Cosmological observations and their information gains

Author(s):
Seehars, Sebastian

Publication Date:
2016

Permanent Link:
https://doi.org/10.3929/ethz-a-010742533

Rights / License:
In Copyright - Non-Commercial Use Permitted

This page was generated automatically upon download from the ETH Zurich Research Collection. For more information please consult the Terms of use.
COSMOLOGICAL OBSERVATIONS AND THEIR INFORMATION GAINS

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

SEBASTIAN SEEHARS

MSc in Theoretical and Mathematical Physics
Ludwig Maximilian University of Munich

born on 11.07.1986

citizen of Germany

accepted on the recommendation of

Prof. Dr. Alexandre Refregier
Dr. Adam Amara
Prof. Dr. Jochen Weller

2016
Abstract

The standard model of cosmology describes the current epoch of our universe as dominated by dark matter and dark energy, two phenomenological components with unknown microscopic origin. Improving our understanding of those components is one of the main motivations of observational cosmology.

One promising way to make progress is to combine and compare the constraints from observations of individual cosmological probes. To this end, we develop a mathematical framework for quantifying the concordance between different experiments. It is based on the information gain as measured by the relative entropy, or Kullback-Leibler divergence, between two multivariate posterior distributions. We first apply our method to a historical sequence of Cosmic Microwave Background (CMB) experiments, focussing on the comparison between constraints from the Wilkinson Microwave Anisotropy Probe and Planck satellite missions. Additionally, we study the concordance between CMB data and other cosmological probes such as distance ladder, supernova type Ia, baryon acoustic oscillations, and weak lensing observations. While, according to our measure, not all data are concordant, the disagreements can be mostly traced back to systematics in the data rather than deviations from the standard model.

A complementary approach to improving our understanding of cosmology is to develop new observational techniques that reveal additional cosmological probes. A particularly promising future probe is the technique of neutral hydrogen intensity mapping. While this technique is able to map larger volumes than any other cosmological probe, extracting cosmological information from these observations is challenging and requires precise models for both the cosmological component and its foregrounds. We thus first propose a method to simulate maps of the large-scale structure of neutral hydrogen that uses a phenomenological prescription for assigning neutral hydrogen to N-body simulations of the dark matter density. These maps are sufficiently realistic to study the impact of non-linear effects on the clustering of neutral hydrogen. We conclude the thesis with a progress report on ongoing observational efforts in the Bleien Galactic Survey, which is designed to test strategies for single-dish drift scan surveys and will produce maps of the galactic foregrounds to intensity mapping.
Zusammenfassung

Laut dem Standardmodell der Kosmologie wird unser Universum von dunkler Materie und dunkler Energie beherrscht. Da die mikroskopischen Ursachen dieser dunklen Komponenten jedoch unbekannt sind, ist die Verbesserung unseres Verständnisses des dunklen Sektors eines der wichtigsten Ziele der beobachtenden Kosmologie.


Contents

Abstract (English/Deutsch) i

1 Introduction 1
   1.1 Homogeneous and Isotropic Universe 3
   1.2 Evolution of Perturbations 5
   1.3 Formation of Large-Scale Structure 10
   1.4 Cosmic Microwave Background 12
   1.5 Cosmological Probes 14

2 Quantifying Concordance with Information Gains 19
   2.1 Inference in Cosmology 20
   2.2 Inference for Gaussian Distributions 21
   2.3 Information Gains 23
   2.4 Information Gains for Gaussian Distributions 26
   2.5 Surprise 27
   2.6 Surprise for Gaussian Distributions 28
   2.7 Other Measures of Concordance 32
   2.8 Other Measures for Gaussian Distributions 33
   2.9 Comparison of Measures for One-Dimensional Toy-Model 36
   2.10 Discussion 39

3 Current Observations and their Information Gains 41
   3.1 Inference with Monte Carlo Markov Chains 41
   3.2 Estimating Information Gains and Surprise 46
   3.3 CMB Surveys 50
   3.4 Concordance of CMB Surveys before Planck 53
   3.5 Concordance of WMAP and Planck 56
   3.6 Cosmological Probes at Low Redshifts 62
   3.7 Concordance at Low Redshifts 63
   3.8 Discussion 65
## Contents

### 4 Future Observations and HI Intensity Mapping  
4.1 HI Intensity Mapping ........................................ 68  
4.2 Simulating HI Intensity Maps ............................. 71  
4.3 Testing the Sub-Grid Halo Model ......................... 80  
4.4 The Angular Power Spectrum of HI Intensity Maps .... 83  
4.5 Covariance of the Angular Power Spectrum ............. 86  
4.6 The Bleien Galactic Survey ................................. 101  
4.7 Discussion .................................................. 106

### 5 Conclusions ................................................. 111

Bibliography .................................................. 115

List of Symbols ............................................. 131

Index .......................................................... 133
1 Introduction

Our understanding of physical cosmology evolved significantly since its modern origination from Einstein’s theory of generalised relativity (GR). Driven by progress in both observational and theoretical cosmology, we are now in an era where increasingly precise predictions from cosmological models can be tested against increasingly powerful observations. The model that emerged from those analyses describes a universe that originated from a phase of exponential expansion (called inflation) and is currently dominated by dark matter and dark energy. These two components have been detected through their gravitational effect, but are seemingly not interacting electromagnetically (hence the term dark). Described by five main parameters—the Hubble constant, the distribution of the energy budget, and the properties of the initial perturbations from inflation—predictions from $\Lambda CDM$ provide a good fit to most current observations. Inflation, dark matter, and dark energy are however mostly phenomenologically motivated and their connection to the standard model of particle physics remains mysterious. Some aspects of the simplest model for dark energy and inflation even have conceptual issues (e.g. the cosmological constant problem and eternal inflation) that motivate theorists to look for other possible mechanisms which are able to explain the observations.

As we collect and analyse ever more precise cosmological data, it is hence an important question whether or not the different available probes are consistent with each other within a cosmological model. To answer this question, one focus of this thesis is the development of a framework in which the cosmological constraints from different probes and surveys can be compared in a quantitative manner. We will introduce such a measure, the relative entropy or Kullback–Leibler divergence (Kullback and Leibler, 1951), in chapter 2. Using the relative entropy to measure the differences between constraints, we can quantify the information gains from adding further observations to constraints on $\Lambda CDM$. We will show how we can further distinguish contributions to this gain that are expected within the model (the expected information gain) from contributions that could not be anticipated a priori (the Surprise). It is the Surprise measure which enables us to quantify the consistency of datasets. We will contrast the Surprise with other methods for comparing cosmological observations that have
been used in the past.

In chapter 3 we will then study how information gain and Surprise can be applied to current datasets. As the constraints from cosmological probes are rarely provided in analytic form, we need to estimate the information gain and the Surprise numerically. The standard algorithm for numerical inference in cosmology are Monte Carlo Markov chains (MCMCs). We will show that for most of the applications we are interested in, MCMC algorithms are sufficient for estimating information gain and Surprise. We will then proceed and apply our method to current observations. The most powerful cosmological observations to date are coming from the anisotropies of the Cosmic Microwave Background (CMB) and we will first analyse the concordance of a historical sequence of CMB observations. We will conclude our study of information gains and Surprise with an investigation of the agreement of CMB observations with current data from other cosmological probes at low redshifts.

Successfully improving our understanding of cosmology in the future will require improved current observations as well as newly developed techniques that reveal additional cosmological probes. A future cosmological probe that has the potential to generate information gains from volumes that are left unobserved by the CMB and low redshift probes is the technique of intensity mapping of molecular and atomic lines. One of the most promising lines for this approach is the 21 cm line of neutral hydrogen (HI) and the corresponding observational technique of HI intensity mapping has been developed through the last years (see e.g. Bharadwaj et al., 2001; Battye et al., 2004; Wyithe et al., 2007; Loeb and Wyithe, 2008; Chang et al., 2010). While intensity mapping is able to efficiently map larger volumes than any other cosmological probe, it is a very challenging probe from both the theoretical perspective of modelling the signal and the observational perspective of extracting the cosmological information. In chapter 4, we will therefore study two aspects (theoretical and observational) of HI intensity mapping: a method for simulating the large-scale structure of HI intensity maps and observations of the foreground emission to the redshifted 21 cm line with a 7 m radio telescope in Bleien, Switzerland.

The thesis is structured as follows: We will first review the basic theoretical concepts of $\Lambda$CDM in chapter 1. In chapter 2 we will introduce the notions of information gain and Surprise for comparing cosmological observations. We will apply these measures to a selection of current observations in chapter 3. The simulations of the large-scale structure of HI and our foreground observations will be discussed in chapter 4. We discuss our conclusions in chapter 5.

Most of the material in this thesis has been assembled from published work (Akeret et al., 2013; Seehars et al., 2014; Seehars et al., 2016; Seehars et al., 2016; Grandis et al., 2016b), and we give the corresponding references at the beginning of each chapter.
1.1 Homogeneous and Isotropic Universe

The theoretical foundations of ΛCDM discussed in this chapter are well known and we will follow the classic textbook by Dodelson (2003) in most of our discussion and give further references wherever appropriate. We will first recapitulate some basic notions of the homogeneous and isotropic Friedmann-Lemaître-Robertson-Walker (FLRW) universe (Friedmann, 1922, 1924; Lemaître, 1927). For the purpose of our discussion, we will describe the evolution of our universe after inflation (Guth, 1981). This initial phase, in which the vacuum energy of a quantum field causes an exponentially growing universe, was introduced to solve conceptual problems such as the horizon (Misner, 1969), flatness (Dicke, 1969), and magnetic monopole (Zel’dovich and Khlopov, 1978) problem while also serving as seeding mechanism for perturbations to a homogeneous and isotropic universe (Mukhanov and Chibisov, 1981). A detailed discussion is however going beyond the scope of this thesis.

As stated earlier, our starting point is the cosmological principle which states that our universe on large scales is statistically homogeneous and isotropic in space but can evolve in time. We will also assume that GR holds, i.e. that the evolution of our universe is described by Einstein’s equations:

\[ G_{\mu\nu} = 8\pi G T_{\mu\nu}, \]  

(1.1)

where \( G_{\mu\nu} \) is the Einstein tensor—a functional of the metric \( g_{\mu\nu} \) which describes the geometry of the universe—and \( T_{\mu\nu} \) is the stress-energy tensor describing the content of the universe.

It can then be shown (see e.g. Carroll, 1997, for a detailed discussion) that the most general solution for the metric \( g_{\mu\nu} \) in the case of a homogeneous and isotropic universe is given by the FLRW metric:

\[ ds^2 = g_{\mu\nu}^{\text{FLRW}} dx^\mu dx^\nu = -dt^2 + a^2(t) \left( \frac{dr^2}{1-kr^2} + r^2 (d\theta^2 + \sin^2 \theta d\phi^2) \right). \]  

(1.2)

The parameter \( k \) specifies the curvature of space. In this thesis, we will only consider models with a spatially flat background, i.e. with \( k = 0 \). The variable \( a(t) \) contains the time-dependence of the metric and is called the scale-factor. The metric is invariant under transformations that rescale \( a \) and \( r \) such that \( \frac{d}{dt} \) remains constant. Conventionally, one uses this freedom to fix the scale factor at present time to \( a_0 = 1 \).

Substituting the FLRW metric into Einstein’s equations, we find the Friedmann equations which describe the evolution of the scale factor \( a \) for a universe containing an ideal fluid with energy density \( \rho \) and pressure \( p \):

\[ H^2(t) \equiv \left( \frac{\dot{a}}{a} \right)^2 = \frac{8\pi G}{3} \rho, \]

(1.3)

\[ \frac{\ddot{a}}{a} = -\frac{4\pi G}{3} (\rho + 3p). \]

(1.4)
Chapter 1. Introduction

For a ΛCDM universe, the dominant fluids are described by their equation of state parameter $w$ through

$$p = w\rho.$$ (1.5)

The conservation of the energy-momentum tensor directly implies that the energy density of a homogeneous fluid with equation of state parameter $w$ scales with the scale factor $a$ as

$$\rho \propto a^{-3(1+w)}.$$ (1.6)

Relativistic fluids such as photons have an equation of state parameter of $w = \frac{1}{3}$ and hence their energy density falls off as $\rho_\gamma \propto a^{-4}$. Non-relativistic fluids such as dark matter obey $w = 0$ and therefore dilute at a slower rate with $\rho_m \propto a^{-3}$. Finally, we can treat a cosmological constant as an ideal fluid with $w = -1$ and a constant energy density $\rho_\Lambda \propto \text{const.}$

As a ΛCDM universe contains all of the aforementioned species, the possible solutions for the scale factor are parametrised by the energy density of each of the components at present time. For convenience, the energy densities are often given in units of the critical density $\rho_c$ which is simply given by solving equation (1.3) for $\rho$:

$$\rho_c = \frac{3H^2}{8\pi G}.$$ (1.7)

The density parameter $\Omega$ is then defined in terms to the energy density $\rho$ variable as:

$$\Omega \equiv \frac{\rho}{\rho_c} = \frac{8\pi G}{3H^2} \rho.$$ (1.8)

For a species $X$ with equation of state parameter $w_X$, we can now parametrise its energy density $\rho_X$ by its present day value at $a = 1$ and write it as a function of scale factor as:

$$\rho_X(a) = \rho_X(a = 1)a^{-3(1+w_X)}.$$ (1.9)

Instead of $\rho_X(a = 1)$, we use the notation that

$$\Omega_X \equiv \frac{\rho_X(a = 1)}{\rho_c(a = 1)}$$ (1.10)

denotes the corresponding density parameter value at present time. Using this definition and also denoting the Hubble parameter today as $H_0$, we can rewrite the first Friedmann equation (1.3) for ΛCDM as:

$$H = H_0\sqrt{\Omega_m a^{-3} + \Omega_\gamma a^{-4} + \Omega_\Lambda}.$$ (1.11)

We will also use the dimensionless notation $h$ for the Hubble parameter which is defined by
1.2 Evolution of Perturbations

\[ H_0 \equiv 100h \text{ km s}^{-1} \text{ Mpc}^{-1}. \]

Very soon after the discovery of the FLRW solution, observations by Hubble (1929) indicated that the universe was expanding, i.e. that \( \dot{a} > 0 \). If one neglects the inflationary period and the cosmological constant, equation (1.4) implies that the FLRW universe is decelerating, i.e. that it was expanding even faster in the past. Overall, this means that the scale factor \( a \) was growing from a starting value of 0, the Big Bang, to the value of 1 which we chose for \( a \) at present time.

As the scale factor grows, the different species dilute at different rates. When including inflation and dark energy, the \( \Lambda \)CDM universe starts with an inflationary phase that ends in a radiation dominated era. This radiation dominated era than transitions into matter domination when \( \rho_m \approx \rho_\gamma \). The last transition happened only recently with the matter energy density becoming comparable to the dark energy density \( \rho_m \approx \rho_\Lambda \).

An observationally important consequence of the Friedmann equations is the relation between distances and the scale factor. As light travels at a finite speed, objects that are observable today emitted their light in the past. Furthermore, the wavelength of the light will redshift as the universe expands, leading to the following relation between observed wavelength \( \lambda_{\text{obs}} \) and emitted wavelength \( \lambda_{\text{emit}} \):

\[ \lambda_{\text{obs}} = \frac{\lambda_{\text{emit}}}{a} \equiv (1 + z) \lambda_{\text{emit}}, \quad (1.12) \]

where we defined the redshift \( z = \frac{1}{a} - 1 \) ranging from 0 today to infinity at the big bang.

As light travels along null-geodesics, i.e. \( ds^2 = 0 \), the maximum distance it can travel radially since \( t = 0 \) defines the co-moving horizon or conformal time \( \eta \):

\[ \eta(t) = \int_0^t \frac{dt'}{a(t')}. \quad (1.13) \]

More generally, the co-moving distance of a source at scale factor \( a \) that is observable today at \( t = t_0 \) is given by:

\[ \chi(a) = \int_{t(a)}^{t_0} \frac{dt'}{a(t')} = \int_a^1 \frac{da'}{a'^2 H(a')} = \frac{1}{H_0} \int_a^1 \frac{da'}{\sqrt{\Omega_{m,0}a'^4 + \Omega_{\gamma,0} + \Omega_{\Lambda,0}a'^4}}. \quad (1.14) \]

We used the first Friedmann equation (1.3) to transform the integral over time into an integral over the scale factor.

1.2 Evolution of Perturbations

Until now, we covered the evolution of a completely homogeneous and isotropic universe. At small scales, however, the universe is composed of large volumes with very low density and dense clumps at a variety of scales such as planets, stars, galaxies, and galaxy clusters.
In ΛCDM, this structure is understood to originate from small perturbations to a completely homogeneous universe imprinted during the end of inflation. While a discussion of the details of this mechanism goes beyond the scope of this section (see e.g. Mukhanov et al., 1992, for more details), we will use it as a motivation for a perturbative analysis of the Einstein equations (1.1) around the FLRW solution.

