University of The Western Cape

Doctoral Thesis

Probing the Expansion History of the Universe using Supernovae and Baryon Acoustic Oscillations

Author:

Sahba Yahya Hamid Ali

Supervisor:

Prof. R. Maartens

Co-Supervisors:

Prof. C. Clarkson

& Prof. M.G. Santos

A thesis submitted in fulfilment of the requirements

for the degree of Doctor of Philosophy

in the

Department of Physics

March 2016

Abstract

The standard model of cosmology (the ΛCDM model) has been very successful and is compatible with all observational data up to now. However, it remains an important task to develop and apply null tests of this model. These tests are based on observables that probe cosmic distances and cosmic evolution history. Supernovae observations use the so-called 'standard candle' property of SNIa to probe cosmic distances $D(z)$. The evolution of the expansion rate $H(z)$ is probed by the baryon acoustic oscillation (BAO) feature in the galaxy distribution, which serves as an effective 'standard ruler'. The observables $D(z)$ and $H(z)$ are used in various consistency tests of ΛCDM that have been developed. We review the consistency tests, also looking for possible new tests.

Then the tests are applied, first using existing data, and then using mock data from future planned experiments. In particular we use data from the recently commissioned Dark Energy Survey (DES) for SNIa. Gaussian Processes, and possibly other non-parametric methods, used to reconstruct the derivatives of $D(z)$ and $H(z)$ that are needed to apply the null tests of the standard cosmological model. This allow us to estimate the current and future power of observations to probe the ΛCDM model, which is the foundation of modern cosmology.

In addition, we present an improved model of the HI galaxy number counts and bias from semi-analytic simulations, and we use it to calculate the expected yield of HI galaxies from surveys with a variety of phase 1 and 2 SKA configurations. We illustrate the relative performance of the different surveys by forecasting errors on the radial and transverse scales of the BAO feature. We use the Fisher matrix method to estimate the error bars on the cosmological parameters from future SKA HI galaxy surveys. We find that the SKA phase 1 galaxy surveys will not contend with surveys such as the Baryon Oscillation Spectroscopic Survey (BOSS) whereas the full "billion galaxy survey" with SKA phase 2 will deliver the largest dark energy Figure of Merit of any current or future large-scale structure survey.

Declaration of Authorship

I, Sahba Yahya Hamid Ali, I declare that *"Probing the Expansion History of the Universe using Supernovae and Baryon Acoustic Oscillations"* is my own work, that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.The following papers have been published, based on the work presented in this thesis:

- S. Yahya, P. Bull, M. G. Santos, M. Silva, R. Maartens, P. Okouma and B. Bassett, " Cosmological performance of SKA HI galaxy surveys," *Mon. Not. Roy. Astron. Soc.*, 450, 3, 2251, March 2015.
- M. G. Santos, D. Alonso, P. Bull, M. Silva and S. Yahya, "HI galaxy simulations for the SKA: number counts and bias," arXiv:1501.03990, January 2015.
- S. Yahya, M. Seikel, C. Clarkson, R. Maartens and M. Smith, " Null tests of the cosmological constant using supernovae," *Phys. Rev. D*, 89:023503, January 2014.
- M. Seikel, S. Yahya, R. Maartens and C. Clarkson, "Using H(z) data as a probe of the concordance model," *Phys. Rev. D*, 86:083001, October 2012.

Full name.....Sahba Yahya Hamid Ali..... Date.....9 March 2016.....

Signed.....Sahba Yahya.....

Acknowledgements

I would like to thank my family and friends for their understanding and support. Without them, I would not have been able to complete my PhD.

Many thanks to the Astrophysics group at the University of the Western Cape, my supervisor Roy Maartens and co-supervisors, Chris Clarkson and Mario G. Santos, for their provided support.

I would like to thank the South African Square Kilometre Array Project (SKA) and the German Academic Exchange Service (DAAD) for their financial support during my PhD.

I am thankful to the National Astrophysics and Space Science programme (NASSP) for their financial support throughout my Honor and MSc degree, without it I would not have gotten to this point.

Finally I would like to thank the researchers that I consistently learned from, Phil Bull, Mathew Smith, Patrice Okouma, Marina Seikel and all the people I had a productive or inspiring conversation during these years...

Contents

List of Figures

List of Tables

Chapter 1

Introduction

One of the key problems in Cosmology is understanding the dark energy that is driving the late-time acceleration of the Universe. Dark energy affects the expansion history of the universe, and probes of this history thus provide tests of dark energy models [4–6]. Determining the time dependence of the equation of state, as well as its present density is an essential step in identifying the physical origin of dark energy [7].

The simplest dark energy candidate is the cosmological constant suggested by the standard model (ΛCDM). Alternative models such as the Quintessence models suggest scenarios where the dark energy density evolves with time. Although recent data and observations are supportive to the cosmological constant scenario, the data can not exclude the dynamical dark energy models. Therefore an important goal for current and future observations is to measure the dark energy parameters in many different ways and detect any departure from the constant value of Λ that has been suggested by the ΛCDM model [8, 9].

Methods to detect time variation in the dark energy density have been developed. It appears those methods are not efficient in detecting the variation in time even with high quality simulated data [10]. One approach that has been followed is to use the luminosity distance fitting formulas from simulated data to reconstruct the assumed time varying dark energy model. The issue with this approach is the unphysical parameters introduced by the luminosity distances fitting formulas, these make recovering the attributes of the assumed model problematic [11–13].

Another approach proposed to distinguish models of dark energy is to use expansions of the equation of state of the dark energy, w. This can be achieved using a polynomial fit in redshift space. This method takes account the slow variation of w. A disadvantage of this fitting approach is that it introduces unphysical parameters. Instead of the polynomial fitting, a logarithmic expansion in the redshift space has also been suggested [14, 15].

In this thesis, we consider a complementary approach to test the consistency of the standard model itself independent of the values of Ω_m and Ω_K. This is achieved by designing specific tests to probe the deviation from ΛCDM model. This approach is a model independent technique, requiring that the data to be from model independent measurements.

Cosmological observations such as Supernovae type Ia (SNIa) apparent magnitude versus redshift data and Baryon Acoustic Oscillation (BAO) scale data provide strong evidence for dark energy.

Two upcoming ground based survey missions, Pan-STARRS (The Panoramic Survey Telescope and Rapid Response System) and LSST(The Large Synoptic Survey Telescope), will add valuable knowledge about the nature of SNIa. The major goal for Pan-STARRS is characterizing Earth-approaching objects, including asteroids and comets, however measuring cosmological objects such as Supernovae is also on of the important scientific goals of this mission. The LSST is a 6.7 meter telescope that will be fully operational by 2022 and will observe all-sky providing millions of SNIa observations during the planned 10 years of the LSST mission.The measurements from these large surveys should substantially reduce the statistical errors in the SN Hubble diagram.

Another future observations of SNIa from upcoming surveys such as the Dark Energy Survey (DES), will provide accurate data that can improve our understanding of the dark energy. While the Baryon Oscillation Spectroscopic Survey (BOSS) already observe the spatial distribution of luminous red galaxies and quasars to detect the BAO signal out to redshift ~ 0.7. Also, Euclid is a Europe Space Agency (ESA) mission will investigate the distance-redshift relationship and the evolution of cosmic structures by measuring shapes and redshifts of galaxies and clusters of galaxies out to redshifts ~ 2. The Square Kilometer Array (SKA) is expected to be the largest and most sensitive radio telescope in the world by 2030, and it will provide accurate measurements of the BAO scale.

This thesis is a combination of interconnected projects discussed in different chapters. Chapters 2 and 3 are introductory chapters; their role is to discuss the theory behind the standard model in cosmology and the large scale structure of the universe respectively.

In Chapter 4, we introduce model independent tests for the standard model. Those tests require measurements of luminosity distance. We consider the luminosity distance of SNIa from the Sloan Digital Sky Survey (SDSS) [16]. Then we use SDSS measurements to compute the derivatives of the luminosity distance.

In Chapter 5 we discuss the Hubble rate data from different probes; the BAO and the galaxy age measurements. We compute the derivatives of the Hubble rate. Then we use the results to perform consistency tests for the ΛCDM model. Some of these tests we develop and introduce in this chapter.

The null tests we examine, do not rely on parameterizations of observables, but focus on quantities that are constant only if dark energy is a cosmological constant.

Gaussian Processes (GP) is a powerful supervised machine learning technique that can be applied to regression problems or classification tasks. As a non-parametric reconstruction technique, GP is used as a smoothing technique in order to extract Hubble rate in Chapter 5 or distance derivatives in Chapter 4 in a model independent way. A detailed description of this method provided in Appendix A.

In Chapter 6, we explore some of the potential uses of the SKA. For neutral hydrogen (HI) redshift galaxy surveys the key inputs are the flux sensitivity ($S_{\rm rms}$), the detection threshold, the telescope field of view and the assumed model for HI evolution. We use semi-analytical simulations (SAX-S^3) of HI to predict HI galaxy number density (dN/dz) and the galaxy bias. We explore different $S_{\rm rms}$'s associated with the different stages of the SKA telescope. We use the Fisher matrix to forecast the performance of SKA HI galaxy surveys in measuring the BAO components.

Finally we conclude in Chapter 7.

Chapter 2

The Expansion History of the Universe

There is no special place in the Universe. Cosmology works with the assumption, known as the cosmological principle, that at any given time the distribution of matter in the universe is homogeneous and isotropic [17].

The homogeneous isotropic universe on large scale is supported by the cleanest observations to date, which is the Cosmic Microwave Background (CMB) observations [18]. Although the early universe matter distribution was homogeneous, the late universe grows inhomogeneous matter distribution with time due to gravitational instability. The local regions of the universe show inhomogeneities, such as stars and galaxies.

We know that from galaxy spectra, if the galaxy is moving away/toward us the frequency shifts to a higher/lower frequency which appears bluer/redder. The galaxy will be referred to as being blue/red shifted. All galaxies observed happen to be redshifted, except for some nearby galaxies that are affected by the gravitational field of the local group [9, 19]. Redshift is defined as

$$z = \frac{\lambda_{\text{obs}} - \lambda_{\text{em}}}{\lambda_{\text{em}}}, \tag{2.1}$$

where λ_{obs} is the observed wavelength and λ_{em} is the emitted wavelength from the galaxy. For small redshifts we have the Doppler approximation to the redshift:

$$z = \frac{v}{c}, \tag{2.2}$$

where v is the receding velocity of the galaxy and c is the speed of light.

In 1929, Edwin Hubble also used the redshift concept to measure the recession velocity and the distance of 1355 galaxies. These measurements showed that the two quantities were proportional to each other. This discovery led to the famous formula known as Hubble's law:

$$\vec{v} = H_0 \vec{r}, \tag{2.3}$$

where H_0 is the proportionality constant, known as Hubble constant. This work by Hubble was the first evidence of an expanding universe [20].

2.1 The Standard Model

The standard model or Λ cold dark matter model (ΛCDM) based on a Friedmann-Lemaitre-Robertson-Walker (FLRW) metric with baryonic matter, cold dark matter, radiation (photons and neutrinos) and a cosmological constant. The ΛCDM model best fits the following measurements:

- The CMB temperature anisotropy and polarization [18, 21].
- The large scale structure in the distribution of galaxies [22].
- The abundance of hydrogen, helium and lithium [23, 24].
- The accelerated expansion of the universe [25].

The model is based on the assumption that General Relativity is the correct theory of gravity on large scales. From a first look, one might think ΛCDM just needs more observations to narrow down the uncertainties on the model parameters to reach perfection. However, the discovery of the accelerating expansion of the universe introduces an extra challenge to the fundamental theory of gravity and ΛCDM model.

The biggest challenge with ΛCDM model is the cosmological constant Λ problem, where Λ is responsible for the acceleration of the Universe, with

$$\rho_\Lambda \text{ (cosmological constant)} \approx \left(10^{-47}\,\text{GeV}^4\right). \tag{2.4}$$

The problem is the very small and highly fine-tuned value of Λ, where the observed value of ρ_Λ is smaller by a factor of 10^{56} than $(1\ \text{TeV})^4$, which is the smallest possible value predicted by current particle physics. The standard model provides no physical cause for Λ, which led to thinking that w might be differing from -1 or to vary with time [8, 9].

The ΛCDM model is based on the Friedmann-Lemaitre-Robertson-Walker (FLRW) metric, the Friedmann equations and the cosmological equations of state to describe the observable universe.

Friedmann-Lemaitre-Robertson-Walker metric

The general metric to describe the space-time of the universe is given by

$$\mathrm{d}s^2 = g_{\mu\nu}(x)\mathrm{d}x^\mu \mathrm{d}x^\nu. \tag{2.5}$$

where μ and ν are indices from 0 to 3, (x^0, x^1, x^2, x^3), where 0 represents time and 1, 2 and 3 represent the spatial coordinates, and $g_{\mu\nu}$ is the metric tensor. The metric is a fundamental quantity to describe the geometric properties of the universe at a given time. In a background universe assuming homogeneity and isotropy, FLRW space-time metric is given by

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\ \left[\frac{\mathrm{d}r^2}{1-\mathrm{K}r^2} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\ \mathrm{d}\phi^2\right)\right], \tag{2.6}$$

where $a(t)$ is the scale factor. The scale factor encapsulates the expansion of the space with respect to time, and it relates to redshift as $1/a = 1+z$ [26]. The terms in the square brackets represent the spatial metric in three dimensions spherical polar coordinates, (r, θ, ϕ). The value of K determines the geometry of the space (or the curvature). In this work, the normalization of the scale factor is to set $a = 1$ at the present time, and the speed of light to $c = 1$, therefore the time and the comoving distance have the same units [9, 26].

Notice that now we described the FLRW homogeneous expanding universe using only the metric, i.e the scale factor, and the geometry. How the scale factor changes with time depends on the energy density of the universe. The Einstein theory of General Relativity is then required. According to General Relativity, the equations that relate the geometry to the energy density are Einstein equations,

given by

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} \mathcal{R} = 8\pi G T_{\mu\nu}. \tag{2.7}$$

where $G_{\mu\nu} \equiv$ Einstein tensor, $\mathcal{R} \equiv g^{\mu\nu} R_{\mu\nu}$, known as the Ricci scalar, $R_{\mu\nu} \equiv$ Ricci tensor. $T_{\mu\nu}$, the energy-momentum tensor, describes the density and the flux of energy and momentum in space-time. Essentially it is a generalization of the stress tensor of Newtonian physics. G is the gravitational constant [9, 26].

The Friedmann equation

The Hubble rate is $H = \dot{a}/a$ which describes the expansion rate of the universe. The $R_{i0} = R_{0i}$ terms in (2.7) vanish due to the assumption of isotropy in Robertson–Walker metric. The 0-0 terms of the Einstein equations in (2.7) given the isotropic FLRW space-time metric give directly

$$H^2(t) = \frac{8\pi G}{3}\rho - \frac{K}{a^2}. \tag{2.8}$$

This is the Friedmann equation: the energy densities of all the various cosmic components are encapsulated in ρ; these components include matter, radiation and dark energy. The Hubble rate at the present time is denoted by $H_0 \equiv H(t = 0)$ which is called the Hubble constant. Hubble constant can be written as $H_0 = 100\ h\ \mathrm{km} s^{-1}\mathrm{Mpc}^{-1}$, where $h = H_0/100$ is a dimensionless number.

The acceleration equation

The i-j term of the Einstein equation, (2.7), gives

$$\frac{\ddot{a}}{a} + \frac{4\pi G}{3}(3P + \rho) = 0, \tag{2.9}$$

which is called the acceleration equation. ρ and P are the total density and pressure, respectively.

The fluid equation

While the Friedmann equation describes the scale factor evolution with time, the fluid equation describes the evolution of the density with time. If we consider the universe to be composed by a homogeneous and isotropic fluid, then from (2.8) and (2.9), the fluid equation of an expanding universe can be written as

$$\dot{\rho} + 3\frac{\dot{a}}{a}[\rho + P] = 0 \tag{2.10}$$

In cosmology it is common to define $w \equiv P/\rho$, this relation known as the equation of state (EoS) where w is assumed to be constant. Therefore, by knowing the pressure of the components the densities can be evaluated by solving (2.10), hence

$$\rho_i \propto a^{-3-3w_i}. \tag{2.11}$$

ρ_i represents all the densities, matter, radiation and dark energy. Where:

- Matter, $P_{\rm m} = 0,\ w_m = 0\ \Rightarrow \rho_{\rm m} \propto a^{-3}$
- Radiation , $P_r = \rho_r/3,\ w_r = 1/3\ \Rightarrow \rho_r \propto a^{-4}$
- Vacuum energy, $P_\Lambda = -\rho_\Lambda$, as a negative pressure driving the expansion $\rho_\Lambda \propto a^0$. Where $w_\Lambda = -1$ for dark energy only if dark energy assumed to be the vacuum energy (or Λ).

The standard model also known as the concordance model, is based on solving Einstein equation, (2.7), as above. The total density at the present time in a flat universe, is known as the critical density and is defined as $\rho_{cr,0} \equiv 3H_0^2/8\pi G$, while $\rho > \rho_{cr,0}$ for closed universe and $\rho < \rho_{cr,0}$ for an open universe [9, 19, 26].

Density parameters

Friedmann equation for all the species could be compressed in one equation,

$$H^2(z) = H_0^2 \sum_i \Omega_i\ (1+z)^{3(1+w_i)}, \tag{2.12}$$

where w_i is constant, the sum is over i which represent the components: matter, radiation, curvature and dark energy (m, r, k, de). In ΛCDM model, those components are known as the *density parameters* with fractional distributions:

$$\begin{aligned} \Omega_m &= \frac{8\pi G\rho_m}{3H(t)^2}, \\ \Omega_r &= \frac{8\pi G\rho_r}{3H(t)^2}, \\ \Omega_\Lambda &= \frac{\Lambda}{3H(t)^2}, \\ \Omega_K &= \frac{-K}{a(t)^2H(t)^2}. \end{aligned} \tag{2.13}$$

A key condition in this model is that $\Omega_K = 1 - \Omega_r - \Omega_{\rm m} - \Omega_\Lambda$ for a flat universe.

Matter: The matter content consists of two types. The baryonic matter (Ω_b) which accounts for all the visible matter; and cold dark matter ($\Omega_{\rm cdm}$) or non-baryonic matter which accounts for all the non-detectable matter that we only know about by its gravitational effects, such as the rotational curves of galaxies and the gravitational lensing of light by galaxy clusters. Thus we have $\Omega_m = \Omega_b + \Omega_{\rm cdm}$. In ΛCDM model, cold dark matter is considered collision-less, it does interact with other components only through gravity. The recent observations from Planck mission measures $\Omega_m \approx 31\%$ of the total (mass) energy density of the universe [27].

Radiation: The radiation consist of photons and neutrinos. Neutrinos contribution to the total density is rather theoretical. Thus, the present density parameter of radiation, which is the sum of photons and neutrinos, Ω_r, counts only 8.051×10^{-5} of the total mass energy density, therefore it is commonly ignored [28].
Curvature: The universe in ΛCDM model can be flat with $K = 0$, open with $K < 0$ or closed with $K > 0$.

Dark energy and EoS parameterization: In standard cosmology, the dark energy field does not exchange energy with other components of the universe [29]. The dark energy equation of state, $w(z) = P_{de}/\rho_{de}$[1] can express the evolution of dark energy. By integrating (2.10) for dark energy with a general equation of state, we get

$$\frac{\rho_{de}(z)}{\rho_{de}(z=0)} = \exp\left[3\int_0^z [1+w(z')]\frac{dz'}{1+z}\right]. \tag{2.14}$$

Thus given (2.11), (2.12) and (2.14), we have

$$H(z)^2 = H_0^2\left[\Omega_m(1+z)^3 + \Omega_K(1+z)^2 + \Omega_{de}\exp(3\int_0^z \frac{1+w(z')}{1+z'}dz')\right], \tag{2.15}$$

where we used the fact that $a = (1+z)^{-1}$. Equation (2.15) gives the Hubble parameter as a function of redshift, in terms of the (energy) density parameters of the various cosmic components.

Note that for a constant w, this equation reduces to the form that has been introduced in (2.11), i.e. $\rho_{de} \propto (1+z)^{3(1+w)}$. In ΛCDM model, the dark energy

[1] In the rest of this thesis, $w(z)$ will refer to the dark energy equation of state.

component, Ω_Λ, is estimated to be between $65 - 72\%$ of the total matter-energy density of the universe. Recent Planck measurements estimate $\Omega_\Lambda \approx 69\%$ of the total mass-energy density of the universe [27].

2.2 Distances

In the nearby (or local) universe, i.e. for $z < 0.1$, the effects of the space time curvature and the cosmic expansion are negligible. At low redshift the distance to objects are measured through kinematic methods; the most known and used methods are: the apparent luminosity, the trigonometric parallax and the proper motion methods, respectively. While at much higher redshift, $z > 0.1$, bright sources are used to measure distance such as galaxies, SNIa and quasars. Such sources enable us to measure distances at higher redshift using methods that are not affected by the curvature and cosmic expansion, such as Tully-Fisher method, Faber-Jackson relation and also Type Ia supernovae method, for further details about these methods see [26, 30].

In general, distances in the universe are expressed in term of the scale factor $a(t)$ (or redshift) to account for the cosmic expansion. A distance between an object and an observer along the line-of-sight is given by

$$\chi(z) \equiv \int_{t(a)}^{t_0} \frac{dt'}{a(t)} = \int_0^z \frac{dz'}{H(z')}, \tag{2.16}$$

which is called the co-moving distance, assuming a flat universe. To account for universes with different curvature, a general form for the co-moving distance will be

$$D_c = \frac{1}{H_0\sqrt{|\Omega_K|}} C_K\left(H_0\sqrt{|\Omega_K|}\chi(z)\right), \tag{2.17}$$

where D_c is known as the transverse co-moving distance; where $C_K(x) = x$ for a flat universe, $\sinh(x)$ for an open universe and $\sin(x)$ for a closed universe [30].

2.2.1 Luminosity distance and standard candles

Consider an object with a distance $D_L(z)$ from the observer with bolometric luminosity (L) of the object. Then, the bolometric flux (F) is given by[2]

$$\mathrm{F} = \frac{\mathrm{L}}{4\pi D_L^2(z)}. \tag{2.18}$$

The quantity D_L is the known as the luminosity distance, which is related to the transverse co-moving distance by

$$D_L(z) = (1+z)D_c(z). \tag{2.19}$$

Equation 2.19 shows that the flux in (2.18) will be reduced by a factor $(1+z)^{-2}$ is due to cosmic expansion. That indicates the photons at the observer lose energy relative to the source of emission. In fact, one factor of $(1+z)$ due to the photon energy loss. The second factor comes from the changed rate of arrival of photons.

To use the absolute luminosity method, the object has to be very luminous and be standardize-able to increase the accuracy of the measurement. In general, such luminous objects are known as the *standard candles*. A widely used probe, which we will consider in this work is SNIa. The standard candles magnitude relates to the distance as

$$\mu(z) + \log H_0 - 25 = 5\log\left(\frac{D_L(z)}{10\ \mathrm{pc}}\right), \tag{2.20}$$

distance modulus, $\mu \equiv m - M$, which is the difference between the apparent magnitude m and the absolute magnitude M of the object at a distance 10pc.

The concept of distance indicators come from the fact that some sources have known luminosity. By knowing the luminosity of a source, then we can estimate its distance using (2.18). Almost all distance measurements in astronomy are based on measurements of the distance of an objects within our own galaxy; these objects are known as primary distance indicators. One of the most famous primary distance indicators is RR Lyrae stars. RR lyrae are variable stars that have been used as distance indicators for long time, with pulsation period between 0.2-0.8 days. In 1912 Cepheids variables took over from the other primary distance indicators because they are so bright. They are the most important stars used to

[2]This formula holds if flux and luminosity are integrated over all frequencies.

measure distances outside our galaxy, with pulsation period between 2-45 days. The primary distance indicators are dim and can not be used to measure distances outside the local group and some nearby groups (more than one Mpc) [9].

Consequently, the second distance indicators, such as galaxies and SNIa can be calibrated using primary distance indicators such as Cepheids. Second distance indicators are brighter than Cepheids and they are used to measure large distances outside the local universe (at high redshift). The SNIa is the only supernovae type that has little variation in the absolute luminosity, which makes them ideal distance indicators and usually they are known as the standard candles in cosmology.

2.2.2 Angular diameter distance and standard rulers

Another method to measure cosmic distances is to use objects with known physical size l. The angular diameter distance of an object is given by

$$D_A = \frac{l}{\theta}, \tag{2.21}$$

where θ is the angle subtended by the object, and $l \ll D_A$. In an expanding universe the comoving size is given by l/a, therefore

$$D_A = aD_c(a) = \frac{D_c(z)}{1+z}. \tag{2.22}$$

Hence, the angular diameter distance will become

$$D_A(z) = \frac{c}{H_0}\frac{(1+z)^{-1}}{\sqrt{-\Omega_K}}\sin\left(\sqrt{-\Omega_K}\int_0^z \frac{dz'}{E(z')}\right). \tag{2.23}$$

The BAO in the matter power spectrum, is developing into an important probe of dark energy, since the scale of the BAO peak in the power spectrum acts as a *standard ruler*, expanding along with the universe from the time of the CMB to the present day. From the BAO scale, D_A is measured using the plane across the line-of-sight. And the measurements of the BAO scale along the line-of-sight determine the Hubble rate, $H(z)$ (see Fig. 2.1).

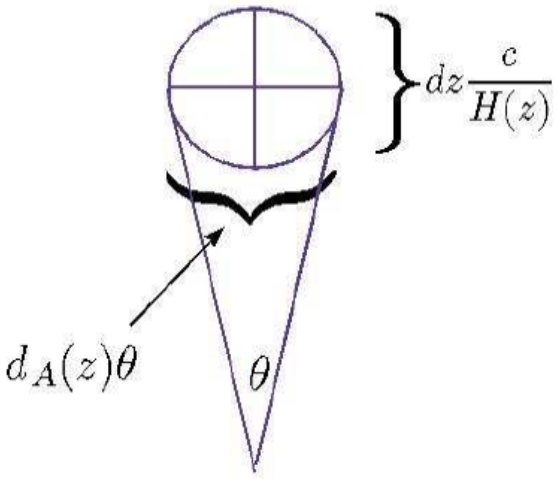

FIGURE 2.1: The sketch shows the BAO scale, where along the line-of-sight we can measure $H(z)$, and across the line-of-sight we can measure D_A. Given dz and θ are the redshift interval and the subtended angle, respectively [1].

Moreover, forthcoming surveys are being designed specially to measure the BAO, such as BINGO [31], BOSS and eBOSS [32]. BAO measurements are also complementary to the SNIa data, since it can be measured at high redshift, $z \gtrsim 1.2$, whereas SNIa current data are best measured at $z \lesssim 1.2$.

The origin of the BAO peak is further discussed in section 3.4.

Chapter 3

Large Scale Structure

Our galaxy the Milky Way, and M31 galaxy are members of the local group which contain over 40 galaxies. Those galaxies interact gravitationally with each other. For instance, our galaxy is falling towards Andromeda which is the nearest galaxy. Also, the local group interacts with other galaxy groups. Galaxy groups are also bound by gravitational force. When the mass of the group is greater than $10^{14} M_{\odot}$, then the group of galaxies is known as a galaxy cluster. Virgo is the nearest cluster to our local group, both are part of a larger structure known as the local super-cluster. Zooming out further, strong clustering of clusters and super-clusters form over-dense regions known as filaments, see Fig. 3.1, leaving under-dense regions or voids in the space.

There are two commonly known effects on large scale structure that are important, Fingers of God (FoG) effect and Kaiser effect. The galaxies in the cluster have the redshift of the cluster in addition to its velocity dispersion. FoG effect make the large scale structure seem to point at the observer (see Fig. 3.1). The second effect is the Kaiser effect which makes the galaxies near to the cluster, on the observer's side, looks further away due to the galaxies falling into the cluster (due to the peculiar velocity), and the opposite happens (they look closer to us) if the galaxies are behind the cluster.

The two effects, FoG and Kaiser effect, combined are known by the redshift space distortions (RSD). In fact, FoG is RSD on small scales whereas Kaiser is RSD on linear scales.

3.1 The growth of structure

The universe looks smooth on large scales and clumpy on small scales. The distribution of density fluctuations are given by

$$\delta = \frac{\rho - \bar{\rho}}{\bar{\rho}}. \tag{3.1}$$

where $\bar{\rho}$ is the expected mean density. The evolution of structure is understood using the linear perturbation theory. Small perturbations grow to form the inhomogeneous universe with stars, galaxies and clusters of galaxies, which are known as the non-linear regime, where $\delta \gg 1$. Therefore more complicated techniques are used to describe the structure on small scales such as N-body simulations and semi-analytical models.

For an expanding universe the growth of fluctuations can be calculated using the perturbed Einstein and energy-momentum conservation equations. Since cold dark matter has no pressure, the growth of linear matter perturbations are described by [33]

$$\ddot{\delta} + 2H\dot{\delta} = 4\pi G\bar{\rho}\delta, \tag{3.2}$$

A general solution for this equation can be written as

$$\delta(x) = f_1(x)D(t) + f_2(x)E(t), \tag{3.3}$$

the first term on the right is the decaying mode, while the left term is the growing mode. The decaying mode corresponds to the small-scale mode and can be neglected, while the growing mode coressponds to the large-scale mode. Since we are interested in the large-scale mode we adopt the growing mode solution, where the growth equation is defined by (3.2) it can be written as [34]

$$\frac{d^2D}{d\ln a} + \left(4 + \frac{d\ln H}{d\ln a}\right)\frac{dD}{d\ln a} + \left[3 + \frac{d\ln H}{d\ln a} - \frac{3}{2}\Omega_m(z)\right]D = 0, \tag{3.4}$$

where D is known as the growth factor and $\Omega_m(z) = \Omega_m H_0^2(1+z)^3/H^2(z)$. This equation holds true in General Relativity theory where the scales are much smaller than the horizon. An approximation for (3.4) is given by

$$f(z) \equiv \frac{d\ln D}{d\ln a} \approx [\Omega_m(z)]^\gamma, \tag{3.5}$$

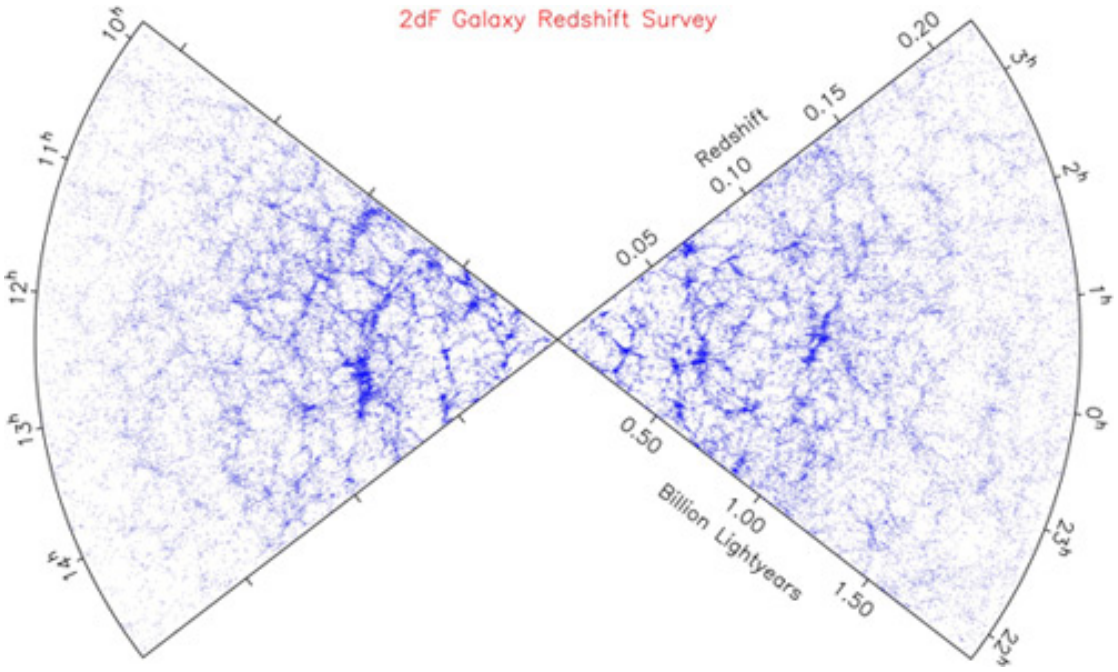

FIGURE 3.1: The 2dF galaxy redshift survey map revealing the structure of the universe with respect to redshift [2].

where f is the growth rate and the growth index $\gamma \approx 0.55 - 0.6$, and it is found to be 0.55 for ΛCDM. That indicates γ can be affected mildly by the cosmological parameters [35]. Using (3.5), we have

$$\frac{D(z)}{D(z_0)} \approx \exp\left[-\int_0^z \frac{dz'}{1+z'}[\Omega_m(z)]^\gamma\right]. \tag{3.6}$$

Theoretically, inflation models predict the primordial matter fluctuation to be Gaussian. Therefore, these primordial matter fluctuations can be fully described by the power spectrum ($P(k)$) or its Fourier transform, the correlation function ($\xi(r)$), we now discuss these two approaches briefly.

