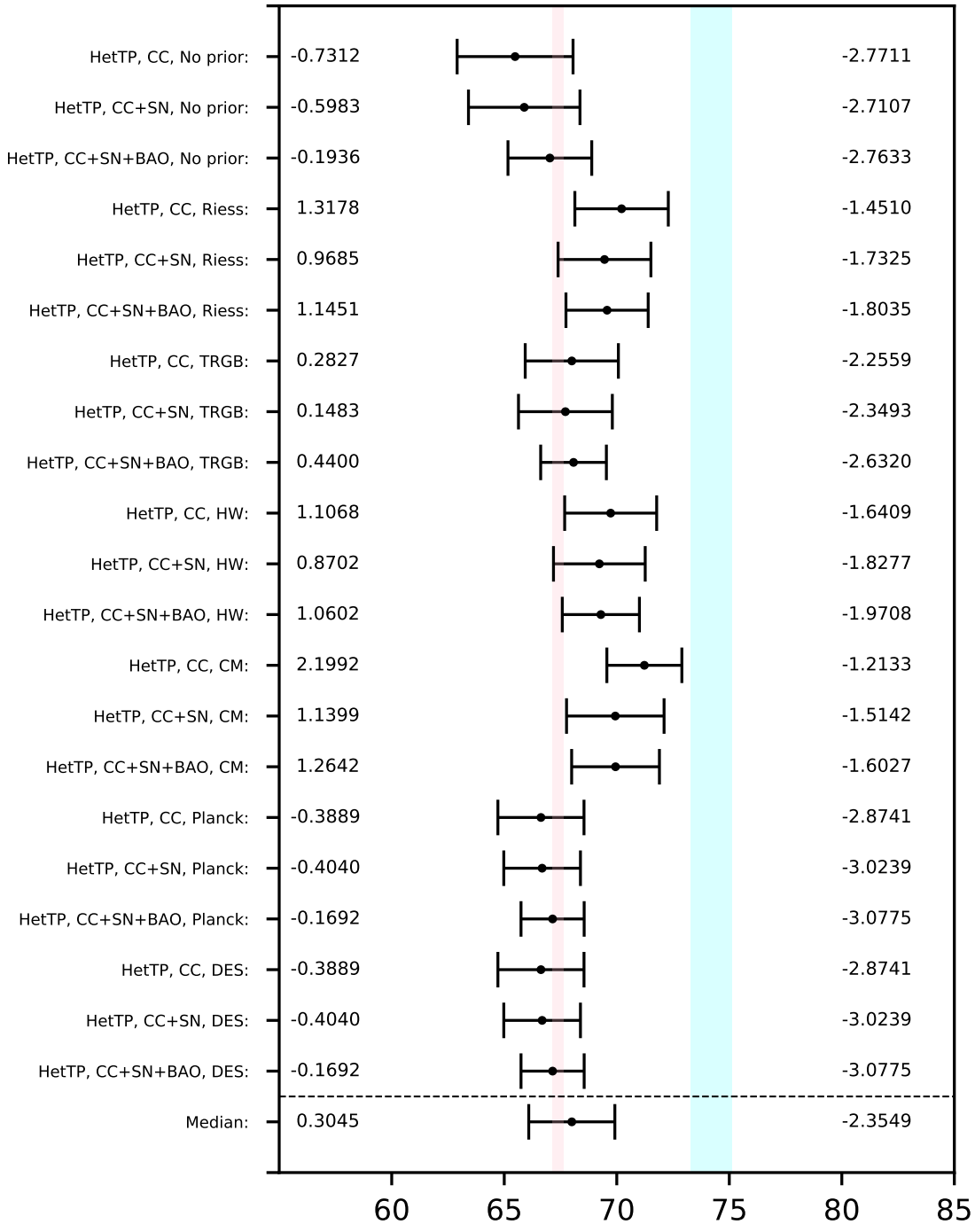



Figure K.13: Comparison of estimates obtained using 'heteroscedastic' TPR method

# References


Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from `www.tensorflow.org`.