We will start by perturbing the FLRW metric $g_{\mu \nu}^{\text{FLRW}}$ given in equation (1.2) with a small perturbation $h_{\mu \nu} \ll g_{\mu \nu}^{\text{FLRW}}$:

$$g_{\mu \nu} = g_{\mu \nu}^{\text{FLRW}} + h_{\mu \nu}. \quad (1.15)$$

As the metric needs to be symmetric, the perturbation $h_{\mu \nu}$ has 10 degrees of freedom. Those degrees of freedom can be split into four scalar, two vector, and one tensor part which all evolve independently. For the purpose of this thesis, the relevant perturbations are the scalar modes and we will only consider those in this section (see e.g. Mukhanov et al., 1992, for a discussion of vector and tensor modes). Two of the remaining scalar degrees of freedom can be fixed by choosing a specific coordinate system, i.e. they are gauge degrees of freedom. As in Dodelson (2003), we will restrict our discussion to the Newtonian gauge (see e.g. Ma and Bertschinger, 1994; Mukhanov et al., 1992, for other choices) in which the scalar perturbations are given by:

$$h_{00} = -2\dot{\Psi}(x_\mu), \quad (1.16)$$
$$h_{ii} = 2a^2\Phi(x_\mu), \quad (1.17)$$
$$h_{ij} = 0 \text{ for } i \neq j. \quad (1.18)$$

One can rewrite Einstein’s equations (1.1) to linear order in the perturbations as

$$\bar{G}^\mu_\nu + \delta G^\mu_\nu = \bar{T}^\mu_\nu + \delta T^\mu_\nu, \quad (1.19)$$

where the terms with the bar are the unperturbed FLRW terms solved by $g_{\mu \nu}^{\text{FLRW}}$ while the perturbed terms will dictate the dynamics of the perturbations. The dependence of $\delta G^\mu_\nu$ on the perturbations $\Psi$ and $\Phi$ can be determined from the Einstein tensor for the perturbed metric (1.15) (see e.g. Dodelson, 2003, chapter 5.2 for details). In Fourier space, the time-time component yields:

$$\delta G^0_0 = -\frac{6\dot{a}}{a^3}\Phi + \frac{6\dot{a}^2}{a^4}\Psi - \frac{2k^2}{a^2}\Phi \quad (1.20)$$

while the longitudinal, traceless part of the spatial components $G^i_j$ is given by:

$$\left(\hat{k}_i \hat{k}_j - \frac{1}{3} \delta^i_j\right)G^i_j = \frac{2}{3a^2}k^2(\Phi + \Psi). \quad (1.21)$$

Here and in the following, dots denote derivatives with respect to conformal time $\eta$ defined in
1.2. Evolution of Perturbations

equation (1.13). We also use \( \hat{k} \) and \( k \) to separately denote the unit vector and the magnitude of the wave vector.

The dependence of the perturbed fluids on the perturbed metric is more involved. For the homogeneous and isotropic universe, all we had to know was the dependence of the density on the scale factor for ideal fluids with different equations of state\(^1\). To understand the dynamics of a perturbed fluid, we need to study the evolution of its distribution function \( f(t, x^i, p, \hat{p}^i) \), where \( t \) is cosmic time, \( x^i \) are the spatial coordinates, \( p \) is the momentum, and \( \hat{p}^i \) is the direction of the momentum. The evolution of \( f \) is given by its Boltzmann equation:

\[
\frac{df}{dt} = C[f],
\]  

(1.22)

where \( C[f] \) is a collision term accounting for self-interactions as well as interactions between species. The perturbative analysis of the Boltzmann equation (1.22) for a perturbed fluid is different for relativistic and non-relativistic species.

**Relativistic species**

For relativistic species we assume a distribution function that is a perturbed version of the equilibrium distribution. For photons, in particular, the perturbed photon distribution function is given by the following modification of a Bose-Einstein distribution:

\[
f(x^i, p, \hat{p}^i, t) = \left( \exp\left( \frac{p}{T(t) \left( 1 + \Theta(x^i, \hat{p}^i, t) \right)} \right) - 1 \right)^{-1}.
\]  

(1.23)

In equation (1.23), \( T \) is the temperature of the uniform solution while \( \Theta \) describes the perturbations to it. To find the differential equation which governs the evolution of \( \Theta \), we need to insert equation (1.23) into equation (1.22) with a Compton-scattering collision term that accounts for interactions with free electrons. The resulting Boltzmann equation for \( \Theta \) (in Fourier space) is then given by (see e.g. Dodelson, 2003, chapter 4 for details)\(^2\):

\[
\dot{\Theta} + i k \mu \Theta + \dot{\Phi} + i k \mu \dot{\Psi} = -\dot{\tau} \left( \Theta_0 - \Theta + \mu v_b \right).
\]  

(1.24)

In equation (1.24), \( \mu \equiv \frac{k \cdot p^i}{k} \) describes the direction of the photon propagation, \( v_b \) is the velocity of the ordinary matter fluid which will be covered in more detail later, and \( \Theta_0(x^i, t) \equiv \frac{1}{4\pi} \int d\Omega \Theta(\hat{p}^i, x^i, t) \) is the monopole of the photon perturbation. In general, we define multipole moments of the photon perturbations as \( \Theta_\ell = \frac{1}{(\ell^\ell)\ell!} \int_1^1 d\mu P_\ell(\mu) \Theta(\mu) \), where \( P_\ell \) are the Legendre polynomials. The optical depth \( \tau(\eta) \) is defined as the integral over the free electron

---

\(^1\)We assumed the equation of state parameters for the different species to be given. They can, however, be derived from the unperturbed Boltzmann equation.

\(^2\)This expression neglects the effect of the polarization field and the angular dependence of Compton scattering.
density $n_e$ via:

$$\tau(\eta) \equiv \int_{\eta}^{\eta_0} d\eta' n_e \sigma_T a.$$  (1.25)

The photon contributions to the perturbed stress-energy tensor can be derived from the distribution function in equation (1.23) by taking moments:

$$T^0_0 = -g_\gamma \int \frac{d^3 p}{(2\pi)^3} E(p) f(x_i, p, \hat{p}_i, t) = -\rho_\gamma (1 + 4\Theta_0)$$  (1.26)

$$\left(\hat{k}_i \hat{k}_j - \frac{1}{3} \delta^j_i\right) T^i_j = g_\gamma \int \frac{d^3 p}{(2\pi)^3} \frac{p^2 \mu^2 - p^2/3}{E(p)} f(x_i, p, \hat{p}_i, t) = -\frac{8}{3} \rho_\gamma \Theta_2.$$  (1.27)

Similar arguments can be used for deriving the evolution of massless neutrino perturbations. Massive neutrinos that transition from relativistic to non-relativistic as the universe expands behave differently (Lesgourgues et al., 2013).

### Non-relativistic species

To study the evolution of non-relativistic species, i.e. dark matter and ordinary matter, we directly analyse moments of the Boltzmann equation (1.22). Most importantly, we study perturbations $\delta X$ to the number density $n_X$ of species $X$ (either dark matter or ordinary matter for $\Lambda$CDM):

$$n_X(x_i, t) = \int \frac{d^3 p}{(2\pi)^3} f_X(x_i, p, \hat{p}_i, t) \equiv \bar{n}_X(t) (1 + \delta_X(x_i, t)),$$  (1.28)

where $\bar{n}_X$ is the number density of the unperturbed solution. As the energy density of the non-relativistic fluid is simply mass times number density, $\delta_X$ is also the fractional overdensity $\delta \rho_X/\bar{\rho}_X$ and hence directly enters the perturbed stress energy tensor as $\delta \rho_X = \delta_X \bar{\rho}_X$.

The main difference between the Boltzmann equations for dark matter and ordinary matter are their interactions terms. In $\Lambda$CDM, dark matter is modelled as a collisionless fluid, i.e. $C[f] = 0$ for equation (1.22). Ordinary matter is composed of electrons and protons which self-interact through Coulomb scattering and also interact with the photons (dominantly through Compton scattering of photons and electrons). Working through the algebra for the zeroth moment of the Boltzmann equation (1.22)$^3$, the resulting differential equation for the number density is given by:

$$\dot{\delta}_X + i k \nu_X = -3 \dot{\Phi}.$$  (1.29)

$^3$The zeroth moment of the Boltzmann equation is given by $\int \frac{d^3 p}{(2\pi)^3} \frac{df_X}{dt} = \frac{dn_X}{dt} = \int \frac{d^3 p}{(2\pi)^3} C[f]$. For interactions that conserve number density, i.e. $\frac{dn}{dt} = 0$, the collision term is hence irrelevant for the zeroth moment.
1.2. Evolution of Perturbations

We can see that the velocity of the non-relativistic fluid enters the Boltzmann equation (1.29) for the number density perturbations. The velocity field is defined as:

\[ v_X(x_i, t) \equiv \frac{1}{n_X} \int \frac{d^3 p}{(2\pi)^3} f_X(x_i, p, \hat{p}_i, t) \frac{p\hat{p}_i}{E}. \]  

To close the set of equations for the non-relativistic species (in linear perturbation theory), it is sufficient to additionally consider the first moment of the Boltzmann equation (1.22)\(^4\):

\[ \dot{v}_X + \frac{a}{a} v_X + i k \Psi = \begin{cases} 0 & \text{for dark matter} \\ \frac{4\rho_v}{3n_X} \dot{t} (v_X + 3i\Theta_1) & \text{for ordinary matter} \end{cases}. \]  

Einstein-Boltzmann equations

We will now collect what is known as the Einstein-Boltzmann equations for ΛCDM. They describe how perturbations to the background solution of section 1.1 evolve over time and are composed of the perturbed Einstein equations and the perturbed Boltzmann equations.

The perturbed Einstein equations lead to two independent differential equations for Ψ and Φ, one from the time-time component of equation (1.19) and one from the longitudinal, traceless part of the spatial components:

\[ 2k^2 \Phi + 3 \frac{\dot{a}}{a} \left( \frac{\dot{a}}{a} - \frac{\dot{\Psi}}{a} \right) = 4\pi G a^2 \left[ 4\rho_\gamma \Theta_0 + \rho \delta + \rho_b \delta_b \right], \]  

\[ k^2 (\Phi + \Psi) = -32\pi G a^2 \rho_\gamma \Theta_2, \]  

where \( \rho, \rho_b, \rho_\gamma \) and \( \delta, \delta_b, \Theta \) respectively denote the energy density of the unperturbed solution and the perturbations of the dark matter, ordinary matter, and photons\(^5\).

The Boltzmann equations for the perturbations to the distributions of the different species\(^6\)

\(^4\)The first moment of the Boltzmann equation is given by \( \int \frac{d^3 p}{(2\pi)^3} \frac{d_f}{E} = \int \frac{d^3 p}{(2\pi)^3} C \left( \frac{P\hat{p}_i}{E} \right) \). For interactions that conserve the momentum of the species \( X \), the collision term would again be irrelevant for the evolution of \( X \). Yet, as ordinary matter interacts with photons, the collision term appears in the first order Boltzmann equation for ordinary matter.

\(^5\)We neglect the effect of neutrinos. Including massless neutrinos in the Einstein equations is as simple as splitting the radiation component into a photon and a neutrino contribution (see e.g. Dodelson, 2003, chapter 5.6).

\(^6\)We are again ignoring neutrinos, polarization and the angular dependence of Compton scattering. For the full set of equations see Dodelson (2003, chapter 4.7), for example.
are given by:

\[ \dot{\Theta} + i k \mu \Theta = -\dot{\Phi} - i k \mu \dot{\Psi} - i \left( \Theta_0 - \Theta + \mu v_b \right), \]  
(1.34)

\[ \dot{\delta} + i k v = -3\Phi, \]  
(1.35)

\[ \ddot{v} + \frac{\dot{a}}{a} v = -i k \dot{\Psi}, \]  
(1.36)

\[ \dot{\delta}_b + i k v_b = -3\Phi, \]  
(1.37)

\[ \ddot{v}_b + \frac{\dot{a}}{a} v_b = -i k \dot{\Psi} + \frac{4\rho_\gamma}{3\rho_b} \dot{\tau} (v_b + 3i \Theta_1). \]  
(1.38)

Overall, the evolution of perturbations in linear perturbation theory is therefore governed by a set of coupled, first-order differential equations. In some asymptotic limits, these equations simplify and can be solved analytically. To get the evolution over all times and all scales, however, the differential equations have to be solved numerically. Several codes that perform the integration of the Boltzmann-Einstein equations are publicly available, such as CAMB (Lewis et al., 2000) and CLASS (Blas et al., 2011).

### 1.3 Formation of Large-Scale Structure

We are interested in predicting the large-scale structure of the matter distribution of the universe at late times when it is accessible to observations. In linear perturbation theory, the formation of large-scale structure from the initial perturbations after inflation is given by the solution to the Einstein-Boltzmann equations for \( \delta \) in section 1.2. The solution for \( \delta(k, \eta) \) at late times is approximately separable into a \( k \) and \( \eta \) dependent function. Conventionally, it is therefore written as

\[ \delta(k, \eta) \propto k^2 \Phi_p(k) T(k) D(a), \]  
(1.39)

where \( \Phi_p(k) \) is the initial condition set by inflation, \( T(k) \) is the transfer function encoding the scale dependence, and \( D(a) \) is the growth function encoding the time dependence.

Perturbations on scales much larger than the co-moving horizon \( \eta \ (k\eta \ll 1) \) are set by the initial conditions and stay (almost) constant as time evolves (see e.g. Dodelson, 2003, chapter 7.2 for details). The transfer function describes how scales evolve through horizon crossing which depends on whether scales cross in matter or radiation dependence. Matter-radiation equality happens when \( \rho_m = \rho_\gamma \), corresponding to a scale factor \( a_{eq} \) given by:

\[ a_{eq} = \frac{\rho_\gamma,0}{\rho_m,0}. \]  
(1.40)

The corresponding Fourier mode is given by the inverse co-moving Hubble scale at equality:

\[ k_{eq} = a_{eq} H(a_{eq}). \]  
(1.41)
From solving the Boltzmann equation for small scale perturbations \((k\eta \gg 1)\) during radiation domination \((\rho \approx \rho_\gamma)\), we can see that the modes start growing logarithmically once they enter the horizon (see e.g. Dodelson, 2003, chapter 7.3 for details):

\[
\delta(k, \eta) \propto \log(k\eta).
\]  

During matter domination \((a \gg a_{\text{eq}})\), all the sub-horizon modes in \(\Lambda\text{CDM}\) are described by Meszaros equation (Meszaros (1974); see also Dodelson (2003, chapters 7.4 and 7.5) for details):

\[
\ddot{\delta} + \left( a \frac{d \log H}{da} + 3 \right) \dot{\delta} - \frac{3}{2} \frac{\Omega_m H_0^2}{a^3 H^2} \delta = 0,
\]  

which is independent of \(k\). Hence, all sub-horizon scales grow at the same rate given by the solution to equation (1.43), the growth function:

\[
D(a) \propto H(a) \int_0^a \frac{da'}{(a' H(a'))^3}.
\]  

Looking back at equation (1.39) for \(\delta\), we see that we also need to understand the initial conditions \(\Phi_p\) for making predictions about \(\delta\). Inflation, however, does not predict specific values for \(\Phi_p(k_i)\) but rather predicts that they are a Gaussian random field with a specific power spectrum (see e.g. Dodelson, 2003, chapter 6 for more details):

\[
\langle \Phi_p(k_i) \Phi_p(k'_j) \rangle = (2\pi)^3 P_\Phi(k) \delta^3(k_i - k'_j),
\]  

where the primordial power spectrum \(P_\Phi\) can be parametrised as:

\[
\frac{k^3}{2\pi^2} P_\Phi(k) = A_s \left( \frac{k}{k_0} \right)^{n_s - 1}.
\]  

\(A_s\) and \(n_s\) are free parameters that have to be inferred from observations and \(k_0\) is a reference scale upon which the definition of \(A_s\) depends.

The matter power spectrum predicted by linear perturbation theory is then given by

\[
\langle \delta(k_i, \eta) \delta(k'_j, \eta) \rangle = (2\pi)^3 P(k, a) \delta^3(k_i - k'_j),
\]  

where \(P(k, a)\) is given by combining equations (1.39) and (1.47):

\[
P(k, a) = \frac{2\pi^2}{k_0} A_s \left( \frac{k}{k_0} \right)^{n_s} (T(k) D(a))^2.
\]  

As we have seen, the matter perturbations \(\delta\) grow over time and eventually form non-linear structures such as galaxies and galaxy clusters. The evolution of those structures is no longer
within the reach of perturbation theory but needs to be solved by numerical simulations of the matter field. To increase the number of modes available through perturbation theory, approximating the matter distribution at scales close to the non-linear scales through higher order perturbation theory has received much attention in the past (see e.g. Bernardeau et al., 2002, for a review). The matter density field, at least on linear scales, is dominated by dark matter. From an observational perspective, a further issue is hence that dark matter is not directly observable. The two main approaches for mapping the matter distribution on large scales are therefore indirect methods either based on gravitational lensing of light by mass (see e.g. Refregier, 2003, for a review) or through the distribution of luminous matter, i.e. through the distribution of galaxies (see e.g. Anderson et al., 2014, for recent observational results).

The latter has the additional complication that galaxies are not directly tracing the matter field. Motivated by the evolution of a spherical overdensity in ΛCDM (Press and Schechter, 1974), we know that a perturbation collapses roughly when linear perturbation theory predicts an overdensity of \(\delta_{sc} \approx 1.686\). These collapsed structures virialise and form what is known as dark matter halos. These dark matter halos host the galaxies we observe thanks to their luminosity. For a given scale, the probability of an overdensity to reach the critical value \(\delta_c\) however depends on the background overdensity on larger scales (Mo and White, 1996). Consequently, the number of halos for the same overdensity on one scale depends on the overdensity of the matter field at larger scales. In the large-scale structure community, this dependence is known as the conditional mass function of dark matter halos and we will later make use of its formulation by Sheth and Tormen (2002).