3.2 The power spectrum

Equation 3.1 determines the density fluctuations as a function of position. One way to characterize the density fluctuations is to estimate the variance in the Fourier transform coefficients as a function of the length of the wave number k thus,

$$< \delta_{\mathbf{k}} \delta_{\mathbf{k}'} >= (2\pi)^3 \delta_D(\mathbf{k}+\mathbf{k}')P(k). \tag{3.7}$$

This is known as the power spectrum where $\delta_{\mathbf{k}}$ is the Fourier transform of δ and δ_D is the Dirac delta function. $P(k)$ is only a function of the wave number $k = |\mathbf{k}|$ due to the assumption of isotropic universe. The variance is defined as

$$\sigma^2 = \frac{1}{2\pi^2}\int P(k)k^2 dk = \int \frac{\Delta^2(k)}{k} dk, \tag{3.8}$$

where $\Delta^2(k)$ is the variance per unit logarithmic interval. To get a more practical expression of the power spectrum, consider a cubic box of the universe with a length L, hence the volume of the universe, V, is proportional to L^3. The wavelengths will be $\lambda = L/n$ where n is the number of modes. Therefore the number of modes from k to $k + dk$ will be given by the density of modes times the volume of the shell,

$$\left(\frac{L}{2\pi}\right)^3 \times \left(4\pi k^2 dk\right). \tag{3.9}$$

It is clear that the power spectrum has units of volume. Although, to avoid that the value of the power spectrum depends on the choice of k, commonly the power spectrum is expressed in dimensionless units such as

$$\Delta^2(k) = \frac{V}{2\pi^2}k^3\langle|\delta_k^2|\rangle = \frac{V}{2\pi^2}k^3 P(k), \tag{3.10}$$

Also, the dimensionless power spectrum of the primordial curvature perturbation from Inflation is given by

$$\Delta^2_\Phi \propto k^{n_s - 1}, \tag{3.11}$$

In terms of $\Delta^2(k)$ we have [36],

$$\Delta^2(k) \propto k^{3+n_s}, \tag{3.12}$$

where n_s is known as the spectral index of scalar perturbations with $n_s = 0.96$ [37].

The power spectrum that we defined in (3.7) depends only on k. In redshift space we also care about the direction of the wave number, $\vec{k}$, therefore we include the effects that we introduced earlier, the redshift space distortion (RSD). Taking the RSD effect in cognizance, the power spectrum will be given by

$$P_s(\vec{k}) = \left[1 + \beta\mu^2\right]^2 P(k). \tag{3.13}$$

where μ is the cosine of the angle between the line of sight and $\vec{k}$. The second term in 3.13 accounts for the RSD presence; In other words, this term is introduced to compensate for the fact that apparent over-density in redshift space is larger than in real space. Due to the undetectable dark matter, the mass over-density δ is not equal to the galaxies over-density δ_g. The linear bias (b) can be defined as

$$b \equiv \frac{\delta_g}{\delta}. \tag{3.14}$$

The linear bias impose that galaxies trace the distribution of dark matter. Consequently, the power spectrum of dark matter can be determined from the galaxy power spectrum. Then the β parameter can be defined as

$$\beta = \frac{f}{b} \simeq \frac{\Omega_m^{0.55}}{b}, \tag{3.15}$$

which is a measurable quantity. This is done by measuring the ratio of the quadrupole and monopole of the power spectrum [26].

In linear perturbation theory, the linear power spectrum is proportional to the growth factor $D^2(z)$, thus

$$P_{\rm lin}(k) = \frac{D^2(0)}{D^2(z)} P_{\rm lin}(k, z). \tag{3.16}$$

3.3 The correlation function

To measure the clustering of the galaxies on large scale, the correlation function is given by

$$\xi(r) = \frac{V}{2\pi^2} \int_0^\infty P(k) \frac{\sin(kr)}{kr} k^2 dk, \tag{3.17}$$

where $\xi(r)$ is the probability of having two galaxies separated by a distance r. The correlation function is related to the probability of finding a galaxy in the volume

dV_1 and another in dV_2 separated by a vector r. Hence

$$d^2Pr(r) = [1 + \xi(r)]\, n^2 dV_1 dV_2, \tag{3.18}$$

where n represent the number of galaxies per unit volume. The correlation function is usually measured by comparing the observational results to a random sample, this method has the advantage of avoiding the computation of Fourier series.

3.4 Large scale structure surveys

We can not observe the large scale structure of the matter distribution directly, since the adopted assumption in CDM model that most of the matter in the universe is dark matter. Observations of the spatial distribution of visible galaxies can be provided, with the assumption that the matter density and the galaxies are related as in (6.7) hence, we can estimate the power spectrum of the density fluctuations.

The study of galaxy distributions provide information about the cosmic density field. There are two common ways that have been developed to study the galaxy distribution; photometric sky surveys and spectroscopic sky surveys.

The photometric sky surveys provide approximate redshifts, while the spectroscopic sky surveys provide accurate redshift information. Spectroscopic surveys provide knowledge about the statistical properties of galaxies. Therefore a growing interest has been developed in redshift surveys.

Redshift surveys

Redshift surveys are very time consuming since a spectrum has to be recorded for each object. The rapid development of technology makes this task easier than before but also current and future surveys are getting more challenging. To design a redshift survey, firstly an area or a region on the sky has to be selected geometrically. Secondly the minimum brightness of the objects that the survey can capture has to be defined, the brightness threshold. The brightness threshold defines the number density of the galaxies in the survey as well as the exposure time. The determination of the power spectrum depends on the relation between the number

density and the galaxies over-density in a given survey volume V, which is given by

$$\delta_g := \left(\frac{\Delta n}{\bar{n}}\right)_V = b\left(\frac{\Delta \rho}{\rho}\right)_V, \tag{3.19}$$

where $\bar{n}$ is the average number density of the galaxies, and Δn is the difference between the local number density and the average number density.

While the power spectrum is theoretically predicted, the normalization of the power spectrum can only be determined from observations [38]. The normalization is measured by determining the fluctuation amplitude, σ_8, which is done by averaging over the selected galaxies on multiple spheres of radius $R = 8\ h^{-1}$ Mpc in the local universe, therefore

$$\sigma_8^2 = \left\langle \left(\frac{\Delta n}{\bar{n}}\right)^2 \right\rangle_8 \approx 1. \tag{3.20}$$

Given that b's value depends on how galaxies have formed in the mass density field, from (3.1), (3.19) and (3.20), it is straightforward to relate the matter density fluctuations to σ_8. Then

$$\sigma_8 = \frac{\sigma_{8,g}}{b} \approx \frac{1}{b}. \tag{3.21}$$

The recent update from Planck mission estimate $\sigma_8 = 0.83$ [27].

Spectrum errors

The acoustic oscillations in the power spectrum are primarily at wavenumbers 0.1–0.2hMpc^{-1}, hence surveys are usually designed with $nP\left(k = 0.2h\text{Mpc}^{-1}\right) \geq 1$. Therefore, the error on the power spectrum is given by

$$\frac{\delta P}{P} = \frac{1}{\sqrt{m}}\left(1 + \frac{1}{nP_{0.2}}\right), \tag{3.22}$$

where m is the number of Fourier modes, and $P_{0.2}$ is the power spectrum at $k = 0.2\ h\text{Mpc}^{-1}$ [1]. The first term and the second term represent two types of errors known as cosmic variance and shot noise respectively. Shot noise reflects the limits on reconstructing the matter distribution from galaxy surveys, and is inversely proportional to the number of galaxies at a given survey volume. Upcoming surveys such as the SKA will limit shot noise on galaxy surveys. Cosmic variance which is the uncertainty that results from observing only part of the

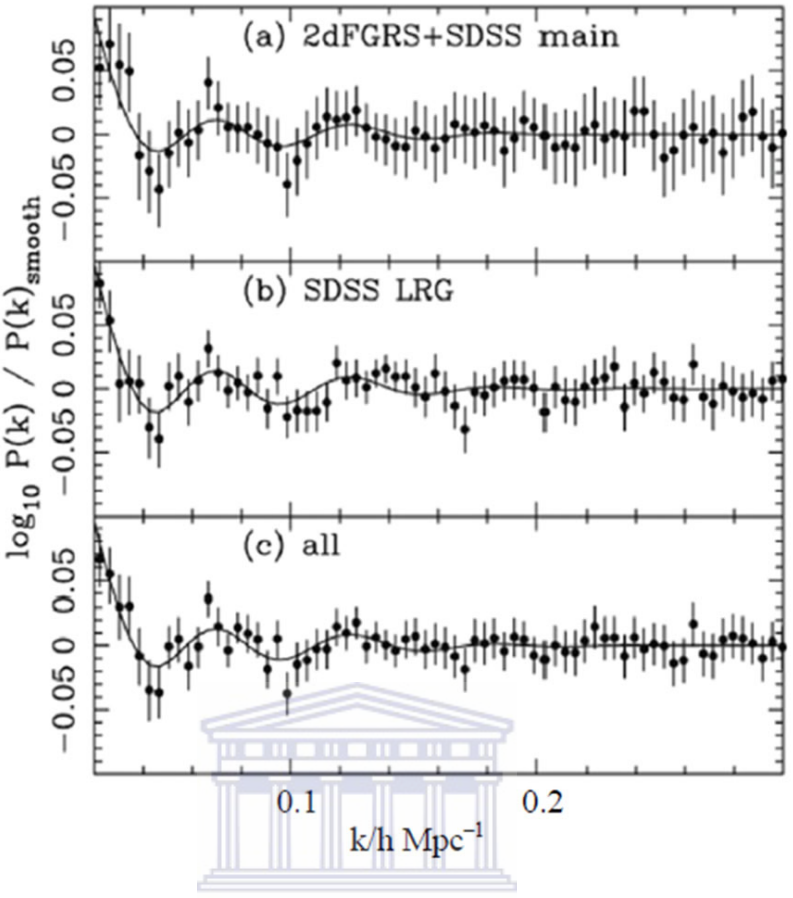

FIGURE 3.2: The predicted baryon wiggles in the galaxy power spectrum (line) and the measurements from the galaxy surveys (dots) of 2dFGRS +SDSS (top), SDSS DR5 LRGS sample (middle) and all of the data combined (bottom) [3].

universe at a specific time limits our statistical knowledge on a cosmological scale. Shot noise and cosmic variance are equal when $nP_{0.2} = 1$.

CMB constraints

CMB measurements provide tight constraints on Ω_m and the normalization of the matter fluctuation, since the CMB angular power spectrum depend on the matter densities. Also they provide constraints on Ω_K from the locations of the peaks. As a consequence of using the CMB constraints we consider additional nuisance parameters such as spectral index (n_s), the electron-scattering optical depth τ, Hubble constant (h), $dn_s/d\ln k$ of the scalar fluctuation spectrum, as well as the amplitude and the slope of the tensor fluctuations spectrum. In Chapter 6, we will see how the constraints on the parameters change if one assumes a flat universe.

Another important value that we can determine from the CMB measurements is the physical scale of the acoustic peaks in the CMB which is equivalent to the scale of the BAO in large scale structure. This is known as the sound horizon (r_s). The sound horizon at drag epoch, $r_s(z_d)$, at which baryons were released from Compton drag of the photons can be used to locate the position of the BAO peak. $r_s(z_d)$ is the distance that sound waves can travel from the big bang moment $t = 0$, to drag epoch t_d. This is given as

$$r_s(z_d, \Omega_b, \Omega_r) = \int_0^{t_d} c_s dt \tag{3.23}$$

where z_d is the redshift at the drag epoch, Ω_r is the radiation density parameter and c_s is the speed of sound given by

$$c_s = \left[3\left(1 + \frac{3\Omega_b}{4\Omega_r}\right)\right]^{-\frac{1}{2}}. \tag{3.24}$$

Baryon wiggles

In the period before recombination the universe consisted of hot plasma and baryons which were tightly coupled. A density perturbation sphere in a tightly coupled baryon-photon plasma will propagate outwards as an acoustic wave with a speed c_s. At recombination the baryons and the photons were decoupled. The Photon wave freely propagated away to form the CMB, while the baryon wave stalls. The radius of the stalled baryon wave is imprinted on the distribution of the baryons; The gravitational interaction between the baryons and dark matter causes clustering of dark matter halos and thus galaxies at this radius, the sound horizon. Therefore the probability of finding a galaxy in the high density region of the stalled baryon wave increases. This probability is evident as a bump in the two point correlation function at radius s, it appears due to the formation of galaxies at the center of the density perturbation sphere.

The acoustic peaks are the imprint of the baryonic matter density fluctuations in the galaxy power spectrum. These fluctuations are known as the BAOs. The BAO features have been detected from the two-degree Field Galaxy Redshift Survey (2dFGRS) and SDSS Data Release 5 galaxies as shown in Fig. 3.2 [3]. The detection of the BAO peak position across the redshift plane is one of the most promising approaches to study the expansion rate of the universe and reconstruct

the dark energy equation of state. The accurate detection of this peak is challenging. Nonetheless, the SDSS and the 2dF galaxy redshift surveys were successful in detecting the BAO peak. In the near future, larger telescopes will be available with the aim to measure the BAO accurately (among other scientific goals), over wider redshift range and to detect the radial and transverse BAO separately thus, determine the nature of dark energy. In Chapter 6 we will forecast how well future redshift surveys with a large telescope such as the SKA will detect the BAO peak.

Chapter 4

Null tests of the cosmological constant using supernovae

SNIa are the best distance indicators to probe the expansion history of the Universe. These 'standardizable candles' can be observed to high redshift, and have produced convincing evidence that the Universe has undergone a recent phase of accelerated expansion. Current samples of SNIa (e.g. [39–43]) comprise several hundred SNIa with $z < 1.8$. Forthcoming surveys of SNIa, such as DES [44], will produce well-measured light-curves for over 4000 SNIa, improving the cosmological constraints by an order of magnitude.

In this chapter we use luminosity distances $D_L(z)$ determined from SNIa observations to test the consistency of the standard model, through a set of null tests. Reconstructing the expansion history of the Universe in a model-independent fashion is essential for these tests. To do this, we use the GP, which have previously been used to reconstruct $w(z)$ from SNIa luminosity distances [6, 45–47].

We use GaPP (Gaussian Processes in Python)[1], a package developed by Seikel and introduced in [45].

4.1 Null tests of ΛCDM – theory

In Chapter 2, we introduced the equation of state parameter of dark energy, $w = p_{\rm de}/\rho_{\rm de}$, and (2.19) which can be used to express the dimensionless comoving

[1] http://www.acgc.uct.ac.za/~seikel/GAPP/index.html

luminosity distance,

$$D(z) \equiv H_0(1+z)^{-1} D_L(z), \tag{4.1}$$

It is typical to parametrize $w(z)$ in order to differentiate between various dark energy models, or to parametrize background and perturbation variables to test classes of modified gravity models, see section 2.1. A complementary approach is to test the consistency of the standard model itself, independent of the values of Ω_m and Ω_K. A range of null tests designed specifically to probe various aspects of the concordance model have been introduced (see e.g. [48–54] and [55] for a review and Appendix B for detailed calculations), such as

$$\begin{aligned} w(z) &= \Big\{2(1+z)(1+\Omega_K D^2)D'' - [(1+z)^2\Omega_K D'^2 + 2(1+z)\Omega_K D D' \\ &- 3(1+\Omega_K D^2)]\Big\} / \Big\{3\{(1+z)^2[\Omega_K + (1+z)\Omega_m]D'^2 - (1+\Omega_K D^2)\}D'\Big\}. \end{aligned} \tag{4.2}$$

Given an observed distance-redshift relationship $D(z)$, it is possible to reconstruct the equation of state of dark energy and test the ΛCDM model [45]. However, a disadvantage of this method is that it depends on the values of the density parameters, Ω_m and Ω_K, which must be measured independently [45].

To avoid this problem and test ΛCDM using SNIa data, we use the consistency tests introduced in [51] (see also [50, 52]). Following this approach, we test the null hypothesis that the expansion of the universe can be described by a flat or a curved ΛCDM model.

The assumptions underlying the consistency tests and the null hypothesis are: (1) the universe is homogeneous and isotropic on large scales; (2) gravity is described by General Relativity; (3) the universe contains cold matter (with $w = 0$) and dark energy with $w = -1$. Photons and neutrinos can be included (Ω_γ, Ω_ν are known independently, from CMB data), but it is reasonable to neglect radiation at the low redshifts probed by SNIa data. Detection of a deviation from the consistency tests would imply a violation of at least one of these assumptions: (1) large-scale nonlinear inhomogeneity or anisotropy; (2) modified gravity; (3) dynamical dark energy ($w \neq -1$), or alternatively, a cosmological constant plus an unknown additional species with equation of state which deviates from that of cold matter, curvature or vacuum energy. Any of these possibilities imply that the standard ΛCDM is ruled out. Note that the tests cannot identify which of these possibilities applies.

For a flat ΛCDM model, i.e. $w = -1$ and $\Omega_K = 0$, by manipulating (2.19), we find that

$$\Omega_m \left[(1+z)^3 - 1\right] D'^2 = 1 - D'^2. \tag{4.3}$$

If we define

$$\mathcal{O}_m^{(1)}(z) = \frac{1 - D'(z)^2}{[(1+z)^3 - 1]D'(z)^2}, \tag{4.4}$$

then

$$\text{flat } \Lambda\text{CDM implies } \mathcal{O}_m^{(1)}(z) = \Omega_m. \tag{4.5}$$

Thus we obtain a null test of the concordance model:

$$\mathcal{O}_m^{(1)}(z) \neq \Omega_m \text{ falsifies flat } \Lambda\text{CDM}. \tag{4.6}$$

Any variation of $\mathcal{O}_m^{(1)}(z)$ with redshift reflects an inconsistency between the flat ΛCDM model and observations. To detect evolution of $\mathcal{O}_m^{(1)}$ with redshift we can differentiate $\mathcal{O}_m^{(1)}(z)$, from which we define the additional diagnostic:

$$\mathcal{L}^{(1)}(z) = (1+z)^{-6} \left\{2\left[(1+z)^3 - 1\right] D''(z) + 3(1+z)^2 D'(z)\left[1 - D'(z)^2\right]\right\}, \tag{4.7}$$

which vanishes if and only if $d\mathcal{O}_m^{(1)}/dz = 0$. The factor $(1+z)^{-6}$ added to ensures stability of the errors (see below). If $\mathcal{L}^{(1)}$ is nonzero at any redshift, then observations are incompatible with ΛCDM:

$$\mathcal{L}^{(1)} \neq 0 \text{ falsifies flat } \Lambda\text{CDM}. \tag{4.8}$$

We can extend this approach to include spatial curvature, and derive null tests for general (curved) ΛCDM. Using (2.12), (2.19) and (4.2) with $w(z) = -1$, and solving for Ω_m and Ω_K, we find [53, 55]:

$$\begin{aligned} \Omega_m &= 2\Upsilon(z)\left\{\left[(1+z)^2 - D^2 - 1\right] D'' - \left(D'^2 - 1\right)\left[(1+z)D' - D\right]\right\}, \\ \Omega_m &\equiv \mathcal{O}_m^{(2)}(z), \end{aligned} \tag{4.9}$$

and

$$\begin{aligned} \Omega_K &= \Upsilon(z)\left\{2\left[1 - (1+z)^3\right] D'' + 3D'\left(D'^2 - 1\right)(1+z)^2\right\}, \\ \Omega_K &\equiv \mathcal{O}_K^{(2)}(z). \end{aligned} \tag{4.10}$$

Here $\Upsilon(z)$ is defined by

$$\begin{aligned}\Upsilon^{-1} &= -2\left[1-(1+z)^3\right]D^2D'' - \Big\{(1+z)\Big[(1+z)^3 \\ &\quad - 3(1+z)+2\Big]D'^2 - 2\left[1-(1+z)^3\right]DD' - 3(1+z)^2D^2\Big\}D'. \end{aligned} \tag{4.11}$$

Then we have

$$\mathcal{O}_m^{(2)}(z) \neq \Omega_m \text{ falsifies curved } \Lambda\text{CDM}, \tag{4.12}$$

$$\mathcal{O}_K^{(2)}(z) \neq \Omega_K \text{ falsifies curved } \Lambda\text{CDM}. \tag{4.13}$$

These are not independent tests: the derivative of $\mathcal{O}_K^{(2)}$ vanishes if and only if the derivative of $\mathcal{O}_m^{(2)}$ vanishes. Hence we need only a single diagnostic for vanishing derivative. We use the derivative of $\mathcal{O}_m^{(2)}$ to define

$$\begin{aligned}\mathcal{L}^{(2)} &= (1+z)^{-6}D'^2\Big\{D\Big[-3\,(1+z)\left(D'^2-1\right)(2D'+3(1+z)D'') \\ &\quad + 2zD'''\Big(3+z(3+z)\Big)\Big] + 9(1+z)^2D^2D''^2 + 3(1+z)D^2D' \\ &\quad \times\Big(2D''-(1+z)D'''\Big) + 6(1+z)^2D'^2\left(D'^2-1\right) \\ &\quad - \Big[3z^2(3+z)D''^2 + zD'\Big(z(3+z)D''' - 6(2+z)D''\Big)\Big](1+z)\Big\}, \end{aligned} \tag{4.14}$$

which vanishes if and only if $d\mathcal{O}_m^{(2)}/dz = 0$. (Again we use the pre-factor to stabilize the errors.) Then we have the null test for curved ΛCDM:

$$\mathcal{L}^{(2)}(z) \neq 0 \text{ falsifies curved } \Lambda\text{CDM}. \tag{4.15}$$

In principle, $\mathcal{L}^{(1)}$ and $\mathcal{L}^{(2)}$ provide no additional information compared to $\mathcal{O}_m^{(1)}$ and $\mathcal{O}_m^{(2)}$. However, it is easier to detect a deviation from zero than to confirm that a quantity is constant, especially since the exact value of this constant is not known a priori. The disadvantage of $\mathcal{L}^{(1)}$ and $\mathcal{L}^{(2)}$ is that they require higher derivatives than $\mathcal{O}_m^{(1)}$ and $\mathcal{O}_m^{(2)}$, which are more challenging to constrain.

Another problem with $\mathcal{L}^{(1)}$ and $\mathcal{L}^{(2)}$ is the degeneracy between w and Ω_m: a model with redshift dependent w can be formally consistent with ΛCDM within the error bars of the reconstruction if the value of Ω_m is adjusted accordingly. Such cases can only be identified with the $\mathcal{O}_m$ tests, but not with $\mathcal{L}$ (see section 4.3 for details).

Note that $\mathcal{L}^{(1)}$ and $\mathcal{L}^{(2)}$ are not identical to $d\mathcal{O}_m^{(1)}/dz$ and $d\mathcal{O}_m^{(2)}/dz$, respectively. Starting from these two derivatives, we have neglected the denominators, which add significant noise to the tests without adding extra information, and used a pre-factor $(1+z)^{-6}$ to obtain $\mathcal{L}^{(1)}$ and $\mathcal{L}^{(2)}$. We are free to do this without loss of generality, since we are testing the equality of these quantities with zero. As a consequence, the error bands of the reconstructions do not necessarily increase with redshift as one might expect, and the size of the errors of $\mathcal{L}^{(1)}$ and $\mathcal{L}^{(2)}$ are not directly comparable. In addition, the errors added from extra redshift factors are small when we have spectroscopic redshift measurements.

4.2 Null tests using SNIa data

To apply these null tests using current datasets, it is essential to choose a model-independent method to reconstruct $D(z)$ and its derivatives. For this purpose, we use GP (via the GaPP code [45]) to smooth the data and reconstruct the derivatives.

4.2.1 Gaussian Processes

GP provide a distribution over functions that are suitable to describe the data. At each point z_i, the distribution of function values $f(z_i)$ is a Gaussian. Thus the reconstruction consists of a mean function with Gaussian error bands. The function values at different points are correlated by a covariance function $k(z,\tilde{z})$, which depends on a set of hyperparameters (e.g. the characteristic length scale ℓ and the signal variance σ_f). This also provides a robust way to estimate derivatives of the function in a stable manner. See Appendix A for a detailed description of GP. In contrast to parametric methods, GP do not assume a specific form for the reconstructed function. Instead only typical changes of the function are considered. The hyperparameter ℓ corresponds roughly to the distance one needs to move in input space before the function value changes significantly, while σ_f describes typical changes in the function value.

There are various formulas of covariance functions. The choice of covariance function affects the reconstruction to some extent therefore a careful choice must be made for each particular problem. A general purpose covariance function is the

squared exponential covariance function $k(z,\tilde{z}) = \sigma_f^2 \exp\left[-(z-\tilde{z})^2/(2\ell^2)\right]$. The Matérn class has a peak width governed by the value of ν

$$k(z,\tilde{z}) = \sigma_f^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu(z-\tilde{z})^2}}{\ell}\right)^{\nu} \times K_\nu \left(\frac{\sqrt{2\nu(z-\tilde{z})^2}}{\ell}\right), \tag{4.16}$$

where K_ν is a modified Bessel function. The Matérn class, along with other covariance functions, has been intensely tested by [56]. Table II in [56] shows that Matérn ($\nu = 9/2$) is the best function to use to test ΛCDM model. Their results indicate that GP in general tend to prefer smooth covariance reconstruction functions if there is no rapid variation on the reconstructed quantities. Therefore, we use the Matérn ($\nu = 9/2$) covariance function [57]:

$$k(z,\tilde{z}) = \sigma_f^2 \exp\left(-\frac{3\,|z-\tilde{z}|}{\ell}\right) \times \left[1 + \frac{3\,|z-\tilde{z}|}{\ell} + \frac{27(z-\tilde{z})^2}{7\ell^2} + \frac{18\,|z-\tilde{z}|^3}{7\ell^3} + \frac{27(z-\tilde{z})^4}{35\ell^4}\right]. \tag{4.17}$$

For a given covariance function, the probability distribution of the hyperparameters depends only on the data. It is necessary either to marginalize over the hyperparameters σ_f and ℓ, or to fix the hyperparameters to their maximum likelihood values. Here we choose the latter approach, which is a good approximation and computationally much less expensive than marginalization. This has been tested by [56], a technical paper which discusses the critical issues when working with GPs, namely the choice of the covariance matrix and the optimization of the hyperparameters.

As we mentioned above, we choose the Matérn ($\nu = 9/2$) covariance function because it leads to the most reliable results amongst the covariance functions that we have tested. Where"reliable" means the following: For various assumed cosmological models and many realizations of mock data sets, the assumed model *on average* lies within the reconstructed 1-σ limits for approximately 68% of the redshift range (and within the reconstructed 2-σ limits for $\sim 95\%$ of the redshift range). These values are theoretically expected, thus making Matérn ($\nu = 9/2$) a reliable covariance function for our purposes. A detailed analysis regarding the

optimal choice of covariance function can be found in [56]. (Note that these results only apply to GP reconstructions using D measurements. When applying GP to other data, another covariance function might be more reliable.)

We follow Appendix A, which contain a summary of the technical details of GP. Also, for detailed reviews of GP, see MacKay [58].

4.2.2 Application to real data

In this section we will explain how we apply GP method to the Union 2.1 dataset [40] and determine the current constraints on the consistency of ΛCDM.

Union 2.1 dataset comprises 580 SNIa, with $0.015 < z < 1.5$, and includes a covariance matrix which incorporates a systematic uncertainty. The distance modulus, $\mu = m - M$, is the difference between the observed magnitude $m(z)$ and the absolute magnitude of an object M, and is given by (2.20). We choose $H_0 = 70\,\mathrm{kms}^{-1}\mathrm{Mpc}^{-1}$, as in [40]. Note that H_0 and M are degenerate in (2.20) so we can fix H_0 and only consider the uncertainties in M which are included in the covariance matrix of the Union 2.1 dataset [40] – this includes the errors on H_0. We convert μ to D and add the theoretical values $D(z = 0) = 0$ and $D'(z = 0) = 1$ to the dataset. This form of D and D' is what we feed to the GP as training points, see Appendix A.

Fig. 4.1 shows the reconstructed $D(z)$ and its first three derivatives for the Union 2.1 data set, while Fig. 4.2 shows the inferred reconstructions for $\mathcal{O}_m^{(1)}$, $\mathcal{O}_m^{(2)}$ and $\mathcal{O}_K^{(2)}$. Fig. 4.3 shows the reconstruction of $\mathcal{L}^{(1)}$ and $\mathcal{L}^{(2)}$.

The errors on the reconstructed distances in Fig. 4.1 increase with increasing order of derivative. For example, at $z = 1.5$, the standard deviation is 0.05 for the reconstruction of D, 0.12 for D', 0.22 for D'', and 0.29 for D'''. The near-constancy of the errors on D''' reflect the fact that we are unable to constrain rapid variations (carried via higher derivatives) on scales below a typical length scale, which is roughly associated with ℓ. By using GP the scale ℓ and the resulting smoothness of the reconstruction is driven purely by the data. Where there is insufficient *evidence* for rapid variations, a smooth function will result, which we see in the second and third derivatives.

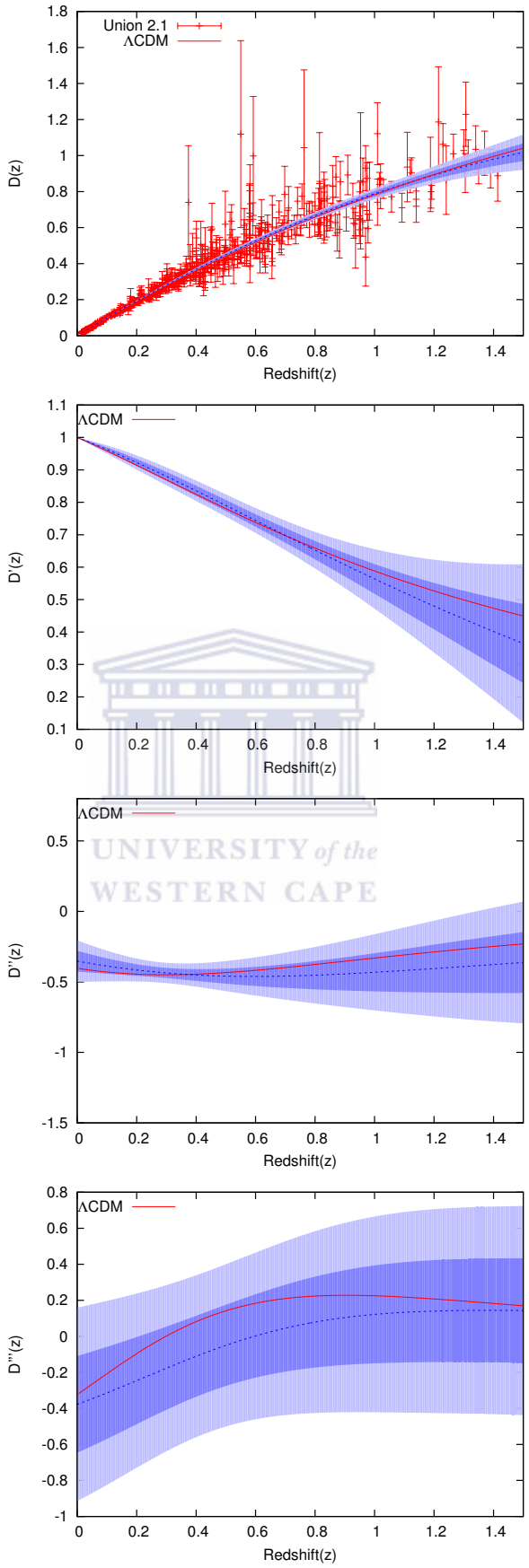

FIGURE 4.1: Gaussian processes reconstruction of D, D' (*top*) and D'', D''' (*bottom*) for Union 2.1 data. The red (solid) line is flat ΛCDM with $\Omega_m = 0.27$. The blue (dashed) line is the mean of the reconstruction. Shaded areas give 95% (light) and 68% (dark) confidence limits of the reconstructed function.