Abbott, T., Abdalla, F. B., Aleksić, J., Allam, S., Amara, A., Bacon, D., et al. (2016). The dark energy survey: More than dark energy – an overview. *Mon Not R Astron Soc*, 460(2):1270–1299.

Aghanim, N., Akrami, Y., Arroja, F., Ashdown, M., Aumont, J., Baccigalupi, C., Ballardini, M., et al. (2020a). Planck 2018 results – I. Overview and the cosmological legacy of Planck. *Astron Astrophys*, 641:A1.

Aghanim, N., Akrami, Y., Ashdown, M., Aumont, J., Baccigalupi, C., Ballardini, M., et al. (2020b). Planck 2018 results – VI. Cosmological parameters. *Astron Astrophys*, 641:A6.

Akarsu, Ö. and Dereli, T. (2012). A comparison of the LVDP and $\Lambda CDM$ cosmological models. *Int J Theor Phys*, 51(10):2995–3000.

Akeret, J., Seehars, S., Amara, A., Refregier, A., and Csillaghy, A. (2013). CosmoHammer: Cosmological parameter estimation with the MCMC hammer. *Astron Comput*, 2:27–39.

Alam, S., Ata, M., Bailey, S., Beutler, F., Bizyaev, D., Blazek, J. A., et al. (2017). The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Cosmological analysis of the DR12 galaxy sample. *Mon Not R Astron Soc*, 470(3):2617–2652.

Ankenman, B., Nelson, B. L., and Staum, J. (2010). Stochastic kriging for simulation metamodeling. *Oper Res*, 58(2):371–382.

Avila, F., Bernui, A., Bonilla, A., and Nunes, R. C. (2022). Inferring $S_8(z)$ and $\gamma(z)$ with cosmic growth rate measurements using machine learning. *Eur Phys J C*, 82(7):594.

Bahcall, N. A. (2015). Hubble's law and the expanding universe. *Proc Natl Acad Sci*, 112(11):3173–3175.

Bazot, M., Bourguignon, S., and Christensen-Dalsgaard, J. (2012). A Bayesian approach to the modelling of $\alpha$ Cen A. *Mon Not R Astron Soc*, 427(3):1847–1866.

Belgacem, E., Foffa, S., Maggiore, M., and Yang, T. (2020). Gaussian processes reconstruction of modified gravitational wave propagation. *Phys Rev D*, 101(6).

Bennett, C. L., Larson, D., Weiland, J. L., Jarosik, N., Hinshaw, G., Odegard, N., Smith, K. M., Hill, R. S., Gold, B., Halpern, M., and et al. (2013). Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Final maps and results. *Astrophys J Suppl Ser*, 208(2):20.

Bernal, J. L., Verde, L., and Riess, A. G. (2016). The trouble with $H_0$. *J Cosmol Astropart Phys*, 2016(10):019.

Binois, M. and Gramacy, R. B. (2021). hetgp: Heteroskedastic Gaussian process modeling and sequential design in R. *J Stat Soft*, 98(13):1—44.

Binois, M., Gramacy, R. B., and Ludkovski, M. (2018). Practical heteroscedastic Gaussian process modeling for large simulation experiments. *J Comput Graph Stat*, 27(4):808–821.

Bond, J. R. and Efstathiou, G. (1987). The statistics of cosmic background radiation fluctuations. *Mon Not R Astron Soc*, 226(3):655–687.

Brewer, B. J. and Stello, D. (2009). Gaussian process modelling of asteroseismic data. *Mon Not R Astron Soc*, 395(4):2226–2233.

Briffa, R., Escamilla-Rivera, C., Said, J. L., and Mifsud, J. (2023). $f(T, B)$ gravity in the late Universe in the context of local measurements. *Phys Dark Universe*, 39:101153.

Briffa, R., Escamilla-Rivera, C., Said, J. L., Mifsud, J., and Pullicino, N. L. (2022). Impact of $H_0$ priors on $f(T)$ late-time cosmology. *Eur Phys J Plus*, 137(5):532.

Brooks, S., Gelman, A., Jones, G., and Meng, X.-L. (2011). *Handbook of Markov Chain Monte Carlo*. CRC Press, Boca Raton, US-FL.