If we write the number of halos of a given mass \(M\) as \(n(M)\), a consequence of the conditional mass function is that the counts of dark matter halos of a given mass are a biased version of the matter overdensity field \(\delta\):

\[
\frac{\delta n(M)}{n(M)} \approx (1 + b(M)) \delta,
\]

(1.49)

where \(\delta n(M)\) is the expected deviation of the number of halos living in the overdensity \(\delta\) from the average \(n(M)\) and \(b(M)\) is called the halo bias. Even though the expressions for the halo mass function and halo bias derived from this scheme can give good fits to the values inferred from numerical simulations, deriving precise models of biasing is an active branch of research in cosmology.

### 1.4 Cosmic Microwave Background

We will now review how the Einstein-Boltzmann equations can be used to propagate the perturbations to the photon distribution \(\Theta\) from their initial conditions set by inflation to their present day values. Due to the pressure of the photon fluid, the perturbations do not grow significantly over cosmic time and the predictions of linear perturbation theory are expected to be a good approximation over all relevant scales.

In the early universe, the photons were tightly coupled to electrons (and hence ordinary
1.4. Cosmic Microwave Background

matter) through Compton scattering. During this phase, photons and ordinary matter behave like a fluid with a characteristic speed of sound

\[ c_s \equiv \left( 3 \left( 1 + \frac{3 \rho_b}{4 \rho_\gamma} \right) \right)^{-1/2} \]  

(1.50)

which defines the sound horizon

\[ r_s(\eta) \equiv \int_0^\eta d\eta' c_s(\eta') \]  

(1.51)

e.i. the distance a soundwave travelled since its origination from a perturbation after inflation. The solution of the Einstein-Boltzmann equations in the tightly coupled regime roughly takes the form (see e.g. Dodelson, 2003, chapter 8.3 for details):

\[ \Theta_0(\eta) + \Phi(\eta) = (\Theta_0(0) + \Phi(0)) \cos(kr_s) + \cdots \]  

(1.52)

hence we expect peaks of the perturbations to form at \( k_p = \frac{n \pi}{r_s} \).

During recombination, electrons and protons formed neutral atoms, leading to a decoupling of photons and ordinary matter. Since the photons are no longer interacting with the electrons and protons, the sound horizon at decoupling imprints a particular scale in the photon and matter distribution given by \( r_* = r_s(\eta_*) \), where \( \eta_* \) is the conformal time at decoupling. Observations further show that the universe is reionised at later times. Photons and electrons, however, remain uncoupled because of the diluted energy densities of electrons and photons. As the photons stream from recombination to present time, the perturbations are damped by the interaction with the free electrons through an \( \exp(-\tau_{\text{reion}}) \) term\(^{7} \), where \( \tau_{\text{reion}} = \tau(\eta_{\text{reion}}) \) is the optical depth (1.25) to reionisation evaluated at the conformal time at which reionisation occurred. This time is not predictable within perturbation theory, and \( \tau_{\text{reion}} \) is treated as a free parameter in most CMB analyses. Other effects on the photon perturbations include diffusion damping (Silk, 1968) and the Sachs-Wolfe effect (Sachs and Wolfe, 1967), and we refer to Dodelson (2003) for a more extensive discussion.

The initial conditions for \( \Theta \) are set by inflation, so again we are interested in the power spectrum of the temperature fluctuations rather than the specific values in a given direction. We can decompose the Fourier transformed perturbations \( \Theta(k, \hat{p}) \) into a spatial component proportional to the matter overdensity \( \delta(k) \) and a second component that only contains the magnitude of \( k \) and the product \( \hat{k} \cdot \hat{p} \):

\[ \Theta(k, \hat{p}) = \frac{\Theta}{\delta}(k, \hat{k} \cdot \hat{p}) \]  

(1.53)

Only the \( \delta \) part depends on the initial conditions while the fraction contains the evolution of the temperature perturbation relative to the matter perturbation and is independent of \( \Phi_p \).

\(^{7}\)This assumes a simple reionisation model where the medium is reionized instantaneously at some redshift \( \eta_{\text{reion}} \).
The power spectrum can hence be written as
\[
\langle \Theta(k_i, \hat{p}_i, \eta_0) \Theta(k_i', \hat{p}_i', \eta_0) \rangle = (2\pi)^3 \delta^3(k_i - k_i') P(k) \frac{\Theta(k, \hat{k} \cdot \hat{p})}{\delta(k)} \frac{\Theta(k', \hat{k} \cdot \hat{p}')}{\delta(k')}. \tag{1.54}
\]

Using the definition of \( \Theta \) in equation (1.23) as the perturbation to the temperature and a solution \( \Theta(x_i, \hat{p}_i, \eta) \) of the Einstein-Boltzmann equations, the temperature of the photons is given by:
\[
T(x_i, \hat{p}_i, \eta) = T(\eta)(1 + \Theta(x_i, \hat{p}_i, \eta)), \tag{1.55}
\]
where \( T(\eta) \) is the evolution of the homogeneous solution set by \( \rho_\gamma = \frac{\pi^2}{15} T^4 \). We observe \( T \) from earth (i.e. at \( x_i,0 \)) and at present time (i.e. at \( \eta_0 \)), so we can measure \( T(x_i,0, \hat{p}_i, \eta_0) \) for different directions \( \hat{p}_i \). In order to relate the observed temperature fluctuations in different directions to the power spectrum of the field \( \Theta(x_i,0, \hat{p}_i, \eta_0) \), it is conventionally expanded in terms of spherical harmonics \( Y_{\ell m} \):
\[
\Theta_0(x_i,0, \hat{p}_i, \eta_0) = \sum_{\ell=1}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m}(x_i,0, \eta_0) Y_{\ell m}(\hat{p}_i), \tag{1.56}
\]
where the angular dependence is encoded in \( Y_{\ell m} \) and the coefficients \( a_{\ell m} \) are the observable perturbations related to the spherical harmonic transform of the observed sky. The statistic of interest is then given by the angular power spectrum \( C_\ell \):
\[
\langle a_{\ell m} a_{\ell' m'}^* \rangle = \delta_{\ell \ell'} \delta_{mm'} C_\ell. \tag{1.57}
\]

Using the definition of the \( a_{\ell m} \) values in equation (1.56), it can be shown that the \( C_\ell \) values are related to the power spectrum of \( \Theta \) through
\[
C_\ell = \frac{2}{\pi} \int_0^{\infty} k^2 P(k) \left| \frac{\Theta(\ell)}{\delta(k)} \right|^2. \tag{1.58}
\]

### 1.5 Cosmological Probes

Since its observational discovery by Penzias and Wilson (1965), the CMB evolved into the most important cosmological probe today. In Figure 1.1 (top panel), we show the angular power spectrum of the temperature fluctuations in the CMB as predicted by perturbation theory in \( \Lambda \)CDM for the best fit cosmology of data from the Wilkinson Microwave Anisotropy Probe (WMAP) (Hinshaw et al., 2013) as computed with CLASS. As expected, the angular scale of the peaks is related to the sound horizon at decoupling through \( \ell_p \approx \pi/\theta \equiv k_p \chi(\eta_*) \). The middle panel shows the relative difference in \( C_\ell^{TT} \) for 100 other cosmologies allowed by the WMAP data (blue lines). We can see that WMAP 9 pins the angular power spectrum.

\[\text{This holds for an isotropic Gaussian random field as expected from inflation. General random fields can have more complicated correlation structures.}\]
Figure 1.1: Angular power spectrum of the temperature fluctuations in the CMB $C_{\ell}^{TT}$ for the cosmological parameters that best fit WMAP 9 data (top panel) as computed with CLASS. The middle panel shows the relative difference between the best-fit model and power spectra for other parameters drawn from the WMAP 9 constraints (blue). We also show the relative difference for a 10% increase in $H_0$ (blue), $\Omega_m$ (green), $\Omega_b$ (red), $n_s$ (violet), and $A_s$ (yellow) in the bottom panel.

down to percent level at the first and second acoustic peak and allows of order 5% variation otherwise. The bottom panel shows that 10% changes in the cosmological parameters result in larger variations in the spectrum (more than 10%). Naïvely, we would hence expect that the WMAP 9 constraints on those parameters from temperature data alone should be much tighter than 10%. The changes in the spectrum for different parameters are however often very similar\(^9\), and it is crucial to include further data from CMB polarization, for example, to achieve this precision. Furthermore, including additional physical and astrophysical phenomena like neutrinos or spatial curvature into the model degrades the constraints on $\Lambda$CDM at the current precision. Complementing the CMB with further cosmological probes can alleviate this problem.

Many of the current probes other than the CMB use observations of the distance-redshift relation given in equation (1.14) in order to constrain the density parameters and $H_0$. Standard candles such as supernovae type Ia rely on a relation between apparent brightness of objects with the same intrinsic brightness at different redshifts (see e.g. Perlmutter and Schmidt, 2003, for a review). Standard rulers like Baryon Acoustic Oscillations (BAOs) rely on observations

\(^9\)The change in the spectrum for variations in $\tau_{\text{reion}}$ and $A_s$, for example, both result in an amplitude shift.
Figure 1.2: Angular diameter (left) and luminosity (right) distance for the cosmological parameters that best fit WMAP 9 data (top panel) as calculated with CLASS. The bottom panel shows the relative difference between the best-fit model and distances for other parameters drawn from the WMAP 9 constraints (blue). We also show the relative difference for a 10% increase in the $H_0$ parameter (green) and the $\Omega_m$ parameter (red).

of a fixed physical angular scale at different redshifts (see e.g. Bassett and Hlozek, 2010). In Figure 1.2, we show two variants of the co-moving distance, the angular diameter distance $d_A = a\chi$ and the luminosity distance $d_L = \chi / a$, as a function of redshift $z$ and for different cosmologies as calculated with CLASS. In the top panel, we show the best fit cosmology to CMB data from WMAP. The bottom panel shows the relative difference between the best fit solution and 100 other cosmologies allowed by the WMAP data (blue lines). We can see that, from CMB data alone, the distance redshift relation is constrained to better than 10% within $\Lambda$CDM. We finally show the relative differences for cosmologies with a 10% increase in $H_0$ (green) or $\Omega_m$ (red).

As already briefly mentioned in section 1.3, the matter distribution in the universe is dominated by dark matter, and can hence only be observed indirectly through luminous objects such as galaxies. Currently, the most important probes of the matter power spectrum are the distribution of galaxies (see e.g. Anderson et al., 2014, for recent observational results) and weak gravitational lensing (see e.g. Refregier, 2003, for a review). In Figure 1.3, we show the matter power spectrum today for different cosmologies (again computed with CLASS). The top panel shows the power spectrum for the best fit cosmology of WMAP data and we can see two distinct features of $P$: The linear matter power spectrum has a distinct peak that roughly corresponds to the equality scale $k_{eq}$ which defined small and large scales in our previous
1.5. Cosmological Probes

Figure 1.3: Matter power spectrum $P(k)$ at redshift $z = 0$ as computed with CLASS for the cosmological parameters that best fit WMAP 9 data (top panel). The middle panel shows the relative difference between the best-fit model and power spectra for other parameters drawn from the WMAP 9 constraints (blue). We also show the relative difference for a 10% increase in $H_0$ (blue), $\Omega_m$ (green), $\Omega_b$ (red), $n_s$ (violet), and $A_s$ (yellow) in the bottom panel.

discussion. The second feature are the small acoustic peaks imprinted by the fluctuations of ordinary matter before it decoupled from the photons. We will see in the discussion of the CMB in section 1.4 that the scale of the peak of those oscillations corresponds to the sound-horizon of the matter-photon plasma at recombination $r_\ast$. It is this scale that is used as a fixed angular scale for BAO observations. The middle panel shows the relative difference between the best fit solution and 100 other cosmologies allowed by the WMAP data (blue lines). We can see that the constraints from WMAP allow changes of up to 20% in the matter power spectrum. This implies that LSS observations have the potential to greatly improve our understanding of $\Lambda$CDM.
2 Quantifying Concordance with Information Gains

The current standard model of cosmology discussed in chapter 1 is in good agreement with a wealth of cosmological probes. Using Bayesian inference, the different probes can be combined to place constraints on the parameters of $\Lambda$CDM. These constraints are represented by probability distributions in parameter space. To test the concordance between different probes, we hence need a framework for comparing probability distributions.

One method to quantify the difference between the constraints from different surveys is the relative entropy, or Kullback-Leibler divergence (Kullback and Leibler, 1951), between the respective distributions. Initially motivated from information theory, the relative entropy has been proposed in the cosmology literature for forecasting and experiment design (Paykari and Jaffe, 2013; Amara and Refregier, 2014; March et al., 2011) as well as for parameter estimation and model selection (Verde et al., 2013; Kunz et al., 2006). In this chapter, we will introduce the relative entropy as a tool for measuring the information gained from individual experiments by applying it to their posteriors on the full cosmological parameter space. Two distinct cases of data combinations are analysed: adding complementary data to existing constraints and replacing data with a more accurate but correlated measurement.

The relative entropy between two posteriors measures gains in statistical precision and shifts of confidence regions at the same time. Disentangling these contributions is of great interest for detecting tensions between datasets. We will show that the relative entropy can indeed be separated into an expected part measuring the improvements in precision and a contribution from shifts in the distributions that we will call ‘Surprise’. In the limit of linear models and Gaussian likelihoods, we will derive explicit expressions for the relative entropy and its decomposition into expected relative entropy and Surprise that can be evaluated from moments of the posteriors. Motivated by the functional form of the Surprise, we analyse how we can identify directions in parameter space that are causing it.

The Gaussian limit is also useful in order to compare the Surprise to other measures of agreement between cosmological datasets that have been introduced in the past (Hobson
et al., 2002; Marshall et al., 2006; Amendola et al., 2013; Heneka et al., 2014; Martin et al., 2014; Karpenka et al., 2015; Raveri, 2015; Verde et al., 2013; MacCrann et al., 2015). We derive expressions for these different measures in the Gaussian limit and illustrate the results using a one-dimensional toy model.

This chapter is based on Seehars et al. (2014) and Seehars et al. (2016).

### 2.1 Inference in Cosmology

A typical problem in cosmology is the inference of the parameters of a cosmological model from astronomical data. In most cases this amounts to comparing observables, such as correlation functions or power spectra, that can be both measured by surveys and predicted from a model. Given such observables and a model with parameters \( \Theta = (\theta_1, \cdots, \theta_d) \), we can typically construct a likelihood function for the parameters, i.e. the probability distribution \( p(\mathcal{D}|\Theta, \mathcal{M}) \) of the data \( \mathcal{D} \) given model \( \mathcal{M} \) and its parameters \( \Theta \).

When prior information on the model parameters is available in the form of a probability density \( p(\Theta) \), Bayes’ theorem describes how to update the knowledge on \( \Theta \) by accounting for the data:

\[
p(\Theta|\mathcal{D}, \mathcal{M}) = \frac{p(\mathcal{D}|\Theta, \mathcal{M}) p(\Theta)}{\int d\Theta p(\mathcal{D}|\Theta, \mathcal{M}) p(\Theta)}, \tag{2.1}
\]

where \( p(\Theta|\mathcal{D}, \mathcal{M}) \) is called the posterior distribution of the parameters of model \( \mathcal{M} \). The denominator is often called the evidence, \( p(\mathcal{D}|\mathcal{M}) \), of the data and is equivalent to the distribution of the data anticipated from prior knowledge on the model, evaluated at the actual measurement:

\[
p(\mathcal{D}|\mathcal{M}) = \int d\Theta p(\mathcal{D}|\Theta, \mathcal{M}) p(\Theta). \tag{2.2}
\]

The evidence is often used in model comparisons as a tool to distinguish the abilities of different models \( \mathcal{M} \) to describe the data \( \mathcal{D} \). For most parts of this thesis, however, we will be dealing with only one model \( \mathcal{M} \), a flat \( \Lambda \)CDM cosmology. We will therefore suppress the dependence on the model \( \mathcal{M} \) to make the notation more compact. Our notation for likelihood, prior, posterior, and evidence is summarised in Table 2.1.
2.2 Inference for Gaussian Distributions

It is instructive to study the consequences of the earlier definitions for the special case of Gaussian distributions. Univariate Gaussian distributions are a family of probability density functions with parameters \( \bar{\theta} \) and \( \sigma \) defined as:

\[
N(\theta; \bar{\theta}, \sigma) = \frac{1}{\sqrt{2\pi} \sigma} \exp \left( -\frac{(\theta - \bar{\theta})^2}{2\sigma^2} \right).
\]  

(2.3)

Mean and variance of the normal distribution (2.3) are related to its parameters as expected:

\[
\langle \theta \rangle \equiv \int_{-\infty}^{\infty} d\theta \theta N(\theta; \bar{\theta}, \sigma) = \bar{\theta}
\]

(2.4)

\[
\langle (\theta - \bar{\theta})^2 \rangle = \int_{-\infty}^{\infty} d\theta (\theta - \bar{\theta})^2 N(\theta; \bar{\theta}, \sigma) = \sigma^2
\]

(2.5)

As we are mostly concerned with multidimensional parameter spaces, we will consider the multidimensional generalisation of (2.3), the multivariate Gaussian distributions. They are parametrised by the multivariate mean \( \bar{\Theta} = (\bar{\theta}_1, \cdots, \bar{\theta}_d) \) and the \( d \times d \) covariance matrix \( \Sigma \) and defined as:

\[
N(\Theta; \bar{\Theta}, \Sigma) = \frac{1}{\sqrt{(2\pi)^d \det|\Sigma|}} \exp \left( -\frac{1}{2} (\Theta - \bar{\Theta})^T \Sigma^{-1} (\Theta - \bar{\Theta}) \right).
\]

(2.6)

where \( d \) is the dimensionality of the parameter space. Again, it is straightforward to show that mean and covariance of \( N(\Theta; \bar{\Theta}, \Sigma) \) are given by the parameters:

\[
\langle \theta_i \rangle \equiv \int_{-\infty}^{\infty} d\Theta \theta_i N(\Theta; \bar{\Theta}, \Sigma) = \bar{\theta}_i
\]

(2.7)

\[
\langle (\theta_i - \bar{\theta}_i)(\theta_j - \bar{\theta}_j) \rangle = \int_{-\infty}^{\infty} d\Theta (\theta_i - \bar{\theta}_i)(\theta_j - \bar{\theta}_j) N(\Theta; \bar{\Theta}, \Sigma) = \Sigma_{ij}
\]

(2.8)

For the purpose of this section, we will define the likelihood as a normal distribution in the \( d_{\mathcal{D}} \) dimensional data vector \( \mathcal{D} \) with the mean given by the model prediction \( M(\Theta) \) and a covariance \( C \):

\[
p(\mathcal{D}|\Theta) \equiv N(\mathcal{D}; M(\Theta), C).
\]

(2.9)

Equivalently, we choose the prior in the \( d \) dimensional parameter space \( \Theta \) as a Gaussian with covariance \( \Sigma_p \) and mean \( \bar{\Theta}_p \):

\[
p(\Theta) \equiv N(\Theta; \bar{\Theta}_p, \Sigma_p).
\]

(2.10)

From the definition of the normal distribution in equation (2.6), we can infer that the posterior can only be normally distributed in \( \Theta \) if the model is linear, as all other functional forms
would not be compatible with the quadratic form in the exponent. We therefore consider the following model:

\[ M(\Theta) = M_0 + M\Theta, \]  

(2.11)

where \(M_0\) is a \(d_3\) dimensional vector and \(M\) is a \(d_3 \times d\) dimensional matrix that elevates the parameter vector \(\Theta\) into the \(d_3\) dimensional data space. Within cosmological applications, we can think of the linear model as a first-order Taylor expansion of the real, non-linear model around the best fit value of the prior. Whenever the prior is tight, i.e. it allows only small deviations in the parameters, this approach is expected to be a reasonable approximation.