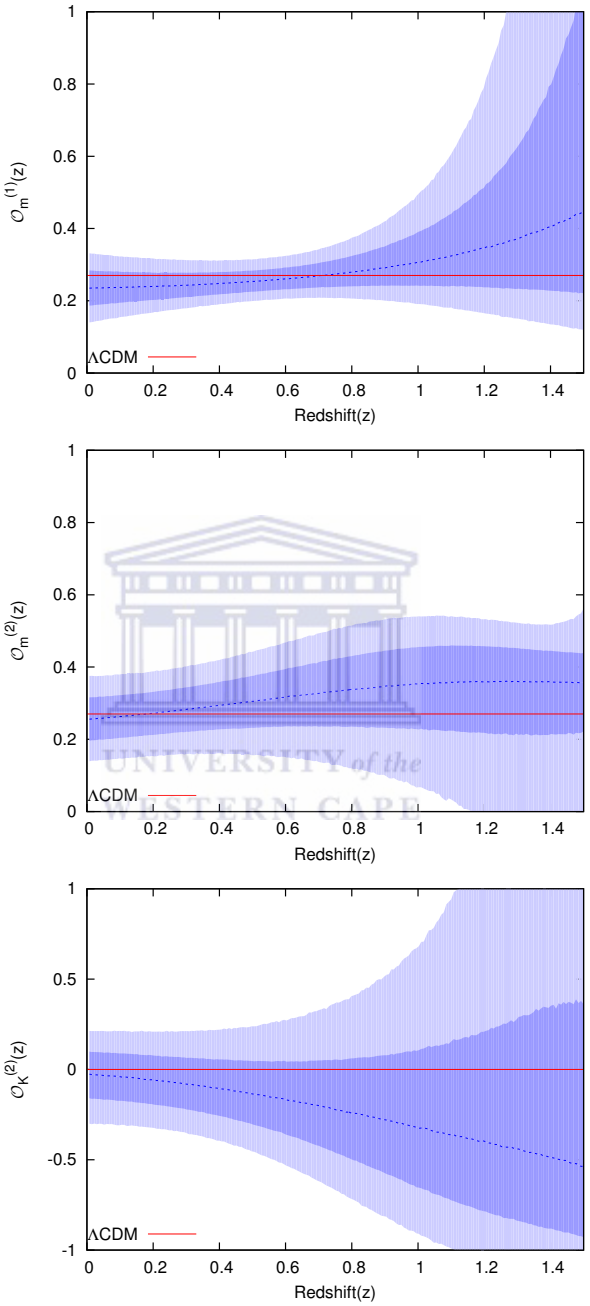

FIGURE 4.2: Reconstruction of $\mathcal{O}_m^{(1)}$ (*top*), $\mathcal{O}_m^{(2)}$ (*middle*) and $\mathcal{O}_K^{(2)}$ (*bottom*) for Union 2.1 data. Lines and shadings are as in Fig. 4.1.

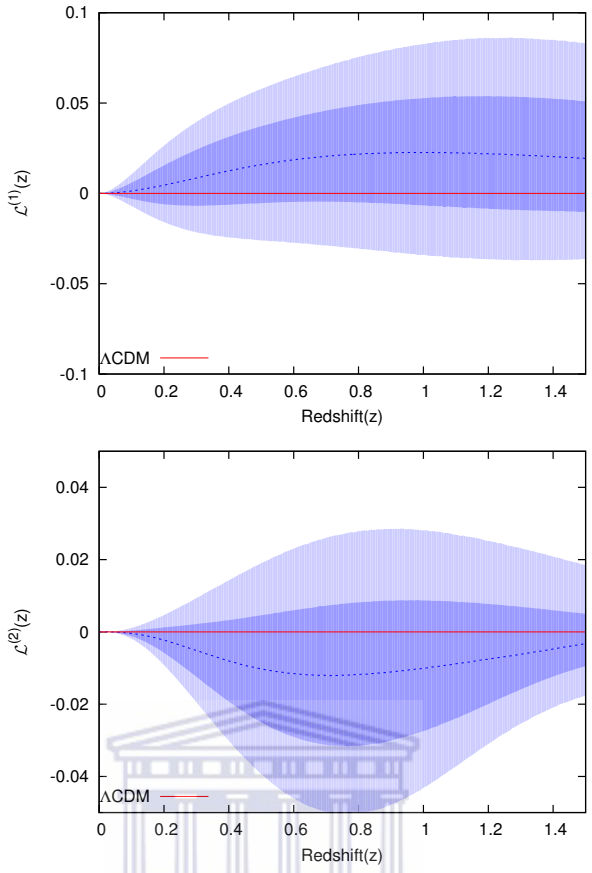

FIGURE 4.3: Reconstruction of $\mathcal{L}^{(1)}$ (*top*) and $\mathcal{L}^{(2)}$ (*bottom*) for Union 2.1 data. Lines and shadings are as in Fig. 4.1.

4.2.3 Mock data

To demonstrate the ability of the null tests to distinguish between different cosmological models when applied to future SNIa datasets, we produce mock catalogues for two fiducial models:

- Flat ΛCDM
- Dynamical dark energy model with $\Omega_K = 0$ and

$$w(z) = \frac{1}{2}\Big\{ -1 + \tanh\Big[3\Big(z - \frac{1}{2}\Big)\Big]\Big\}. \tag{4.18}$$

The choice of this form of $w(z)$ is motivated by our interest in testing a slow evolving function of dark energy.

To create the mock catalouges, we take $\Omega_m = 0.3$. Using the redshift distribution and scatter anticipated by DES [44], we simulate ~ 4000 data points in the redshift range $0 < z < 1.2$. Note that the scatter only includes statistical errors.

For each of the two simulated datasets, we reconstruct $D(z)$ and its derivatives and apply the null tests. Fig. 4.4 and 4.5 shows the constraints and uncertainties on $\mathcal{O}_m^{(1)}$, $\mathcal{O}_m^{(2)}$ and $\mathcal{O}_K^{(2)}$ for ΛCDM model and evolving w model, respectively, while Fig. 4.6 shows the results for $\mathcal{L}^{(1)}$ and $\mathcal{L}^{(2)}$.

4.3 Discussion

We have introduced an approach to applying null tests of the ΛCDM models (flat and curved). Using a GP technique to reconstruct the distance-redshift relationship and its derivatives from SNIa data sets in a model-independent fashion, we have shown that the flat concordance model is consistent with current data, falling within the 1σ limits. The null tests are stronger if we assume flatness, as expected.

For the Union 2.1 dataset, the consistency tests are in good agreement with a constant, indicating no evidence of a deviation from a flat ΛCDM model (see Figs. 4.2, 4.4 and 4.5). For the $\mathcal{O}_m^{(1)}$ and $\mathcal{O}_m^{(2)}$ tests we find a value for $\Omega_m \sim 0.27$. $\mathcal{O}_K^{(2)}$ is consistent with zero, as expected for flat ΛCDM. Due to the limited number of SNIa in the Union 2.1 sample and the model-independent method we use, the reconstructed uncertainties are significant.

For a mock data set based on the DES supernova survey, we find that our approach can distinguish between competing cosmological models. Using a simulated sample drawn from a flat ΛCDM model, the recovered distribution of $\mathcal{O}_m^{(1)}$ is constant over the redshift range considered (Fig. 4.4 and 4.5), consistent with $\mathcal{O}_m^{(1)} = \Omega_m$. For the evolving w model of (4.18), $\mathcal{O}_m^{(1)}$ deviates strongly from a constant value, so that flat ΛCDM would be disfavoured. This is confirmed by the deviation of $\mathcal{L}^{(1)}$ from zero in Fig. 4.6.

When spatial curvature is allowed, the constraints from the null tests tend to be weakened, as would be expected by the degeneracy introduced by the extra degree of freedom [59]. For a flat ΛCDM fiducial model, the reconstructed distribution of $\mathcal{O}_m^{(2)}$ and $\mathcal{O}_K^{(2)}$ are consistent with being constant and equal to Ω_m and Ω_K

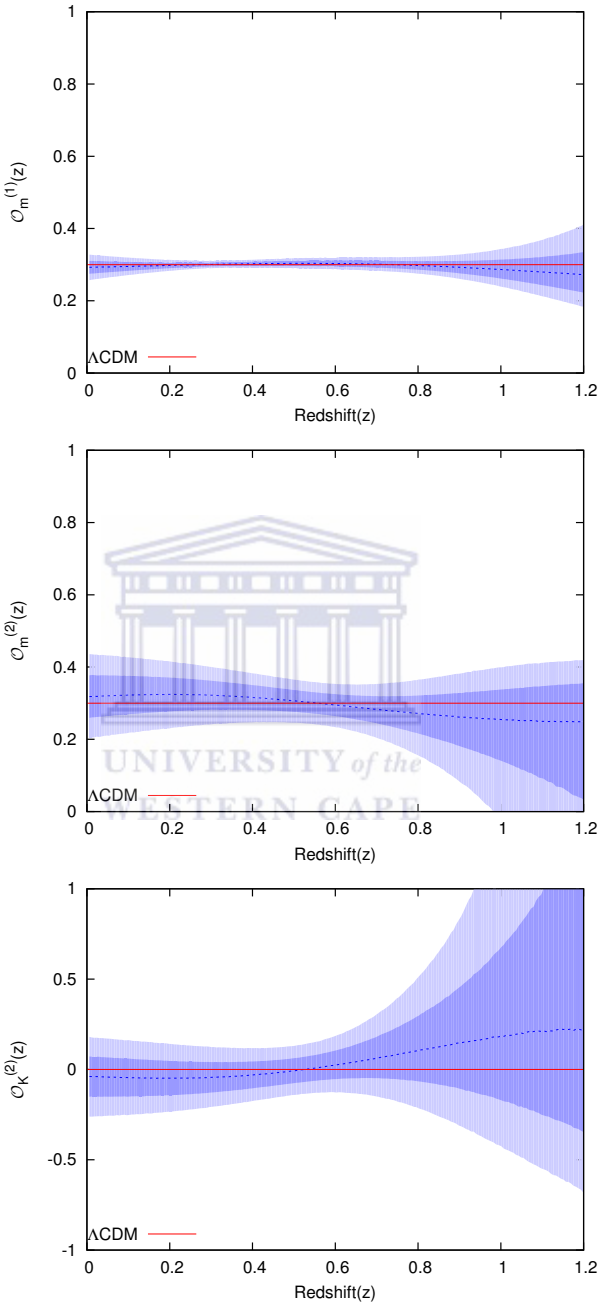

FIGURE 4.4: $\mathcal{O}_m^{(1)}$ (*top*), $\mathcal{O}_m^{(2)}$ (*middle*) and $\mathcal{O}_K^{(2)}$ (*bottom*) reconstructed using simulated DES data, and assuming ΛCDM.

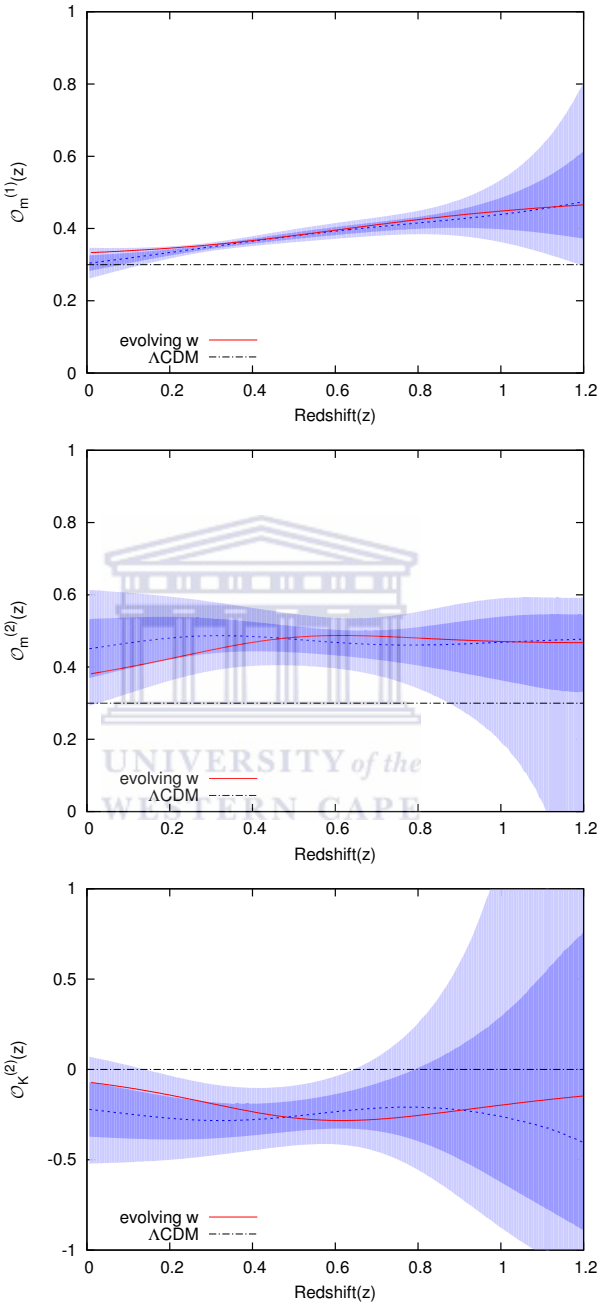

FIGURE 4.5: $\mathcal{O}_m^{(1)}$ (*top*), $\mathcal{O}_m^{(2)}$ (*middle*) and $\mathcal{O}_K^{(2)}$ (*bottom*) reconstructed using simulated DES data, and assuming evolving w in (4.18).

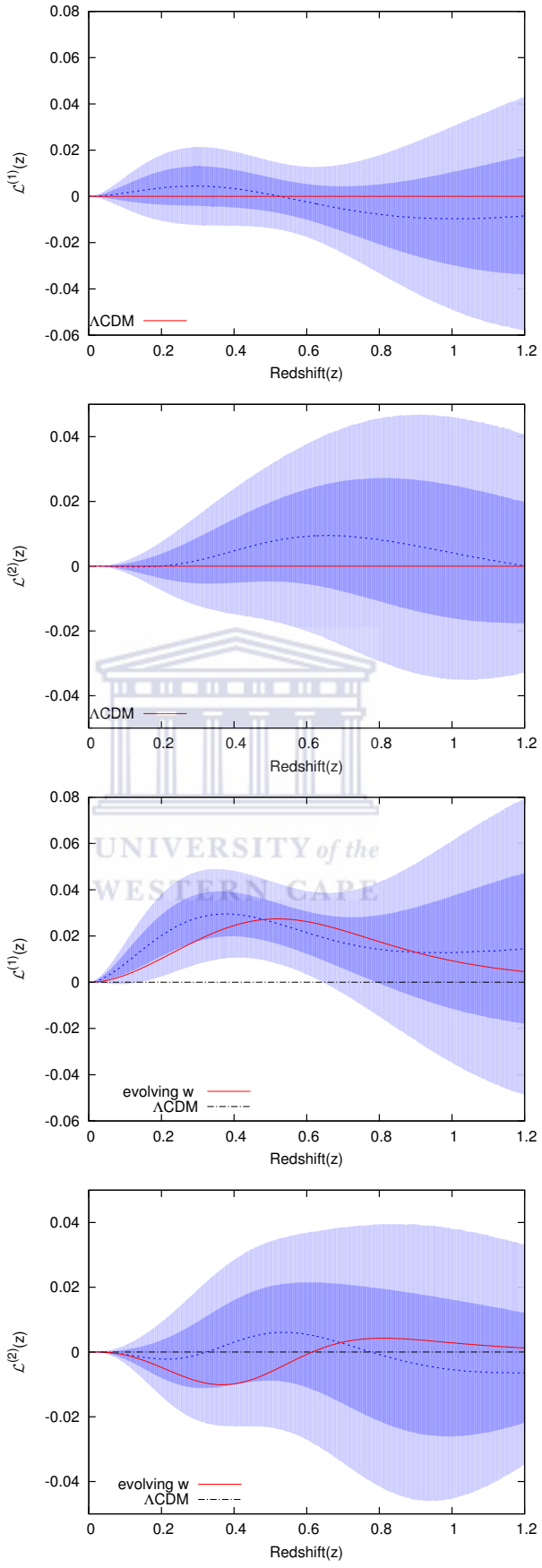

FIGURE 4.6: Reconstruction of $\mathcal{L}^{(1)}$ and $\mathcal{L}^{(2)}$ for simulated DES data, and assuming ΛCDM (*top*) and the evolving w in (4.18) (*bottom*). Due to the degeneracy between w and Ω_m, the reconstruction of $\mathcal{L}^{(2)}$ for the model with evolving dark energy is consistent with ΛCDM. However, the inferred values of Ω_m and Ω_K differ significantly from the input value as can be seen in Figs. 4.4 and 4.5.

(Fig. 4.4), respectively, confirming that the model does not deviate from ΛCDM, as anticipated. But the errors are significantly larger when curvature is allowed.

For the evolving w fiducial model, the reconstructions of $\mathcal{O}_m^{(2)}$ and $\mathcal{O}_K^{(2)}$ are consistent with constants (Fig. 4.5) – but these constant values differ significantly from the input values of Ω_m and Ω_K, respectively. The evolving w model can erroneously be interpreted as a ΛCDM model with a large matter density Ω_m and negative curvature Ω_K. Consequently, the reconstruction of $\mathcal{L}^{(2)}$ (Fig. 4.6) is consistent with a constant, indicating that ΛCDM is not disfavoured. In both cases, the errors are large and the null tests are degraded.

This problem reflects the degeneracy between the density parameters and the dark energy equation of state (see also [45, 59, 60]). The reconstructions are formally consistent with a constant, and thus with ΛCDM, due to their incorrectly inferred values. Additional constraints on the value of Ω_m and Ω_K from, for instance, BAO or CMB measurements, are needed to break this degeneracy.

4.4 Conclusions

In this Chapter, we described a series of null tests that can be applied to SNIa data to determine the consistency of observations with a (flat) ΛCDM model – without the need to parametrize the equation of state of dark energy. The tests require that the distance D and the diagnostics $\mathcal{O}_m^{(1)}$, $\mathcal{O}_m^{(2)}$, $\mathcal{O}_K^{(2)}$, $\mathcal{L}^{(1)}$ and $\mathcal{L}^{(2)}$ are reconstructed in a model-independent way. We used GP to perform these reconstructions.

We applied the null tests to the Union 2.1 SNIa data set. The results were consistent with a flat ΛCDM model (Figs 4.2 and 4.3).

Using the anticipated redshift distribution for the DES supernova survey, we produced mock data sets of 4000 SNIa, with two competing fiducial cosmological models: flat ΛCDM and an evolving w model. The reconstructed distributions of $\mathcal{O}_m^{(1)}$ for these datasets show that the consistency tests are able to distinguish between different cosmological models, and can correctly identify deviations from ΛCDM, in the case when spatial flatness is assumed. However, allowing for spatial curvature degrades the null tests in general (although not always – see Fig. 4.2). The inherent degeneracy between the equation of state of dark energy and the

density parameters (Ω_m, Ω_K) reduces our ability to distinguish between various models. The distributions of $\mathcal{O}_m^{(2)}$, $\mathcal{O}_K^{(2)}$ and $\mathcal{L}^{(2)}$ were consistent with a constant for the evolving w model (Fig. 4.6), but the inferred values of Ω_m and Ω_K from the $\mathcal{O}_m^{(2)}$ and $\mathcal{O}_K^{(2)}$ distributions were unrealistic (Fig. 4.4 and 4.5). The degeneracy needs to be broken using other data.

For future data sets which will have the power to probe ΛCDM at high precision, the null tests we have introduced will require further refinement. In particular, we need to develop a method of quantifying the significance of any possible deviation. This is left for future work.

Chapter 5

Using $H(z)$ data as a probe of the concordance model

In Chapter 4 we discussed the fact that one of the most direct ways to reconstruct w is via SNIa observations that give the luminosity distance D_L. Model-independent approaches to reconstructing w have been developed [5, 6, 45, 47, 61–74]. SNIa observations lead indirectly to $H(z)$ via the derivative $D'_L(z)$. Then we need the second derivative of $D_L(z)$ to reconstruct w. This is very challenging for any reconstruction technique since any noise on the measured $D_L(z)$ will be magnified in the derivatives. The problem can be lessened if direct $H(z)$ data are used because only the first derivative needs to be calculated to determine $w(z)$.

In this Chapter we focus on observations that directly give $H(z)$. This may be derived from differential ages of galaxies ('cosmic chronometers') and from the radial BAO scale in the galaxy distribution. Compared to SNIa observations, less $H(z)$ observational data are needed to reconstruct w with the same accuracy. For the cosmic chronometer data, it has been estimated [75] that 64 data points with the accuracy of the measurements in [76] are needed to achieve the same reconstruction accuracy as from the Constitution SNIa data [77].

We use GP method that we introduced in Chapter 4 for smoothing $H(z)$ data to also perform consistency tests of the flat ΛCDM model and of curved ΛCDM models. These consistency tests are formulated as functions of $H(z)$ and its derivatives which are constant or zero in ΛCDM, independently of the parameters of the model (see [78] for a review). Accordingly, deviations from a constant function indicate problems with our assumptions about dark energy, theory of gravity, or perhaps

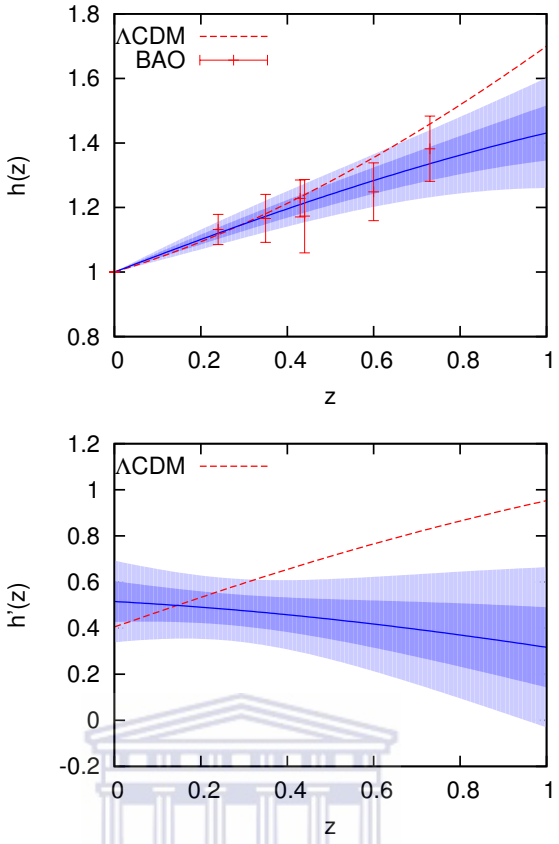

FIGURE 5.1: $h(z) = H(z)/H_0$ (top) and $h'(z)$ (bottom) reconstructed from BAO data, using Gaussian processes. Shaded areas represent 68% and 95% confidence levels. The dashed (red) curve is flat ΛCDM with $\Omega_m = 0.27$; the solid (blue) curve is the GP mean. Note that while the BAO data appear to give an inconsistent $h'(z)$, this is driven by the two highest redshift points both of which happen to lie below the flat ΛCDM curve.

something else, but without the usual problems of postulating an alternative to ΛCDM. Some of the tests we use here are given for the first time.

5.1 Testing ΛCDM

It is convenient to express $w(z)$ as a function of $h(z) = H(z)/H_0$ using (2.15)

$$h^2(z) = \Omega_m(1+z)^3 + \Omega_K(1+z)^2 + \Omega_{de}\exp[3\int_0^z \frac{1+w(z')}{1+z'}dz'], \tag{5.1}$$

where $\Omega_{de} = 1 - \Omega_m - \Omega_K$. To find a formula for $w(z)$ in terms of $h(z)$, we can rearrange (5.1)

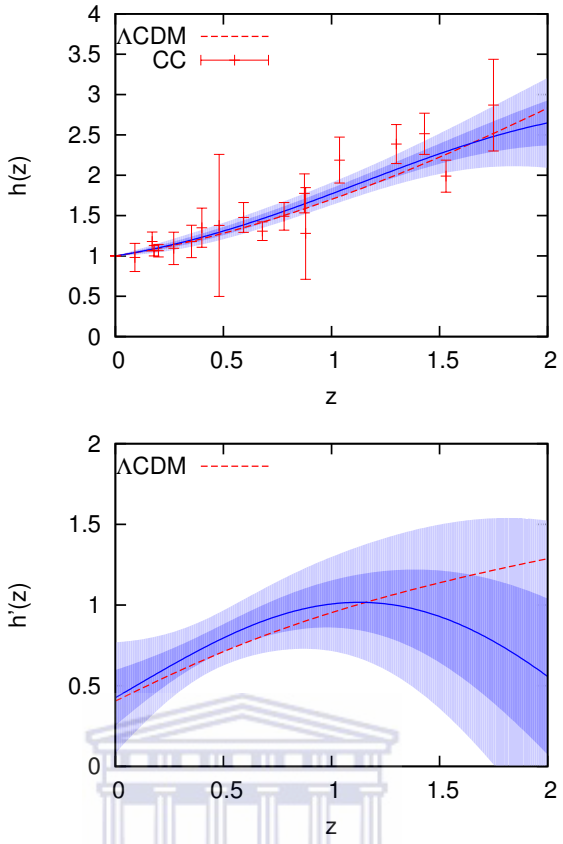

FIGURE 5.2: $h(z) = H(z)/H_0$ (top) and $h'(z)$ (bottom) reconstructed from cosmic chronometer data, using Gaussian processes. Lines and shadings are as in Fig. 5.1.

$$\exp \int_0^z \frac{1+w(z')}{1+z'} dz' = \frac{1}{3\Omega_{de}} [h(z)^2 - \Omega_m (1+z)^3 - \Omega_K (1+z)^2]. \tag{5.2}$$

Taking the logarithm of both sides of (5.2), then differentiate one time, we get

$$\frac{1+w(z)}{1+z} = \frac{1}{3} \frac{2hh' - 3\Omega_m (1+z)^2 - 2\Omega_K (1+z)}{h^2 - \Omega_m (1+z)^3 - \Omega_K (1+z)^2}, \tag{5.3}$$

thus, $w(z)$ in terms of $h(z)$ can be written as

$$w(z) \equiv \frac{p_{\rm de}}{\rho_{\rm de}} = \frac{2(1+z)hh' - 3h^2 + \Omega_K (1+z)^2}{3\left[h^2 - \Omega_m (1+z)^3 - \Omega_K (1+z)^2\right]}. \tag{5.4}$$

In principle, given $h(z)$ data we can smooth it, attempt to estimate its derivative, and reconstruct $w(z)$. Same as in $w(z)$ in terms of $D(z)$ case, the reconstruction of $w(z)$ in terms of Hubble rate is compromised by various difficulties. It depends on the values of Ω_m and Ω_K, so we need independent information about these

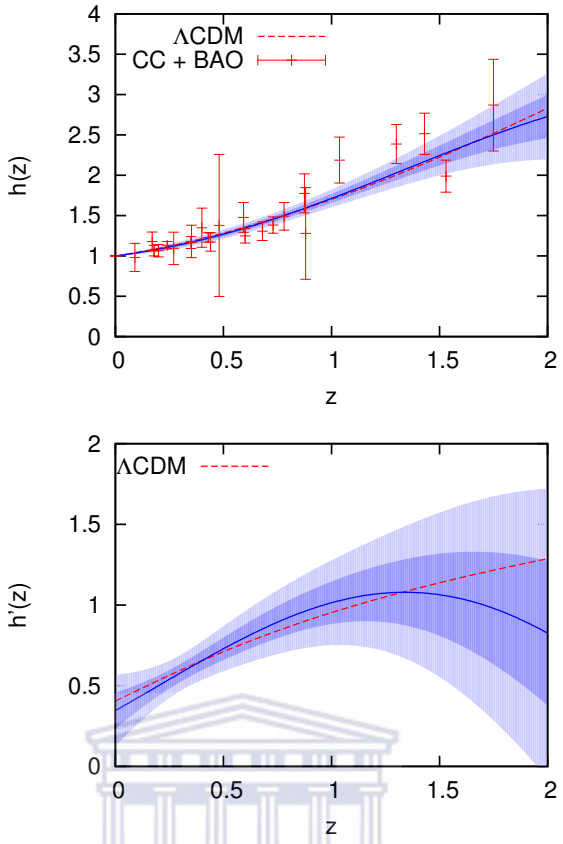

FIGURE 5.3: $h(z) = H(z)/H_0$ (top) and $h'(z)$ (bottom) reconstructed from CC+BAO data, using Gaussian processes. Lines and shadings are as in Fig. 5.1.

parameters when we reconstruct $w(z)$ from $H(z)$ data. These are difficult to estimate without assuming a form for $w(z)$ [59, 60, 79].

These difficulties reflect the fact that we cannot use data to construct physical models – rather, we need to use data to test physical models. The ΛCDM model could be tested by looking for deviations from $w = -1$. However, there is a more focused approach: to develop null hypotheses for ΛCDM, independently of the parameters Ω_m and Ω_K [78].

To test the concordance model – i.e. flat ΛCDM – we can use (5.1) to define a diagnostic function of redshift [50–52]:

$$\mathcal{O}_m^{(1)}(z) \equiv \frac{h^2 - 1}{z(3 + 3z + z^2)}. \tag{5.5}$$

By measuring $h(z)$ and calculating the right hand side of this equation, we should obtain the same value of Ω_m if the assumed model is true, regardless of the redshift

of the measurements. Therefore

$$\mathcal{O}_m^{(1)}(z) \quad = \quad \Omega_m \quad \text{implies the concordance model.}$$

This test is equivalent to the test in (4.4) that has been introduced in Chapter 4 in terms of $D(z)$. Similarly, if $\mathcal{O}_m^{(1)}(z)$ is not a constant, this is a signal of an alternative dark energy or modified gravity model. Given observed $h(z)$ data, we can estimate confidence limits for $\mathcal{O}_m^{(1)}$. If these are not consistent with a constant value, we can rule out the concordance model.

Therefore it is easier to measure deviations from zero than from a constant. The more effective diagnostic is thus the vanishing of the derivative $\mathcal{O}_m^{(1)\prime}(z)$. This is equivalent to $\mathcal{L}^{(1)} = 0$, where [51]

$$\mathcal{L}^{(1)} \quad \equiv \quad 3(1+z)^2(1-h^2) + 2z(3+3z+z^2)hh'. \tag{5.6}$$

The null test is therefore

$$\mathcal{L}^{(1)} \quad \neq \quad 0 \quad \text{falsifies the concordance model.}$$

To apply this test, we need to reconstruct $h'(z)$ from the data. Note that this test is equivalent to (4.7).