Browne, M. W. (1973). Generalized least squares estimators in the analysis of covariance structures. *ETS Res Bull Ser*, 1973(1):i–36.

Bélisle, C. J. P., Romeijn, H. E., and Smith, R. L. (1993). Hit-and-run algorithms for generating multivariate distributions. *Math Oper Res*, 18(2):255–266.

Camarena, D. and Marra, V. (2020). Local determination of the Hubble constant and the deceleration parameter. *Phys Rev Res*, 2:13–28.

Cassidy, D. T. (2016). A multivariate Student's $t$-distribution. *Open J Stat*, 06(03):8.

Christen, J. and Fox, C. (2007). *A general purpose scale-independent MCMC algorithm*. Technical Report I-07-16, Centro de Investigación en Matemáticas, Guanajuato, MX.

Conley, A., Guy, J., Sullivan, M., Regnault, N., Astier, P., Balland, C., et al. (2010). Supernova constraints and systematic uncertainties from the first three years of the Supernova Legacy Survey. *Astrophys J Suppl Ser*, 192(1):1.

Cook, J. D. (2009). Student-$t$ as a mixture of normals. Retrieved from `www.johndcook.com/t_normal_mixture.pdf` on July 4, 2022.

Courbin, F., Saha, P., and Schechter, P. L. (2002). Quasar lensing. In *Gravitational Lensing: An Astrophysical Tool*, pages 1–54. Springer, Heidelberg, DE.

Cuceu, A., Farr, J., Lemos, P., and Font-Ribera, A. (2019). Baryon acoustic oscillations and the Hubble constant: Past, present and future. *J Cosmol Astropart Phys*, 2019(10):044.

Daemi, A., Kodamana, H., and Huang, B. (2019). Gaussian process modelling with Gaussian mixture likelihood. *J Process Control*, 81:209–220.

Damianou, A. and Lawrence, N. D. (2013). Deep Gaussian processes. In *AISTATS '13*, pages 207–215. PMLR.

Dawid, A. P. (1981). Some matrix-variate distribution theory: Notational considerations and a Bayesian application. *Biometrika*, 68(1):265–274.

di Valentino, E., Mena, O., Pan, S., Visinelli, L., Yang, W., Melchiorri, A., et al. (2021). In the realm of the Hubble tension – a review of solutions. *Class Quantum Gravity*, 38(15):153001.

du Mas des Bourboux, H., Le Goff, J.-M., Blomqvist, M., Busca, N. G., Guy, J., Rich, J., et al. (2017). Baryon acoustic oscillations from the complete SDSS-III Ly$\alpha$-quasar cross-correlation function at $z = 2.4$. *Astron Astrophys*, 608:A130.

Dutordoir, V., Wilk, M., Artemev, A., and Hensman, J. (2020). Bayesian image classification with deep convolutional Gaussian processes. In *AISTATS '21*, pages 1529–1539. PMLR.

Duvenaud, D. (2014). Kernel cookbook. Retrieved from `www.cs.toronto.edu/duvenaud/cookbook/index.html` on May 23, 2021.

Eisenstein, D. (2005). Dark energy and cosmic sound. *New Astron Rev*, 49(7):360–365. Wide-Field Imaging from Space.

Escamilla-Rivera, C., Said, J. L., and Mifsud, J. (2021). Performance of non-parametric reconstruction techniques in the late-time universe. *J Cosmol Astropart Phys*, 2021(10):016.

Fan, Z. (2020). Precision cosmology in the era of large surveys. *J Instrum*, 15(10):C10019.

Feast, M. and Walker, A. (1987). Cepheids as distance indicators. *Annu Rev Astron*, 25(1):345–375.

Feigelson, E. D. and Babu, G. J. (2012). *Modern Statistical Methods for Astronomy: With R Applications*. Cambridge University Press, Cambridge, UK.

Foreman-Mackey, D., Hogg, D. W., Lang, D., and Goodman, J. (2013). emcee: The MCMC hammer. *Publ Astron Soc Pac*, 125(925):306–312.