The posterior is then defined by equation (2.1). To show that the posterior is normally distributed it is useful to define \(\delta \mathcal{D} = \mathcal{D} - M_0 \in \mathbb{R}^{d_3}\), the linear subspace \(W \subset \mathbb{R}^{d_3}\):

\[ W \equiv \{ O \in \mathbb{R}^{d_3} : \exists \Theta \in \mathbb{R}^d \text{ s.th. } O = M\Theta \}, \]  

(2.12)

and its orthogonal complement with the respect to the bilinear form \(B(x, y) \equiv x^T C^{-1} y\):

\[ W^\perp \equiv \{ O \in \mathbb{R}^{d_3} : B(O, O') = 0 \ \forall \ O' \in W \}. \]  

(2.13)

One can now decompose \(\delta \mathcal{D}\) as \(\delta \mathcal{D} = \delta \mathcal{D}^\perp + \delta \mathcal{D}^\parallel\) with \(\delta \mathcal{D}^\perp \in W^\perp\) and \(\delta \mathcal{D}^\parallel \in W\). As \(\delta \mathcal{D}^\parallel \in W\), there exists a \(\Theta_{\mathcal{D}}\) such that \(\delta \mathcal{D}^\parallel = M\Theta_{\mathcal{D}}\). In qualitative terms, we hence split up the data \(\mathcal{D}\) into a part \(\delta \mathcal{D}^\parallel\) that can be explained by the model and a part \(\delta \mathcal{D}^\perp\) which is not within the model space.

One can hence rewrite the likelihood as follows:

\[-2 \log p_L(\mathcal{D}|\Theta) \propto (\delta \mathcal{D} - M\Theta)^T C^{-1} (\delta \mathcal{D} - M\Theta) \]
\[= (\Theta_{\mathcal{D}} - \Theta)^T M^T C^{-1} M (\Theta_{\mathcal{D}} - \Theta) + (\delta \mathcal{D}^\perp)^T C^{-1} \delta \mathcal{D}^\perp \]
\[\propto \mathcal{N}(\Theta; \Theta_{\mathcal{D}}, (M^T C^{-1} M)^{-1}), \]  

(2.14)

showing that the likelihood is indeed proportional to a Gaussian in \(\Theta\) with mean \(\Theta_{\mathcal{D}}\) and covariance matrix \((M^T C^{-1} M)^{-1}\). As \(p(\Theta|\mathcal{D}) \propto p(\mathcal{D}|\Theta) p(\Theta)\), it is straightforward to show that the posterior is a normal distribution. Using that the Fourier transform of a Gaussian is of the form

\[ \mathcal{N}(X; \mu, \Sigma) = \int \frac{d^d k}{(2\pi)^d} \exp \left( -i k^T (X - \mu) \right) \exp \left( \frac{1}{2} k^T \Sigma k \right), \]

(2.15)

it is furthermore easy to calculate the evidence \(p(\mathcal{D})\) as well as mean \(\bar{\Theta}\) and covariance matrix
2.3. Information Gains

\[ \Sigma \text{ of the posterior:} \]

\[ p(\Xi) = N(\Xi; M(\bar{\Theta}_p), C + M\Sigma_p M^T), \]  
\[ \Theta = \bar{\Theta}_p + \Sigma M^T C^{-1}(\Xi - M(\bar{\Theta}_p)), \]  
\[ \Sigma_{ij} = \left(\Sigma_p^{-1} + M^T C^{-1} M\right)_{ij}^{-1}. \]  

Equation (2.18) shows how the covariance matrix of the parameters \( \Theta \) behaves in a Bayesian update: It is the reciprocal sum of the prior covariance \( \Sigma_p \) and the inverted Fisher matrix \( I = M^T C^{-1} M \). As \( I \) is positive semi-definite, the covariance of the posterior is always smaller than the prior-covariance. This motivates us to define a wide prior as the limit where \( \Sigma_p^{-1} \ll M^T C^{-1} M \). For such a wide prior, the impact of the prior on the posterior is negligible and \( \Sigma \approx \left( M^T C^{-1} M \right)^{-1} \). This means that for this definition of a wide prior, the posterior covariance attains the Cramér-Rao bound (Cramér, 1945; Rao, 1945) for an unbiased estimator with Fisher matrix \( I \).

The evidence, finally, is a Gaussian in the data with the mean given by the prior model \( M(\bar{\Theta}_p) \) and a covariance that is composed of the error on the data \( C \) and the prior uncertainty on the parameters \( \Sigma_p \). It is interesting to see that the limit of a wide prior is not useful for the evidence, as it always depends on the prior mean.

2.3 Information Gains

First defined by Kullback and Leibler (1951), the relative entropy or Kullback-Leibler divergence is an important concept in information theory. It aims at measuring differences between two probability densities. In data compression, for example, it has a clear interpretation as the number of extra bits needed when assuming a wrong distribution of the data’s alphabet (see e.g. Cover and Thomas, 2012, for more details).

In cosmology, the relative entropy has been proposed as a tool for experiment design, forecasting, and model selection. March et al. (2011) constructed a figure of merit that is related
Chapter 2. Quantifying Concordance with Information Gains

to the relative entropy in order to study the robustness of parameter constraints to possible systematic errors. Paykari and Jaffe (2013) employed a special case of the relative entropy to forecast the constraints from different survey strategies with and without sparse sampling of the sky. Amara and Refregier (2014) used the relative entropy between distributions in data space to compare the model breaking potential of different surveys. Kunz et al. (2006) applied a Bayesian measure of complexity related to relative entropy to the problem of model selection in cosmology. Verde et al. (2013) used relative entropy to study the one dimensional marginals of WMAP and Planck constraints for the parameters of $\Lambda$CDM and extensions to the basic model.

In this thesis, we focus on applying the relative entropy to the full multivariate posterior distributions in order to develop a new tool for comparing the constraints from different datasets. In order to define relative entropy, let $X$ be a continuous, $d$ dimensional random variable with probability density functions $p_1(X)$ and $p_2(X)$. The relative entropy $D(p_2||p_1)$ between $p_2$ and $p_1$ is then given by

$$D(p_2||p_1) = \int_\mathcal{S} dX \frac{p_2(X)}{p_1(X)} \log \frac{p_2(X)}{p_1(X)},$$

where $\mathcal{S}$ is the support of $p_2$. Note that the base of the logarithm in equation (2.20) has to be chosen to equal $2$ in order to interpret $D(p_2||p_1)$ as an information gain measured in bits. When evaluated for the natural logarithm the unit is called nats and can be simply transformed into bits by dividing $D$ by $\log(2)$. $D$ is finite only if $\mathcal{S}$ is contained in the support of $p_1$. Although not being symmetric in $p_1$ and $p_2$, $D(p_2||p_1)$ is often interpreted as a distance between the two distributions as it is non-negative, $D(p_2||p_1) \geq 0$, and zero if and only if $p_1 = p_2$ almost everywhere (Cover and Thomas, 2012). It is furthermore easy to see that the relative entropy is invariant under invertible transformations in $X$: as probability distributions satisfy $p(Y) = p(X)|dX/dY|$ for an invertible transformation $Y(X)$, the Jacobian matrices $dX/dY$ cancel in the logarithm and $p_2(X)dX = p_2(Y)dY$. Finally, the relative entropy is additive if $X$ can be decomposed into independent sets of variables for both $p_1$ and $p_2$ (Kullback and Leibler, 1951). Those properties make the relative entropy a convenient measure for comparing posteriors.

We will apply the relative entropy to measure the difference between posteriors and call it the ‘information gain’ when going from one posterior to another. The following section introduces the information gain in two important cases of updating the constraints on cosmological parameters. As we are dealing with posteriors from different datasets, we also introduce the shorthand notation $p_X(\Theta)$ to indicate the posterior from some dataset $\mathcal{D}_X$:

$$p_X(\Theta) \equiv p(\Theta|\mathcal{D}_X).$$

Adding complementary data

In this section, we consider the case of a sequential updating of the parameter constraints with uncorrelated or very weakly correlated datasets that complement each other. As an example,
2.3. Information Gains

One might think of updating the constraints from a CMB survey with supernova data or from low-\(\ell\) multipoles of the CMB power spectrum as measured by WMAP with high-\(\ell\) data from SPT. The problem is hence to update prior knowledge \(p_1(\Theta) = p(\Theta|\mathcal{D}_1)\) from one dataset \(\mathcal{D}_1\) using new data \(\mathcal{D}_2\) with likelihood \(p(\mathcal{D}_2|\Theta)\) via equation 2.1. When focusing on measuring information gains in such a sequential updating scheme, the quantity of interest is given by the relative entropy between prior and posterior, defined by

\[
D(p_{1,2}||p_1) = \int d\Theta \ p_{1,2}(\Theta) \log \frac{p_{1,2}(\Theta)}{p_1(\Theta)},
\]

where \(p_{1,2}(\Theta) \equiv p(\Theta|\mathcal{D}_1, \mathcal{D}_2)\). \(D\) quantifies the difference between the parameter distributions before and after updating with the new data. It can hence be interpreted as a measure for the amount by which the constraints on the model have to be changed when accounting for the new data. Due to the invariance under parameter transformations of the relative entropy, this measure does not depend on the particular parametrisation of a given model.

Replacing data

Another important case is parameter estimation from correlated datasets. A typical example of correlated datasets in cosmology are power spectra on large scales because of cosmic variance. Obviously, the likelihood of two correlated datasets cannot be described by two independent functions, preventing a joint analysis with Bayes‘ theorem as in the previous section.

Whenever a new dataset is strongly correlated but superior to old data, thinking of WMAP and Planck CMB power spectra for example, a joint analysis is typically too complex compared to the expected effects on the precision. The more accurate new data is then usually simply used to replace the correlated older data in the parameter estimation step. But there are also more complex situations of parameter estimation from correlated datasets, for example a partial replacement of data or a joint likelihood function for both datasets that correctly takes correlations into account. An example for the former is the replacement of the WMAP temperature power spectrum with new Planck data, while the latter case can be illustrated with the successive WMAP releases after one, three, five, seven, and nine years of data collection. In any of these cases, the relative entropy between the two separately analysed datasets is a useful diagnostic for measuring changes in the posteriors and detecting inconsistencies.

Usually starting from the same prior \(p(\Theta)\), the posteriors from the old likelihood \(p(\mathcal{D}_1|\Theta)\) and the new likelihood \(p(\mathcal{D}_2|\Theta)\) result in the posteriors \(p_1\) and \(p_2\) as given by equation (2.1). The quantity of interest in this case is given by

\[
D(p_2||p_1) = \int d\Theta \ p_2(\Theta) \log \frac{p_2(\Theta)}{p_1(\Theta)},
\]
Chapter 2. Quantifying Concordance with Information Gains

2.4 Information Gains for Gaussian Distributions

In the following we will study the behaviour of the information gain for two Gaussian posteriors as described in section 2.2. We will therefore assume that \( p_1 \) and \( p_2 \) are given by the following distributions:

\[
\begin{align*}
    p_1(\Theta) &= \mathcal{N}(\Theta; \hat{\Theta}_1, \Sigma_1), \\
    p_2(\Theta) &= \mathcal{N}(\Theta; \hat{\Theta}_2, \Sigma_2).
\end{align*}
\] (2.24)

Plugging the definition of the multivariate Gaussian (2.6) into the definition of the relative entropy (2.20), we find:

\[
D(p_2||p_1) = -\frac{1}{2} \left( \langle (\Theta - \hat{\Theta}_2)^T \Sigma_2^{-1} (\Theta - \hat{\Theta}_2) \rangle_{p_2} - \langle (\Theta - \hat{\Theta}_1)^T \Sigma_1^{-1} (\Theta - \hat{\Theta}_1) \rangle_{p_2} + \log \frac{\det(\Sigma_2)}{\det(\Sigma_1)} \right),
\] (2.26)

where \( \langle \cdot \rangle_{p_2} \) indicates that we average the expression within the brackets over the distribution \( p_2 \). When we write out the matrix multiplication for the first term within the bracket we find:

\[
\langle (\Theta - \hat{\Theta}_2)^T \Sigma_2^{-1} (\Theta - \hat{\Theta}_2) \rangle_{p_2} = \sum_{i,j=1}^{d} (\Sigma_2^{-1})_{ij} \langle (\Theta_i - \hat{\Theta}_{2,i})(\Theta_j - \hat{\Theta}_{2,j}) \rangle_{p_2} = \sum_{i,j=1}^{d} (\Sigma_2^{-1})_{ij} \Sigma_{2,ij} = \sum_{i} 1_{ii} = d,
\] (2.27)

where \( \mathbb{I} \) is the \( d \)-dimensional identity matrix and we used that \( \langle (\Theta_i - \hat{\Theta}_{2,i})(\Theta_j - \hat{\Theta}_{2,j}) \rangle_{p_2} \) is the definition of the covariance \( \Sigma_{2,ij} \). For the second term, we find:

\[
\langle (\Theta - \hat{\Theta}_1)^T \Sigma_1^{-1} (\Theta - \hat{\Theta}_1) \rangle_{p_2} = (\hat{\Theta}_2 - \hat{\Theta}_1)^T \Sigma_1^{-1} (\hat{\Theta}_2 - \hat{\Theta}_1) + \langle (\Theta - \hat{\Theta}_2)^T \Sigma_1^{-1} (\Theta - \hat{\Theta}_2) \rangle_{p_2},
\] (2.28)

where the first term has no averaging because it does not depend on \( \Theta \) and mixed terms like \( \langle (\Theta - \hat{\Theta}_2)^T \Sigma_1^{-1} \hat{\Theta}_1 \rangle_{p_2} \) average to zero. Using a similar calculation as in equation (2.27), we finally find:

\[
\langle (\Theta - \hat{\Theta}_2)^T \Sigma_1^{-1} (\Theta - \hat{\Theta}_2) \rangle_{p_2} = \sum_{i,j=1}^{d} (\Sigma_1^{-1})_{ij} \Sigma_{2,ji} = \text{tr}(\Sigma_1^{-1} \Sigma_2)
\] (2.29)

Putting everything together, the relative entropy between two Gaussians is then simply given by:

\[
D(p_2||p_1) = \frac{1}{2} \left( (\hat{\Theta}_2 - \hat{\Theta}_1)^T \Sigma_1^{-1} (\hat{\Theta}_2 - \hat{Theta}_1) + \text{tr}(\Sigma_1^{-1} \Sigma_2) - d - \log \frac{\det(\Sigma_2)}{\det(\Sigma_1)} \right). \tag{2.30}
\]

Depending on whether we are adding complementary data or replacing correlated data, the
distribution \( p_2 \) is either given by the joint posterior of both datasets or the posterior of data 2 alone.

The same relation is also used in Amara and Refregier (2014), while Paykari and Jaffe (2013) and March et al. (2011) restrict themselves to aligned means. As can be seen from equation (2.30), the relative entropy contains ratios between the covariance matrices \( \Sigma_2 \) and \( \Sigma_1 \) as well as a weighted difference between their means \( \bar{\Theta}_1 \) and \( \bar{\Theta}_2 \). From a parameter estimation point of view, this can be intuitively understood as contributions from an increase in the precision of the measurement and from the significance of the shifts in the central values of the constraints, respectively. Next, this distinction is made more explicit by separating the relative entropy into an expected and a Surprise part with the former measuring gains in precision and the latter quantifying the significance of the shifts in parameter space.