If the concordance model is ruled out, it is still possible that a curved ΛCDM model describes the Universe. Equations 5.1 and 5.4 (with $w = -1$) form a linear system for Ω_m and Ω_K. Solving for these parameters we can define

$$\mathcal{O}_m^{(2)}(z) \quad \equiv \quad 2\frac{(1+z)(1-h^2) + z(2+z)hh'}{z^2(1+z)(3+z)}, \tag{5.7}$$

$$\mathcal{O}_K(z) \quad \equiv \quad \frac{3(1+z)^2(h^2-1) - 2z(3+3z+z^2)hh'}{z^2(1+z)(3+z)}, \tag{5.8}$$

and we have

$$\mathcal{O}_m^{(2)}(z) \quad = \quad \Omega_m \quad \text{implies ΛCDM,}$$
$$\mathcal{O}_K(z) \quad = \quad \Omega_K \quad \text{implies ΛCDM.}$$

These quantities are equivalent to those in (4.9) and (4.10) in terms of $D(z)$, the dimensionless comoving luminosity distance [53]. The $D(z)$ forms contain second derivatives D'' whereas the $h(z)$ forms above contain only first derivatives h'.

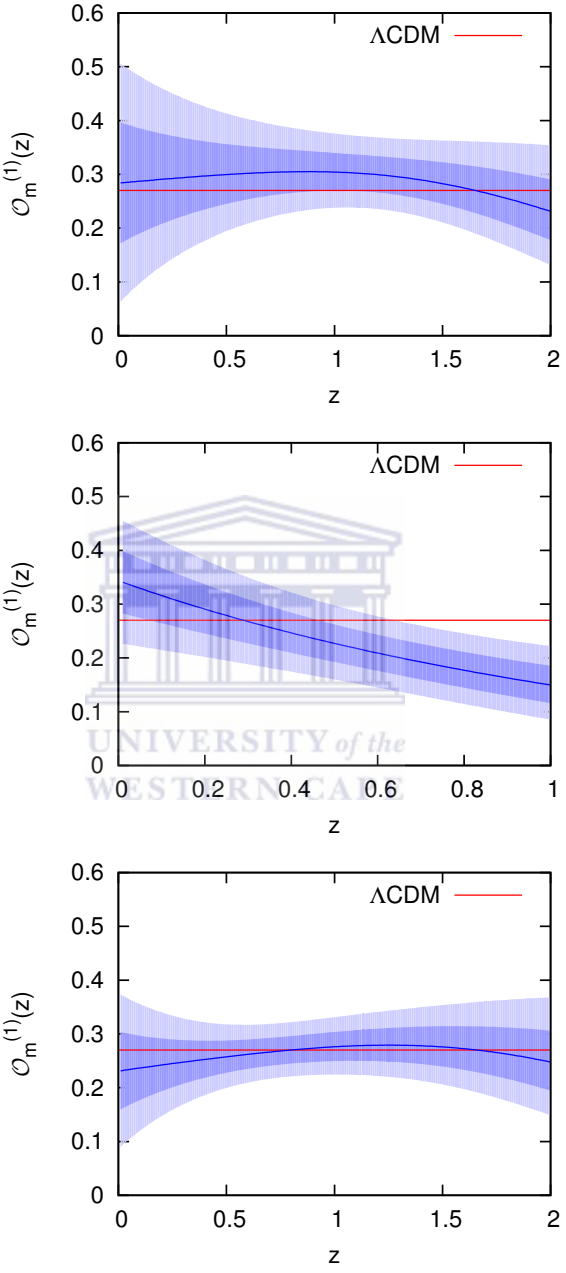

FIGURE 5.4: $\mathcal{O}_m^{(1)}(z)$ reconstructed from cosmic chronometers (top), BAO (middle) and CC+BAO (bottom). Where the dashed (red) curve is flat ΛCDM.

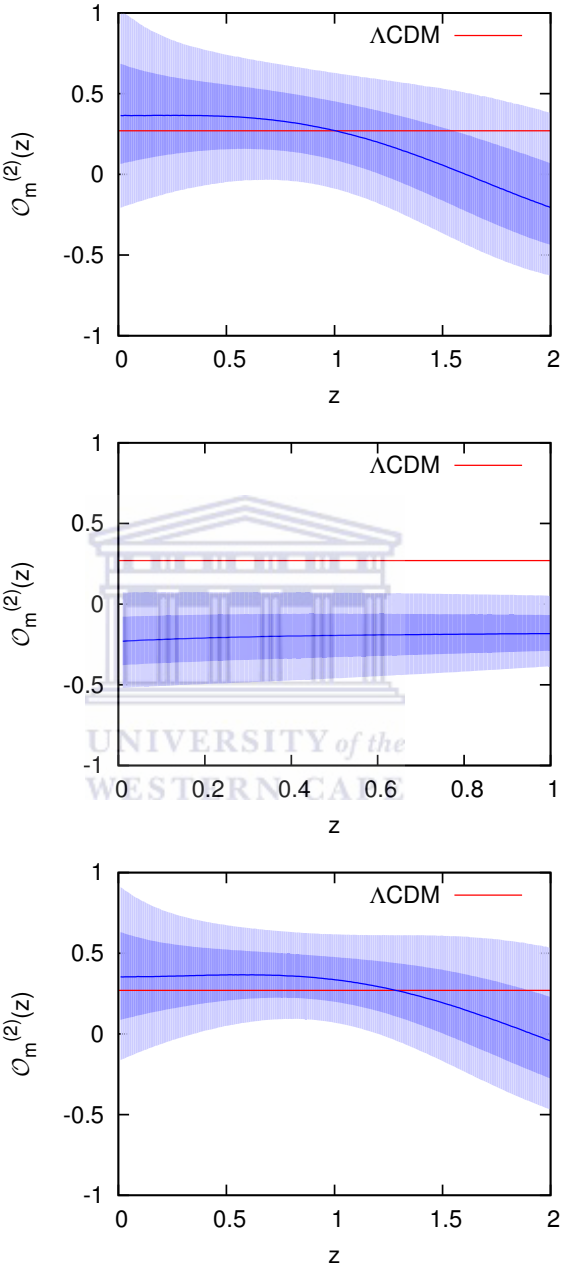

FIGURE 5.5: $\mathcal{O}_m^{(2)}(z)$ reconstructed from cosmic chronometers (top), BAO (middle) and CC+BAO (left). The dashed (red) curve is a curved ΛCDM model.

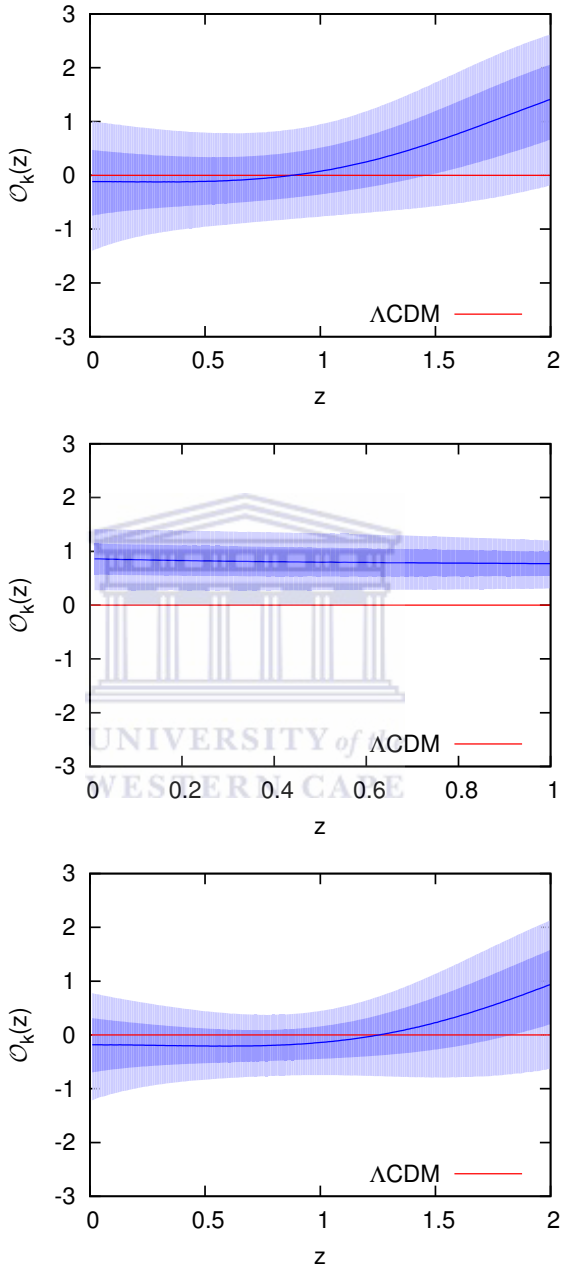

FIGURE 5.6: $\mathcal{O}_K(z)$ reconstructed from cosmic chronometers (top), BAO (middle) and CC+BAO (bottom). The dashed (red) curve is a curved ΛCDM model.

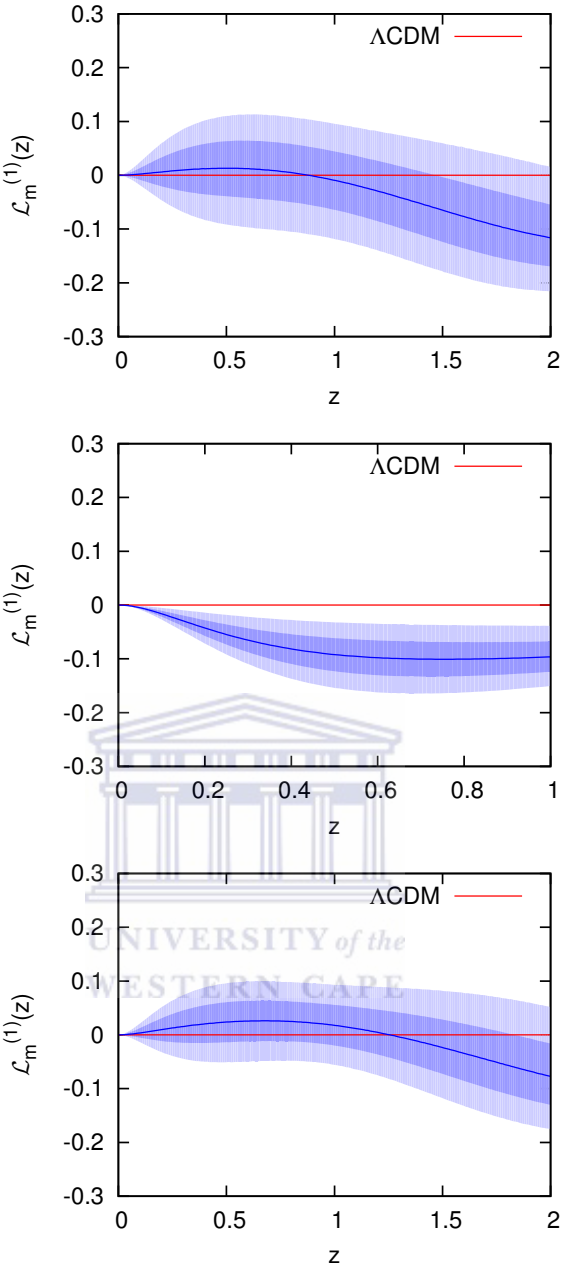

FIGURE 5.7: $\mathcal{L}_m^{(1)} = \mathcal{L}^{(1)}/(1+z)^6$ reconstructed from cosmic chronometers (top), BAO (middle) and CC+BAO (bottom). The dashed (red) curve is a ΛCDM model.

Given observed Hubble rate data from which we can estimate the derivative $h'(z)$, we can then estimate confidence limits for $\mathcal{O}_m^{(2)}(z)$ and $\mathcal{O}_K^{(2)}(z)$. If these are not consistent with a constant value, we can rule out ΛCDM in general, and conclude that dark energy has $w \neq -1$ (or there is modified gravity). The more effective diagnostic of these consistency tests is the vanishing of the derivatives of (5.7) and

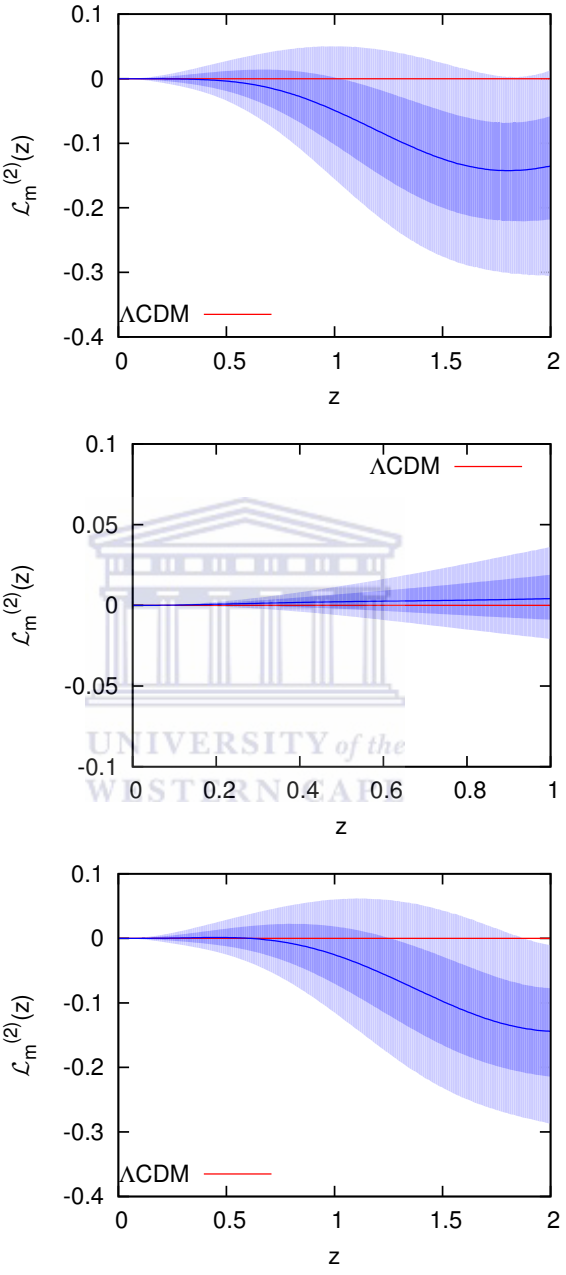

FIGURE 5.8: $\mathcal{L}_m^{(2)} = \mathcal{L}^{(2)}/(1+z)^6$ reconstructed from cosmic chronometers (top), BAO (middle) and CC+BAO (bottom). The dashed (red) curve is a ΛCDM model.

(5.8). The vanishing of $\mathcal{O}_m^{(2)\prime}$ is equivalent to $\mathcal{L}^{(2)} = 0$, where

$$\mathcal{L}^{(2)}(z) \equiv 3(1+z)^2(h^2-1) - 2z(3+6z+2z^2)hh' + z^2(3+z)(1+z)(h'^2+hh''). \quad (5.9)$$

Then

$$\mathcal{L}^{(2)}(z) \neq 0 \quad \text{falsifies } \Lambda\text{CDM}.$$

The vanishing of $\mathcal{O}_K^{(2)\prime}$ does not give any independent information – it is also equivalent to $\mathcal{L}^{(2)} = 0$.

Given observations of $h(z)$, we can construct this function independently of the parameters of the model and test ΛCDM by measuring consistency with zero.

Remember, this has the advantage that it is easier to detect deviations from zero rather than a constant, but at the expense of requiring an extra derivative in the observable. This is akin to detecting deviations from constant in w, but without reliance on the parameters of the model.

For the application of these consistency tests in term of $h(z)$, it is crucial to use a model-independent method to reconstruct $\mathcal{O}_m^{(1)}$, $\mathcal{O}_m^{(2)}$, $\mathcal{O}_K$, $\mathcal{L}^{(1)}$ and $\mathcal{L}^{(2)}$. Model-dependent approaches have the problem that they affect or even determine the outcome of the consistency test: While fitting a ΛCDM model to the data would always lead to a result that is consistent with ΛCDM, fitting a model that does not include ΛCDM as a special case would result in inconsistencies with ΛCDM. The only model-dependent approaches that do not entirely determine the outcome of the test are those assuming a model which includes ΛCDM as a special case. Nevertheless, they affect the result by forcing the data into a specific parametrisation, which might not reflect the true model. The only way to avoid this problem is to use a non-parametric approach. As in Chapter 4, we use GPs, which are described in Appendix A.

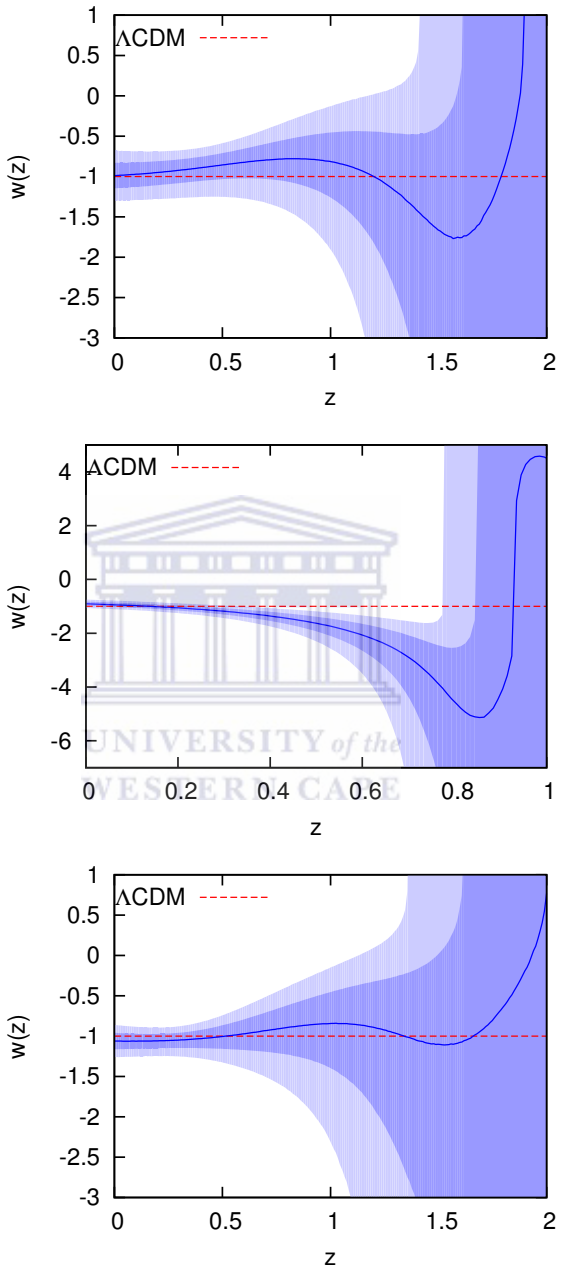

FIGURE 5.9: $w(z)$ reconstructed from cosmic chronometers (top), BAO (middle – note the different z range) and CC+BAO (bottom) by marginalizing over $\Omega_m = 0.275 \pm 0.016$. The dashed (red) curve is a ΛCDM model.

5.2 Reconstruction and consistency tests from $H(z)$ data

Cosmic chronometers are based on observations of the differential ages of galaxies [76, 80–82]. The Hubble rate at an emitter with redshift z is

$$H(z) = -\frac{1}{1+z}\frac{dz}{dt_e}, \tag{5.10}$$

where t_e is the proper time of emission. The differential method uses passively evolving galaxies formed at the same time to determine the age difference Δt_e in a small redshift bin Δz, assuming a Friedmann background. To find old galaxies sharing the same formation time, we have to look for the oldest stars in both galaxies and show that they have the same age. This method is effective; but while the differential approach significantly reduces the systematics that would be present when determining the absolute ages of galaxies, it still faces uncertainties due to the assumptions that are made to estimate the age.

The second way to measure $H(z)$ is the observed line-of-sight redshift separation Δz of the BAO feature in the galaxy 2-point correlation function [83–85],

$$H(z) = \frac{\Delta z}{r_s(z_d)}, \tag{5.11}$$

where $r_s(z_d)$ is the sound horizon at the baryon drag epoch.

Results: real data

We use the following $H(z)$ datasets:

CC: 18 cosmic chronometer data points [86].
BAO: 6 radial BAO data points [83–85].
CC+BAO: Combination of CC and BAO sets.

We normalize $H(z)$ using $H_0 = 70.4 \pm 2.5\,\mathrm{km\,s^{-1}Mpc^{-1}}$. The uncertainty in H_0 is transferred to $h(z)$ as

$$\sigma_h^2 = \frac{\sigma_H^2}{H_0^2} + \frac{H^2}{H_0^4}\sigma_{H_0}^2. \tag{5.12}$$

The reconstructed functions $h(z)$ and $h'(z)$ for the three datasets are shown in Fig. 5.1, 5.2 and 5.3. The shaded regions correspond to the 68% and 95% confidence levels (CL). The true model is expected to lie 68% of the plotted redshift range within the 68% CL. Note that this is only an expectation value. The actual value for a specific function may deviate from the expectation. The dependence of the actual percentage on the smoothness of the function has been tested and analysed in [45].

Fig. 5.4 shows the reconstruction of $\mathcal{O}_m^{(1)}$. The reconstruction of $\mathcal{O}_m^{(2)}$ and $\mathcal{O}_K$ is shown in Figs. 5.5 and 5.6, respectively. While Figs. 5.7 and 5.8 gives $\mathcal{L}^{(1)}$ and $\mathcal{L}^{(2)}$, respectively. We actually plot a modified $\mathcal{L}_m = \mathcal{L}/(1+z)^6$ which stabilises the errors at high redshift without affecting the consistency condition. To reconstruct $w(z)$, we use (5.4), and implement the smooth h and h'. The reconstructed $w(z)$, is shown in Fig. 5.9, where we assume the concordance values $\Omega_m = 0.275 \pm 0.016$ and $\Omega_K = 0$ [87].

Results: mock data

To demonstrate how a larger number of data will affect our results when reconstructing w and testing ΛCDM, we simulated a data set of 64 points for $H(z)$, drawing the error from a Gaussian distribution $\mathcal{N}(\bar{\sigma}, \epsilon)$ with $\bar{\sigma} = 10.64z + 8.86$ and $\epsilon = 0.125(12.46z + 3.23)$, where $\bar{\sigma}$ are the uncertainties we want to produce while ϵ is a parameter that guarantees the probability of $\bar{\sigma}$ lies within the 94% confidence level. A clear explanation of the adopted methodology can be found in [75].

We simulated data points for two different models:
Concordance model, $\Omega_K = 0$, $\Omega_m = 0.27$.
A model with slowly evolving equation of state, introduced in (4.18), and the same concordance density parameters.

The GPs reconstructions are shown in Figs. 5.10, 5.11–5.16.

Discussion

Fig. 5.4 shows that for the CC and CC+BAO data (18 and 24 points), we get good reconstructions when there is no differentiation of $h(z)$ involved. The BAO

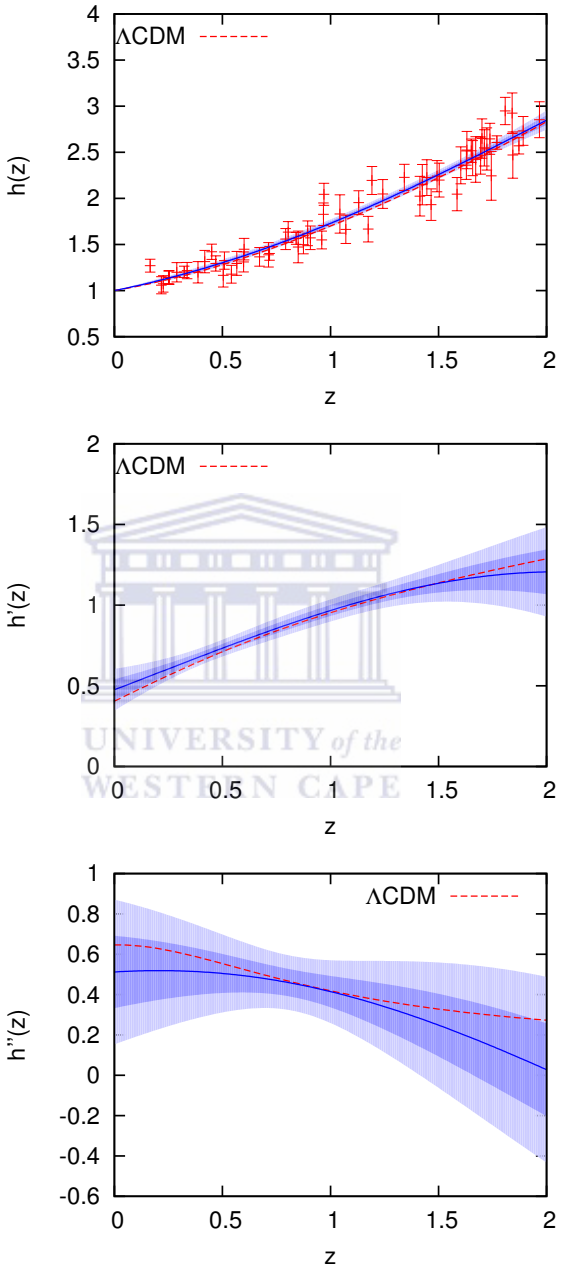

FIGURE 5.10: $h(z)$ (top), $h'(z)$ (middle) and $h''(z)$ (bottom) reconstructed from simulated data, assuming a concordance model.

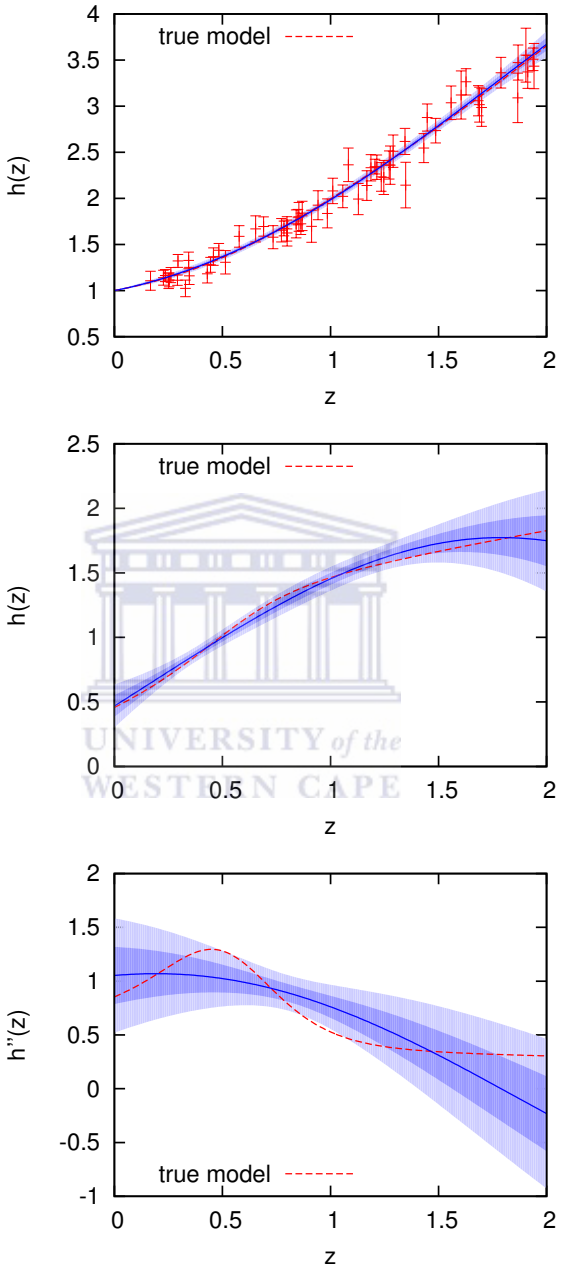

FIGURE 5.11: $h(z)$ (top), $h'(z)$ (middle) and $h''(z)$ (bottom) reconstructed from simulated data, assuming a model (4.18) with slowly evolving $w(z)$.

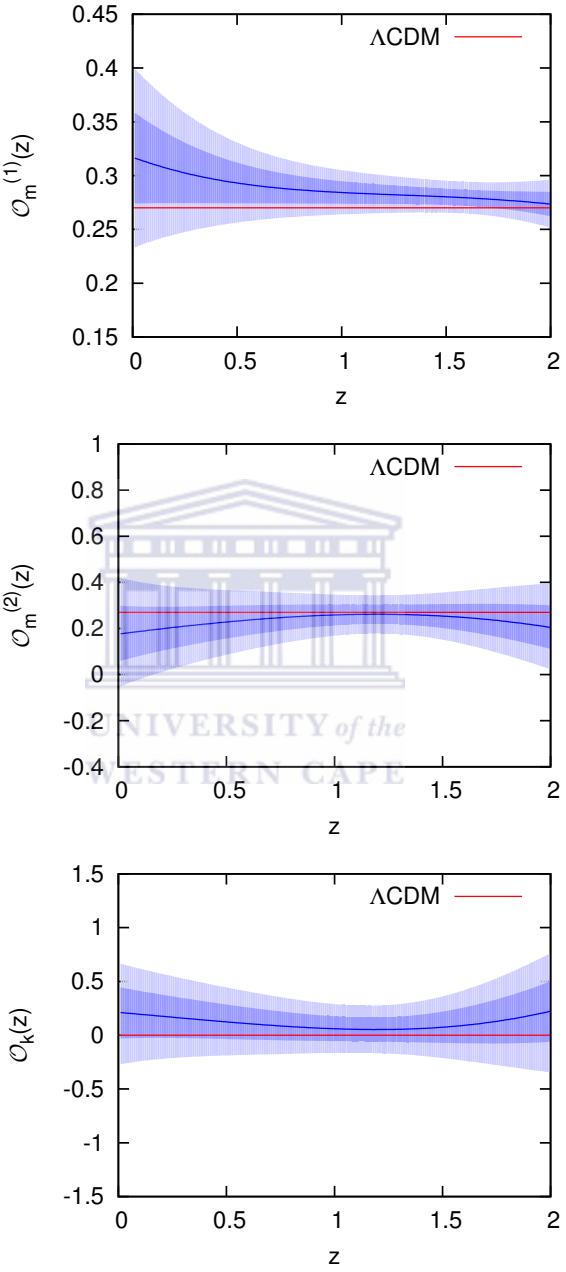

FIGURE 5.12: $\mathcal{O}_m^{(1)}(z)$ (top), $\mathcal{O}_m^{(2)}(z)$ (middle) and $\mathcal{O}_K(z)$ (bottom) reconstructed from simulated data, assuming a concordance model.

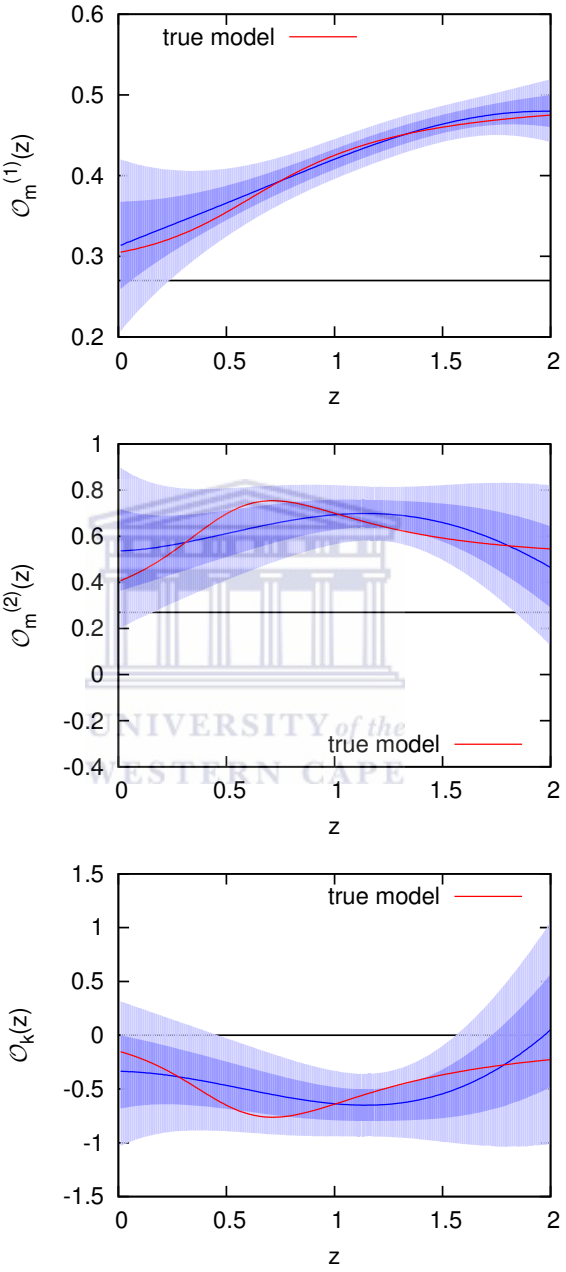

FIGURE 5.13: $\mathcal{O}_m^{(1)}(z)$ (top), $\mathcal{O}_m^{(2)}(z)$ (middle) and $\mathcal{O}_K(z)$ (bottom) reconstructed from simulated data, assuming a model (4.18).