Foreman-Mackey, D., Hogg, D. W., Lang, D., and Goodman, J. (2021). Moves – emcee. Retrieved from `emcee.readthedocs.io/en/stable/user/moves/` on January 29, 2023.

Freedman, W. L., Madore, B. F., Hatt, D., Hoyt, T. J., Jang, I. S., Beaton, R. L., et al. (2019). The Carnegie-Chicago Hubble program. VIII. An independent determination of the Hubble constant based on the tip of the red giant branch. *Astrophys J*, 882(1):34.

Friedmann, A. (1922). Über die Krümmung des Raumes. *Z Phys*, 10(1):377–386. English translation available in *Gen Relativ Gravit*, 31(12): 1991—2000.

Gal, Y. and Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In *ICML '16*, pages 1050–1059. PMLR.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans Pattern Anal Mach Intell*, PAMI-6(6):721–741.

Gibson, N. P., Aigrain, S., Roberts, S., Evans, T. M., Osborne, M., and Pont, F. (2012). A Gaussian process framework for modelling instrumental systematics: Application to transmission spectroscopy. *Mon Not R Astron Soc*, 419(3):2683–2694.

Goodman, J. and Weare, J. (2010). Ensemble samplers with affine invariance. *Comm App Math Comp Sci*, 5(1):65–80.

Gramacy, R. B. (2020). Gaussian process regression. In *Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences*, chapter 5, pages 143–221. Chapman and Hall/CRC, London, UK.

Green, P. J. and Han, X.-L. (1992). Metropolis methods, Gaussian proposals and antithetic variables. In *Stochastic Models, Statistical Methods, and Algorithms in Image Analysis*, pages 142–164. Springer, Heidelberg, DE.

Guo, R.-Y., Zhang, J.-F., and Zhang, X. (2019). Can the $H_0$ tension be resolved in extensions to $\Lambda CDM$ cosmology? *J Cosmol Astron Phys*, 2019(2):054.

Gómez-Valent, A. (2019). Current status of the $H_0$-tension. Retrieved from `www.mpi-hd.mpg.de/lin/seminar_theory/talks/Talk_Gomez_211019.pdf` on April 12, 2021.

Gómez-Valent, A. and Amendola, L. (2018). $H_0$ from cosmic chronometers and Type Ia

supernovae, with Gaussian processes and the novel weighted polynomial regression method. *J Cosmol Astropart Phys*, 2018(4).

Harpaz, A. (1993). *Stellar Evolution*. CRC Press, Boca Raton, US-FL.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.

Hinshaw, G., Larson, D., Komatsu, E., Spergel, D. N., Bennett, C. L., Dunkley, J., et al. (2013). Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Cosmological parameter results. *Astrophys J*, 208(2):19.

Hollow, R. (2006). Concepts for the cosmic engine. Retrieved from `www.atnf.csiro.au/outreach/education/teachers/resources/conceptscosmicenginestw2006.pdf` on December 4, 2022.

Hubble, E. (1929). A relation between distance and radial velocity among extra-galactic nebulae. *P Natl Acad Sci USA*, 15(3):168–173.

Huijser, D., Goodman, J., and Brewer, B. J. (2022). Properties of the affine-invariant ensemble sampler's 'stretch move' in high dimensions. *Aust N Z J Stat*, 64(1):1–26.

Hult, H. and Lindskog, F. (2002). Multivariate extremes, aggregation and dependence in elliptical distributions. *Adv Appl Probab*, 34(3):587–608.

International Astronomical Union (2018). IAU members vote to recommend renaming the Hubble law as the Hubble–Lemaître law. Retrieved from `www.iau.org/news/pressreleases/detail/iau1812` on November 30, 2021.

Jähnichen, P., Wenzel, F., Kloft, M., and Mandt, S. (2018). Scalable generalized dynamic topic models. *arXiv preprint arXiv:1311.6678*.

Kelly, B. C., Becker, A. C., Sobolewska, M., Siemiginowska, A., and Uttley, P. (2014). Flexible and scalable methods for quantifying stochastic variability in the era of massive time-domain astronomical data sets. *Astrophys J*, 788(1):33.