### 2.5 Surprise

To distinguish contributions to the information gain from gains in precision and shifts in parameter space, we can use constraints from one dataset to forecast the information gain that is expected for another dataset assuming that both are well described by the same model. Whenever the observed gains strongly differ from prior expectations, we interpret this as a systematic effect that is not explained by the model or included in the likelihood. From the posterior \( p(\Theta|D_1) \), and for consistent datasets, data \( D_2 \) is anticipated to be a realisation from

\[
p(D_2|D_1) = \int d\Theta \ p(\Theta|D_1) p(D_2|\Theta),
\]

which is sometimes called the posterior predictive distribution of \( D_1 \). In light of the posterior predictive, we can treat the relative entropy \( D \) as a function of \( D_2 \) which is a random variable with distribution \( p(D_2|D_1) \). The posterior predictive hence induces a distribution in \( D \) which is anticipated from the knowledge on the model inferred from \( D_1 \). On average, the posterior predictive \( p(D_2|D_1) \) predicts a relative entropy of

\[
\langle D \rangle = \int dD_2 \ p(D_2|D_1) D(p(\Theta|D_2)||p(\Theta|D_1)).
\]

The concept of expected relative entropy is also well known from Bayesian experimental design, where the aim is to design an experiment which maximises \( \langle D \rangle \) (Lindley, 1956).

We can hence compare the observed relative entropy \( D \) to the expected relative entropy \( \langle D \rangle \). A measure of consistency, the Surprise of the constraints derived from \( D_2 \), can consequently be defined as:

\[
S \equiv D(p(\Theta|D_2)||p(\Theta|D_1)) - \langle D \rangle.
\]

By construction, we expect \( S \) to scatter around zero. If \( S \) is positive, the posteriors are more different than expected a priori. If \( S \) is negative, the constraints are more consistent than
expected a priori. Next, we will take a look at the precise distribution of $S$ in the limit of Gaussian distributions, where all the mentioned quantities are analytic.

### 2.6 Surprise for Gaussian Distributions

In this section, we will discuss the Surprise and its distribution for two Gaussian posteriors:

$$p_1(\Theta) = p(\Theta|\mathcal{D}_1) = \mathcal{N}(\Theta; \bar{\Theta}_1, \Sigma_1),$$  \hspace{1cm} (2.34)

$$p_2(\Theta) = p(\Theta|\mathcal{D}_2) = \mathcal{N}(\Theta; \bar{\Theta}_2, \Sigma_2).$$  \hspace{1cm} (2.35)

Using the result from section 2.4, it is straightforward to calculate the relative entropy between the posteriors:

$$D(p_2||p_1) = \frac{1}{2} \left( \text{tr} \left( \Sigma_1^{-1} \Sigma_2 \right) - d - \log \frac{\det \Sigma_2}{\det \Sigma_1} + (\bar{\Theta}_2 - \bar{\Theta}_1)^T \Sigma_1^{-1} (\bar{\Theta}_2 - \bar{\Theta}_1) \right).$$  \hspace{1cm} (2.36)

We already know from section 2.2 that the posterior $p_2$ derived from prior $p$ and likelihood $p(\mathcal{D}_2|\Theta) = \mathcal{N}(\mathcal{D}_2; M(\Theta), C)$ is given by

$$p_2(\Theta) = \mathcal{N}(\Theta; \tilde{\Theta}_2, \Sigma_2),$$  \hspace{1cm} (2.37)

with

$$\Sigma_2 = \left( \Sigma_p^{-1} + M^T C^{-1} M \right)^{-1},$$  \hspace{1cm} (2.38)

$$\tilde{\Theta}_2 = \Theta_p + M^T C^{-1} (\mathcal{D}_2 - M(\tilde{\Theta}_p)).$$  \hspace{1cm} (2.39)

Hence, in the expression for the relative entropy in equation (2.36), only the last part depends on the data $\mathcal{D}_2$ via $\tilde{\Theta}_2$ from equation (2.39):

$$\Delta \equiv (\bar{\Theta}_2 - \bar{\Theta}_1)^T \Sigma_1^{-1} (\bar{\Theta}_2 - \bar{\Theta}_1).$$  \hspace{1cm} (2.40)

The distribution of $D(p_2||p_1)$ induced by the posterior predictive $p(\mathcal{D}_2|\mathcal{D}_1)$ is then equivalent to the distribution of $\Delta$ up to a $\mathcal{D}_2$ independent constant. For the Surprise, defined in equation (2.33), these constant terms cancel and it is given by the following expression in the Gaussian limit:

$$S = \frac{1}{2} \left( (\bar{\Theta}_2 - \bar{\Theta}_1)^T \Sigma_1^{-1} (\bar{\Theta}_2 - \bar{\Theta}_1) - \langle (\bar{\Theta}_2 - \bar{\Theta}_1)^T \Sigma_1^{-1} (\bar{\Theta}_2 - \bar{\Theta}_1) \rangle_{p(\mathcal{D}_2|\mathcal{D}_1)} \right),$$

$$= \frac{1}{2} \left( \Delta - \langle \Delta \rangle_{p(\mathcal{D}_2|\mathcal{D}_1)} \right).$$  \hspace{1cm} (2.41)

Equation (2.41) is a quite intuitive measure, as it measures the shift in the means $(\bar{\Theta}_2 - \bar{\Theta}_1)$ relative to the covariance matrix $\Sigma_1$ of the prior information and is hence a natural generalisation of the typical one-dimensional notion of $\Delta \mu/\sigma$ tensions.
To understand the distribution of $\Delta$, we need to look into the posterior predictive for $D_2$ derived from the posterior $p_1$ and the likelihood $p(D_2|\Theta)$. It is given by a slight modification of the evidence (2.16) from section 2.2:

$$p(D_2|D_1) = \int d\Theta p(D_2|\Theta)p_1(\Theta) = N(D_2; M(\bar{\Theta}_1), C + M\Sigma_1M^T). \quad (2.42)$$

To find the function form of $\langle \Delta \rangle$, we first find:

$$\bar{\Theta}_2 - \bar{\Theta}_1 \equiv \delta \Theta = \hat{\Theta}_p + \Sigma_2 M^T C^{-1} (D_2 - M(\hat{\Theta}_p)) - \bar{\Theta}_1$$

$$= \Sigma_2 \left( M^T C^{-1} (D - F(\hat{\Theta}_p)) + \Sigma_2^{-1}(\hat{\Theta}_q - \bar{\Theta}_1) \right)$$

$$= \Sigma_2 \left( M^T C^{-1} (D - F(\hat{\Theta}_p)) + (\Sigma_2^{-1} + M^T C^{-1} M)(\hat{\Theta}_p - \bar{\Theta}_1) \right)$$

$$= \Sigma_2 M^T C^{-1} (D - F(\bar{\Theta}_1) + MQ^{-1} T), \quad (2.43)$$

with

$$Q \equiv M^T C^{-1} M = \Sigma_2^{-1} - \Sigma_p^{-1}, \quad (2.44)$$

$$T \equiv \Sigma_p^{-1}(\hat{\Theta}_q - \bar{\Theta}_1). \quad (2.45)$$

Plugging $\delta \Theta$ into $\Delta$, one finds

$$\Delta = \delta \Theta \Sigma_1^{-1} \delta \Theta = X^T AX, \quad (2.46)$$

with

$$X \equiv D_2 - M(\bar{\Theta}_1) + MQ^{-1} T, \quad (2.47)$$

$$A \equiv C^{-1} MW M^T C^{-1}, \quad (2.48)$$

$$W \equiv \Sigma_2 \Sigma_1^{-1} \Sigma_2. \quad (2.49)$$

As the posterior predictive for $D_2$ is a normal distribution with mean $M(\bar{\Theta}_1)$ and covariance matrix $C + M\Sigma_1M^T$, the new variable $X$ is also distributed as a normal distribution with mean $\mu$ and covariance matrix $\Sigma$ now given by

$$\mu = MQ^{-1} T, \quad (2.50)$$

$$\Sigma = C + M\Sigma_1M^T. \quad (2.51)$$

The distribution of $\Delta$ is therefore the distribution of a quadratic form in $X$ where $X$ is normally distributed and is usually called a generalised $\chi^2$ distribution which is completely defined by $A$, $\Sigma$, and $\mu$. In fact, the distribution of $\Delta$ can be completely described by the eigenvalues of the matrix $A\Sigma$ and $\mu$. 


Using textbook results for generalised $\chi^2$ distributions, the moments of $\Delta$ can be easily derived. According to Mathai and Provost (1992, chapter 3.2b), mean and variance of $\Delta$ are given by

$$\langle \Delta \rangle_{p(D_2|D_1)} = \text{tr}(A\Sigma) + \mu^T A\mu,$$

$$\langle (\Delta - \langle \Delta \rangle)^2 \rangle_{p(D_2|D_1)} = 2\text{tr}((A\Sigma)^2) + 4\mu^T A\Sigma A\mu.$$  (2.52) (2.53)

Using definitions (2.48), (2.50), and (2.51) for $A$, $\mu$, and $\Sigma$ one finds

$$E(\Delta) = \text{tr}(QW + QWQ\Sigma_1) + T^TWT,$$  (2.54)

and

$$\text{Var}(\Delta) = 2\text{tr}((QW + QWQ\Sigma_1)^2) + 4T^TW(Q + Q\Sigma_1 Q)WT.$$  (2.55)

Note that while $A$ and $\Sigma$ are matrices in data space, $Q$, $W$, $T$, and $\Sigma_1$ are all in parameter space and hence functions of the first two moments of $p_1$, $p_2$, and $p$ only. Finally, one finds for $\langle D \rangle$ and $\sigma^2(D)$:

$$\langle D \rangle = \frac{1}{2} \left( \text{tr}\{\Sigma_1^{-1}\Sigma_2\} - d - \log\frac{\det\Sigma_2}{\det\Sigma_1} + E(\Delta) \right),$$  (2.56)

$$\sigma^2(D) = \frac{1}{4} \text{Var}(\Delta).$$  (2.57)

**Adding complementary data**

When updating constraints with complementary data as outlined in section 2.3, the relative entropy between prior and posterior is considered. In this case $p$ is identical with $p_1$, resulting in $T = 0$ and $\Sigma_p = \Sigma_1$. As mentioned before, the generalised $\chi^2$ distribution of $S$ is defined through the eigenvalues of $A\Sigma$ and $\mu$. The non-centrality parameter $\mu = 0$ as it depends linearly on $T$. The matrix $A\Sigma$ is living in data space. However, since the eigenvalues of a product of matrices are independent under cyclic permutations of the product, we find that the eigenvalues of the matrix $A\Sigma$ are given by the eigenvalues of the following matrix:

$$\text{eig}(A\Sigma) = \text{eig}(I - \Sigma_1^{-1}\Sigma_2).$$  (2.58)

Note that the matrices on the left hand side of equation (2.58) are living in parameter space. The properties of the distribution of the relative entropy hence only depend on the covariances of the posteriors $p_1$ and $p_2$.

The expected value and variance of the relative entropy when adding complementary data are
then given by:

\[
\langle D \rangle = -\frac{1}{2} \log \det \Sigma_2 / \det \Sigma_1, \\
\sigma^2(D) = \frac{1}{2} \text{tr} \left( (\Sigma_1^{-1} \Sigma_2 - \mathbb{1})^2 \right).
\]

(2.59)  
(2.60)

As the square root of the determinant of a covariance matrix describes the volume of the 1-\(\sigma\) confidence region for a multivariate Gaussian, the expected relative entropy behaves just as expected: The smaller the determinant of \(\Sigma_2\) is relative to the determinant of \(\Sigma_1\), the more information we expect to gain from the new data.

To get some intuition for the relative entropy, we consider two special cases. In the extreme case where \(\mathcal{D}_2\) does not add any information, i.e. its Fisher matrix \(M^T C^{-1} M\) is identically zero, \(\Sigma_2\) is equal to \(\Sigma_1\) according to equation (2.38) and the expected information gain becomes zero. The second interesting case is the combination of two experiments with equal constraining power. In this case, the Fisher matrix is given by \(M^T C^{-1} M = \Sigma_1^{-1}\) and \(\Sigma_2 = \Sigma_1/2\). The expected relative entropy from such a repetition is therefore given by \(\langle D \rangle = \log 2/2\) in nats or by 1/2 in bits.

Using the result for \(\langle D \rangle\), the Surprise is given by:

\[
S = \frac{1}{2} \left( (\tilde{\Theta}_2 - \tilde{\Theta}_1)^T \Sigma_1^{-1} (\tilde{\Theta}_2 - \tilde{\Theta}_1) - \text{tr}(\mathbb{1} - \Sigma_1^{-1} \Sigma_2) \right)
\]

(2.61)

The first part measures the deviation between the mean \(\tilde{\Theta}_1\) of \(p_1\) and the mean \(\tilde{\Theta}_2\) of the joint posterior \(p_{1,2}\). The second part gives the amount of deviation we expect a priori for consistent datasets. In the limits considered earlier, we see that we do not expect any deviations in the means when adding non-informative data such that \(\Sigma_2 = \Sigma_1\). When combining two experiments with equal constraining power, i.e. when \(\Sigma_2 = \Sigma_1/2\), the expected deviations are given by \(d/2\), where \(d\) is the dimensionality of the parameter space.

**Replacing data**

When comparing the results of two datasets that replace each other, the relative entropy between the two separately analysed posteriors is of interest. Considering a wide prior for the derivation of \(p_1\) and \(p_2\), all terms containing \(\Sigma_p^{-1}\) are small compared to the terms independent of \(p\), resulting in \(Q = \Sigma_2^{-1}\) and \(T \to 0\). The eigenvalues of the matrix \(A \Sigma\) are then given by the eigenvalues of the following matrix:

\[
\text{eig}(A \Sigma) = \text{eig}(\mathbb{1} + \Sigma_1^{-1} \Sigma_2),
\]

(2.62)

which again lives in parameter space.
Chapter 2. Quantifying Concordance with Information Gains

For the expected relative entropy and variance we find for this scenario:

\[
\langle D \rangle \approx -\frac{1}{2} \log \det \Sigma_2 \det \Sigma_1^{-2} + \text{tr}(\Sigma_2 \Sigma_1^{-1}),
\]

\[
\sigma^2(D) \approx \frac{1}{2} \text{tr}((\Sigma_1^{-1} \Sigma_2 + 1)^2).
\]

We again consider the special case of two experiments with equal constraining power for illustration. When comparing the two separately derived posteriors, this means that both \(p_1\) and \(p_2\) have the same covariance \(\Sigma_1 = \Sigma_2\). The expected relative entropy from such a comparison is therefore given by \(\langle D \rangle = d\) in nats, where \(d\) is the dimensionality of the parameter space.

The Surprise for two separately analysed posteriors is given by:

\[
S = \frac{1}{2} \left( (\bar{\Theta}_2 - \bar{\Theta}_1)^T \Sigma_1^{-1} (\bar{\Theta}_2 - \bar{\Theta}_1) - \text{tr}(\mathbb{1} + \Sigma_1^{-1} \Sigma_2) \right)
\]

In the limit of two experiments with equal constraining power, i.e. when \(\Sigma_2 = \Sigma_1\), the expected deviations are hence given by \(2d\), where \(d\) is the dimensionality of the parameter space.

Reparameterisations motivated by the Surprise

As already mentioned earlier, the relative entropy, and hence the Surprise, are independent of the parametrisation of the model. Indeed, equation (2.41) suggests a simple reparameterisation of the parameter space when comparing constraints. An intuitive choice is a space \(\Psi\) in which \(p(\Theta|\mathcal{D}_1)\) has parameters that are uncorrelated, have unit variance, and mean zero. In this space, \(\tilde{\Sigma}_1 = 1\) by construction and the Surprise is given by

\[
S = \frac{1}{2} \left( \tilde{\mu}_2^T \tilde{\mu}_2 - \text{tr}(\mathbb{1} + \tilde{\Sigma}_2) \right),
\]

where \(\sim\) indicates that the moments are now taken in the new space \(\Psi\). Each shift away from zero in \(\Psi\) independently contributes to \(S\). As the constraints \(p(\Psi|\mathcal{D}_1)\) are very simple, tensions between Gaussian constraints can then be simply visualised by shifts in the one-dimensional marginal distributions of the new parameters \(\Psi\).

2.7 Other Measures of Concordance

The agreement between datasets within a given model can be estimated from the evidence in data space or the posterior in model space. In this section, we summarise two evidence and one posterior based method that have been applied to cosmology in the past.

The most common tool (Marshall et al., 2006; Amendola et al., 2013; Heneka et al., 2014; Martin et al., 2014; Karpenka et al., 2015; Raveri, 2015) for assessing the agreement between
cosmological observations was introduced by Marshall et al. (2006). It is defined as

\[ R \equiv \frac{p(D_1, D_2)}{p(D_1)p(D_2)}, \tag{2.67} \]

with the idea that it compares the evidence for \( D_1 \) and \( D_2 \) when both have to be described by the same parameters of the model (numerator) to the evidence when each dataset is allowed to be described by different parameter values (denominator). The value \( R \) is then interpreted on Jeffrey’s scale, indicating consistent datasets when \( R > 1 \) and inconsistencies otherwise. It can be used to compare multi-dimensional, non-Gaussian constraints and numerical techniques for estimating the Bayesian evidence are well established.