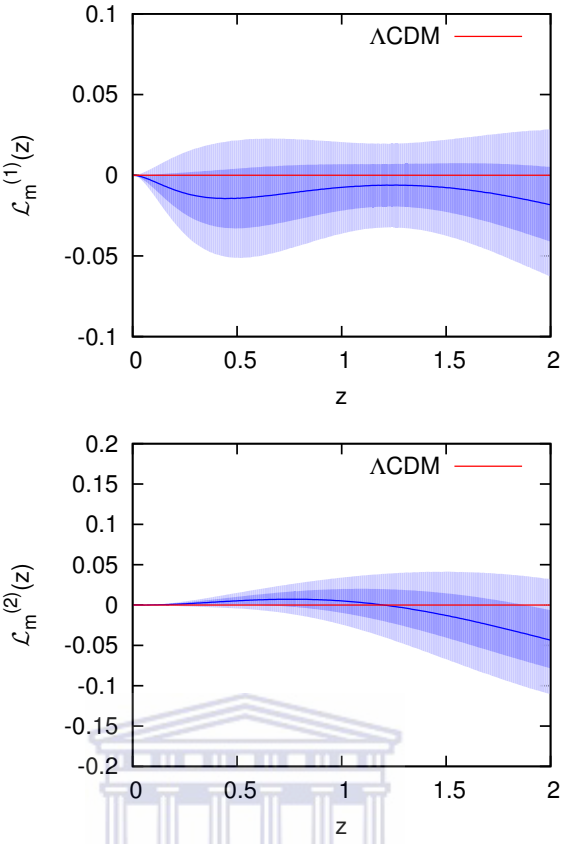

FIGURE 5.14: $\mathcal{L}_m^{(1)} = \mathcal{L}^{(1)}/(1+z)^6$ (top) and $\mathcal{L}_m^{(2)} = \mathcal{L}^{(2)}/(1+z)^6$ (bottom) reconstructed from simulated data, assuming a concordance model.

dataset only contains 6 data points up to redshift 0.73. Beyond that redshift, the reconstruction differs significantly from ΛCDM. The results from the CC and CC+BAO sets are however in very good agreement with ΛCDM.

The BAO data appear to be inconsistent with the concordance model. However, 6 data points are not sufficient for a reliable reconstruction. The two data points with highest redshift happen to be below the concordance curve, which pulls the reconstructed curve down. However, this illustrates the importance of having the derivative of the data consistent with the model, as well as the data itself. Current and upcoming large-volume surveys, such as BOSS [88], EUCLID [89] and SKA [90], will provide radial BAO measurements of increasing number and precision.

The reconstruction of $\mathcal{O}_m^{(2)}$ and $\mathcal{O}_K$ shown in Figs. 5.5 and 5.6 is more challenging for the available dataset, since we need the first derivative of h. With present datasets, the uncertainties in the reconstruction are quite large. Using CC and

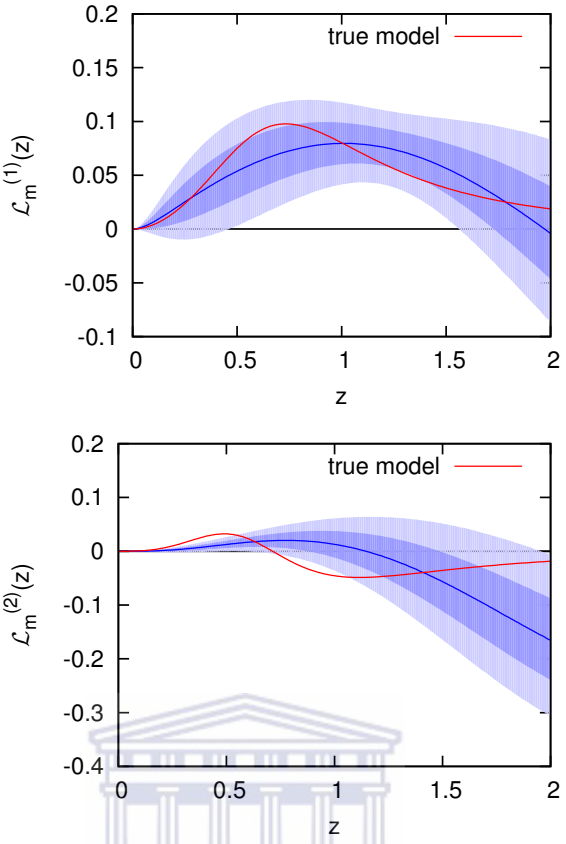

FIGURE 5.15: $\mathcal{L}_m^{(1)} = \mathcal{L}^{(1)}/(1+z)^6$ (top) and $\mathcal{L}_m^{(2)} = \mathcal{L}^{(2)}/(1+z)^6$ (bottom) reconstructed from simulated data, assuming a model (4.18).

CC+BAO, these results as well as the results for $\mathcal{L}^{(1)}$ and $\mathcal{L}^{(2)}$ shown in Figs. 5.7 and 5.8, are consistent with ΛCDM.

For the mock datasets, Figs. 5.10, 5.11, 5.12 and 5.13 show that the GP reconstructions recovers the assumed models very effectively. We can clearly distinguish the model with slowly evolving $w(z)$ from ΛCDM in $\mathcal{O}_m^{(1)}$. For $\mathcal{O}_m^{(2)}$ and $\mathcal{O}_K$, the reconstruction errors are too large to see this difference. The same is true for consistency tests $\mathcal{L}^{(1)}$ and $\mathcal{L}^{(2)}$ shown in Figs. 5.14 and 5.15.

The reconstruction of the equation of state $w(z)$ also shows a clear difference of the two models, assuming we can accurately determine H_0, Ω_m and Ω_K separately from $w(z)$: see Fig. 5.16. GP works very well to recover the assumed w. With less than 100 data points, we can reconstruct a dynamical dark energy model far better than is achievable using thousands of SNIa data – compare to analogous reconstructions in [45].

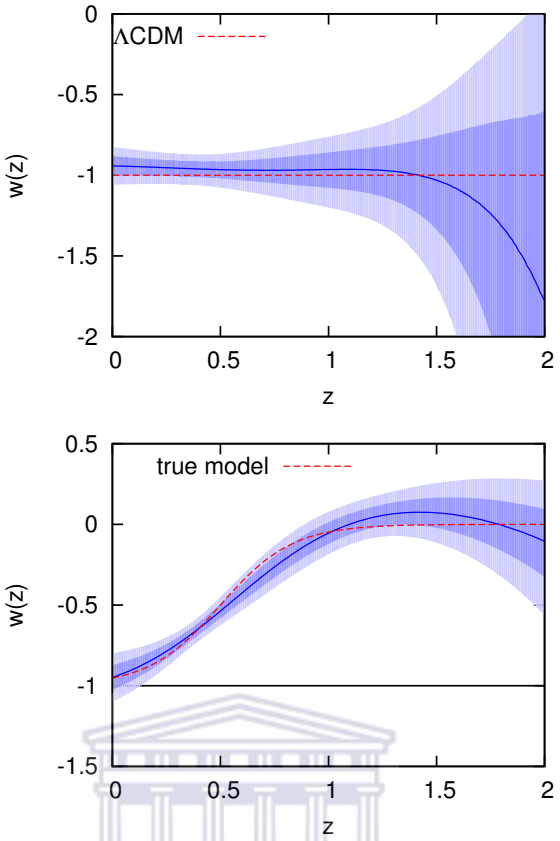

FIGURE 5.16: $w(z)$ reconstructed from simulated data, assuming a concordance model (top) and a model introduced in (4.18) (bottom), by marginalizing over $\Omega_m = 0.275 \pm 0.016$.

5.3 Conclusions

We have considered the information that current and future $H(z)$ data can give us. Currently such data come from cosmic chronometers and BAO data, and is plainly consistent with the concordance model. Future data, however, will provide a powerful discriminator between different models. It is remarkable how few data points are required compared to supernovae: to reconstruct $w(z)$ accurately in our non-parametric way requires many thousands of SNIa, compared to less than 100 $H(z)$ data points.

We have derived and analysed new consistency tests for the ΛCDM model, which we have formulated in terms of $H(z)$ directly, rather than using the more familiar distance function [53, 78]. By smoothing the data points using GP, we have shown that these can be very effective in determining that ΛCDM is the incorrect model, but without having to assume the key parameters Ω_m and Ω_K, which currently

only have constraints derived by assuming ΛCDM or a similar alternative. These tests not only require that the data points themselves are consistent with the model, but that their derivative is also.

Future data which directly measures the expansion history will therefore play an important role in future dark energy studies.

Chapter 6

Future SKA HI galaxy surveys

The SKA is a giant radio telescope array, to be constructed across two sites, in South Africa and Western Australia. The first phase early science stage starts in 2020 while the completion is due in 2023 [91]. The second phase (with about ten times the sensitivity and twenty times the field of view) planned for 2030. One of the key science aims of the SKA is to probe the nature of dark energy by mapping out large-scale structure, primarily using the 21cm emission line of HI to detect galaxies and measure their redshifts with high (spectroscopic) precision.

At present, HI galaxy surveys (e.g. HIPASS [92]) are quite small compared to optical and near-infrared counterparts like BOSS and WiggleZ, limiting their use for precision cosmology. The unprecedented sensitivity and the field of view of the SKA will allow for dramatically faster survey speeds, making it possible to map the galaxy distribution out to high redshifts over most of the sky. The end result will be sample variance-limited observations over a truly gigantic survey volume, allowing HI surveys to outperform other methods in terms of precision cosmological constraints, and making it possible to probe ultra-large scales and novel wide-angle effects [93, 94].

The current best cosmological constraints from large-scale structure surveys come from observations of the BAO. The BAO feature is a preferred clustering scale imprinted in the matter distribution by acoustic oscillations in the coupled photon-baryon fluid around the time of decoupling [1]. The radial and transverse BAO scales depend on the Hubble rate, $H(z)$, and the angular diameter distance, $D_A(z)$, as well as the (comoving) sound horizon in the 'baryon drag' epoch, $r_s(z_d)$ that we introduced in section 3.4. The comoving sizes of the BAO feature along and

across the line of sight are given by

$$s_{\parallel}(z) = \frac{c\Delta z}{H(z)}, \quad s_{\perp}(z) = (1+z)D_A(z)\Delta\theta, \tag{6.1}$$

where the redshift extent Δz and angular size $\Delta\theta$ of the BAO feature in the galaxy correlation function are the observables. In the absence of RSDs and nonlinear effects, we have $s_{\perp} = s_{\parallel} = r_s(z_d)$, which can be precisely estimated from CMB measurements. The BAO scale is therefore a 'standard ruler', with which we can obtain precise constraints on D_A and H, and thus the dark energy equation of state, $w(z) = p_{de}/\rho_{de}$ and other quantities.

The expected performance of SKA HI galaxy surveys in constraining dark energy was previously investigated by Abdalla et al. [93]. In this work we update those results, using improved modelling of the number density and bias of the HI galaxy distribution, as well as the various SKA configurations specified in [95]. We provide the expected galaxy number counts and bias as a function of redshift and raw flux sensitivity, and map these on to specific SKA configurations. We then present Fisher forecasts for the BAO for each configuration, and use these to compare with the performance of other galaxy surveys.

6.1 Telescope and Survey specifications

In this section, we analyse the specifications and expected flux sensitivities of surveys with various SKA configurations.

The SKA will be built in two phases. Phase 1 will consist of three separate sub-arrays: SKA1-MID, SKA1-SUR[1] and SKA1-LOW [96]. MID and SUR are dish arrays equipped with the mid-frequency receivers ($\nu \lesssim 1.4$ GHz) necessary to detect HI emission at low/intermediate redshift, while LOW is an aperture array optimised for lower frequencies (< 350 MHz) and thus higher redshifts. We will concentrate on MID and SUR here, and their corresponding 'precursor' arrays, MeerKAT and ASKAP, which they will be co-sited with, and which can be connected into the final Phase 1 systems. LOW will be capable of detecting HI emission only for $z \geq 3$, which will presumably be done most efficiently using

[1] After this work was completed, SUR was deferred as part of a re-baselining review for SKA1 [91].

Telescope	Band [MHz]	Target freq.$^{(a)}$ [MHz]	$T_{\rm inst}{}^{(b)}$ [K]	$N_{\rm dish}{}^{(c)}$	$D_{\rm dish}{}^{(d)}$ [m]	$A_{\rm e}{}^{(e)}$ [m^2]	$N_b{}^{(f)}$	Beam$^{(g)}$ [deg^2]	$S_{\rm rms}{}^{(h)}$ [μJy]
SKA1-MID	950 – 1670	1355	20	190	15	26,189	1	0.48	247
MID+MeerKAT	950 – 1670	1310	20	254	–	32,144	1	0.51	202
SKA1-SUR	650 – 1670$^{(i)}$	1300	30	60	15	8,482	36	18	1151
SUR+ASKAP	650 – 1670	1300	38/30$^{(j)}$	96	–	11,740	36	18	1050/830$^{(j)}$
SKA2$^{(l)}$	480 – 1290	–	–	–	–	–	–	30	16

TABLE 6.1: Specifications for the various SKA telescope configurations (see text and notes below for explanation). Values quoted at the middle of the band in order to compare directly to the SKA baseline design document [96].

Notes: **(a)** This is the frequency at which frequency-dependent quantities are calculated. For PAFs this is taken to be the critical frequency (below which the beam is constant). **(b)** Instrument temperature. **(c)** Number of dishes. **(d)** Dish diameter. **(e)** Total effective collecting area of the interferometer. **(f)** Total number of beams (feeds). **(g)** Total FOV (primary beam, for one pointing), calculated at the target frequency. For combined telescopes, the smaller beam of the two is used. For PAFs, it is multiplied by the number of beams above the critical frequency. **(h)** Flux rms for a frequency interval of 10 kHz and 1 hour integration using (6.3). **(i)** Only 500 MHz instantaneous bandwidth assumed. **(j)** The first value takes a weighted average of the SUR+ASKAP temperature, while the second assumes that the ASKAP PAFs are replaced to match SUR system temperature. **(l)** Values here are only indicative (see text).

intensity mapping rather than a galaxy survey, so we will not consider it here (although see e.g. [97]).

The specifications of Phase 2 are less well-defined. While its target sensitivity has been given – around ten times that of MID or SUR at mid-frequency – the receiver technology, field of view, and baseline distribution are not yet decided. As such, we can only speculate on these details here. To "future-proof" our results to some extent, in later sections we will present results for the HI galaxy number counts and bias as a function of raw flux sensitivity, as well as for individual experimental configurations. The former can easily be rescaled for the actual specifications of Phase 2 when they are announced, as well as for any other future radio experiment that targets HI.

6.1.1 Flux sensitivity

We begin by reviewing the basic flux sensitivity equation. The rms (root mean square) noise associated with the flux measured by an interferometer is

$$S_{\rm rms} \approx \frac{2k_B T_{\rm sys}}{A_e\sqrt{2\delta\nu\, t_p}}, \tag{6.2}$$

for a telescope with system temperature $T_{\rm sys}$, total effective collecting area A_e, frequency resolution $\delta\nu$, and observation time per pointing t_p (k_B is the Boltzmann constant). We have assumed that the noise is Gaussian. The extra factor of $1/\sqrt{2}$ comes from assuming a dual-polarisation receiver system. For a dish reflector, the effective collecting area is typically about 70% of its total geometrical area.

The expression above gives the flux sensitivity for the telescope psf (point spread function); that is, the noise rms for an "angular" pixel set by the resolution of the interferometer (not to be confused with its field of view or primary beam). This calculation corresponds to the so called "natural array" sensitivity.

In this work, we will consider a range of values when analysing the cosmological performance to allow for differences in the final line-processed sensitivity.

The total system temperature is given by $T_{\rm sys} = T_{\rm inst} + T_{\rm sky}$, where the contribution from the sky is $T_{\rm sky} \approx 60\,(300\,{\rm MHz}/\nu)^{2.55}$ K, where the coefficient 60 comes from the system noise. $T_{\rm inst}$ is the instrument temperature (which is usually higher than

the sky temperature for $\nu \gg 300$ MHz). For typical instrumental specifications, the noise rms for the array can be written as

$$S_{\rm rms} = 260\,\mu{\rm Jy}\left(\frac{T_{\rm sys}}{20\,{\rm K}}\right)\left(\frac{25,000\,{\rm m}^2}{A_e}\right)\times\left(\frac{10\,{\rm kHz}}{\delta\nu}\cdot\frac{1{\rm hr}}{t_{\rm p}}\right)^{1/2}.$$

We will assume that the interferometer, *in a single pointing*, can observe the following sky area, corresponding to the primary beam or field of view of a dish:

$$\theta_{\rm B}^2 \approx \frac{\pi}{8}\left(\frac{1.3\lambda}{D}\right)^2\ [{\rm sr}], \tag{6.3}$$

where any efficiency factor has already been taken into account. This is valid for dishes with single feeds (single pixels) like MeerKAT and SKA1-MID. The ASKAP and SKA1-SUR dishes are equipped with Phased Array Feeds (PAFs), however, for which the situation is slightly more complicated. PAF systems are able to observe a total of N_b beams, depending on the number of feeds, so that the total field of view should be $N_b \times \theta_{\rm B}^2$. While $\theta_{\rm B}^2$ increases with wavelength, the total effective PAF beam will remain constant above a certain critical wavelength, corresponding to where the individual sub-beams begin to overlap with one another.

The specifications for each SKA configuration are summarised in Table 6.1, along with the expected flux rms for a one hour integration in a single pointing with a frequency resolution of 10 kHz. For SKA1-MID/SUR and their combination with MeerKAT/ASKAP, only Band 2 is considered, as the lower-frequency Band 1 will provide insufficient sensitivity for a HI galaxy survey. For the combined telescopes, only the overlapping band is given. Note however that the SKA1 baseline specifications suggest that the ASKAP PAFs should be replaced to match the SKA1-SUR band and instrumental temperature (taken to be 30 K).

For SKA2, as mentioned above, we just assume 10 times the sensitivity of the Phase 1 configurations, leaving other aspects of the specification (e.g. system temperature, number of dishes) undefined. We must still choose a field of view (FOV) and bandwidth, however; reasonable estimates are a FOV about 20 times that of the Phase 1 configurations, and a bandwidth sufficient to cover $0.1 \leq z \leq 2.0$ (i.e. $1290 \geq \nu \geq 480$ MHz). The significantly larger FOV can be supported by various proposed technologies for Phase 2, e.g. MFAA,[2] while the Phase 1 dish arrays will already possess the technology required to cover the specified frequency

[2] https://www.skatelescope.org/mfaa/

range (albeit with lower sensitivity, limiting the useful minimum frequency for HI galaxy surveys).

6.1.2 Survey specifications

To maximise its effectiveness, a balance must be found between the sensitivity of a survey and its area. In principle, wide surveys can probe larger volumes and thus sample a greater number of Fourier modes, but this comes at the cost of reducing sensitivity per pointing (for a fixed total survey time), thus increasing shot noise and reducing the maximum redshift that can be reached.

For a 10,000 hour survey, and the sensitivities given in Table 6.1, SKA2 will essentially be able to achieve a sample variance-limited "full sky" survey, so we set its survey area to 30,000 deg^2. For SKA1, however, the situation is less obvious. Using the Fisher matrix analysis described in the following sections, we searched for the optimal SKA1 survey area for our target science – in this case, whatever maximises the dark energy figure of merit (FOM). FOM is discussed briefly below. We also considered two possible frequency intervals: the current SKA1 MID specification (950-1670 MHz) and a slightly "deeper" band (800-1300 MHz) with a maximum redshift of ~ 0.8. The results of the optimisation procedure are shown in Fig. 6.1.

For the lower frequency range, Fig. 6.1 shows that the FOM is maximised for a survey area of around 5,000 deg^2. This is where a balance between depth and width is reached – the information gain from detecting the BAO at higher redshifts is traded-off against the larger sample variance due to the smaller area. Conversely, the higher frequency band is restricted to lower redshifts, limiting the maximum depth that can be achieved. This leads to a preference for larger areas ($\sim 25,000$ deg^2), although note that this survey would not be sample variance-limited as in the SKA2 case.

These results are specific to the figure of merit that we are optimising for. If we instead required a strong detection of the BAO at the highest redshift for the 950-1670 MHz band, for example, the optimal area would again be around 5,000 deg^2. Other issues can also be considered. For instance, we might want to "piggy-back" the HI galaxy survey on top of other surveys to optimise the overall observing

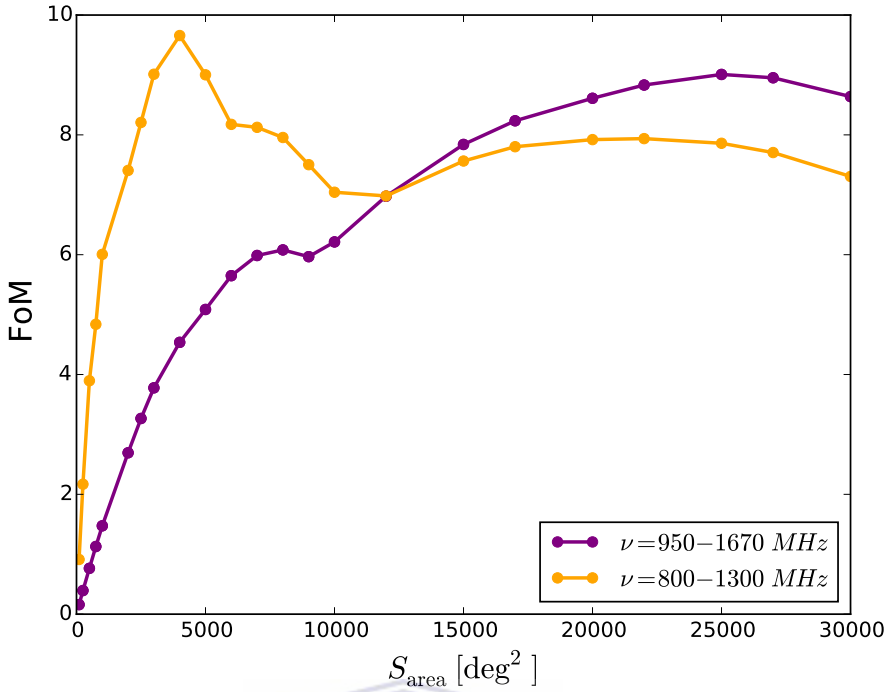

FIGURE 6.1: Dark energy figure of merit versus survey area for SKA1, with different frequency ranges. A Planck CMB prior has been included in the FOM calculation, but the optimal survey areas are the same even if the prior matrix is not included.

time, which could drive us to use the 10,000 hours over $\sim 25,000$ deg^2. A "full sky" survey would also have the advantage of probing wide-angle effects.

In this work, we have opted to assume a 5,000 deg^2 area for SKA1 using the 800-1300 MHz band. Although the current specifications for, say, SKA1-MID Band 2, specify a minimum frequency of 950 MHz, the numbers are still under review, so that it is acceptable to assume that such a change could happen. Using a smaller 5,000 deg^2 area for SKA1 allows the survey to be sample variance-limited in every redshift bin, which also brings advantages in terms of dealing with possible systematics (e.g. it will be easier to deal with a 5-sigma detection threshold). The final specifications that we assumed are summarised in Table 6.2.

For a given survey area, $S_{\rm area}$ we will need approximately $S_{\rm area}/(\theta_{\rm B})^2$ pointings. The time per pointing t_p is then related to the total integration time $t_{\rm tot}$ through

$$t_p = t_{\rm tot}\frac{(\theta_{\rm B})^2}{S_{\rm area}}. \tag{6.4}$$

Telescope	Redshift	Target freq.	Beam [deg^2]	$S_{\rm area}$ [deg^2]	$t_{\rm p}$ [hours]	$S^{\rm ref}_{\rm rms}$ [μJy]
MID+MeerKAT$^{(a)}$	0.0 – 0.78	1.0 GHz	0.88	5,000	1.76	152
SUR+ASKAP$^{(b)}$	0.0 – 1.19	1.3 GHz	18	5,000	36	175/140$^{(c)}$
SKA2$^{(d)}$	0.1 – 2.0	1.0 GHz	30	30,000	10	5.14

TABLE 6.2: Survey specifications. We assume a total observation time of 10,000 hours. For MID+MeerKAT, a modification of band 2 is assumed (800-1300 MHz) in order to achieve the target redshift range. Flux rms is calculated for a frequency interval of 10 kHz. Values for the beam and flux sensitivity are quoted at the target frequency.

Notes: (a) Beam and time per pointing (t_p) are assumed to change as $(1\,{\rm GHz}/\nu)^2$ across the band, and the flux rms is assumed to change as $\nu/(1\,{\rm GHz})$. **(b)** Values calculated at the PAF critical frequency. Below that frequency, the values are assumed constant. Above it, the beam and t_p are assumed to go as $1/\nu^2$, and the flux rms as ν. **(c)** The first value takes a weighted average of the SUR+ASKAP temperature while the second value assumes that the ASKAP PAFs are replaced to match SUR system temperature. **(d)** Indicative; the beam and flux rms are assumed constant across the band.

Since $(\theta_B)^2$ goes as $1/\nu^2$, this will increase the available integration time per pointing at lower frequencies; the flux rms is therefore proportional to the frequency, and so decreases for lower frequencies. The flux rms will remain constant below the critical frequency for PAFs, however, as explained above. In order to cover the required survey area, we assume that the mosaicking (how we pack the pointings/beams) is done at the highest frequency used for the HI survey; that is, the telescope pointings are packed side by side at the highest frequency. This ensures that the full survey area is covered at the highest frequency, but means that the beams will overlap at lower frequency, reducing the survey efficiency.

6.2 HI galaxy simulations

Crucial ingredients to any cosmological calculation using galaxy surveys are the galaxy number density as a function of redshift and detection threshold, and the corresponding bias with respect to the underlying dark matter distribution. Analytical calculations, though possible, would have to rely on some relation between the HI luminosity for a given galaxy and its host dark matter halo. As such, they might fail to emulate the actual distribution unless properly calibrated to full simulations, as the HI luminosity can depend on other factors besides the halo mass.

Instead, to calculate the HI galaxy number density and bias as a function of the survey rms sensitivity $S_{\rm rms}$ we have used the S^3-SAX simulation[3]. This simulation consists of a galaxy catalogue containing the position and several astrophysical properties for objects in a mock observing cone. It was produced by Obreschkow and Rawlings [98] by adding HI and CO properties to the galaxies obtained by De Lucia and Blaizot [99] through the post-processing of the Millennium dark matter simulation [100]. Since each galaxy in the simulation has associated with it a HI luminosity and line profile, as well as a redshift, we can proceed to calculate the number of galaxies that one could expect to detect with a given survey.

[3] http://s-cubed.physics.ox.ac.uk/s3_sax

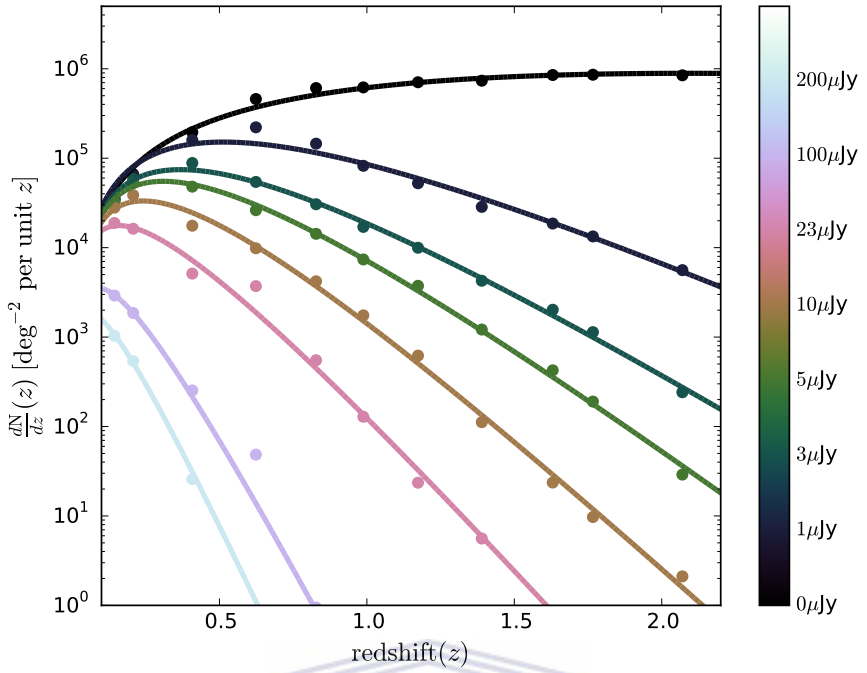

FIGURE 6.2: HI galaxy redshift distribution, dN/dz, calculated from simulations (solid circles) and the corresponding fitting function, (6.5). From top to bottom, the curves shown correspond to flux sensitivities $S_{\rm rms} = (0, 1, 3, 5, 10, 23, 100, 200)$ μJy (colour-coded according to the panel on the right).

6.2.1 HI galaxy number densities

Detection of a HI galaxy relies on the measurement of its corresponding HI line profile. This is usually set by the galaxy rotation curve and the inclination angle at which the galaxy is observed. The largest line width will be obtained if we observe the spiral galaxy edge-on and the smallest when it is observed face-on. The choice of detection algorithm is crucial to the success of any large HI galaxy survey campaign, as it will determine the total number of galaxies detected and how clean that detection is, i.e. how well spurious detections (due to RFI, for instance), can be rejected. As such, the expected galaxy number density for a given survey is not simply a function of the flux sensitivity.

In this Chapter, we take the simple approach that at least two points on the HI line are required to be measured in order for a galaxy to be detected. That is, the width of the line has to be larger than twice the assumed frequency resolution of the survey. The idea is to obtain information on the typical line double peak (double horn) expected from HI galaxies due to their rotation. This will remove

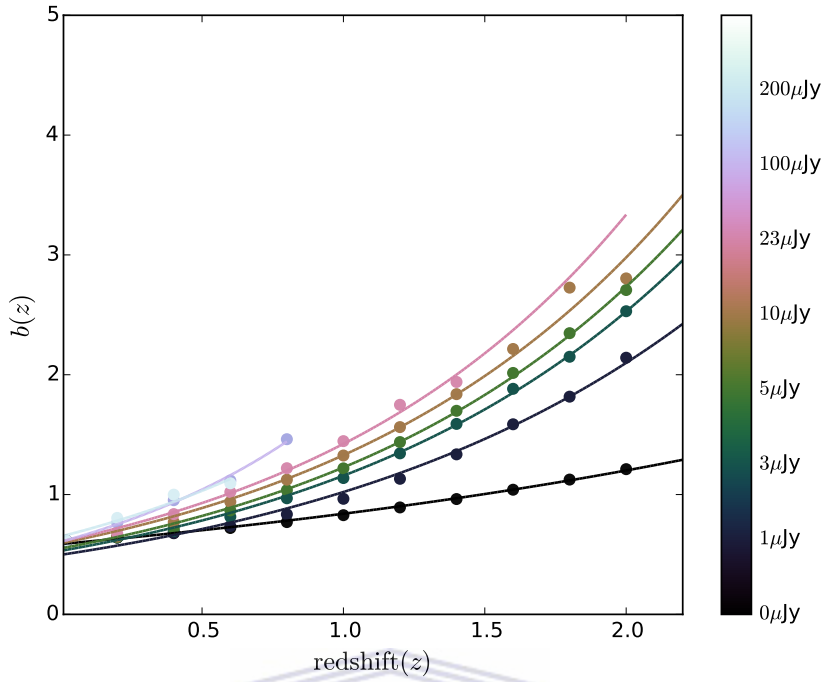

FIGURE 6.3: HI galaxy bias for different $S_{\rm rms}$. Note that bias values for high flux rms are uncertain. This has little impact, however, as shot noise will dominate at these sensitivities.

any galaxy that is seen face-on since it would just show as a narrow peak, which could be confused with RFI. Typical line profiles have widths of tens of kilometres per second, which is fine for the radio telescopes we are considering, as resolutions of 10 kHz are easily achievable (corresponding to ~ 2 km/s in the rest frame).