Kirkby, J. L., Nguyen, D., and Nguyen, D. (2019). Moments of Student's $t$-distribution: A unified approach. *arXiv preprint arXiv:1912.01607*.

Krige, D. G. (1979). Lognormal-de Wijsian geostatistics for ore evaluation. *Math Geol*, 11(6).

Lemaître, G. (1927). Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques. *Ann Soc Sci Bruxelles*, 47:49–59.

Lewis, A. (2013). Efficient sampling of fast and slow cosmological parameters. *Phys Rev D*, 87(10).

Li, E.-K., Du, M., Zhou, Z.-H., Zhang, H., and Xu, L. (2020). Testing the effect of $H_0$ on $f\sigma_8$ tension using a Gaussian process method. *Mon Not R Astron Soc*,

501(3):4452–4463.

Liu, H., Ong, Y.-S., Shen, X., and Cai, J. (2020). When Gaussian process meets big data: A review of scalable GPs. *IEEE Trans Neural Netw Learn Syst*, 31(11):4405–4423.

Maeder, A. (2017). An alternative to the $\Lambda CDM$ model: The case of scale invariance. *Astrophys J*, 834(2):194.

Matheron, G. (1973). The intrinsic random functions and their applications. *Adv Appl Probab*, 5(3):439–468.

Melia, F. and Yennapureddy, M. K. (2018). Model selection using cosmic chronometers with Gaussian processes. *J Cosmol Astropart Phys*, 2018(2):034.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *J Chem Phys*, 21(6):1087–1092.

Mohammed, R. O. and Cawley, G. C. (2017). Over-fitting in model selection with Gaussian process regression. In *MLDM '17*, pages 192–205. Springer.

Moresco, M., Pozzetti, L., Cimatti, A., Jimenez, R., Maraston, C., Verde, L., et al. (2016). A 6% measurement of the Hubble parameter at $z \sim 0.45$: Direct evidence of the epoch of cosmic re-acceleration. *J Cosmol Astropart Phys*, 2016(5):14.

Mueller, T. (2012). Kernel approximations for efficient SVMs (and other feature extraction methods). Retrieved from `peekaboo-vision.blogspot.com/2012/12/kernel-approximations-for-efficient.html` on May 22, 2021.

Nave, C. R. (2016). Density parameter, Ω. Retrieved from `hyperphysics.phy-astr.gsu.edu/hbase/Astro/denpar.html` on January 19, 2022.

Nemiroff, R. J. and Patla, B. (2008). Adventures in Friedmann cosmology: A detailed expansion of the cosmological Friedmann equations. *Am J Phys*, 76(3):265—-276.

Odderskov, I., Hannestad, S., and Haugbølle, T. (2014). On the local variation of the Hubble constant. *J Cosmol Astropart Phys*, 2014(10):28.

Oddo, A., Rizzo, F., Sefusatti, E., Porciani, C., and Monaco, P. (2021). Cosmological parameters from the likelihood analysis of the galaxy power spectrum and bispectrum in real space. *J Cosmol Astropart Phys*, 2021(11):038.

Owen, A. and Tribble, S. D. (2005). A quasi-Monte Carlo Metropolis algorithm. *P Natl Acad Sci USA*, 102(25):8844–8849.

Padilla, L. E., Tellez, L. O., Escamilla, L. A., and Vazquez, J. A. (2021). Cosmological parameter inference with Bayesian statistics. *Universe*, 7(7):213.

Pandey, K. L., Karwal, T., and Das, S. (2020). Alleviating the $H_0$ and $\sigma_8$ anomalies with a decaying dark matter model. *J Cosmol Astropart Phys*, 2020(07):026.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). scikit-learn: Machine learning in Python. *J Mach Learn Res*, 12(85):2825–2830.

Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R. A., Nugent, P., Castro, P. G., et al. (1999). Measurements of $\Omega$ and $\Lambda$ from 42 high-redshift supernovae. *Astrophys J*, 517(2):565–586.

Perlmutter, S. and Schmidt, B. P. (2003). Measuring cosmology with supernovae. *Lect Notes Phys*, pages 195–217.