To analyse the agreement between constraints on the Hubble constant and the age of the universe from low-redshift measurements and CMB observations, Verde et al. (2013) use a modified version of the \( R \) measure. The idea is to perform a translation in the parameters \( \Theta \) of one of the likelihoods \( p(D_i|\Theta) \), with \( \tilde{p}(D_i|\Theta) \) being the shifted distribution. For the translation where the maxima of the two distributions coincide, the evidence is written as

\[ \tilde{p}(D_1, D_2)\big|_{\max D_1=\max D_2} = \int d\Theta \tilde{p}(D_2|\Theta)p(\Theta|D_1), \tag{2.68} \]

where we chose to shift \( \tilde{p}(D_2|\Theta) \) in this example. The measure of tension by Verde et al. (2013) is then defined as

\[ T \equiv \frac{\tilde{p}(D_1, D_2)\big|_{\max D_1=\max D_2}}{p(D_1, D_2)}. \tag{2.69} \]

The intuition is that \( \tilde{p}(D_1, D_2)\big|_{\max D_1=\max D_2} \) is the evidence for maximally consistent data \( D_1 \) and \( D_2 \), to which the actual evidence is compared. \( T \) is interpreted on a slightly modified Jeffrey’s scale indicating tensions when \( \log T > 1 \).

MacCrann et al. (2015) define a measure of consistency between two datasets in parameter space by looking at the best-fit point \( \Theta_{\text{joint}} \) of the joint constraints and its likelihood \( p(D_i|\Theta_{\text{joint}}) \) given either \( D_1 \) or \( D_2 \). Within a sample \( \{\Theta_i\}_{i=1}^N \) from the posterior \( p(\Theta|D_i) \) one can compute the percentile of likelihoods \( p(D_i|\Theta_i) \) that are smaller than \( p(D_i|\Theta_{\text{joint}}) \). This percentile is then interpreted as the confidence for the two datasets being consistent. This technique is able to compare multi-dimensional, non-Gaussian constraints. The \( p \)-value of the best-fit point can be estimated from Monte Carlo Markov chains (MCMCs) of the individual posteriors.

### 2.8 Other Measures for Gaussian Distributions

Formally, we again consider two experiments that are described by data \( D_i \) and likelihood \( p(D_i|\Theta) \) where \( \Theta \) are the parameters of a model for the data and \( i = 1, 2 \). As pointed out in section 2.1, both experiments can put individual constraints on the parameters of the model...
Chapter 2. Quantifying Concordance with Information Gains

via Bayes' theorem

\[ p_i \equiv p(\Theta | \mathcal{D}_i) = \frac{p(\mathcal{D}_i | \Theta) p(\Theta)}{p(\mathcal{D}_i)}, \]  

(2.70)

Both sets of data can also be used to put joint constraints on the model parameters. This either requires knowledge of the joint likelihood \( L(\Theta; \mathcal{D}_1, \mathcal{D}_2) = p(\mathcal{D}_1, \mathcal{D}_2 | \Theta) \) or independence of the measurement in the sense that the joint likelihood can be factorised \( L(\Theta; \mathcal{D}_1, \mathcal{D}_2) = L_1(\Theta; \mathcal{D}_1) L_2(\Theta; \mathcal{D}_2) \). If independence is assumed, the joint posterior is given by:

\[ p_{1,2} \equiv p(\Theta | \mathcal{D}_1, \mathcal{D}_2) = \frac{L_1(\Theta; \mathcal{D}_1) L_2(\Theta; \mathcal{D}_2) p(\Theta)}{p(\mathcal{D}_1, \mathcal{D}_2)}. \]  

(2.71)

It is worth noting that this is formally equivalent to using the constraints from one set of data as a prior for the analysis of the other. For the evidence, a similar statement holds:

\[ p(\mathcal{D}_1, \mathcal{D}_2) = \int d\Theta p(\mathcal{D}_1 | \Theta) p(\mathcal{D}_2 | \Theta) p(\Theta) = p(\mathcal{D}_1) \int d\Theta p(\mathcal{D}_2 | \Theta) p(\Theta | \mathcal{D}_1) \equiv p(\mathcal{D}_1) p(\mathcal{D}_2 | \mathcal{D}_1). \]  

(2.72)

For the case of a Gaussian prior with mean \( \bar{\Theta}_p \) and covariance \( \Sigma_p \), a linear model for the data \( M_i(\Theta) = M_{0,i} + M_i \Theta \), and a Gaussian likelihood with covariance \( C_i \), we showed in section 2.2 that the posteriors are Gaussian with moments:

\[ \Sigma_i = \left( \Sigma_p^{-1} + M_i^T C_i^{-1} M_i \right)^{-1}, \]  

(2.73)

\[ \bar{\Theta}_i = \bar{\Theta}_p + \Sigma_i M_i^T C_i^{-1} (\mathcal{D}_i - M(\bar{\Theta}_p)) \]  

(2.74)

Also \( p(\mathcal{D}_i) \) is Gaussian (in the data) with mean \( M(\bar{\Theta}_p) \) and covariance \( C_i + M_i \Sigma_p M_i^T \) and we again use the following notation for this:

\[ p(\mathcal{D}_i) = \mathcal{N}(\mathcal{D}_i; M(\bar{\Theta}_p), C_i + M_i \Sigma_p M_i^T) \]  

(2.75)

We now have all the ingredients to study the various techniques outlined in section 2.7 in more detail.

We start with the R measure by Marshall et al. (2006), defined as

\[ R = \frac{p(\mathcal{D}_1, \mathcal{D}_2)}{p(\mathcal{D}_1) p(\mathcal{D}_2)}. \]  

(2.76)

Using equation (2.72), we can rewrite R as

\[ R = \frac{p(\mathcal{D}_2 | \mathcal{D}_1)}{p(\mathcal{D}_2)}. \]  

(2.77)
2.8. Other Measures for Gaussian Distributions

i.e. as the ratio between the evidence for $\mathcal{D}_2$ given the posterior from $\mathcal{D}_1$ to the evidence for $\mathcal{D}_2$ from the prior. In the linear Gaussian model, we know both distributions:

$$p(\mathcal{D}_2|\mathcal{D}_1) = \mathcal{N}(\mathcal{D}_2; M_2(\hat{\Theta}_1), C_2 + M_2\Sigma_1M_2^T)$$ (2.78)

$$p(\mathcal{D}_2) = \mathcal{N}(\mathcal{D}_2; M_2(\hat{\Theta}_p), C_2 + M_2\Sigma_pM_2^T)$$ (2.79)

Using this, we can rewrite the logarithm of $R$ as

$$\log R = -\frac{1}{2} \left[ \log \frac{\det(C_2 + M_2\Sigma_1M_2^T)}{\det(C_2 + M_2\Sigma_pM_2^T)} + (\mathcal{D}_2 - M_2(\hat{\Theta}_1))^T(C_2 + M_2\Sigma_1M_2^T)^{-1}(\mathcal{D}_2 - M_2(\hat{\Theta}_1)) 
\quad - (\mathcal{D}_2 - M_2(\hat{\Theta}_p))^T(C_2 + M_2\Sigma_pM_2^T)^{-1}(\mathcal{D}_2 - M_2(\hat{\Theta}_p)) \right] .$$ (2.80)

Equation (2.80) hence depends on the difference between the covariances of $p(\mathcal{D}_2|\mathcal{D}_1)$ and $p(\mathcal{D}_2)$ as well as on the difference between the model of prior and posterior means $\hat{\Theta}_p$ and $\hat{\Theta}_1$, and the data $\mathcal{D}_2$. Note in particular that $\log R$ depends on the prior moments $\hat{\Theta}_p$ and $\Sigma_p$ because of $p(\mathcal{D}_2)$. This is a well known problem of the evidence when used with a wide prior that is supposed to be uninformative. On average, i.e. when calculating the expected value of (2.80) under $p(\mathcal{D}_1, \mathcal{D}_2)$, all the data dependent terms cancel each other and we find:

$$\langle \log R \rangle \equiv \int d\mathcal{D}_1 d\mathcal{D}_2 p(\mathcal{D}_1, \mathcal{D}_2) \log R = -\frac{1}{2} \log \frac{\det(C_2 + M_2\Sigma_1M_2^T)}{\det(C_2 + M_2\Sigma_pM_2^T)} .$$ (2.81)

We point out that even though the expression for $\langle \log R \rangle$ has a similar form as $\langle D \rangle$, its matrices live in data space rather than parameter space.

The second evidence-based measure by Verde et al. (2013) is defined as:

$$T = \frac{\tilde{p}(\mathcal{D}_1, \mathcal{D}_2)|_{\max\mathcal{D}_1=\max\mathcal{D}_2}}{p(\mathcal{D}_1, \mathcal{D}_2)} ,$$ (2.82)

where $\tilde{p}(\mathcal{D}_1, \mathcal{D}_2)|_{\max\mathcal{D}_1=\max\mathcal{D}_2}$ is defined as the evidence when shifting $\Theta$ in $p(\mathcal{D}_2|\Theta)$ such that the maximum of $p(\Theta|\mathcal{D}_1)$ is at the same position in parameter space as the maximum of $p(\Theta|\mathcal{D}_2)$ (see section 2.7 for more details).

In the linear Gaussian model, shifting the parameter in the likelihood $p(\mathcal{D}_2|\Theta)$ is equivalent to changing the linear model for $\mathcal{D}_2$ to $M'_2(\Theta) = M_{0,2} + M_2(\Theta + \Theta')$. We have to choose $\Theta'$ such that the mean of the new posterior $\hat{\Theta}'_2$ is equal to $\hat{\Theta}_1$:

$$\hat{\Theta}'_2 = \hat{\Theta}_1 + \Sigma_2M'_2C^{-1}_2(\mathcal{D}_2 - M'_2(\hat{\Theta}_1)) \overset{1}{=} \hat{\Theta}_1 .$$ (2.83)

Solving for $\Theta'$, we find:

$$\Theta' = \Sigma_2M'_2C^{-1}_2(\mathcal{D}_2 - M_2(\Theta_1))$$ (2.84)
Chapter 2. Quantifying Concordance with Information Gains

Plugging $\Theta'$ it into (2.82), we find that $T$ evaluates to:

$$\log T = \frac{1}{2}(\mathcal{D}_2 - M_2(\Theta_1))^T K(\mathcal{D}_2 - M_2(\Theta_1)), \quad (2.85)$$

where

$$K = C_2^{-1} M_2 \Sigma_2 (\Sigma_1 + \Sigma_2)^{-1} \Sigma_2 M_2^T C_2^{-1}. \quad (2.86)$$

The measure $T$ hence measures the difference between the data $\mathcal{D}_2$ and the prediction for the data from the posterior of $\mathcal{D}_1$. Averaging (2.85) over $p(\mathcal{D}_1, \mathcal{D}_2)$, we find that this measure is expected to vary around the following value for consistent datasets:

$$\langle \log T \rangle = \frac{d}{2}, \quad (2.87)$$

where $d$ is the dimensionality of the parameter space $\Theta$.

Finally, we take a look at the measure by MacCrann et al. (2015) which is the p-value of the best-fit point of the joint analysis on the posteriors of the individual constraints. Without loss of generality, we consider the distribution of the best-fit point of the joint distribution as compared to the distribution of the posterior $p(\Theta|\mathcal{D}_1)$. In the linear Gaussian model, the best fit point is simply given by the mean of $p_{1,2}$:

$$\hat{\Theta}_{1,2} = \hat{\Theta}_1 + \Sigma_{1,2} M_2^T C_2^{-1} (\mathcal{D}_2 - F_2(\hat{\Theta}_1)) \quad (2.88)$$

with $\Sigma_{1,2} = (\Sigma_1^{-1} + M_2^T C_2^{-1} M_2)^{-1}$ and $\hat{\Theta}_1$ and $\Sigma_1$ being mean and covariance of $p(\Theta|\mathcal{D}_1)$. Under $p(\mathcal{D}_2|\mathcal{D}_1)$, $\hat{\Theta}_{1,2}$ is expected to be normally distributed with mean $\hat{\Theta}_1$ and covariance $\Sigma_1 - \Sigma_{1,2}$:

$$\hat{\Theta}_{1,2} \sim N(\Theta; \hat{\Theta}_1, \Sigma_1 - \Sigma_{1,2}). \quad (2.89)$$

Using the p-value of $\Theta_{1,2}$ under $p_1$ is hence a good approximation for the significance of the shift only if $\Sigma_{1,2} \ll \Sigma_1$, i.e. when the joint posterior is much tighter than the individual constraints, and is otherwise underestimating it.

### 2.9 Comparison of Measures for One-Dimensional Toy-Model

As we have seen, it is instructive to study the details of the different measures for the agreement between constraints in the limit of a linear model and a Gaussian likelihood and prior. In this case, posteriors and evidences can be evaluated analytically based on a relation between the data covariance and parameter covariance given by the Fisher matrix. For comparing the measures in this section, we will illustrate our results with an even simpler, one-dimensional toy model. For this we assume that we collect two independent data points $d_1$ and $d_2$ which
measure the same parameter $\theta$ and follow a Gaussian likelihood:

$$p(d_i|\theta) = \frac{1}{\sqrt{2\pi s_i^2}} \exp \left\{ -\frac{1}{2} \left( \frac{d_i - \theta}{s_i} \right)^2 \right\} \equiv \mathcal{N}(d_i; \theta, s_i),$$

(2.90)

where $s_i$ is the standard deviation corresponding to the Gaussian error on the $d_i$ measurement.

We assume that we start with some Gaussian prior distribution for $\theta$ with mean $\bar{\theta}_p$ and variance $\sigma_p^2$. We also assume that $\sigma_p^2 \gg \sigma_i^2$, i.e. that the prior is weak compared to the measurements.

In this case, the posterior is again a univariate Gaussian and given by

$$p(\theta|d_i) \approx \mathcal{N}(\theta; \bar{\theta}_i, \sigma_i),$$

(2.91)

where $\bar{\theta}_i = d_i$ and $\sigma_i = s_i$ for this simple toy model. Even though posterior and likelihood are identical in this case, we will keep distinguishing between $\bar{\theta}_i$ and $d_i$ in order to make the difference between data and parameter space more apparent.

**Surprise**

For the Surprise, this one-dimensional toy model is now a simple limit of equation (2.41). In the case when $d_2$ is used to update $p(\theta|d_1)$, $S$ is given by

$$S = \frac{1}{2} \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \left( \frac{(\bar{\theta}_1 - \bar{\theta}_2)^2}{\sigma_1^2 + \sigma_2^2} - 1 \right).$$

(2.92)

When $d_2$ is used to derive the posterior separately from $d_1$, the toy model results in:

$$S = \frac{1}{2} \frac{(\bar{\theta}_1 - \bar{\theta}_2)^2}{\sigma_1^2} - \frac{\sigma_2^2 + \sigma_1^2}{\sigma_1^2}. $$

(2.93)

In this toy model, the Surprise hence quantifies the difference between $d_1$ and $d_2$ by comparing their mean constraints $\bar{\theta}_1$ and $\bar{\theta}_2$ on the model. The error by which the difference is weighted depends on the scenario: deviations of $\bar{\theta}_1$ and $\bar{\theta}_2$ are compared to $\sigma_1^2 + \sigma_2^2$ for separately analysed posteriors and to $\sigma_1^2$ only for joint analyses. The second summand of both expressions ensures that the Surprise is 0 on average.

**Measure $R$ by Marshall et al. (2006)**

For the one-dimensional toy model, $R$ evaluates to the following relation:\

$$\log R = -\frac{1}{2} \left[ \log \left( \frac{s_2^2 + \sigma_1^2}{\sigma_p^2} \right) + \frac{(d_2 - \bar{\theta}_1)^2}{s_2^2 + \sigma_1^2} - \frac{(d_2 - \bar{\theta}_p)^2}{\sigma_p^2} \right].$$

(2.94)

\[\text{The asymmetry between $d_1$ and $d_2$ in this expression is an artefact of the limit of a wide prior. As the expressions are much simpler in this limit we decided to anyways show these results. The general expression in 2.8 does not show this asymmetry and all conclusions are not affected by the limit.}\]
The terms in the bottom row of equation (2.94) compare $d_2$ to the mean model $\tilde{\theta}_1$ from the posterior of $d_1$ and to the mean model $\tilde{\theta}_p$ of the prior. For consistent datasets, i.e. when $d_1$ and $d_2$ are drawn from $p(d_1, d_2)$, we expect that those cancel each other on average. The term in the top row, however, is data independent and affected only by changes in the variance. For this example, $\log R$ is therefore expected to vary about the variance dependent term. Hence, the weaker the prior, the more one tends to underestimate the tensions when interpreting $\log R$ relative to 1. The example shows that interpreting the ratio independent of specific properties of prior and posterior can be non-trivial, as deviations from $R = 1$ can be attributed to inconsistencies between the data and differences between variances at the same time.

**Measure $T$ by Verde et al. (2013)**

Looking at our toy model, we find for $\log T$

$$\log T = \frac{1}{2} \frac{(d_2 - \tilde{\theta}_1)^2}{\sigma_1^2 + \sigma_2^2}. \quad (2.95)$$

Similarly to $R$, this measure compares the deviations between $d_2$ and the mean model for the data from the posterior of $d_1$. It however does neither depend on the prior nor show the data independent variance term, implying that $\log T$ is fairly problem independent for this toy model. On average, we expect this expression to be of order 1, or of order $d$ for general, $d$-dimensional parameter spaces, for consistent data. We hence find that $\log T$ tends to larger values as the size of the parameter space increases. Interpreting it relative to 1 will lead to an overestimation of tensions for parameter spaces with dimensionality greater than 2.