Using the S^3-SAX database, we applied the following "detection" pipeline:

1. Take $z_{\rm A}$ (the apparent redshift, including Doppler correction) from the database.
2. Set the spectral resolution to $\delta V = 2.1(1 + z_{\rm A})$ km/s, corresponding to a frequency resolution of 10 kHz (which was assumed for the sensitivity calculations).
3. Take $w_{\rm P}$ (the line width between the two horns of the HI line profile, corrected for galaxy inclination) from the database, and select only galaxies with $w_{\rm P} > 2\delta V$.

4. Take $v_{\rm HI}$ (the velocity-integrated line flux of the HI line) from the database and select only galaxies where the flux $= v_{\rm HI}/w_{\rm P} > N_{\rm cut} \times S_{\rm rms}/\sqrt{(w_{\rm P}/\delta V)}$. This corresponds to a detection threshold of $N_{\rm cut} \times 1\sigma$ for the HI line.

Note that $S_{\rm rms}$ is only the flux sensitivity – the survey flux cut will be a factor of several above that (usually five or ten, depending on the chosen threshold), although the actual value is not straightforward to specify since it depends on the detection algorithm.

In order to be as general as possible, we give results for a range of $S_{\rm rms}$ values so that a simple interpolation can be used if there is a change in the survey specifications. We use the formula of Obreschkow and Rawlings [98] to fit the dN/dz data points from S^3-SAX:

$$\frac{dN(z)/dz}{1\,\mathrm{deg}^2} = 10^{c_1} z^{c_2} \exp\left(-c_3 z\right), \tag{6.5}$$

where c_i are free parameters. Note that $\frac{dN}{dz}$ is the number of galaxies per square degree and per redshift interval. Fig. 6.2 shows the fitted curves and the simulated data points, and the fitted parameters are given in Table 6.4.

6.2.2 HI galaxy bias

To calculate the galaxy bias using the SAX simulation, two approaches were considered. The most direct was to put the extracted HI galaxies in a box according to their redshift and position, and to then calculate the galaxy power spectrum. The bias squared is then the ratio of this power spectrum to the dark matter one at a given scale k. Ideally we would target large scales, to avoid non-linearities and shot noise contamination. The initial box for the simulation was $500h^{-1}\,$Mpc, but this was further reduced along the line of sight to avoid cosmic evolution, which raises a problem for the bias extraction since linear modes with $k \lesssim 0.1h/\,$Mpc will be affected by cosmic variance.

The other option was to calculate the HI galaxy bias using the dark matter halo bias. To that end, we need to extract from the simulation box, at a given redshift, the dark matter halo hosting each HI galaxy above the target flux cut. The HI

bias can then be calculated using a weighted sum of the dark matter halo bias,

$$b_{\rm HI}(z, S_{\rm rms}) \approx \sum_i b(z, M_i) \frac{N_i}{N_{\rm tot}}, \tag{6.6}$$

where $b(z, M_i)$ is the halo bias for mass M_i [101], N_i is the number of halos in the box with mass M_i hosting HI galaxies above the detection threshold, and $N_{\rm tot} = \sum_i N_i$. This method is less affected by shot noise and does not suffer from the cosmic variance issues of the previous method. As such, in this Chapter we opted to calculate the bias following this second prescription. The data points obtained from the simulation are shown in Fig. 6.3 as a function of redshift for different $S_{\rm rms}$ sensitivities, and numerical values are given in Table. 6.3. We fit the simulated data using

$$b_{\rm HI}(z) = c_4 \exp(c_5 z), \tag{6.7}$$

and give the values of the best-fit parameters in Table 6.4.

The galaxies used in the bias calculation are contained in small volumes between $\sim (60/h)^3$ Mpc^3 (for $z \approx 0$) and $(175/h)^3$ Mpc^3 (for $z \approx 2$) due to the size of the redshift bins considered. Given the much larger volumes probed by an experiment like the SKA, one would expect to find a number of halos larger than those contained in the simulation boxes. However, this should only have an impact for large flux cuts, which are dominated by shot noise anyway and so will have little consequence in terms of cosmological constraints.

For halos of a given mass, there is significant variation in the HI mass of the galaxies residing within them. This implies that some galaxies with considerably higher HI masses than the average will be found. The number of halos rapidly decreases with halo mass and redshift, however, and so the majority of galaxies with high HI masses will be found in modest halos with modest bias. The fraction $M_{\rm HI}/M_{\rm halo}$ has also been shown to rapidly decrease with increasing halo mass for halos with masses above $10^{12} M_\odot$ [102], so even very massive halos are likely to have modest HI masses of the order of $10^9 M_\odot$ on average. This has the effect of introducing an effective upper limit to the bias at each redshift, which we estimated to be only slightly higher than the maximum values we were able to obtain from the simulation. As such, at each redshift one can assume that the bias remains constant for values of $S_{\rm rms}$ higher than the maximum that could be extracted from the simulation.

$S_{\rm rms}$ [μJy]	Redshift bin										
	0.02	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
0	0.614	0.641	0.678	0.721	0.770	0.828	0.892	0.963	1.04	1.12	1.21
1	0.614	0.642	0.680	0.738	0.836	0.965	1.13	1.34	1.59	1.82	2.14
3	0.614	0.643	0.695	0.815	0.969	1.14	1.34	1.59	1.88	2.15	2.53
5	0.614	0.644	0.718	0.868	1.04	1.22	1.44	1.70	2.02	2.34	2.71
6	0.614	0.645	0.730	0.886	1.06	1.25	1.47	1.73	2.07	2.46	2.86
7.3	0.614	0.646	0.745	0.907	1.09	1.28	1.50	1.78	2.12	2.55	2.86
10	0.614	0.650	0.770	0.940	1.12	1.33	1.57	1.84	2.22	2.73	2.80
23	0.614	0.675	0.837	1.021	1.22	1.45	1.75	1.95	–	–	–
40	0.614	0.706	0.879	1.08	1.25	1.48	1.75	2.01	–	–	–
70	0.614	0.742	0.924	1.11	1.13	1.61	1.86	–	–	–	–
100	0.615	0.764	0.953	1.12	1.46	–	–	–	–	–	–
150	0.614	0.787	0.982	1.11	–	–	–	–	–	–	–
200	0.614	0.805	0.999	1.094	–	–	–	–	–	–	–

TABLE 6.3: Bias values calculated in each redshift bin of the simulation, as a function of flux rms.

$S_{\rm rms}$	c_1	c_2	c_3	c_4	c_5
0.0	6.21	1.72	0.79	0.5874	0.3577
1.0	6.55	2.02	3.81	0.4968	0.7206
3.0	6.53	1.93	5.22	0.5302	0.7809
5.0	6.55	1.93	6.22	0.5504	0.8015
6.0	6.58	1.95	6.69	0.5466	0.8294
7.3	6.55	1.92	7.08	0.5623	0.8233
10	6.44	1.83	7.59	0.5928	0.8072
23	6.02	1.43	9.03	0.6069	0.8521
40	5.74	1.22	10.58	0.6280	0.8442
70	5.62	1.11	13.03	0.6094	0.9293
100	5.63	1.41	15.49	0.6052	1.0859
150	5.48	1.33	16.62	0.6365	0.9650
200	5.00	1.04	17.52	—	—

TABLE 6.4: Best-fit parameters for the number density and bias fitting functions, (6.5) and (6.7), for different flux limits. $S_{\rm rms}$ is measured in μJy.

For HI masses below $10^9 M_\odot$, locally-measured HI luminosity functions seem to imply many more galaxies than predicted by the simulation, suggesting that low mass galaxies are more HI rich than previously thought [102]. If this is the case, the bias will be smaller than predicted here for small values of $S_{\rm rms}$ (e.g. $\lesssim 1$ μJy). This result is subject to completeness uncertainty and cosmic variance, however, and is yet to be confirmed [103]. Conversely, the Damped Lyman-Alpha (DLA) observations (though model-dependent, and suffering from several uncertainties) are so far consistent with our predictions for the HI bias [104].

6.3 Cosmological Performance

In this section, we use Fisher forecasts to compare the ability of the proposed SKA HI galaxy surveys to constrain various cosmological quantities. Our focus is on the detection of the BAO feature, which we use as a figure of merit owing to its status as arguably the cleanest [105, 106] and most 'standard' observable targeted by cosmological large-scale structure surveys. In order to parameterize deviations from the standard model of DE ($w = -1$), we can parameterize the equation of state,

$$w(z) = w_0 + w_a \frac{z}{(1+z)}. \tag{6.8}$$

Constraints on the dark energy equation of state parameters, w_0 and w_a, are also presented. We take the Planck best-fit flat ΛCDM model [37] as our fiducial cosmology, with $h = 0.67$, $\Omega_{\rm cdm} = 0.267$, $\Omega_{\rm b} = 0.049$, $n_s = 0.962$, and $\sigma_8 = 0.834$.

6.3.1 SKA assumed sensitivities

Our forecasts follow the specifications given in Table 6.2, with the sensitivities obtained for a total observation time of 10,000 hours, and a survey area of 5,000 deg^2 for SKA1 and 30,000 deg^2 for SKA2. For each configuration we also considered 'optimistic' and 'pessimistic' variations, which are intended to bracket the possible range of flux sensitivities once HI modelling uncertainties and possible changes to the instrumental design are taken into account.

For SKA1, we take the flux rms at the target frequency of 1 GHz to be

$$S_{\rm rms}^{\rm ref} = 70/150/200\,\mu{\rm Jy}({\rm opt./ref./pess.}). \tag{6.9}$$

The optimistic scenario is roughly equivalent to taking the reference flux for SKA1-MID+MeerKAT (152 μJy), but assuming that the detection threshold would be set at the 5σ level. For SKA2, in lieu of any other information about its design we take the flux rms to be constant across the band, with

$$S_{\rm rms}^{\rm ref} = 3.0/5.4/23\,\mu{\rm Jy}({\rm opt./ref./pess.}). \tag{6.10}$$

The frequency/redshift interval for SKA1 is taken to be compatible with SKA1-SUR + ASKAP Band 2 or a modification of SKA1-MID + MeerKAT Band 2 as

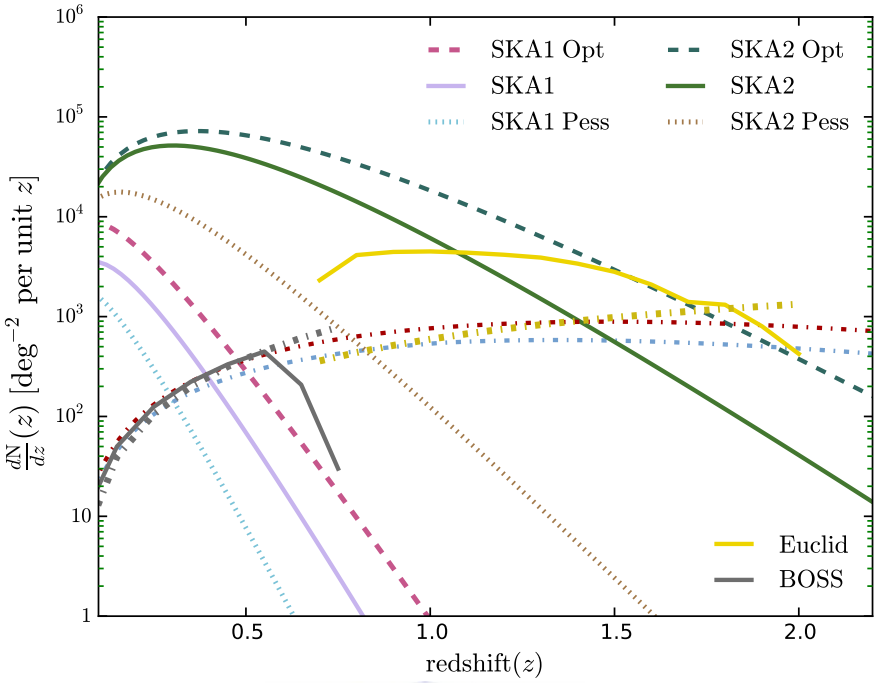

FIGURE 6.4: Number densities for the optimistic, reference, and pessimistic cases of SKA1 and 2, compared with Euclid. Dashed-dotted lines show the number density at which $n(z)b^2(z)P(z, k_{\rm max}) = 1$ for the various surveys, with $k_{\rm max} \approx 0.2\,h$/Mpc. When dN/dz is above this line, sample variance dominates the shot noise for all $k < k_{\rm max}$; the point at which it dips below the line is effectively the maximum redshift of the survey.

explained in section 6.1.2, such as 800-1300 MHz. We ignore Band 1, since above $z \sim 0.8$ one cannot detect enough galaxies for cosmological purposes with SKA1 sensitivities anyway. Note that both MID and SUR have similar sensitivities for the HI galaxy survey we are describing, although the current SUR band 2 definition is more optimal for this. For SKA2, we take the z range given in Table 6.2.

The number density and bias scale with frequency/redshift, as explained in Section 6.1.1. We take this into account by interpolating between the best-fit sensitivity curves shown in Figs. 6.2 and 6.3, as a function of redshift. The interpolation also allows us to factor in possible changes to the flux cut (galaxy detection threshold). For a given survey, the flux rms therefore scales as

$$S_{\rm rms} = S^{\rm ref}_{\rm rms} \frac{N_{\rm cut}}{10} \frac{\nu_{21}}{\nu_c} (1+z)^{-1}, \tag{6.11}$$

where ν_{21} is the rest frame frequency of the 21 cm line, $S^{\rm ref}_{\rm rms}$ is the reference flux sensitivity quoted in the tables, $N_{\rm cut}$ is the threshold above which galaxies are taken to be detected, in multiples of the noise rms, and ν_c is the target/critical

frequency at which $S_{\rm rms}^{\rm ref}$ was calculated (1.0 GHz for MID and 1.3 GHz for SUR). Note that for SUR (PAFs), the flux $S_{\rm rms}$ will remain constant for frequencies below ν_c.

As mentioned above, we assume that the reference experiment for SKA1 has non-PAF receivers (i.e. SKA1-MID + MeerKAT). For SKA2 we take the flux to be constant with redshift, also as discussed above. Then we correct for number density and bias by interpolating (6.5) and (6.7) using the values in Table 6.4. The resulting best-fit parameters for the number density and bias functions are given in Table 6.5. The redshift distribution for the target surveys is shown in Fig. 6.4, and compared to the limit below which the survey becomes shot noise-dominated.

6.3.2 Fisher forecasts

We use the Fisher forecasting technique to estimate how well the SKA surveys will be able to measure the BAO scale, and thus the various cosmological parameters. For a qualitative description of what the Fisher matrix is, see chapter 11 in [26]. The first step is to construct the Fisher matrix, which is derived from a Gaussian approximation of the likelihood, evaluated for a set of fiducial parameters. For a spectroscopic galaxy redshift survey, the Fisher matrix in a single redshift bin is

$$F_{ij} = \frac{1}{2} V_{\rm sur} \int \frac{d^3k}{(2\pi)^3} \frac{\partial \log P^T(\mathbf{k}, z)}{\partial \theta_i} \frac{\partial \log P^T(\mathbf{k}, z)}{\partial \theta_j}, \tag{6.12}$$

where $\{\theta_i\}$ are the cosmological parameters of interest, and $V_{\rm sur}$ is the comoving volume of the redshift bin given by,

$$V_{\rm sur} = \left(\frac{\pi}{180}\right)^2 S_{\rm area} \int_{z_{\rm min}}^{z_{\rm max}} (1+z)^2 D_A^2(z) \frac{c}{H(z)} dz \tag{6.13}$$

in units of Mpc^3/h^3. This definition neglect the redshift evolution within the bin. while the subscript in $S_{\rm area}$ represents the survey area.

The total variance of the measured fluctuations in the galaxy distribution is

$$P^T(\mathbf{k}, z) = P(\mathbf{k}, z) + 1/n(z), \tag{6.14}$$

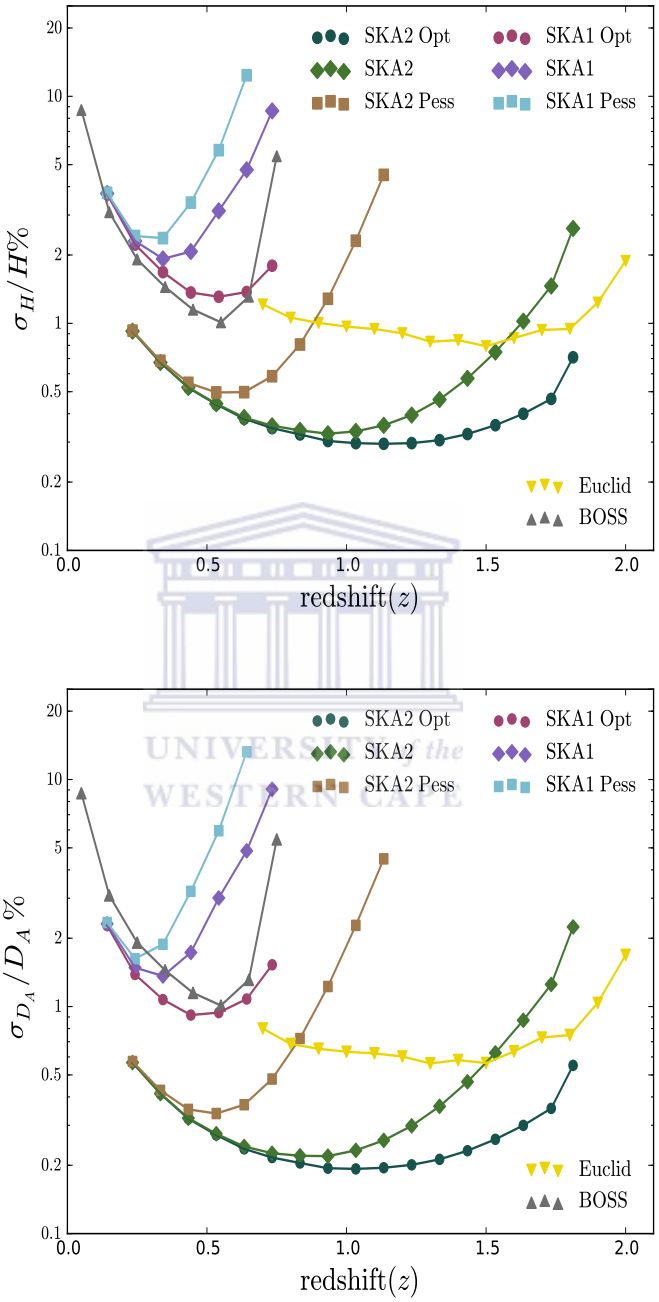

FIGURE 6.5: Forecast fractional errors on the expansion rate, $H(z)$, and angular diameter distance, $D_A(z)$, from BAO measurements with the various surveys. The redshift binning is fixed at $\Delta z = 0.1$ for all experiments.

where $P(\mathbf{k}, z)$ is the redshift-space galaxy power spectrum, and $1/n(z)$ is the inverse of the galaxy number density, which acts as a shot noise term. Only the power spectrum depends on the cosmological parameters, so we can write

$$F_{ij} = \frac{V_{\rm sur}}{8\pi^2} \int_{-1}^{1} d\mu \int_{k_{\rm min}}^{k_{\rm max}} k^2 dk \left(\frac{nP}{1+nP} \right)^2 \frac{\partial \log P}{\partial \theta_i} \frac{\partial \log P}{\partial \theta_j},$$

where $\mu = \cos\theta$ is the cosine of the angle between the line of sight and the Fourier mode $\mathbf{k}$. We fix the lower integration limit to $k_{\rm min} = 10^{-3} h\,{\rm Mpc}^{-1}$, and discard all information from modes beyond a non-linear cutoff scale,

$$k_{\rm max} = k_{\rm NL,0}\,(1+z)^{2/(2+n_s)}\,, \tag{6.15}$$

where $k_{\rm NL,0} \simeq 0.2h\,{\rm Mpc}^{-1}$ [107].

We adopt a simplified 'wiggles-only' approach to deriving BAO constraints, where only derivatives of the (Fourier-space) BAO feature are included in the Fisher matrix calculation. We first calculate the full (isotropic) power spectrum, $P(k, z)$, for the fiducial cosmology using CAMB (see Appendix C.1 for further details) [108], and then separate it into smooth and wiggles-only components such that [109]

$$P(k, z) = [1 + f_{\rm BAO}(k)]\, P_{\rm smooth}(k, z). \tag{6.16}$$

If the actual cosmology differs from the fiducial cosmology, the observed wavenumber, k, of a feature in the isotropic power spectrum will be shifted according to [110]

$$k = \sqrt{k_\perp^2 (D_A^{\rm (fid.)}/D_A)^2 + k_\parallel^2 (H/H^{\rm (fid.)})^2}, \tag{6.17}$$

where $D_A^{\rm (fid.)}$ and $H^{\rm (fid.)}$ are the values of the angular diameter distance and Hubble rate given the fiducial model, respectively. Since our aim is to provide a consistent comparison of the performance of various surveys, rather than to give high-precision forecasts on a large set of parameters, we make a number of simplifying assumptions: we ignore redshift-space distortions, non-linear effects, and uncertainty in both the bias and acoustic scale, and assume that the cosmological information encoded by the BAO feature comes entirely from the shift in k. We can then write

$$\frac{\partial \log P}{\partial \theta} \approx [1 + f_{\rm BAO}(k)]^{-1} \frac{df_{\rm BAO}}{dk} \frac{dk}{d\theta} \tag{6.18}$$

where, Appendix C.1 discusses further details on how to calculate this derivative.

		$S_{\rm rms}^{\rm ref}$	c_1	c_2	c_3	c_4	c_5
SKA1	opt.	70	5.253	0.901	7.536	0.628	0.819
	ref.	150	5.438	1.332	11.837	0.625	0.881
	pess.	200	5.385	1.278	14.409	0.646	0.896
SKA2	opt.	3.0	6.532	1.932	5.224	0.530	0.781
	ref.	5.4	6.555	1.932	6.378	0.549	0.812
	pess.	23.0	6.020	1.430	9.028	0.607	0.852

TABLE 6.5: Fitted parameters for the galaxy number density and bias, for the frequency-corrected $S_{\rm rms}$ of the various experiments. The flux rms at the reference frequency, $S_{\rm rms}^{\rm ref}$, is in μJy, while the fitted coefficients are dimensionless.

We work in terms of the parameters $\theta \in \{\log D_A, \log H\}$, so that the Fisher integral factorises into a simple 2×2 matrix of analytic angular integrals multiplied by the (scalar) k integral. This calculation includes the cross-correlation between D_A and H (Further details in C.2). Because we are neglecting a number of nuisance parameters and other effects, our forecasts could be interpreted as somewhat optimistic – although note that we are using only the information encoded in the BAO wiggles, which is quite insensitive to such effects (e.g. [111]).

Using the definition of the expansion rate, H, given in (2.15), and the angular diameter distance, D_A, for a flat universe given in (2.23). We can project the constraints on D_A and H to various basic cosmological parameters. Given $H_0 = 100h$ km/s/Mpc, and adopt the commonly used parametrisation of the dark energy equation of state, given in (6.8). The full set of parameters that we consider is then

$$\theta' = \{w_0, w_a, \Omega_{\rm cdm}, \Omega_{\rm b}, \Omega_K, h\}, \tag{6.19}$$

with the Fisher matrix found by projecting from the original 2×2 matrix and summing over redshift bins,

$$F'_{\alpha\beta} = \sum_{ij,n} \frac{\partial\theta_i}{\partial\theta'_\alpha}\frac{\partial\theta_j}{\partial\theta'_\beta}\bigg|_{z_n} F_{ij}(z_n). \tag{6.20}$$

Finally, we add the Planck CMB prior Fisher matrix from Amendola et al. [35] to represent the high-z constraints that will be available. See Appendix C.3 for detailed analytical derivatives of D_A and H with respect to the cosmological parameters of interest and how we combine Planck CMB prior to the SKA Fisher matrix.

6.3.3 Comparison with previous results and future experiments

The results of our Fisher forecasts are shown in Figs. 6.5 – 6.7 and Table 6.6. For comparison, we have also included forecasts for

(a) a future optical/near-infrared Hα galaxy survey with similar specifications to Euclid, using the number counts and bias model for the reference case described in Amendola et al. [35],

(b) the BOSS LRG galaxy survey, using the specifications in [112], with a total of 1.5 million galaxies out to $z \lesssim 0.75$, and with a bias of $b \approx 2$.

As can be seen from Fig. 6.5, an SKA1 galaxy survey will offer – at best – only slight improvements over existing experiments at low redshift ($z \lesssim 0.7$). Indeed, from Table 6.6 it can be seen that BOSS outperforms SKA1, although this is predominantly due to the larger assumed bias.

SKA1 should still significantly improve the cosmological constraints at low redshift, however, for the simple reason that it will cover a mostly independent survey area to existing experiments like BOSS and WiggleZ, thus increasing the total volume surveyed overall.

The picture is considerably more interesting for SKA2, which will be capable of performing a sample variance-limited survey over ¾ of the sky from $0.3 \lesssim z \lesssim 1.5$ in the reference case (increasing to $z \approx 2.0$ in the optimistic case). This will constitute the final word in spectroscopic redshift surveys in this redshift range, as there is little prospect of covering a greater survey area in the future. As shown in Fig. 6.5, the SKA2 reference case is forecast to provide measurements of $H(z)$ and $D_A(z)$ to better than 0.5% and 0.3% precision respectively, out to $z \approx 1.3$. This significantly outperforms future Hα surveys such as Euclid, which has half

the survey area (and approximately double the errors) over the same range. This is contingent on performing at least as well as the reference case, however; the pessimistic case would only be competitive with Euclid out to $z \simeq 0.8$.

Even in the reference case, measurements above $z \sim 1.5$ would be difficult, as the HI source density falls too low (contrary to what has been forecast for Euclid, for instance). Note that the HI source density at $z > 1$ flattens as $S_{\rm rms} \to 0$, however (Fig. 6.2), suggesting that a sufficiently deep HI survey could produce precision constraints out to substantially higher redshift, at least in principle.

Figs. 6.6 and 6.7 show forecasts for the equation of state and spatial curvature parameters for the reference cases of the various surveys. These were derived by projecting the (H, D_A) Fisher matrices, including the cross-correlation terms, to the parameter set described in section 6.3.2, and then adding a Planck CMB Fisher matrix prior. Corresponding marginal errors are given in Table 6.6 for the same parameters, for all cases. As before, SKA2 outperforms Euclid by a factor of around 2, reflecting its having double the survey area, as well as a further improvement due to its 4 additional redshift bins below Euclid's minimum redshift. In terms of the dark energy figure of merit, defined as [113, 114]

$$\mathrm{FOM} = 1/\sqrt{\det(F^{-1}|_{w_0,w_a})} \tag{6.21}$$

(equivalent to the inverse of the area of the 1σ (w_0, w_a) ellipse), the SKA2 reference case performs around $4\times$ better than Euclid, and some $60\times$ better than SKA1 (opt. case).

Note that our forecasts are only intended for comparison of the various surveys. In reality, systematic effects (radio interference, the efficiency of source extraction algorithms, contamination by foreground emission, non-linearities, modelling errors etc.) should further affect the survey performance. We have concentrated exclusively on the BAO wiggles in our forecasts, however, which are hoped to give constraints more insensitive to such systematics. On the other hand, other observables (e.g. redshift space distortions) can also be measured, significantly improving the constraints on some parameters.

Leaving these issues aside, our calculations predict that the SKA2 (reference case) survey will be sample variance-limited over a significant fraction of the redshift range that is important for dark energy (i.e. $z \lesssim 2$). As a result, it can come remarkably close to what would be possible with a 'perfect' noise-free HI survey

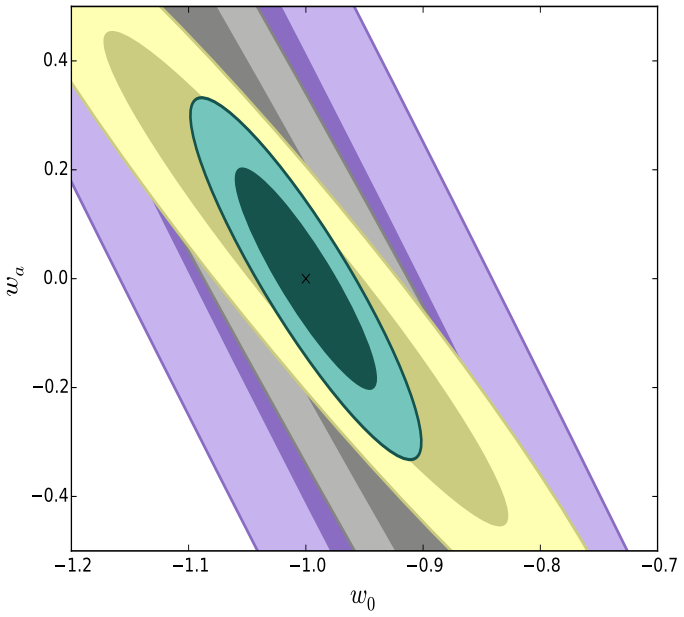

FIGURE 6.6: Forecast constraints (1 and 2σ contours) on the dark energy equation of state parameters, w_0 and w_a, for the reference cases of both SKA1 (purple, largest), BOSS (grey, second largest) and SKA2 (green, smallest), compared to a Euclid-like $H\alpha$ galaxy survey (yellow, intermediate). A Planck CMB prior has been included for all three experiments.

over the same area (represented by the $S_{\rm rms} = 0$ entry in Table 6.6); the 1σ errors on w_0 and w_a are only $\sim 1.5\times$ larger than their 'noise-free' values, for example, and even in the pessimistic case they are still only $\sim 3\times$ larger.

Abdalla et al. [93] also investigated how well the SKA can measure the BAO scale and dark energy parameters. Our work differs from theirs in various aspects. They used an analytical HI evolution model relying on prior knowledge of the star formation rate (SFR) and overall mass density of neutral hydrogen at a specific redshift, functions which depend on fitting formulas. We use a more realistic simulation to estimate the number counts, which we consider to be an improvement as our simulation relies on more physical properties, making our predictions more reliable. The difference between the two sets of results can be seen by comparing the number counts (Fig. 6.2). Although the number count estimate at high redshift is consistent in both cases, where they have a sharp curve as a function of redshift, ours decreases more gradually[4]. The second important difference is that while they assumed $b = 1$, the bias in our simulation was a function of redshift, and was dependant on the frequency-corrected $S_{\rm rms}$ value (see Fig. 6.3).

[4]See Fig. 3 of Abdalla et al. [93].