Prokopec, T. (2014). Lecture notes for cosmology: Part II: The standard cosmological model. Retrieved from `webspace.science.uu.nl/proko101/2scm.pdf` on April 19, 2021.

Ranganathan, A., Yang, M.-H., and Ho, J. (2010). Online sparse Gaussian process regression and its applications. *IEEE Trans Image Process*, 20(2):391–404.

Refsdal, S. (1964). On the possibility of determining Hubble's parameter and the masses of galaxies from the gravitational lens effect. *Mon Not R Astron Soc*, 128(4):307–310.

Riess, A. G., Casertano, S., Yuan, W., Macri, L. M., and Scolnic, D. (2019). Large Magellanic Cloud Cepheid standards provide a 1% foundation for the determination of the Hubble constant and stronger evidence for physics beyond $\Lambda CDM$. *Astrophys J*, 876(1):85.

Riess, A. G., Filippenko, A. V., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P. M., et al. (1998). Observational evidence from supernovae for an accelerating universe and a cosmological constant. *Astron J*, 116(3):1009–1038.

Riess, A. G., Rodney, S. A., Scolnic, D. M., Shafer, D. L., Strolger, L.-G., Ferguson, H. C., et al. (2018). Type Ia supernova distances at redshift $> 1.5$ from the Hubble Space Telescope multi-cycle treasury programs: The early expansion rate. *Astrophys J*, 853(2):126.

Ritter, C. and Tanner, M. A. (1992). Facilitating the Gibbs sampler: The Gibbs stopper and the Griddy-Gibbs sampler. *J Am Stat Assoc*, 87(419):861–868.

Royal Swedish Academy of Sciences (2011). Scientific background on the Nobel Prize in Physics 2011: The accelerating universe. Retrieved from `www.nobelprize.org/uploads/2018/06/advanced-physicsprize2011-1.pdf` on May 24, 2021.

Ryden, B. (2017). *Introduction to Cosmology*. Cambridge University Press, Cambridge, UK.

Sacks, J., Schiller, S. B., and Welch, W. J. (1989). Designs for computer experiments. *Technometrics*, 31(1):41–47.

Santner, T. J., Williams, B. J., Notz, W. I., and Williams, B. J. (2003). *The Design and Analysis of Computer Experiments*, volume 1. Springer, Heidelberg, DE.

Scolnic, D., Brout, D., Carr, A., Riess, A. G., Davis, T. M., Dwomoh, A., et al. (2022). The Pantheon+ analysis: The full data set and light-curve release. *Astrophys J*, 938(2):113.

Scolnic, D. M., Jones, D. O., Rest, A., Pan, Y. C., Chornock, R., Foley, R. J., et al. (2018). The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from Pan-STARRS1 and Cosmological Constraints from the Combined Pantheon Sample. *Astrophys J*, 859(2):101.

Seikel, M. and Clarkson, C. (2013). Optimising Gaussian processes for reconstructing dark energy dynamics from supernovae. *arXiv preprint arXiv:1311.6678*.

Seikel, M., Clarkson, C., and Smith, M. (2012). Reconstruction of dark energy and expansion dynamics using Gaussian processes. *J Cosmol Astropart Phys*, 2012(6):036.

Shah, A., Wilson, A., and Ghahramani, Z. (2014). Student-$t$ Processes as Alternatives to Gaussian Processes. In Kaski, S. and Corander, J., editors, *AISTATS '14*, volume 33 of *PMLR*, pages 877–885, Reykjavik, IS. PMLR.

Shiryayev, A. N. (1992). Interpolation and extrapolation of stationary random sequences. In *Selected Works of A. N. Kolmogorov: Volume II – Probability Theory and Mathematical Statistics*, pages 272–280. Springer, Dordrecht, NL.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *J R Stat Soc Series B Stat Methodol*, 64(4):583–639.

Stern, D., Jimenez, R., Verde, L., Kamionkowski, M., and Stanford, S. A. (2010). Cosmic chronometers: constraining the equation of state of dark energy. I: $H(z)$ measurements. *J Cosmol Astropart Phys*, 2010(2):8.