**Measure by MacCrann et al. (2015)**

For the toy model, the best fit point of the joint distribution is simply the mean $\tilde{\theta}_{1,2}$ of the joint posterior. We look at the expected distribution of $\tilde{\theta}_{1,2}$ as predicted by the posterior $p(\theta|d_1)$ of $d_1$ and the likelihood $p(d_2|\theta)$ assuming that the model correctly describes both sets of data. We find that, for consistent data, $\tilde{\theta}_{1,2}$ is expected to be drawn from the following normal distribution:

$$\tilde{\theta}_{1,2} \sim N \left( \theta; \tilde{\theta}_1, \sigma_1 \sqrt{\frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}} \right). \quad (2.96)$$

Comparing the distribution for $\tilde{\theta}_{1,2}$ (2.96) to the posterior $p(\theta|d_1)$ of $d_1$ (2.91), we find that even though the means coincide, the expected standard deviation of $\tilde{\theta}_{1,2}$ is smaller than $\sigma_1$. We hence expect that, on average, the likelihood $p(d_1|\tilde{\theta}_{1,2})$ of the best fit point is larger than the likelihood of a sample from the posterior $p(\theta|d_1)$. The percentile of $p(d_1|\tilde{\theta}_{1,2})$ will therefore underestimate the tension introduced by the update. Only in the limit where $\sigma_1^2 \gg \sigma_2^2$, i.e.
where the constraints from $d_1$ are weak compared to the joint constraints, is $\bar{\theta}_{1,2}$ expected to be distributed approximately according to $p(\theta|d_1)$.

### 2.10 Discussion

In order to compare the cosmological parameter constraints from different experiments, a tool for quantifying changes in posterior distributions on the full parameter space is needed. Motivated from information theory, the concept of relative entropy measures differences between distributions in a parametrisation independent way and is therefore able to quantify the information gained from new data. In this chapter, we used the relative entropy to develop a new tool for comparing the parameter constraints of the $\Lambda$CDM model from different observations. Two ways of combining data from different experiments are discussed: complementary datasets that can be analysed sequentially and correlated measurements that replace earlier datasets. We saw that the relative entropy captures both changes in confidence volumes and locations of the posteriors. These contributions can be distinguished as an expected relative entropy measuring differences in confidence volume and a Surprise coming from shifts in parameter space. The notions of expected relative entropy and Surprise turn the relative entropy into a powerful diagnostic for the consistency of datasets.

In a short summary, we have furthermore seen that it is possible to define and estimate a variety of other measures of agreement between general, non-Gaussian constraints. Interpreting them on a fixed, problem-independent scale can however over- or underestimate the tensions in the data. A potential way out is to focus on situations where analytic results from a linear Gaussian model are a good approximation. Later we will see that this limit tends to be a good description for flat $\Lambda$CDM constraints as soon as CMB data from Planck or WMAP is involved. In these cases, the distributions and expectations for most of the reviewed measures can be derived.

The Surprise is particularly useful in this limit. Being derived from a general framework that measures both tensions and gains in precision between posterior distributions in the same units of bits, it can handle Bayesian updates as well as independently analysed data. Operating in parameter space, it furthermore only depends on mean and covariance of prior and posterior. In the next chapter, we will see that those can be robustly estimated from standard MCMC samples. Finally, the Surprise follows a relatively simple distribution given by the weighted sum of $d \chi^2$-distributed variables, where $d$ is the dimension of the parameter space. Given again only the covariance of prior and posterior, the $p$-value of the observed Surprise can be evaluated numerically.

A challenging problem when comparing constraints are correlations between datasets. In the current scheme, we treat correlated datasets by comparing the separately derived posteriors rather than the joint constraints. Since the correlations in the data lead to correlations in the posteriors, this approach however overestimates the expected information gains and underestimates potential tensions. A simulation based approach as in Larson et al. (2015)
could help to understand the effect of correlations in further analyses.

To conclude, we find that the relative entropy is a valuable diagnostic to compare constraints from different measurements. To distinguish contributions from gains in precision and systematic shifts in parameter space, we can decompose the relative entropy into an expected information gain and the Surprise. The strength of the Surprise is its interpretability for well constrained models. As more cosmological probes are able to put tight constraints on $\Lambda$CDM parameters, estimating the Surprise between the constraints can help to detect systematic issues in data or model in the future.
3 Current Observations and their Information Gains

Observational evidence in support of the ΛCDM model has grown steadily. Though some of the key ingredients of the model, including Dark Matter and Dark Energy, are not fully understood, an impressive array of new experiments show findings consistent with predictions of the model. Chief among the datasets are high-precision measurements of the Cosmic Microwave Background (CMB) (Wright et al., 1996; MacTavish et al., 2008; Bennett et al., 2013; Story et al., 2013; Das et al., 2011). This area has received significant attention recently with the release of the cosmological analysis of data from the Planck satellite (Planck Collaboration, 2015c).

In this chapter, we apply the concepts introduced in chapter 2 to a historic sequence of CMB observations as well as additional data from cosmological probes at low redshifts such as the distance ladder (Efstathiou, 2014), supernovae type Ia (Conley et al., 2011), Baryon Acoustic Oscillations (BAOs) (Beutler et al., 2011; Anderson et al., 2014; Kazin et al., 2014; Ross et al., 2015), and weak gravitational lensing (Kilbinger et al., 2013).

We first cover some details on computational inference with Monte Carlo Markov chains and how those algorithms can be used to estimate the information gain from a given dataset. Using our own Monte Carlo Markov chain framework CosmoHammer (Akeret et al., 2013), we will then proceed to estimate the relative entropy, its expected, and its Surprise contributions for different combinations of cosmological datasets.

This chapter is based on Seehars et al. (2014) and Seehars et al. (2016). It further summarises results from Grandis et al. (2016b) and Akeret et al. (2013).

3.1 Inference with Monte Carlo Markov Chains

In section 2.1 we discussed the principles of Bayesian inference in cosmology. We highlighted that the quantity of interest is the posterior distribution $p(\Theta|\mathcal{D})$ of the parameters $\Theta$ of the
Chapter 3. Current Observations and their Information Gains

cosmological model:

\[ p(\Theta|\mathcal{D}) = \frac{p(\mathcal{D}|\Theta) p(\Theta)}{\int d\Theta p(\mathcal{D}|\Theta) p(\Theta)}, \]  

(3.1)

where, as in section 2.1, \( p(\mathcal{D}|\Theta) \) is the likelihood of the data given the model and \( p(\Theta) \) is a prior distribution on the parameters. When both the model for the data and the likelihood are known analytically, we can in principle calculate the posterior for the parameters analytically. One such example was explicitly outlined in section 2.2, where we derived the posterior for a Gaussian likelihood and a linear model.

In cosmology, however, the model is rarely known as an analytical function of the parameters. Rather, there are ways to numerically solve the model for a particular parametrisation. In this case, we need ways to estimate the posterior from a number of numerical evaluations of the likelihood. For small numbers of parameters, for example, the posterior can be estimated by calculating the likelihood on a grid of parameters that fills the interesting region of the parameter space. As the number of evaluations in this case scales exponential in the number of parameters, this method typically becomes too expensive for more than 3 dimensions.

Computational Bayesian inference therefore typically relies on stochastic processes where the likelihood is evaluated on a random set of points rather than a regular grid. These algorithms are called Markov chain Monte Carlo (MCMC) algorithms, and there exist many different flavours for computational Bayesian inference (see for example Brooks et al. (2011) for an overview). Monte Carlo Markov chains have been first applied to cosmology for the analysis of CMB data by Christensen et al. (2001) and Knox et al. (2001). A popular package for MCMC analyses of a wide range of datasets called CosmoMC was made available to the community by Lewis and Bridle (2002). These days, many other codes and frameworks for MCMC analyses in cosmology exist (see e.g. Akeret et al. (2013); Audren et al. (2013); Das and Souradeep (2014); Zuntz et al. (2015)). Adding another level of complication, there are cases where even the likelihood cannot be reasonably approximated by an analytic form. For those cases, there exist promising likelihood-free MCMC methods such as approximate Bayesian computing (ABC) that rely on simulations for the data based on the theoretical models. A discussion of ABC methods goes beyond the scope of this thesis, but we successfully applied ABC to the problem of forward modelling of imaging data in Akeret et al. (2015). In this section, we will briefly review the classic Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970) before discussing our own CosmoHammer framework (Akeret et al., 2013) for parallelised MCMC inference.

**Metropolis-Hastings algorithm**

As mentioned earlier, the basic problem we would like to tackle with MCMC algorithms is the efficient estimation of probability density functions \( p(\Theta) \), called target distribution in the following, that can only be evaluated point wise for specific parameter values \( \Theta \), where the
3.1. Inference with Monte Carlo Markov Chains

Figure 3.1: Illustration of a Monte Carlo Markov chain using the Metropolis-Hastings algorithm with a Gaussian proposal distribution $q(\Theta)$. Also the target distribution is chosen to be Gaussian in this example.

parameter space of $\Theta$ can be high-dimensional. When the target distribution is a posterior $p(\Theta|D)$, yet another complication is that the normalisation of the target, given by the evidence $p(D)$, is unknown as it depends on the likelihood via an integration over the prior.

MCMC algorithms rely on a stochastic process in parameter space, i.e. a sequence of positions $\{\Theta_t\}_{t=0 \ldots N}$, which has a stationary distribution that is identical to the target distribution. This stochastic process is constructed such that each new position $\Theta_t$ only depends on its previous position $\Theta_{t-1}$ and is therefore called a Markov chain. A classic textbook for further information is MacKay (2003). In this section, we focus on a specific construction of such a process called the Metropolis-Hastings algorithm.

Given an initial position $\Theta_0$, the Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970) is an iterative procedure for generating $\Theta_{t+1}$ from $\Theta_t$ such that the stationary distribution is given by the target distribution $p(\Theta)$. In order to determine the next position, the Metropolis-Hastings algorithm employs a proposal distribution $q(\Theta|\Theta_0)$ from which a new position $\Theta_q$ is proposed. The proposed position $\Theta_q$ is accepted as the next position in the Markov chain with probability:

$$P_{\text{accept}} = \min\left(1, \frac{p(\Theta_q) q(\Theta_t|\Theta_q)}{p(\Theta_t) q(\Theta_q|\Theta_t)}\right).$$ (3.2)

If $\Theta_q$ is accepted, the new position $\Theta_{t+1}$ is given by $\Theta_q$. Otherwise, the new position is simply given by the old position: $\Theta_{t+1} = \Theta_t$. The chain $\{\Theta_t\}_{t=0 \ldots N}$ converges to a sample from the target distribution $p(\Theta)$ for large $N$ (Metropolis et al., 1953; Hastings, 1970). It is easy to see from equation (3.2) that this algorithm only depends on ratios of $p(\Theta)$ evaluations and is hence independent of a $\Theta$-independent normalisation.

In Figure 3.1 we show an illustration of such a process for a Gaussian target distribution. The result of the Metropolis-Hastings algorithm is a correlated sample from the posterior distribution. In section 3.2 we will discuss how those samples can be used to estimate properties of the target distribution. We will find that the errors on those estimates depend only on the size
Chapter 3. Current Observations and their Information Gains

Initial positions: Split walkers into 2 groups at random.
Update positions of group 1: First, randomly select a partner from group 2. Second, propose positions on the connecting rays. Accept and reject the positions as in the Metropolis-Hastings algorithm. Apply the same process to group 2 and iterate the algorithm. Distribution of positions after 1'000 iterations.

Figure 3.2: Illustration of a Monte Carlo Markov chain using the Goodman-Weare algorithm as implemented by Foreman-Mackey et al. (2013). The target distribution is chosen to be Gaussian in this example.

Figure 3.2: Illustration of a Monte Carlo Markov chain using the Goodman-Weare algorithm as implemented by Foreman-Mackey et al. (2013). The target distribution is chosen to be Gaussian in this example.

of the sample and the correlation strength within the chain, but not on the dimensionality of the parameter space. The less correlated the chain, the more efficient is the sampling.

It is worth highlighting that the efficiency of the Metropolis-Hastings algorithm crucially depends on the choice of proposal distribution. The closer the proposal is to the target distribution, the less positions have to be rejected by the algorithm and the smaller are the correlations in the sample. At the same time, we need a proposal from which we can easily draw random numbers. A typical choice is hence a Gaussian proposal distribution that is as close to the target as possible. The higher the dimensionality of the parameter space, the more drastic is the effect of the choice of proposal on the efficiency.

**Goodman-Weare algorithm**

It is evident from the previous section that computational inference with the Metropolis-Hastings algorithm is an iterative procedure, typically iterated thousands of times. Whenever the model or the likelihood is computationally expensive, i.e. each step in the chain costs a significant amount of time, this leads to long wall-times. As the inference step in cosmology is typically repeated many times for different combinations of data and a wide range of cosmological models, long wall-times of the inference step can become a limiting factor in the data analysis. We will in the following outline how a new algorithm by Goodman and Weare (2010), implemented by Foreman-Mackey et al. (2013), can be used to greatly decrease the wall-time of the inference step. We will also briefly present our own CosmoHammer framework for parallelised MCMC inference which greatly simplifies the distribution of the inference step over thousands of cores on computer clusters.

The idea of the algorithm by Goodman and Weare (2010) is to use an ensemble of stochastic processes, called walkers in the following. At each step of the chain, each walker proposes a new position in the chain that is randomly generated from the positions of the rest of the walkers. In Figure 3.2 we show an illustration of a slight modification of this algorithm as implemented by Foreman-Mackey et al. (2013).
Instead of a single initial position, the Goodman-Weare algorithm hence starts with an ensemble of positions \( \Theta_0, i \) \( i = 1, \ldots, n_{\text{walkers}} \). To update walker \( i \), we select a partner \( j \) at random and sample another random value \( z \) from the distribution

\[
q(z) = \begin{cases} 
\frac{1}{\sqrt{2\pi}} \text{ if } z \in \left[ \frac{1}{a}, a \right], \\
0 \text{ otherwise}
\end{cases}
\]

where \( a \) is the only tuning parameter of the algorithm. Given walker \( i \) at position \( t \) in the chain, a new step \( \Theta_{q, i} \) is then proposed as

\[
\Theta_{q, i} = \Theta_{t, j} + z(\Theta_{t, i} - \Theta_{t, j})
\]

and accepted with probability

\[
P_{\text{accept}} = \min\left(1, \frac{z^d}{d} \frac{p(\Theta_{q, i})}{p(\Theta_{t, i})}\right),
\]

where \( d \) is the dimensionality of the parameter space. As before, the new position \( \Theta_{t+1, i} \) is given by \( \Theta_{q, i} \). Otherwise, the new position is simply given by the old position: \( \Theta_{t+1, i} = \Theta_{t, i} \).

The algorithm by Goodman and Weare (2010) is still a serial process, but in the implementation by Foreman-Mackey et al. (2013), called \texttt{emcee}, half of the walkers can be updated in parallel. Additionally, it reduces the tuning of the algorithm from finding a full proposal distribution to the problem of choosing a proper value for \( a \). This is achieved by relying on the distribution of the ensemble of walkers, which is by construction expected to be a good approximation to the target distribution.

**Parallelised MCMC inference with CosmoHammer**

In the \texttt{CosmoHammer} framework, we exploited the parallelisability of the implementation by Foreman-Mackey et al. (2013) by combining it with a powerful and modular chaining design for the evaluation of the likelihood. Each inference step is defined by a set of modules that either calculate the theoretical predictions or the likelihood of the parameters proposed by \texttt{emcee}. The user simply adds all relevant modules to a likelihood chain which calls each module and adds up all likelihoods when it is queried with a parameter value. The whole process is illustrated in Figure 3.3. When multiple cores or even multiple compute nodes are available, \texttt{CosmoHammer} distributes the evaluation of the likelihood chain for the individual walkers over all cores at each step in the MCMC process with minimal effort by the user. The code is written in \texttt{Python} and publicly available at https://github.com/cosmo-ethz/CosmoHammer. As we have shown in Akeret et al. (2013), using the \texttt{CosmoHammer} for parameter estimation from CMB data of the WMAP team, for example, can speed up the inference step from a wall time of 30 hours on a notebook to 16 minutes on 2048 cores.
3.2 Estimating Information Gains and Surprise

Once we have a sample of parameter values \( \{\Theta_i\} \) from the posterior, we need to turn our attention to study how we can use them for estimating basic properties such as covariance and mean as well as the information gain and eventually even the Surprise. We start with a brief discussion of the analysis of MCMC chains in general and proceed to the estimation of information gain and Surprise in the Gaussian limit as well as for general distributions.

Analysis of MCMC chains

We want to study the properties of a sample \( \{\Theta_i\} \) from a distribution \( p(\Theta) \) that was generated with an MCMC algorithm. Typically, this means that we have some function \( f \) of the parameter \( \Theta \) and we would like to know its expectation value under the target distribution:

\[
\bar{f} \equiv \int d\Theta \, f(\Theta) \, p(\Theta).
\] (3.6)
3.2. Estimating Information Gains and Surprise

We will again denote the components of the multivariate $\Theta$ by $\theta_i$. Special cases for $f$ are given by $f(\Theta) = \theta_i$ for estimating the mean $\bar{\theta}_i$ of $\theta_i$ and $f(\Theta) = (\theta_i - \bar{\theta}_i)(\theta_j - \bar{\theta}_j)$ for estimating the covariance $\Sigma_{ij}$ of $p(\Theta)$.