	$S_{\rm rms}$ [μJy]	σ_{w_0}	σ_{w_a}	$\sigma_{\Omega_{\rm cdm}}$	$\sigma_{\Omega_{\rm b}}$	σ_{Ω_K}	σ_h	FOM
SKA1 (5,000 deg^2)	70	0.258	0.972	0.00796	0.0476	0.00941	0.00943	28
	150	0.414	1.76	0.0115	0.0861	0.0109	0.0109	9
	200	0.582	2.66	0.0188	0.130	0.0122	0.0122	4
SKA2 (30,000 deg^2)	3.0	0.0328	0.116	0.00328	0.000158	0.00338	0.0034	547
	5.4	0.0407	0.137	0.00357	0.000169	0.00365	0.0040	426
	23.0	0.0912	0.322	0.00464	0.000224	0.00432	0.0070	160
	0.0	0.0273	0.100	0.00288	0.000148	0.00299	0.0029	699
Euclid (15,000 deg^2)	—	0.114	0.299	0.00907	0.000442	0.00944	0.0130	106
BOSS (10,000 deg^2)	—	0.2416	0.9429	0.00810	0.0003958	0.0069091	0.01564	30

TABLE 6.6: Forecast 1σ marginal errors and dark energy FOM for the various SKA1 and 2 reference experiments. A Euclid-like Hα survey, BOSS forecasts, and a noise-free SKA2 configuration, are shown for comparison. (Planck CMB priors have been included in all cases.)

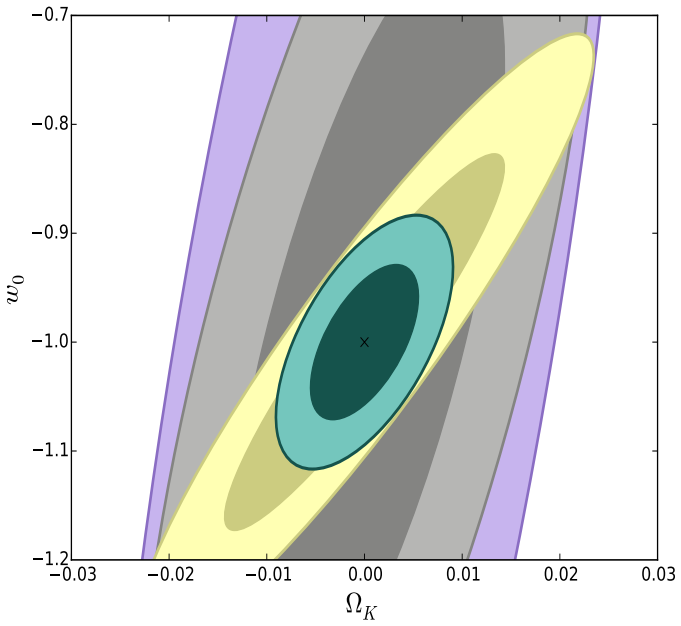

FIGURE 6.7: Forecast constraints (1 and 2σ) on w_0 and Ω_K for the references cases of SKA1 (purple, largest), BOSS (grey, second largest) and SKA2 (green, smallest), compared with a Euclid-like Hα galaxy survey (yellow, intermediate). A Planck CMB prior has been included for all experiments.

6.4 Conclusions

In this Chapter, we analysed the potential for producing precision cosmological constraints with future HI galaxy surveys using the SKA telescope. HI is abundant in the late Universe, making it a prime candidate for detecting large numbers of galaxies which can then be used to trace the underlying dark matter distribution. In particular, modern radio receivers have the high sensitivity and bandwidth to detect the HI emission over an extremely wide redshift range, making it possible to trace the cosmological matter distribution over unprecedentedly large volumes.

Our analysis uses up-to-date simulations to calculate the expected galaxy number density and bias as a function of redshift and flux sensitivity. We have also provided a set of fitting formulas, (6.5) and (6.7), that can be used to convert these results into number density and bias functions for specific experiments, such as the SKA or any other array.

One of our main conclusions is that although SKA1 will already detect a large number of HI galaxies, it will only be useful for cosmological applications up to $z \sim 0.7$ due to the sharp decline of the detected HI galaxy number density with redshift.

This means that first, for a cosmological HI galaxy survey with SKA Phase 1, frequencies above $\sim$ 1 GHz should be enough (i.e. Band 2). Moreover, these arrays will lack the sensitivity to detect enough galaxies to produce constraints that are competitive with contemporary optical and near-infrared galaxy surveys in the early 2020s.

On the other hand, the full SKA will push the HI galaxy detection limit up to $z \sim 2.0$ (requiring a larger band down to 500 MHz), and over the full visible sky, making it a prime cosmological survey instrument. Its sensitivity will allow us to produce an immense galaxy redshift survey over almost ¾ of the sky, surpassing all other planned surveys in terms of precision measurements of the BAO. This should allow it to pin down the equation of state of dark energy with unprecedented precision. Note that, while we have concentrated on the BAO as the most robust large-scale structure observable, redshift space distortions and even the overall shape of the power spectrum contain a great deal of extra information that can also be used to constrain dark energy. In this sense, the forecasts in this work represent conservative estimates of the cosmological constraints that can be achieved with the SKA (although recall that we optimistically neglected several nuisance parameters in our forecasts).

Note that the SKA will also be able to produce competitive cosmological constraints using the HI intensity mapping (IM) technique [109]. IM surveys are sensitive to large-scale fluctuations in the HI brightness temperature, which can be used to recover information about the cosmological matter distribution without requiring high signal-to-noise detections of many individual sources. This means that the flux sensitivity of a telescope is used more efficiently in IM mode, as none of the detected signal need be discarded due to thresholding. Indeed, an IM survey with Phase 1 of the SKA will produce a dark energy figure of merit of at least half that of Euclid+BOSS [115], in stark contrast to the underwhelming performance predicted for a Phase 1 HI galaxy survey (Section 6.3.3).

This is not to say that galaxy surveys should be deprecated in favour of intensity mapping, however. Of the two, galaxy surveys are certainly the more tried-and-tested (and thus less risky) method – the first large cosmological IM surveys are still a few years away, and a number of significant technical challenges (e.g. foreground contamination, polarisation leakage, autocorrelation calibration) remain to be solved [116]. In fact, a galaxy survey may be the preferred choice for a dark energy survey with Phase 2 of the SKA, as it will likely be easier to approach

the sample variance limit for $z \lesssim 2$. This is because IM surveys are subject to a number of effective noise contributions separate from the instrumental noise (e.g. residual foregrounds and calibration errors) that are difficult to reduce to a negligible level, while galaxy surveys do not suffer from such residuals. Realistic simulations informed by experience with Phase 1 surveys will be needed to confirm this, however.

Chapter 7

Conclusion

The lack of theoretical and physical models to explain the dark energy problem led us to think that the data from probes of dark energy should lead us to the correct model or at least exclude the models do not agree with the data. In Chapter 4 we described a series of null tests that can be applied to SNIa data to determine the consistency of observations with flat ΛCDM model without the need to parameterize $w(z)$.

The tests require that the luminosity distance, D, and its derivatives (D' and D'') to be reconstructed using a model independent techniques. Thus, we used the GP method to reconstruct those functions. The reconstructed and smoothed D, D', D'' and D''' were used to perform the diagnostics $\mathcal{O}_m^{(1)}$, $\mathcal{O}_m^{(2)}$, $\mathcal{O}_K^{(2)}$, $\mathcal{L}^{(1)}$ and $\mathcal{L}^{(2)}$.

These diagnostics were applied to SNIa Union 2.1 data set and the results were consistent with flat ΛCDM model as shown in Figs. 4.2 and 4.3. We also simulated DES supernovae survey data. We simulated the data using two different models, flat ΛCDM and evolving w. Using large number of simulated DES data, over 4000 SNIa, the diagnostic $\mathcal{O}_m^{(1)}$ has the ability to distinguish between different cosmological models, and correctly identify deviation from ΛCDM as shown in Figs. 4.4 and 4.5. While the other null tests where the curvature was restricted, $\mathcal{O}_m^{(2)}$ and $\mathcal{O}_K^{(2)}$ showed unrealistic reconstruction of the input values of Ω_m and Ω_K suggesting that the degeneracy between w and the density parameters need to be broken by other data from different probes. The $\mathcal{L}^{(2)}$ diagnostic shows huge uncertainty which made it impossible to distinguish between models, suggesting that the dependence of this test on an accurate reconstruction of D''' is an issue.

The tests we introduced in Chapter 4 need further refinement, also the results suggest that we need a huge amount of SNIa data to reconstruct the model values correctly. Fortunately, future surveys such as DES, promising large amount of data that will help to preform these tests accurately.

Similarly, in Chapter 5 we developed and analyzed consistency tests for ΛCDM model in terms of $H(z)$ instead of $D(z)$ function. We applied $H(z)$ data to those null tests, by combining cosmic chronometers and radial BAO data. The $H(z)$ null tests require less derivatives than the tests in terms of D, therefore the reconstruction of the tests were easier and more accurate. In fact these null tests require an order of magnitude less $H(z)$ data point than the $D(z)$ null tests required of SNIa data points. We expect more from these tests specially that future surveys such as BOSS [88], EUCLID [89] and SKA [90], will provide $H(z)$ data from BAO measurements of increasing number and precision.

Due to the urge of knowing how well the future surveys will contribute to solve the quest of the expansion history of the universe thus, we forecast the potential of the SKA HI galaxy surveys which will indeed revolutionize our knowledge. We forecast for different stages of the SKA telescope. Our investigation was based on the simulations of the galaxy number density and the bias as a function of redshift and flux sensitivity. We produced fitting formulas for these quantities which can be used for other similar galaxy surveys require an estimate of galaxy number density or bias (see (6.5) and (6.7)).

The SKA1 will detect a large number of HI galaxies up to $z \sim 0.7$. This indicates that for HI galaxy surveys with SKA1, frequencies above ~ 1 GHz will be efficient, see Fig. 6.5. These results suggest the SKA1 will only be complementary to the current galaxy surveys at best. Although the full SKA will push the HI galaxy detection to $z \sim 2$, and with coverage of 3/4 of the sky, SKA will be a prime cosmological HI survey instrument. This should allow for high precision BAO measurements and thus high accuracy in recovering the dark energy parameters.

Appendix A

Gaussian Processes

For a data set $\{(z_i, y_i)|i = 1, \ldots, n\}$, where $\boldsymbol{Z}$ represents the training points z_i, i.e. the locations of the observations, we want to reconstruct the function that describes the data at the test input points $\boldsymbol{Z}^*$.

A Gaussian Process is a distribution over functions and is thus a generalization of a Gaussian distribution. It is defined by the mean $\mu(z)$ and covariance $k(z, \tilde{z})$:

$$f(z) \sim \mathcal{GP}\left(\mu(z), k(z, \tilde{z})\right) . \tag{A.1}$$

At each z_i, the value $f(z_i)$ is drawn from a Gaussian distribution with mean $\mu(z_i)$ and variance $k(z_i, z_i)$. $f(z_i)$ and $f(z_j)$ are correlated by the covariance function $k(z_i, z_j)$.

Choosing the covariance function is one of the main points for achieving satisfactory results. The squared exponential is a general purpose covariance function, which we use throughout in Chapter 5:

$$k(z_i, z_j) = \sigma_f^2 \exp\left[-\frac{(z_i - z_j)^2}{2\ell^2}\right] . \tag{A.2}$$

The 'hyperparameters' are σ_f (signal variance) and ℓ (characteristic length scale). ℓ can be thought of as the distance moved in input space before the function value changes significantly. σ_f describes the typical change in y-direction. In contrast to actual parameters, they do not specify the exact form of a function, but describe typical changes in the function value.

For $\boldsymbol{Z}^*$, the covariance matrix is given by $[K(\boldsymbol{Z}^*, \boldsymbol{Z}^*)]_{ij} = k(z_i^*, z_j^*)$. Then the vector $\boldsymbol{f}^*$ with entries $f(z_i^*)$ is drawn from a Gaussian distribution:

$$\boldsymbol{f}^* \sim \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{Z}^*), K(\boldsymbol{Z}^*, \boldsymbol{Z}^*)) \,. \tag{A.3}$$

This can be considered as a prior for the distribution of $\boldsymbol{f}^*$. One needs to add observational information to obtain the posterior distribution.

The observational data have a covariance matrix C. For uncorrelated data, C is a diagonal matrix with entries σ_i. The combined distribution for $\boldsymbol{f}^*$ and the observations $\boldsymbol{y}$ is given by:

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{f}^* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu^*} \end{bmatrix}, \begin{bmatrix} K(\boldsymbol{Z}, \boldsymbol{Z}) + C & K(\boldsymbol{Z}, \boldsymbol{Z}^*) \\ K(\boldsymbol{Z}^*, \boldsymbol{Z}) & K(\boldsymbol{Z}^*, \boldsymbol{Z}^*) \end{bmatrix} \right) \tag{A.4}$$

While the values of $\boldsymbol{y}$ are already known, we want to reconstruct $\boldsymbol{f}^*$. Thus, we are interested in the conditional distribution

$$\boldsymbol{f}^* | \boldsymbol{Z}^*, \boldsymbol{Z}, \boldsymbol{y} \sim \mathcal{N}\left(\bar{\boldsymbol{f}}^*, \mathrm{cov}(\boldsymbol{f}^*)\right) \,, \tag{A.5}$$

where

$$\bar{\boldsymbol{f}}^* = \boldsymbol{\mu^*} + K(\boldsymbol{Z}^*, \boldsymbol{Z}) \left[K(\boldsymbol{Z}, \boldsymbol{Z}) + C\right]^{-1} (\boldsymbol{y} - \boldsymbol{\mu}) \tag{A.6}$$

$$\begin{aligned} \mathrm{cov}(\boldsymbol{f}^*) &= K(\boldsymbol{Z}^*, \boldsymbol{Z}^*) \\ &- K(\boldsymbol{Z}^*, \boldsymbol{Z}) \left[K(\boldsymbol{Z}, \boldsymbol{Z}) + C\right]^{-1} K(\boldsymbol{Z}, \boldsymbol{Z}^*), \end{aligned} \tag{A.7}$$

are the mean and covariance of $\boldsymbol{f}^*$, respectively. The variance of $\boldsymbol{f}^*$ is simply the diagonal of $\mathrm{cov}(\boldsymbol{f}^*)$. Equation A.5 is the posterior distribution of the function given the data and the prior in (A.3).

In order to use this equation, we need to know the values of the hyperparameters σ_f and ℓ. They can be trained by maximizing the log marginal likelihood:

$$\begin{aligned} \ln \mathcal{L} &= \ln p(\boldsymbol{y} | \boldsymbol{Z}, \sigma_f, \ell) \\ &= -\frac{1}{2} (\boldsymbol{y} - \boldsymbol{\mu})^\top \left[K(\boldsymbol{Z}, \boldsymbol{Z}) + C\right]^{-1} (\boldsymbol{y} - \boldsymbol{\mu}) \\ &\quad - \frac{1}{2} \ln \left|K(\boldsymbol{Z}, \boldsymbol{Z}) + C\right| - \frac{n}{2} \ln 2\pi \,. \end{aligned} \tag{A.8}$$

Note that this likelihood only depends on the observational data, but is independent of the locations $\boldsymbol{Z}^*$ where the function is to be reconstructed.

Derivatives of the function can be reconstructed in a similar way. For the first derivative, the conditional distribution is given by [45]:

$$\boldsymbol{f}^{*\prime}|\boldsymbol{Z}^*, \boldsymbol{Z}, y \sim \mathcal{N}\left(\bar{\boldsymbol{f}}^{*\prime}, \mathrm{cov}(\boldsymbol{f}^{*\prime})\right) \ , \tag{A.9}$$

where

$$\bar{\boldsymbol{f}}^{*\prime} = \boldsymbol{\mu}^{*\prime} + K^{\prime}(\boldsymbol{Z}^*, \boldsymbol{Z})\left[K(\boldsymbol{Z}, \boldsymbol{Z}) + C\right]^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) \tag{A.10}$$

$$\mathrm{cov}(\boldsymbol{f}^{*\prime}) = K^{\prime\prime}(\boldsymbol{Z}^*, \boldsymbol{Z}^*) - K^{\prime}(\boldsymbol{Z}^*, \boldsymbol{Z})\left[K(\boldsymbol{Z}, \boldsymbol{Z}) + C\right]^{-1} K^{\prime}(\boldsymbol{Z}, \boldsymbol{Z}^*). \tag{A.11}$$

For the covariance matrices, we use the notation:

$$[K^{\prime}(\boldsymbol{Z}, \boldsymbol{Z}^*)]_{ij} = \frac{\partial k(z_i, z_j^*)}{\partial z_j^*} \tag{A.12}$$

$$[K^{\prime\prime}(\boldsymbol{Z}^*, \boldsymbol{Z}^*)]_{ij} = \frac{\partial^2 k(z_i^*, z_j^*)}{\partial z_i^* \, \partial z_j^*} \ . \tag{A.13}$$

$K^{\prime}(\boldsymbol{Z}^*, \boldsymbol{Z})$ is the transpose of $K^{\prime}(\boldsymbol{Z}, \boldsymbol{Z}^*)$.

To calculate a function $g(f, f^{\prime})$ which depends on f and $f^{\prime}$, we also need to know the covariances between $f^* = f(z^*)$ and $f^{*\prime} = f^{\prime}(z^*)$ at each point z^* where g is to be reconstructed. This covariance is given by:

$$\mathrm{cov}(f^*, f^{*\prime}) = \left.\frac{\partial k(z^*, \tilde{z})}{\partial \tilde{z}}\right|_{z^*} - K^{\prime}(z^*, \boldsymbol{Z})\left[K(\boldsymbol{Z}, \boldsymbol{Z}) + C\right]^{-1} K(\boldsymbol{Z}, z^*). \tag{A.14}$$

$g^* = g(z^*)$ is then determined by Monte Carlo sampling, where in each step f^* and $f^{*\prime}$ are drawn from a multivariate normal distribution:

$$\begin{bmatrix} f^* \\ f^{*\prime} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \bar{f}^* \\ \bar{f}^{*\prime} \end{bmatrix}, \begin{bmatrix} \mathrm{var}(f^*) & \mathrm{cov}(f^*, f^{*\prime}) \\ \mathrm{cov}(f^*, f^{*\prime}) & \mathrm{var}(f^{*\prime}) \end{bmatrix}\right) \ . \tag{A.15}$$

Appendix B

Derivation of the consistency tests of $w(z)$

$w(z)$ in terms of $D(z)$ and derivatives

To get a formula for $w(z)$ in terms of $D(z)$, we can use the general definition of $w(z)$,

$$w(z) = \frac{\rho_{de}(z)}{p_{de}(z)}. \tag{B.1}$$

Equation B.1 shows the dark energy equation of state. Using (2.16) and (2.17), $H(z)$ can be expressed in this form [**?**]:

$$H(z) = \left(\frac{dD_c}{dz}\right)^{-1} \sqrt{1 - D_c^2(z) H_0^2 \Omega_K}, \tag{B.2}$$

where $D_c(z)$ is the transverse co-moving distance introduced in (2.17).

$$\begin{aligned} \rho_{de}(z) \quad &= \tfrac{3}{K}\left[\tfrac{1}{(dD_c/dz)^2} + H_0^2\Omega_K\left\{(1+z)^2 - \tfrac{D_c^2}{(dD_c/dz)^2}\right\}\right] \\ &\quad -\Omega_m(1+z)^3 \tfrac{1}{(dD_c(z)/dz)^2} \end{aligned} \tag{B.3}$$

and

$$\begin{aligned} p_{de}(z) \qquad &= \tfrac{1}{k}\Big[- \tfrac{3}{(dD_c/dz)^2} + (1+z)\tfrac{d}{dz}\tfrac{1}{(dD_c/dz)^2} \\ &\quad -H_0^2\Omega_K\left\{(1+z)^2 - \tfrac{3D_c^2}{(dD_c/dz)^2} + (1+z)\tfrac{d}{dz}\tfrac{D_c^2}{(dD_c/dz)^2}\right\}\Big]. \end{aligned} \tag{B.4}$$

We also have the luminosity distance $D_L(z)$ in term of $D_c(z)$,

$$D_L(z) = a_0(1+z)f(\chi) \equiv (1+z)D_c(z). \tag{B.5}$$

Then the dimensionless luminosity distance $D_L(z)$ is

$$D(z) = H_0(1+z)D_c(z). \tag{B.6}$$

The first derivative of $D_c(z)$ with respect to z,

$$\frac{dD(z)}{dz} = H_0\frac{dD_c(z)}{dz}. \tag{B.7}$$

By substituting the expressions of $D(z)$ introduced in (B.6) and $dD(z)/dz$ introduced in (B.7) into the density and the pressure equations, (B.3) and (B.4), respectively, we get

$$\rho(z) = 3\frac{H_0^2}{D'^2k}\left[(1+D^2\Omega_K) + (1+z)^2D'^2\Omega_K + \Omega_m(1+z)^3D'^2\right], \tag{B.8}$$

and

$$\begin{aligned} p(z) = {} & \frac{1}{D'}\frac{H_0^2}{D'^2k}[2(1+z)(\Omega_K D^2+1)D''] \\ & - [3(1+D^2\Omega_K) + 2(1+z)\Omega_K DD' + (1+z)^2\Omega_K D'^2]D'. \end{aligned} \tag{B.9}$$

By dividing (B.8) by (B.9), we get the expression for $w(z)$ in term of $D(z)$ given in (4.2).

Appendix C

Fisher matrix and propagation of errors

C.1 Wiggles only power spectrum

These are the steps and calculations to follow in order to produce wiggles only power spectrum from the theoretical galaxy matter power spectrum.

Produce the matter power spectrum

- First we need to generate the $P(k)$ using Plank parameters from CAMB with high resolution wiggles, to do that we modify CAMB parameters input file with these values:

```
use_physical   = T
ombh2          = 0.022068
omch2          =0.12029E+00
 transfer_kmax           = 2
 transfer_k_per_logint   = 50
```

- Then we smooth the wiggles by putting $\Omega_b = 0.004$ (the lowest value CAMB can run without crashing) and in this case we add the value of the Ω_b to Ω_c. Therefore, $\Omega_c(\text{new}) = \Omega_c(\text{old}) + \Omega_b$.

```
use_physical   = T
```

```
ombh2           = 0.004
omch2           =0.142358E+00
 transfer_kmax            = 2
 transfer_k_per_logint    = 5
```

To produce the wiggles only power spectrum, we subtract the smoothed from the high resolution wiggles power spectrum. The results will produce the BAO wiggles function $f_{\rm BAO}$. Accordingly, we can define the power spectrum as

$$P(k) = [1 + f_{\rm BAO}(k)]\, P_{\rm smooth}(k) \tag{C.1}$$

Numerical differentiation of $f_{\rm BAO}(k)$

We can differentiate the function $f_{\rm BAO}(k)$ with respect to k numerically.

A numerical and an accurate method that we can use to differentiate the numerical function $f_{\rm BAO}$ with respect to k, is known as the Parabola method[1].

The method works as follow, stepping through all the points from 2 to $n-1$. For each point i there is one on the left $i-1$ and one on the right $i+1$. We can draw an explicit parabola through these three points (just as we can draw a line through two points). The equation for a parabola is $y = Ax^2 + Bx + C$, and for each point the do-loop finds the A, B and C for the parabola that runs through the point and the two on each side. The slope at the center point i is $y' = 2Ax(i) + B$. Note that A, B and C are different for each step. This function has been tested on many functions x^2, $\exp(x)$ and $\cos(x)$.

BAO signal

The basic quantities and parameters we require are

$$k^2 = k_\parallel^2 + k_\perp^2, \tag{C.2}$$

[1] http://mathfaculty.fullerton.edu/mathews//n2003/NewtonPolyMod.html

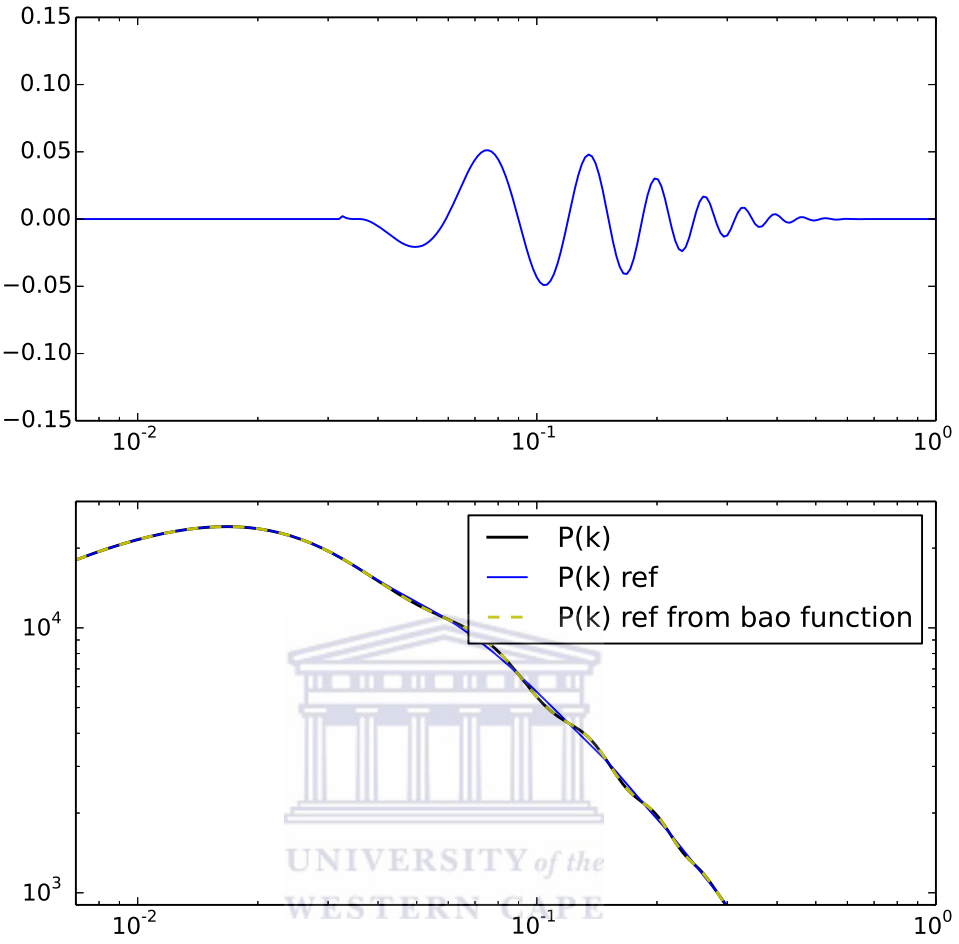

FIGURE C.1: The power spectrum $P(k)$, the smoothed power spectrum and the wiggles only function $f_{\rm BAO}(k)$ on the top panel. On the x-axis k in Mpc^{-1}h, and on the y-axis the units are Mpc$^3h^{-3}$

where $k_\parallel$ and $k_\perp$ are the wave number along and across the line of sight respectively, and the total wave number is $k = \sqrt{k_\parallel^2 + k_\perp^2}$. Where

$$\begin{aligned} k_{\perp \rm ref} &= \frac{k_\perp D_A(z)}{D_A(z)_{\rm ref}} \\ k_{\parallel \rm ref} &= \frac{k_\parallel H(z)_{\rm ref}}{H(z)} \end{aligned} \tag{C.3}$$

The subscript ref means the reference cosmology and the ones without ref are the ones with true cosmology. Note that $k_{\perp \rm ref}$ and $k_{\parallel \rm ref}$ are fixed. $k_\perp$ and $k_\parallel$ are

directly related to μ by

$$k_\perp = k\sqrt{1-\mu^2} \text{ and } k_\parallel = \mu^2 k. \tag{C.4}$$

Fisher method

Taking the derivative of k with respect to both quantities, $H(z)$ and $D(z)$

$$\begin{aligned} \frac{\partial k}{\partial D_A} &= -\frac{k^2_{\perp \mathrm{ref}} D^2_A(z)_{\mathrm{ref}}}{D^3_A(z)} \left[\left(k_{\perp \mathrm{ref}} \frac{D_A(z)_{\mathrm{ref}}}{D_A(z)} \right)^2 + k^2_\parallel \right]^{-\frac{1}{2}} \\ \frac{\partial k}{\partial H} &= \frac{k^2_{\parallel \mathrm{ref}}}{H(z)^2_{\mathrm{ref}}} H(z) \left[\left(k_{\mathrm{ref}\parallel} \frac{H(z)}{H(z)_{\mathrm{ref}}} \right)^2 + k^2_\perp \right]^{-\frac{1}{2}}. \end{aligned} \tag{C.5}$$

Therefore the Fisher formula will be given by:

$$F_{ij} = \int_{-1}^{1} d\mu \int_0^\infty \frac{k^2 dk}{8\pi} \frac{V_{\mathrm{survey}}}{[P^z + n^{-1}]^2} \left[\frac{\partial P_b}{\partial \ln(ks)} \right]^2 \frac{\partial \ln(ks)}{\partial \theta_i} \frac{\partial \ln(ks)}{\partial \theta_j}. \tag{C.6}$$

If the fractional errors on $s_\perp^{-1}$ and $s_\parallel$ are equivalent to measuring the fractional errors on D_A/s and Hs (where s is the true physical value of the sound horizon),

$$\frac{\partial \ln(ks)}{\partial \ln s_\perp^{-1}} = \mu^2 - 1, \quad \frac{\partial \ln(ks)}{\partial s_\parallel} = \mu^2 \tag{C.7}$$

Therefore, Fisher matrix can be expressed as

$$F_{ij} = \int_{-1}^{1} d\mu \int_0^\infty \frac{k^2 dk}{8\pi^2} \frac{V_{\mathrm{survey}}}{[P^z + n^{-1}]^2} \left[\frac{\partial [1 + f_{\mathrm{BAO}}(k)] P_{\mathrm{ref}}}{\partial \ln(ks)} \right]^2 \frac{\partial \ln(ks)}{\partial \theta_i} \frac{\partial \ln(ks)}{\partial \theta_j}. \tag{C.8}$$

We can also rewrite the Fisher matrix where θ_j and θ_i are replaced by the parameters we aim to forecast for, D_A and H, hence

$$F_{ij} = \int_{-1}^{1} d\mu \int_{k_{\min}}^{k_{\max}} \frac{k^2 dk}{8\pi^2} \frac{V_{\text{survey}}}{\left[P^z + n^{-1}\right]^2} \times \left[\frac{\partial f_{\text{BAO}}(k) P_{\text{ref}}}{\partial k}\right] \left[\frac{\partial f_{\text{BAO}}(k) P_{\text{ref}}}{\partial k}\right] \frac{\partial k}{\partial \log D_A} \frac{\partial k}{\partial \log H}, \tag{C.9}$$

assuming that the reference cosmology = true cosmology. Using (C.2), (C.3) and (C.5), the Fisher matrix can be written as

$$F_{ij} = \int_{-1}^{1} d\mu \int_{k_{\min}}^{k_{\max}} \frac{V_{\text{survey}}\, k^2 dk}{8\pi^2} \left[\frac{n(z)}{P(k)n(z)+1}\right]^2 \times \left[\frac{\partial F_{\text{BAO}}}{\partial k}\right]^2 \left[\frac{P(k)}{1+F_{\text{BAO}}}\right]^2 \left[(\mu^2-1)k\right] \left(\mu^2 k\right), \tag{C.10}$$

where the power spectrum is defined as

$$P(k) = P_{\text{lin}} \left[\frac{D(z)}{D(z_{in})}\right]^2 \left[\frac{(1+z_{in})}{(1+z)}\right]^2 R(\mu)^2\, b^2, \tag{C.11}$$

and

$$R(\mu) = 1 + \beta\mu. \tag{C.12}$$

C.2 Errors on D_A and H

Using the 2×2 Fisher matrix defined in (C.9), the errors on $\ln D_A(z)$ and $\ln H(z)$ are computed using [**?**],

$$\sigma_{\ln D_A} = \sqrt{(F^{-1})_{11}}, \quad \sigma_{\ln H} = \sqrt{(F^{-1})_{22}}. \tag{C.13}$$

Fig. 6.5 shows the fractional percentage error on the Hubble rate (σ_H/H) and the angular diameter distance (σ_{D_A}/D_A) for the reference experiments.