Tang, Q., Niu, L., Wang, Y., Dai, T., An, W., Cai, J., et al. (2017). Student-$t$ process regression with Student-$t$ likelihood. In *IJCAI '17*, pages 2822–2828.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *J Am Stat Assoc*, 82(398):528–540.

Tanvir, N. R., Shanks, T., Ferguson, H. C., and Robinson, D. R. T. (1995). Determination of the Hubble constant from observations of Cepheid variables in the galaxy M96. *Nature*, 377(6544):27–31.

Teckentrup, A. L. (2020). Convergence of Gaussian process regression with estimated hyper-parameters and applications in Bayesian inverse problems. *SIAM-ASA J Uncertain*, 8(4):1310–1337.

Thompson, P. D. (1956). Optimum smoothing of two-dimensional fields. *Tellus*, 8(3):384–393.

Tian, Y., Ko, C.-M., and Chiu, M.-C. (2013). Hubble constant lensing and time delay in relativistic modified Newtonian dynamics. *Astrophys J*, 770(2):154.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Ann Stat*, 22(4):1701–1728.

Tsun, A. (2020). Maximum *a posteriori* estimation. Retrieved from `courses.cs.washington.edu/courses/cse312/20su/files/student_drive/7.5.pdf` on December 23, 2021.

Uttley, P., McHardy, I. M., and Vaughan, S. (2005). Non-linear X-ray variability in X-ray binaries and active galaxies. *Mon Not R Astron Soc*, 359(1):345–362.

Verde, L. (2010). Statistical methods in cosmology. In *Lectures on Cosmology*, pages 147–177. Springer, Heidelberg, DE.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in python. *Nat Methods*, 17:261–272.

Vogt, N. (2012). Astronomy 505: Galaxy spectra. Retrieved from `astronomy.nmsu.edu/nicole/teaching/ASTR505/lectures/lecture26/slide01.html` on April 23, 2021.

Whittle, P. (1963). *Prediction and Regulation by Linear Least-Square Methods*. English Universities Press, London, UK.

Wiener, N. (1949). *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. MIT Press, Cambridge, US-MA.

Williams, C. K. I. and Rasmussen, C. E. (1996). Gaussian processes for regression. In Touretzky, D. S., Mozer, M. C., and Hasselmo, M. E., editors, *NIPS '96*, Cambridge, US-MA. MIT Press.

Williams, C. K. I. and Rasmussen, C. E. (2006). *Gaussian Processes for Machine Learning*, volume 2(3). MIT Press, Cambridge, US-MA.

Wong, K. C., Suyu, S. H., Chen, G. C.-F., Rusu, C. E., Millon, M., Sluse, D., et al. (2019). H0LiCOW – XIII. A 2.4 per cent measurement of $H_0$ from lensed quasars: $5.3\sigma$ tension between early- and late-universe probes. *Mon Not R Astron Soc*, 498(1):1420–1439.

Wu, S., Zhou, X., and Gao, Y. (2019). A novel indoor coverage measurement scheme based on FRFT and Gaussian process regression. In *2019 IEEE Glob Commun Conf*, pages 1–6.

Yu, S., Tresp, V., and Yu, K. (2007). Robust multi-task learning with $t$-processes. In

*ICML '07*, pages 1103–1110.

Zeilik, M., Gregory, S., and van Panhuys Smith, E. (1992). *Introductory Astronomy and Astrophysics*. Saunders Golden Sunburst Series. Saunders College, Rochester, US-NY.

Zhang, Y. and Yeung, D. (2010). Multi-task learning using generalized $t$-process. In Teh, Y. W. and Titterington, M., editors, *AISTATS '10*, volume 9 of *PMLR*, pages 964–971, Chia, IT.

Zhao, G.-B., Wang, Y., Saito, S., Gil-Marín, H., Percival, W. J., Wang, D., et al. (2018). The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample: A tomographic measurement of cosmic structure growth and expansion rate based on optimal redshift weights. *Mon Not R Astron Soc*, 482(3):3497–3513.