C.3 Error propagation to w

We use the parameters of $w(z)$ introduced in (6.8). Therefore the cosmological parameters under consideration are given by (6.19). We propagate the errors in

$D_A(z)$ and $H(z)$ to the cosmological parameters using

$$\tilde{F}_{\alpha\beta} = \sum_{ij} \frac{\partial\theta_i}{\partial\theta_\alpha}\frac{\partial\theta_j}{\partial\theta_\beta}F_{ij}. \tag{C.14}$$

where $\theta_i = \ln D_A, \ln H$ and $\theta_\alpha = w_0, w_a, \Omega_{\rm cdm}, \Omega_{\rm b}, \Omega_K, h$. Partial derivatives of D_A and H with respect to w_0 and w_a are given by [**?**]

$$\begin{aligned}
\frac{\partial \ln D_A}{\partial w_0} &= -\frac{3}{2}\Omega_{\rm de}\frac{\int_0^z \ln(1+z')\mathcal{F}(z')E(z')^{-3}dz'}{\int_0^z E(z')^{-1}dz'},\\
\frac{\partial \ln D_A}{\partial w_a} &= -\frac{3}{2}\Omega_{\rm de}\\
&\times \frac{\int_0^z \left\{\ln(1+z') - \frac{z'}{(1+z')}\right\}\mathcal{F}(z')E(z')^{-3}dz'}{\int_0^z E(z')^{-1}dz'},\\
\frac{\partial \ln D_A}{\partial \Omega_{\rm cdm}} &= -\frac{1}{2}\frac{\int_0^z \left\{(1+z')^3 - \mathcal{F}(z')\right\}E(z')^{-3}dz'}{\int_0^z E(z')^{-1}dz'},\\
\frac{\partial \ln D_A}{\partial \Omega_{\rm b}} &= -\frac{1}{2}\frac{\int_0^z \left\{(1+z')^3 - \mathcal{F}(z')\right\}E(z')^{-3}dz'}{\int_0^z E(z')^{-1}dz'},\\
\frac{\partial \ln D_A}{\partial \Omega_K} &= -\frac{1}{2}\frac{\int_0^z \left\{(1+z')^2 - \mathcal{F}(z')\right\}E(z')^{-3}}{\int_0^z E(z')^{-1}dz'}\\
&+ \frac{1}{6}\left(\int_0^z E(z')^{-1}dz'\right)^2,\\
\frac{\partial \ln D_A}{\partial h} &= \frac{1}{h}\\
\frac{\partial \ln H}{\partial w_0} &= \frac{3}{2}\Omega_{\rm de}\ln(1+z)\frac{\mathcal{F}(z)}{E^2(z)},\\
\frac{\partial \ln H}{\partial w_a} &= \frac{3}{2}\Omega_{\rm de}\left\{\ln(1+z) - \frac{z}{(1+z)}\right\}\frac{\mathcal{F}(z)}{g(z)},\\
\frac{\partial \ln H}{\partial \Omega_{\rm cdm}} &= \frac{1}{2}\left\{(1+z)^3 - \mathcal{F}(z)\right\}\frac{1}{E^2(z)},\\
\frac{\partial \ln H}{\partial \Omega_{\rm b}} &= \frac{1}{2}\left\{(1+z)^3 - \mathcal{F}(z)\right\}\frac{1}{E^2(z)},\\
\frac{\partial \ln H}{\partial \Omega_K} &= \frac{1}{2}\left\{(1+z)^2 - \mathcal{F}(z)\right\}\frac{1}{E^2(z)},\\
\frac{\partial \ln H}{\partial h} &= -\frac{1}{h},
\end{aligned}$$

where $\mathcal{F}(z)$ and $H(z) = H_0E(z)$ are given by (2.14) and (2.15) respectively.

We marginalize over H_0, $\Omega_{\rm cdm}$, $\Omega_{\rm b}$ and Ω_K, and we add the distance information from the CMB as follows

$$F_{\alpha\beta}^{\rm total}(z) = F_{\alpha\beta}^{\rm CMB} + F_{\alpha\beta}^{\rm gal}(z), \tag{C.15}$$

where

$$F_{\alpha\beta}^{\rm CMB} = 10^4 \frac{\partial \ln D_A(z=1090)}{\partial q_\alpha} \frac{\partial \ln D_A(z=1090)}{\partial q_\beta}. \tag{C.16}$$

Forecasting the errors for the cosmological parameters w_0, w_a, $\Omega_{\rm cdm}$, $\Omega_{\rm b}$, Ω_K and h, is achieved by adding a diagonal prior matrix, derived from the Planck prior matrix, to the SKA Fisher matrix – see Table C.1. We also add prior information about the angular diameter distance out to last scattering at $z = 1090$, using (C.16):

$$\frac{\sigma_{D_A(1090)}}{D_A(1090)} = 0.001 \tag{C.17}$$

-	n_s	w_0	w_a	Ω_b	Ω_K	Ω_cdm	h	σ_8
n_s	1.995e+05	-3.736e+04	-1.049e+04	1.399e+06	5.586e+05	1.911e+06	-7.651e+04	-2.238e+03
w_0	-3.736e+04	1.839e+05	5.165e+04	-7.420e+06	-3.987e+06	-1.252e+07	1.324e+06	-4.515e+02
w_a	-1.049e+04	5.165e+04	1.450e+04	-2.083e+06	-1.119e+06	-3.517e+06	3.719e+05	-1.268e+02
Ω_b	1.399e+06	-7.420e+06	-2.083e+06	3.649e+08	1.585e+08	5.661e+08	-5.168e+07	3.203e+04
Ω_K	5.586e+05	-3.987e+06	-1.119e+06	1.585e+08	8.705e+07	2.705e+08	-2.917e+07	1.884e+04
Ω_cdm	1.911e+06	-1.252e+07	-3.517e+06	5.661e+08	2.705e+08	9.116e+08	-8.940e+07	6.346e+04
h	-7.651e+04	1.324e+06	3.719e+05	-5.168e+07	-2.917e+07	-8.940e+07	9.889e+06	-1.018e+04
σ_8	-2.238e+03	-4.515e+02	-1.268e+02	3.203e+04	1.884e+04	6.346e+04	-1.018e+04	1.517e+04

TABLE C.1: Planck prior matrix for the cosmological parameters of interest.

Bibliography

[1] B. Bassett and R. Hlozek. *arXiv: 0910.5224*. 2010.

[2] J. A. Peacock et al. A Measurement of the cosmological mass density from clustering in the 2dF Galaxy Redshift Survey. *Nature*, 410:169–173, 2001. doi: 10.1038/35065528.

[3] W. J. Percival, S. Cole, D. J. Eisenstein, R. C. Nichol, J. A. Peacock, A. C. Pope, and A. S. Szalay. Measuring the Baryon Acoustic Oscillation scale using the SDSS and 2dFGRS. *Mon. Not. Roy. Astron. Soc.*, 381:1053–1066, 2007. doi: 10.1111/j.1365-2966.2007.12268.x.

[4] E. Majerotto et al. Probing deviations from General Relativity with the Euclid spectroscopic survey. *Mon. Not. Roy. Astron. Soc.*, 424:1392–1408, 2012. doi: 10.1111/j.1365-2966.2012.21323.x.

[5] C. Clarkson and C. Zunckel. Direct reconstruction of dark energy. *Phys. Rev. Lett.*, 104:211301, 2010. doi: 10.1103/PhysRevLett.104.211301.

[6] T. Holsclaw, U. Alam, B. Sanso, H. Lee, K. Heitmann, S. Habib, and D. Higdon. Nonparametric Dark Energy Reconstruction from Supernova Data. *Phys. Rev. Lett.*, 105:241302, 2010. doi: 10.1103/PhysRevLett.105.241302.

[7] H. Seo and D. J. Eisenstein. Probing dark energy with baryonic acoustic oscillations from future large galaxy redshift surveys. *ApJ*, 598:720–740, 2003. doi: 10.1086/379122.

[8] S. Weinberg. The Cosmological Constant Problem. *Rev. Mod. Phys.*, 61: 1–23, 1989. doi: 10.1103/RevModPhys.61.1.

[9] S. Weinberg. *Cosmology*. 2007.

[10] V. Sahni, T. D. Saini, A. A. Starobinsky, and U. Alam. Statefinder: A New geometrical diagnostic of dark energy. *JETP Lett.*, 77:201–206, 2003. doi: 10.1134/1.1574831. [Pisma Zh. Eksp. Teor. Fiz.77,249(2003)].

[11] D. Huterer and M. S. Turner. Probing the dark energy: Methods and strategies. *Phys. Rev.*, D64:123527, 2001. doi: 10.1103/PhysRevD.64.123527.

[12] P. Astier et al. Search for heavy neutrinos mixing with tau neutrinos. *Phys. Lett.*, B506:27–38, 2001. doi: 10.1016/S0370-2693(01)00362-8.

[13] J. Weller and A. Albrecht. Opportunities for future supernova studies of cosmic acceleration. *Phys. Rev. Lett.*, 86:1939–1942, 2001. doi: 10.1103/PhysRevLett.86.1939.

[14] B. F. Gerke and G. Efstathiou. Probing quintessence: Reconstruction and parameter estimation from supernovae. *Mon. Not. Roy. Astron. Soc.*, 335: 33, 2002. doi: 10.1046/j.1365-8711.2002.05612.x.

[15] G. Efstathiou et al. Evidence for a non-zero lambda and a low matter density from a combined analysis of the 2dF Galaxy Redshift Survey and cosmic microwave background anisotropies. *Mon. Not. Roy. Astron. Soc.*, 330:L29, 2002.

[16] J. A. Frieman et al. The Sloan Digital Sky Survey-II Supernova Survey: Technical Summary. *Astron. J.*, 135:338–347, 2008. doi: 10.1088/0004-6256/135/1/338.

[17] J. D. Barrow. The anthropic cosmological principle. 1986.

[18] A Kogut, DN Spergel, C Barnes, CL Bennett, M Halpern, G Hinshaw, N Jarosik, M Limon, SS Meyer, L Page, et al. First-year wilkinson microwave anisotropy probe (wmap) wmap is the result of a partnership between princeton university and the nasa goddard space flight center. scientific guidance is provided by the wmap science team. observations: Temperature-polarization correlation. *The Astrophysical Journal Supplement Series*, 148(1):161, 2003.

[19] Liddle A. R. *An introduction to modern cosmology*. Wiley, 2003.

[20] E. Hubble. A relation between distance and radial velocity among extragalactic nebulae. *Proc. Nat. Acad. Sci.*, 15:168–173, 1929. doi: 10.1073/pnas.15.3.168.

[21] U. Seljak and M. Zaldarriaga. A line of sight approach to cosmic microwave background anisotropies. *arXiv preprint astro-ph/9603033*, 1996.

[22] Thomas J Broadhurst, Richard S Ellis, DC Koo, and AS Szalay. Large-scale distribution of galaxies at the galactic poles. *Nature*, 343(6260):726–728, 1990.

[23] MH Pinsonneault, Constantine P Deliyannis, and P Demarque. Evolutionary models of halo stars with rotation. ii-effects of metallicity on lithium depletion, and possible implications for the primordial lithium abundance. *The Astrophysical Journal Supplement Series*, 78:179–203, 1992.

[24] John N Bahcall, MH Pinsonneault, and GJ Wasserburg. Solar models with helium and heavy-element diffusion. *Reviews of Modern Physics*, 67(4):781, 1995.

[25] Eric V Linder. Exploring the expansion history of the universe. *Physical Review Letters*, 90(9):091301, 2003.

[26] S. Dodelson. *Modern cosmology*. 2003.

[27] P. A. R. Ade et al. Planck 2015 results. XIII. Cosmological parameters. 2015.

[28] L. Amendola and S. Tsujikawa. *Dark Energy: Theory and Observations*. 2010. ISBN 0521516005. URL http://books.google.co.za/books/about/Dark_Energy.html?id=U0HTngEACAAJ&pgis=1.

[29] D. H. Weinberg, M. J. Mortonson, D. J. Eisenstein, C. Hirata, A. G. Riess, and E. Rozo. Observational Probes of Cosmic Acceleration. *Phys. Rept.*, 530:87–255, 2013. doi: 10.1016/j.physrep.2013.05.001.

[30] D. W. Hogg. Distance measures in cosmology. *arXiv: astro-ph/9905116*, 1999.

[31] R. A. Battye, I. W. A. Browne, C. Dickinson, G. Heron, B. Maffei, and A. Pourtsidou. HI intensity mapping : a single dish approach. *Mon. Not. Roy. Astron. Soc*, 434:1239–1256, 2013. doi: 10.1093/mnras/stt1082.

[32] D. J. Eisenstein et al. SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way Galaxy, and Extra-Solar Planetary Systems. *Astron. J.*, 142:72, 2011. doi: 10.1088/0004-6256/142/3/72.

[33] E. V. Linder. Cosmic growth history and expansion history. *Phys. Rev.*, D72:043529, 2005. doi: 10.1103/PhysRevD.72.043529.

[34] Phillip James Edwin Peebles. *The large-scale structure of the universe*. Princeton university press, 1980.

[35] L. Amendola et al. Cosmology and fundamental physics with the Euclid satellite. *Living Rev. Rel.*, 16:6, 2013. doi: 10.12942/lrr-2013-6.

[36] H. Mo, F. C. van den Bosch, and S. White. *Galaxy Formation and Evolution*. May 2010.

[37] P. A. R. Ade et al. Planck 2013 results. XVI. Cosmological parameters. *Astron. Astrophys.*, 571:A16, 2014. doi: 10.1051/0004-6361/201321591.

[38] S. Serjeant. *Observational Cosmology*. The open university. Cambridge University Press, 2010. ISBN 9780521157155. URL https://books.google.co.za/books?id=LfaZRAAACAAJ.

[39] A. Conley et al. Supernova Constraints and Systematic Uncertainties from the First 3 Years of the Supernova Legacy Survey. *Astrophys. J. Suppl.*, 192: 1, 2011. doi: 10.1088/0067-0049/192/1/1.

[40] N. Suzuki et al. The Hubble Space Telescope Cluster Supernova Survey: V. Improving the Dark Energy Constraints Above z¿1 and Building an Early-Type-Hosted Supernova Sample. *Astrophys. J.*, 746:85, 2012. doi: 10.1088/0004-637X/746/1/85.

[41] H. Campbell et al. Cosmology with Photometrically-Classified Type Ia Supernovae from the SDSS-II Supernova Survey. *Astrophys. J.*, 763:88, 2013. doi: 10.1088/0004-637X/763/2/88.

[42] D. Scolnic et al. Systematic Uncertainties Associated with the Cosmological Analysis of the First Pan-STARRS1 Type Ia Supernova Sample. *Astrophys. J.*, 795(1):45, 2014. doi: 10.1088/0004-637X/795/1/45.

[43] A. Rest et al. Cosmological Constraints from Measurements of Type Ia Supernovae discovered during the first 1.5 yr of the Pan-STARRS1 Survey. *Astrophys. J.*, 795(1):44, 2014. doi: 10.1088/0004-637X/795/1/44.

[44] J. P. Bernstein et al. Supernova Simulations and Strategies For the Dark Energy Survey. *Astrophys. J.*, 753:152, 2012. doi: 10.1088/0004-637X/753/2/152.

[45] M. Seikel, C. Clarkson, and M. Smith. Reconstruction of dark energy and expansion dynamics using Gaussian processes. *JCAP*, 1206:036, 2012. doi: 10.1088/1475-7516/2012/06/036.

[46] T. Holsclaw, U. Alam, B. Sanso, K. Lee, H and. Heitmann, S. Habib, and D. Higdon. Nonparametric Reconstruction of the Dark Energy Equation of State from Diverse Data Sets. *Phys. Rev.*, D84:083501, 2011. doi: 10.1103/PhysRevD.84.083501.

[47] A. Shafieloo. Crossing Statistic: Reconstructing the Expansion History of the Universe. *JCAP*, 1208:002, 2012. doi: 10.1088/1475-7516/2012/08/002.

[48] C. Clarkson, B. Bassett, and T. H. Lu. A general test of the Copernican Principle. *Phys. Rev. Lett.*, 101:011301, 2008. doi: 10.1103/PhysRevLett.101.011301.

[49] J. Uzan, C. Clarkson, and G. F. R. Ellis. Time drift of cosmological redshifts as a test of the Copernican principle. *Phys. Rev. Lett.*, 100:191303, 2008. doi: 10.1103/PhysRevLett.100.191303.

[50] V. Sahni, A. Shafieloo, and A. A. Starobinsky. Two new diagnostics of dark energy. *Phys. Rev.*, D78:103502, 2008. doi: 10.1103/PhysRevD.78.103502.

[51] C. Zunckel and C. Clarkson. Consistency Tests for the Cosmological Constant. *Phys. Rev. Lett.*, 101:181301, 2008. doi: 10.1103/PhysRevLett.101.181301.

[52] Arman Shafieloo and Chris Clarkson. Model independent tests of the standard cosmological model. *Phys. Rev.*, D81:083537, 2010. doi: 10.1103/PhysRevD.81.083537.

[53] C. Clarkson. On the determination of dark energy. *AIP Conf. Proc.*, 1241:784–796, 2010. doi: 10.1063/1.3462717.

[54] S. Nesseris and A. Shafieloo. A model independent null test on the cosmological constant. *Mon. Not. Roy. Astron. Soc.*, 408:1879–1885, 2010. doi: 10.1111/j.1365-2966.2010.17254.x.

[55] C. Clarkson. Establishing homogeneity of the universe in the shadow of dark energy. *Comptes Rendus Physique*, 13:682–718, 2012. doi: 10.1016/j.crhy.2012.04.005.

[56] M. Seikel and C. Clarkson. Optimising Gaussian processes for reconstructing dark energy dynamics from supernovae. 2013.

[57] David J. C. MacKay. *Information Theory, Inference & Learning Algorithms*. Cambridge University Press, New York, NY, USA, 2002. ISBN 0521642981.

[58] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005. ISBN 026218253X.

[59] C. Clarkson, M. Cortes, and B. A. Bassett. Dynamical Dark Energy or Simply Cosmic Curvature? *JCAP*, 0708:011, 2007. doi: 10.1088/1475-7516/2007/08/011.

[60] R. Hlozek, M. Cortes, C. Clarkson, and B. Bassett. Dark energy degeneracies in the background dynamics. *Gen. Rel. Grav.*, 40:285–300, 2008. doi: 10.1007/s10714-007-0548-6.

[61] J. Weller and A. Albrecht. Future supernovae observations as a probe of dark energy. *Phys. Rev.*, D65:103512, 2002. doi: 10.1103/PhysRevD.65.103512.

[62] U. Alam, V. Sahni, T. D. Saini, and A. A. Starobinsky. Exploring the expanding universe and dark energy using the Statefinder diagnostic. *Mon. Not. Roy. Astron. Soc.*, 344:1057, 2003. doi: 10.1046/j.1365-8711.2003.06871.x.

[63] R. A. Daly and S. G. Djorgovski. A model-independent determination of the expansion and acceleration rates of the universe as a function of redshift and constraints on dark energy. *Astrophys. J.*, 597:9–20, 2003. doi: 10.1086/378230.

[64] U. Alam, V. Sahni, and A. A. Starobinsky. The Case for dynamical dark energy revisited. *JCAP*, 0406:008, 2004. doi: 10.1088/1475-7516/2004/06/008.

[65] Y. Wang and M. Tegmark. New dark energy constraints from supernovae, microwave background and galaxy clustering. *Phys. Rev. Lett.*, 92:241302, 2004. doi: 10.1103/PhysRevLett.92.241302.

[66] R. A. Daly and S. G. Djorgovski. Direct determination of the kinematics of the universe and properties of the dark energy as functions of redshift. *Astrophys. J.*, 612:652–659, 2004. doi: 10.1086/422673.

[67] V. Sahni and A. Starobinsky. Reconstructing Dark Energy. *Int. J. Mod. Phys.*, D15:2105–2132, 2006. doi: 10.1142/S0218271806009704.

[68] A. Shafieloo, U. Alam, V. Sahni, and A. A. Starobinsky. Smoothing Supernova Data to Reconstruct the Expansion History of the Universe and its Age. *Mon. Not. Roy. Astron. Soc.*, 366:1081–1095, 2006. doi: 10.1111/j.1365-2966.2005.09911.x.

[69] C. Zunckel and R. Trotta. Reconstructing the history of dark energy using maximum entropy. *Mon. Not. Roy. Astron. Soc.*, 380:865, 2007. doi: 10.1111/j.1365-2966.2007.12000.x.

[70] C. Espana-Bonet and P. Ruiz-Lapuente. Tracing the equation of state and the density of the cosmological constant along z. *JCAP*, 0802:018, 2008. doi: 10.1088/1475-7516/2008/02/018.

[71] C. R. Genovese, P. Freeman, L. Wasserman, R. C. Nichol, and C. Miller. Inference for the Dark Energy Equation of State Using Type Ia Supernova Data. *Ann. Appl. Stat.*, 3:144, 2009. doi: 10.1214/08-AOAS229SUPP.

[72] C. Bogdanos and S. Nesseris. Genetic Algorithms and Supernovae Type Ia Analysis. *JCAP*, 0905:006, 2009. doi: 10.1088/1475-7516/2009/05/006.

[73] R. G. Crittenden, G. Zhao, L. Pogosian, L. Samushia, and X. Zhang. Fables of reconstruction: controlling bias in the dark energy equation of state. *JCAP*, 1202:048, 2012. doi: 10.1088/1475-7516/2012/02/048.

[74] R. Lazkoz, V. Salzano, and I. Sendra. Revisiting a model-independent dark energy reconstruction method. *Eur. Phys. J.*, C72:2130, 2012. doi: 10.1140/epjc/s10052-012-2130-y.

[75] C. Ma and T. Zhang. Power of Observational Hubble Parameter Data: a Figure of Merit Exploration. *Astrophys. J.*, 730:74, 2011. doi: 10.1088/0004-637X/730/2/74.

[76] D. Stern, R. Jimenez, L. Verde, M. Kamionkowski, and S. A. Stanford. Cosmic Chronometers: Constraining the Equation of State of Dark Energy.

I: H(z) Measurements. *JCAP*, 1002:008, 2010. doi: 10.1088/1475-7516/2010/02/008.

[77] M. Hicken, W. M. Wood-Vasey, S. Blondin, P. Challis, S. Jha, P. L. Kelly, A. Rest, and R. P. Kirshner. Improved Dark Energy Constraints from 100 New CfA Supernova Type Ia Light Curves. *Astrophys. J.*, 700:1097–1140, 2009. doi: 10.1088/0004-637X/700/2/1097.

[78] C. Clarkson. Establishing homogeneity of the universe in the shadow of dark energy. *Comptes Rendus Physique*, 13:682–718, 2012. doi: 10.1016/j.crhy.2012.04.005.

[79] M. Kunz. The phenomenological approach to modeling the dark energy. *Comptes Rendus Physique*, 13:539–565, July 2012. doi: 10.1016/j.crhy.2012.04.007.

[80] R. Jimenez and A. Loeb. Constraining cosmological parameters based on relative galaxy ages. *Astrophys. J.*, 573:37–42, 2002. doi: 10.1086/340549.

[81] S. M. Crawford, A. L. Ratsimbazafy, C. M. Cress, E. A. Olivier, S-L. Blyth, and K. J. van der Heyden. Luminous Red Galaxies in Simulations: Cosmic Chronometers? *Mon. Not. Roy. Astron. Soc.*, 406:2569, 2010. doi: 10.1111/j.1365-2966.2010.16849.x.

[82] M. Moresco et al. Improved constraints on the expansion rate of the Universe up to z 1.1 from the spectroscopic evolution of cosmic chronometers. *JCAP*, 1208:006, 2012. doi: 10.1088/1475-7516/2012/08/006.

[83] E. Gaztanaga, A. Cabre, and L. Hui. Clustering of Luminous Red Galaxies IV: Baryon Acoustic Peak in the Line-of-Sight Direction and a Direct Measurement of H(z). *Mon. Not. Roy. Astron. Soc.*, 399:1663–1680, 2009. doi: 10.1111/j.1365-2966.2009.15405.x.

[84] C. Chuang and Y. Wang. Measurements of H(z) and D_A(z) from the Two-Dimensional Two-Point Correlation Function of Sloan Digital Sky Survey Luminous Red Galaxies. *Mon. Not. Roy. Astron. Soc.*, 426:226, 2012. doi: 10.1111/j.1365-2966.2012.21565.x.

[85] C. Blake et al. The WiggleZ Dark Energy Survey: Joint measurements of the expansion and growth history at z ¡ 1. *Mon. Not. Roy. Astron. Soc.*, 425:405–414, 2012. doi: 10.1111/j.1365-2966.2012.21473.x.

[86] M. Moresco, L. Verde, L. Pozzetti, R. Jimenez, and A. Cimatti. New constraints on cosmological parameters and neutrino properties using the expansion rate of the Universe to z 1.75. *JCAP*, 1207:053, 2012. doi: 10.1088/1475-7516/2012/07/053.

[87] E. Komatsu et al. Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. *Astrophys. J. Suppl.*, 192:18, 2011. doi: 10.1088/0067-0049/192/2/18.

[88] D. Schlegel, M. White, and D. Eisenstein. The Baryon Oscillation Spectroscopic Survey: Precision measurements of the absolute cosmic distance scale. 2009.

[89] R. Laureijs et al. Euclid Definition Study Report. 2011.

[90] R. Ansari, J. E. Campagne, P. Colom, J. M. Le Goff, C. Magneville, J. M. Martin, M. Moniez, J. Rich, and C. Yeche. 21 cm observation of LSS at z 1 Instrument sensitivity and foreground subtraction. *Astron. Astrophys.*, 540: A129, 2012. doi: 10.1051/0004-6361/201117837.

[91] A. McPherson et al. Report and options for re-baselining of ska-1. *SKA Organisation*, 2015. URL https://www.skatelescope.org/wp-content/uploads/2014/03/SKA-TEL-SKO-0000229_Report-and-Options-for-Re-Baselining-of-SKA-1-Word_Amm-signed.pdf.

[92] M. J. Meyer et al. The HIPASS Catalog. 1. Data presentation. *Mon. Not. Roy. Astron. Soc.*, 350:1195, 2004. doi: 10.1111/j.1365-2966.2004.07710.x.

[93] F. B. Abdalla, C. Blake, and S. Rawlings. Forecasts for Dark Energy Measurements with Future HI Surveys. *Mon. Not. Roy. Astron. Soc.*, 401:743–758, 2009. doi: 10.1111/j.1365-2966.2009.15704.x.

[94] S. Camera, M. G. Santos, and R. Maartens. Probing primordial non-Gaussianity with SKA galaxy redshift surveys: a fully relativistic analysis. *Mon. Not. Roy. Astron. Soc.*, 448:1035, 2015. doi: 10.1093/mnras/stv040.

[95] R. Braun, T. Bourke, J. A Green, E. Keane, and J. Wagg. Advancing Astrophysics with the Square Kilometre Array. In *Proceedings, Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, volume AASKA14, page 174, 2015.

[96] P. E. Dewdney et al. Ska1 system baseline design. *SKA Organisation*, 2013. URL https://www.skatelescope.org/wp-content/uploads/2013/05/SKA-TEL-SKO-DD-001-1_BaselineDesign1.pdf.

[97] F. Villaescusa-Navarro et al. Cross-correlating 21cm intensity maps with Lyman Break Galaxies in the post-reionization era. 2014.

[98] D. Obreschkow and S. Rawlings. A Virtual Sky with Extragalactic HI and CO Lines for the SKA and ALMA. *Astrophys. J.*, 703:1890–1903, 2009. doi: 10.1088/0004-637X/703/2/1890.

[99] G. De Lucia and J. Blaizot. The hierarchical formation of the brightest cluster galaxies. *Mon. Not. Roy. Astron. Soc.*, 375:2–14, 2007. doi: 10.1111/j.1365-2966.2006.11287.x.

[100] V. Springel et al. Simulating the joint evolution of quasars, galaxies and their large-scale distribution. *Nature*, 435:629–636, 2005. doi: 10.1038/nature03597.

[101] R. K. Sheth and G. Tormen. Large scale bias and the peak background split. *Mon. Not. Roy. Astron. Soc.*, 308:119, 1999. doi: 10.1046/j.1365-8711.1999.02692.x.

[102] G. Popping, P. S. Behroozi, and M. S. Peeples. Evolution of the atomic and molecular gas content of galaxies in dark matter haloes. *Mon. Not. Roy. Astron. Soc.*, 449(1):477–493, 2015. doi: 10.1093/mnras/stv318.

[103] D. Obreschkow, X. Ma, M. Meyer, C. Power, M. Zwaan, L. Staveley-Smith, and M. J. Drinkwater. Confronting Cold Dark Matter Predictions with Observed Galaxy Rotations. *Astrophys. J.*, 766:137, 2013. doi: 10.1088/0004-637X/766/2/137.

[104] A. Font-Ribera et al. The large-scale cross-correlation of Damped Lyman Alpha Systems with the Lyman Alpha Forest: First Measurements from BOSS. *JCAP*, 1211:059, 2012. doi: 10.1088/1475-7516/2012/11/059.

[105] H. Seo, J. Eckel, D. J. Eisenstein, K. Mehta, M. Metchnik, N. Padmanabhan, P. Pinto, R. Takahashi, M. White, and X. Xu. High-precision predictions for the acoustic scale in the non-linear regime. *Astrophys. J.*, 720:1650–1667, 2010. doi: 10.1088/0004-637X/720/2/1650.

[106] K. T. Mehta, H. Seo, J. Eckel, D. J. Eisenstein, M. Metchnik, P. Pinto, and X. Xu. Galaxy Bias and its Effects on the Baryon Acoustic Oscillations Measurements. *Astrophys. J.*, 734:94, 2011. doi: 10.1088/0004-637X/734/2/94.

[107] R. E. Smith et al. Stable clustering, the halo model and nonlinear cosmological power spectra. *Mon. Not. Roy. Astron. Soc.*, 341:1311, 2003. doi: 10.1046/j.1365-8711.2003.06503.x.

[108] A. Lewis, A. Challinor, and A. Lasenby. Efficient computation of CMB anisotropies in closed FRW models. *Astrophys. J.*, 538:473–476, 2000. doi: 10.1086/309179.

[109] P. Bull, P. G. Ferreira, P. Patel, and M. G. Santos. Late-time cosmology with 21cm intensity mapping experiments. 2014.

[110] C. Blake and K. Glazebrook. Probing dark energy using baryonic oscillations in the galaxy power spectrum as a cosmological ruler. *ApJ*, 594:665–673, 2003. doi: 10.1086/376983.

[111] L. Anderson et al. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations in the Data Release 9 Spectroscopic Galaxy Sample. *Mon. Not. Roy. Astron. Soc.*, 427 (4):3435–3467, 2013. doi: 10.1111/j.1365-2966.2012.22066.x.

[112] K. S. Dawson et al. The Baryon Oscillation Spectroscopic Survey of SDSS-III. *Astron. J.*, 145:10, 2013. doi: 10.1088/0004-6256/145/1/10.

[113] B. A. Bassett, Y. Fantaye, R. Hlozek, and J. Kotze. Fisher Matrix Preloaded – Fisher4Cast. *Int. J. Mod. Phys.*, D20:2559–2598, 2011. doi: 10.1142/S0218271811020548.

[114] D. Coe. Fisher Matrices and Confidence Ellipses: A Quick-Start Guide and Software. 2009.

[115] P. Bull, S. Camera, A. Raccanelli, C. Blake, P. G. Ferreira, M. G. Santos, and D. J. Schwarz. Measuring baryon acoustic oscillations with future SKA surveys. In *PoS AASKA14 (2015) 024*, 2015. URL http://inspirehep.net/record/1339611/files/arXiv:1501.04088.pdf.

[116] M. G. Santos et al. Cosmology with a SKA HI intensity mapping survey. 2015.