We will call the following expression the Monte Carlo estimate for the expectation value of $f$:

$$\hat{f} = \frac{1}{n} \sum_{i=1}^{n} f(\Theta_i) \approx \int d\Theta f(\Theta) p(\Theta). \quad (3.7)$$

The autocorrelation function $C(t)$ of the stochastic process $f_i = f(\Theta_i)$ quantifies the correlation between two points in the sample separated by $t$ steps:

$$C(t) \equiv \langle (f_i - \bar{f})(f_{i+t} - \bar{f}) \rangle, \quad (3.8)$$

The autocorrelation function further defines the number of iterations $\tau$ which separate two approximately uncorrelated samples:

$$\tau = 1 + 2 \sum_{t=1}^{\infty} \frac{C(t)}{C(0)}. \quad (3.9)$$

We will call $\tau$ the autocorrelation time of $\{f_i\}$ in the following. In the limit of an uncorrelated process, i.e. $C(t) = 0$ for $t > 0$, we find that $\tau = 1$.

Looking at the variance of the estimator $\hat{f}$ from equation (3.7), we find that its error is given by:

$$\text{Var}(\hat{f}) = \frac{1}{N/\tau} \text{Var}(f). \quad (3.10)$$

In the limit of no correlations, i.e. $\tau = 1$, we hence recover the standard expression for the error in the mean. For $\tau > 1$, however, we see that the correlations in the chain effectively reduce the number of samples by $\frac{1}{\tau}$. The larger $\tau$, the longer do we need to run the MCMC process to achieve a certain error on $\hat{f}$.

In addition to the autocorrelation in the process, a further complication of the analysis of MCMC chains is the initialisation bias. As we typically do not know how to create an unbiased sample from the posterior before we run the MCMC process (this is exactly why we run it after all), we usually start the process at a biased position in parameter space. The initial phase of the sampling process is therefore biased towards the initialisation. To avoid this bias, it is customary to discard the initial samples in the estimation process. To get an idea for how long the initialisation takes, the autocorrelation is again a useful measure: As two steps in the chain become approximately uncorrelated after $\tau$ iterations, the initialisation should be approximately forgotten after a few autocorrelation times. For more details on the analysis of MCMC chains, we refer to the excellent lecture notes by Sokal (1989).
Estimation of information gain and Surprise in the Gaussian limit

In section 2.4, we showed that the relative entropy between two Gaussian distributions \( p_1(\Theta) = \mathcal{N}(\bar{\Theta}_1, \Sigma_1) \) and \( p_2(\Theta) = \mathcal{N}(\bar{\Theta}_2, \Sigma_2) \) is given by

\[
D(p_1 \| p_2) = \frac{1}{2} \left( \text{tr}(\Sigma_2^{-1} \Sigma_1) - d - \log \det(\Sigma_2^{-1} \Sigma_1) + (\bar{\Theta}_1 - \bar{\Theta}_2)^T \Sigma_2^{-1} (\bar{\Theta}_1 - \bar{\Theta}_2) \right). \tag{3.11}
\]

As discussed in the previous section, it is straightforward to estimate mean \( \mu \) and covariance matrix \( \Sigma \) of a distribution \( p(\Theta) \) from an MCMC sample \( \{\Theta_i\}_{i=1}^N \):

\[
\hat{\Theta} = \frac{1}{N} \sum_{i=1}^N \Theta_i \approx \bar{\Theta},
\tag{3.12}
\]

\[
\hat{\Sigma}_{kl} = \frac{1}{N} \sum_{i=1}^N (\theta_{i,k} - \bar{\theta}_k)(\theta_{i,l} - \bar{\theta}_l).
\tag{3.13}
\]

It is left to plug the resulting estimates for mean and covariance of prior and posterior into equation (3.11) to find an estimate for the relative entropy. Similarly, expected relative entropy and Surprise in the Gaussian case are completely determined by mean and covariance of the posteriors (see section 2.6).

As also shown in section 2.6, the Surprise \( S \) follows a generalised \( \chi^2 \) distribution and can be rewritten as a weighted sum of \( \chi^2 \) variables \( Z_i \) with one degree of freedom,

\[
S = \frac{1}{2} \left( \sum_{i=1}^d \lambda_i Z_i - \text{tr} \left( \mathbb{1} \pm \Sigma_2 \Sigma_1^{-1} \right) \right)
= \frac{1}{2} \sum_{i=1}^d \lambda_i (Z_i - 1)
\tag{3.14}
\]

where \( \lambda_i \) are the eigenvalues of \( \mathbb{1} \pm \Sigma_2 \Sigma_1^{-1} \) and \( d \) is the dimensionality of the parameter space. The \( \pm \) depends on whether data was added or replaced in the update.

Given a particular Surprise value \( S \), one can use the cumulative distribution of the generalised \( \chi^2 \) random variable defined in equation (3.14) to calculate the probability for measuring a Surprise that deviates from zero by more than \( S \). This probability is the so-called \( p \)-value for the hypothesis that the posteriors are derived from datasets which are consistent within the model. If the \( p \)-value is small, the datasets are unlikely to be consistent within the model. To compute the \( p \)-value for \( S \) from mean and covariance of the posteriors, one can use an algorithm by Davies (1980), for instance, implemented in the R package CompQuadForm by Duchesne and De Micheaux (2010).

We provide a module for calculating the relative entropy, the Surprise and its \( p \)-value in the linear Gaussian case at https://github.com/seeh/surprise.
3.2. Estimating Information Gains and Surprise

Estimation of information gain and surprise for general distributions

In general, the relative entropy between two distributions $p_1$ and $p_2$ is given by the following integral:

$$D(p_1 || p_2) = \int d\Theta \ p_1(\Theta) \log \frac{p_1(\Theta)}{p_2(\Theta)}.$$  \hspace{1cm} (3.15)

Using a MCMC sample $\{\Theta_i\}_{i=1,\ldots,N}$ from $p_1$, the relative entropy can hence be estimated as

$$\hat{D}(p_1 || p_2) = \frac{1}{N} \sum_{i=1}^{N} \log \frac{p_1(\Theta_i)}{p_2(\Theta_i)}. \hspace{1cm} (3.16)$$

When the distributions of interest are posteriors from cosmological applications, however, they can be numerically evaluated only up to a normalisation factor by calculating the product of prior and likelihood:

$$\tilde{p}_i(\Theta) = p(D_i | \Theta) p(\Theta) \propto p_i(\Theta).$$  \hspace{1cm} (3.17)

Before we are able to calculate the estimate (3.16), it is hence left to estimate the normalisation of $p_1$ and $p_2$ which is given by the evidence:

$$p(D_i) = \int d\Theta p(D_i | \Theta) p(\Theta).$$  \hspace{1cm} (3.18)

Estimating the evidence directly from MCMC samples is notoriously difficult, as the naïve estimator

$$\hat{p}(D_i) = \frac{1}{N} \sum_{i=1}^{N} p(D_i | \Theta_i) p(\Theta_i)$$  \hspace{1cm} (3.19)

typically suffers from large variance. An alternative algorithm for numerically estimating $p_i(D)$ is nested sampling (Skilling, 2004), or brute-force Monte Carlo integration. Once an estimate $\hat{p}_i(D)$ is calculated, the relative entropy for general distributions can hence be estimated by

$$\hat{D}(p_1 || p_2) = \frac{1}{N} \sum_{i=1}^{N} \log \frac{\hat{p}_1(\Theta_i)}{\hat{p}_2(\Theta_i)} + \log \frac{\hat{p}(D_2)}{\hat{p}(D_1)}.$$  \hspace{1cm} (3.20)

For the results in section 3.4, we use CosmoHammer to create MCMC samples of $p_1$. Furthermore, we employ a Monte Carlo integrator to evaluate $\int d\Theta \hat{p}_i$ over a five-sigma region of $p_i$, where the integral boundaries are estimated from covariance matrix and mean of the MCMC samples.

To estimate the expected relative entropy, we can build from expertise from Bayesian experimental design. Many algorithms have been proposed to perform this optimisation in general non-linear cases (see e.g. Drovandi et al., 2013; Long et al., 2013; Huan and Marzouk,
Chapter 3. Current Observations and their Information Gains

2013). Huan and Marzouk (2013), for example, rewrite the expected relative entropy for a joint analysis of $D_1$ and $D_2$ as:

$$
\langle D \rangle = \int dD_2 \int d\Theta \left( \log p(D_2|\Theta) - \log p(D_2|D_1) \right) p(D_2|\Theta) p(\Theta|D_1).
$$

Given a sample of size $N$ from $p(\Theta|D_1)$, $\langle D \rangle$ can then be estimated via

$$
\langle D \rangle \approx \frac{1}{N} \sum_{i=1}^{N} \left( \log p(D_i^2|\Theta_i) - \log p(D_i^2|D_1) \right),
$$

$$
p(D_i^2|D_1) \approx \frac{1}{N} \sum_{j=1}^{N} p(D_j^2|\Theta_j),
$$

where $D_i^2$ is a sample from $p(D_i^2|\Theta_i)$ for each $\Theta_i$ of the sample from $p(\Theta|D_1)$. Together with approaches for calculating the relative entropy between non-Gaussian priors and posteriors (see e.g. Skilling, 2004), this approach allows for efficient calculation of the Surprise in general, non-Gaussian cases. In order to interpret the observed Surprise, one would however also need to estimate the relative entropy for each data $D_i^2$ to estimate the expected distribution of the Surprise.

A final complication arises when the parameter space of the likelihood contains parameters which are not of cosmological interest, so called nuisance parameters. In most cases, marginalisation of nuisance parameters in the Monte Carlo estimate is straightforward and simply amounts to ignoring the additional parameters. In equations (3.20) and (3.23), however, we rely on the fact that we can estimate the likelihood as a function of the cosmological parameters only. This is no longer true in the case of additional nuisance parameters. A particularly promising procedure for estimating $D$ and $S$ in this situation based on the Gaussianisation of general, non-Gaussian posteriors was recently applied in Grandis et al. (2016a).

3.3 CMB Surveys

The anisotropies in the cosmic microwave background (CMB) are a prediction of $\Lambda$CDM and inflationary models and are measured to great accuracy. First observed in the 1960s, precision cosmology from CMB observations started with the Cosmic Background Explorer (COBE) launched in 1989 and found its temporary conclusion with the recent publication of the Planck results. The observable which is of most cosmological interest in CMB observations is the power spectrum of the temperature and polarisation fluctuations on the sky.

Considering the measured CMB temperature anisotropies as an example, one wants to evaluate the power spectrum of $\frac{\delta T}{T}(\hat{n})$, i.e. the deviation $\delta T$ from the average temperature $T$ in direction $\hat{n}$. As already mentioned in section 1.4, it turns out to be useful to expand the temperature fluctuations in terms of spherical harmonics: $\frac{\delta T}{T}(\hat{n}) = \sum_{\ell=1}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\hat{n})$. 

50
The multipoles $C_\ell$ of the temperature power spectrum can then be calculated via

$$\left\langle a_{\ell m} a_{\ell' m'}^* \right\rangle = \delta_{\ell \ell'} \delta_{mm'} C_\ell,$$  \hspace{1cm} (3.24)

where the average is over all possible CMB realisations. In an experiment, the $a_{\ell m}$ values can be estimated from the measured map of temperature anisotropies. The estimator for the observed multipole $C^\text{obs}_\ell$ is given by $C^\text{obs}_\ell = \frac{1}{2\ell+1} \sum_{-\ell \leq m \leq \ell} |a^\text{obs}_{\ell m}|^2$. $C^\text{obs}_\ell$ has two sources of uncertainty: The measurement errors on the $a^\text{obs}_{\ell m}$s and the cosmic variance arising from the fact that we can observe only one CMB realisation and are hence unable to do the averaging from equation (3.24).

Once the estimates for $C^\text{obs}_\ell$ and its errors are obtained, the next step is to construct a likelihood for the cosmological parameters. As the power spectrum is depending on the cosmological parameters in a non-trivial way, the input of such a likelihood function is usually a numerically evaluated power spectrum $C^\text{mod}_\ell$ predicted by the parametrisation of the cosmological model:

$$p(C^\text{obs}_\ell | \Theta) = p(C^\text{obs}_\ell | C^\text{mod}_\ell (\Theta)).$$  \hspace{1cm} (3.25)

The likelihood as a function of the parameters $\Theta$ can hence only be evaluated numerically for a specific choice of parameters and is unknown in its analytic form.

We work in the six-dimensional flat $\Lambda$CDM parameter space given by the Hubble constant $H_0$, the cold dark matter density today $\Omega_c h^2$, the baryonic matter density $\Omega_b h^2$, the optical depth to reionisation $\tau$, and the amplitude $A_s$ and spectral index $n_s$ of the power spectrum of primordial curvature fluctuations.

In this work, the power spectra and likelihoods of four observations are considered: Starting with the BOOMERanG data (Jones et al., 2006; Piacentini et al., 2006; Montroy et al., 2006), the cosmological parameter constraints are updated with Wilkinson Microwave Anisotropy Probe (WMAP) data (Bennett et al., 2013; Hinshaw et al., 2013), South Pole Telescope (SPT) data (Story et al., 2013), and finally Planck data (Planck Collaboration, 2014b, 2015d). The temperature power spectra are shown in Figure 3.4. In the following, all four datasets will be briefly discussed.

**BOOMERanG**

The BOOMERanG data was collected during an antarctic balloon flight in 2003 (Masi et al., 2006). The telescope which was attached to the balloon measured both temperature and polarisation of the CMB over three sky regions of 90, 750, and 300 square degrees and in three wide frequency bands centred at 145, 245, and 345 GHz. The power spectra are estimated over the range $75 \leq \ell \leq 1400$ in temperature and $150 \leq \ell \leq 1000$ in polarisation and cosmic variance limited for $\ell < 375$ (Jones et al., 2006; Piacentini et al., 2006; Montroy et al., 2006). The likelihood is a log-normal distribution in the temperature power spectrum and a normal distribution in polarisation. It is numerically marginalised over a calibration factor and the
Chapter 3. Current Observations and their Information Gains

Figure 3.4: Power spectra of BOOMERanG, WMAP (9-year release), SPT, and Planck (2013 release). The shaded regions show the $\ell$-range in which the data is limited by cosmic variance.

size of the beam with an iterated Gauss-Legendre quadrature.

**WMAP**

The all-sky survey of the CMB by the WMAP satellite had a duration of 9 years, resulting in a measurement of both temperature and polarisation power spectrum of the CMB over the range $2 \leq \ell \leq 1200$ in temperature and $2 \leq \ell \leq 1000$ in polarisation (Bennett et al., 2013; Hinshaw et al., 2013). The measurements are limited by cosmic variance for all $\ell < 457$. Together with the data, the WMAP team published a likelihood code which is used in this work without modifications. The data releases after three (Spergel et al., 2007), five (Dunkley et al., 2009), seven (Larson et al., 2011), and nine years (Bennett et al., 2013; Hinshaw et al., 2013) are considered.

**SPT**

The measurements of the ground based South Pole Telescope focus on the small scale fluctuations of the CMB. Estimated from data of 2500 square degrees, the publicly available
3.4. Concordance of CMB Surveys before Planck

temperature power spectrum ranges from \(650 \leq \ell \leq 3000\) (Story et al., 2013). As the CMB signal is contaminated by foreground effects on these scales, templates have to be removed from the observed power spectrum in order to model the foreground effects. This results in three parameters—accounting for the amplitude of the power from Poisson distributed point sources, clustered point sources, and Sundaev-Zel’dovich clusters—that additionally enter the Gaussian likelihood for the cosmological parameters. As these parameters come with Gaussian priors they can be marginalised theoretically and the marginalised likelihood of the cosmological parameters only is nevertheless easily accessible (see Seehars et al., 2014, Appendix B for more details).

Planck

The final data is coming from the all-sky measurements of the CMB temperature by the Planck satellite (Planck Collaboration, 2014b, 2015d). The temperature power spectrum as observed by Planck covers the range \(2 \leq \ell \leq 2500\). The likelihood function is split into a low-\(\ell\) part and a high-\(\ell\) at \(\ell = 50\). While the low-\(\ell\) part is depending on the cosmological parameters, the high-\(\ell\) likelihood has to model foregrounds just as in the SPT likelihood. The 2015 data additionally contains polarisation data. The Planck team uses 27 nuisance parameters (16 in the 2013 release) with mostly flat priors to describe the foregrounds in great detail. Such a large number of nuisance parameters can be marginalised neither numerically nor analytically when flat priors are used, consequently only the full posterior as a function of nuisance and cosmological parameters is accessible in practice. We use the official MCMC chains published by the Planck team to estimate the Planck posteriors.

3.4 Concordance of CMB Surveys before Planck

Because the CMB power spectra are correlated due to cosmic variance, only datasets that have a small overlap in the measured scale of the power spectra or temperature and polarisation datasets can be combined in a simple sequential analysis. Examples for both joint analyses of complementary data and separate analyses of correlated data are discussed in the following, considering combinations of the datasets introduced in section 3.3. The comparison of Planck and WMAP will be covered separately in section 3.5.

Replacing BOOMERanG with WMAP data

As can be seen in Figure 3.4, BOOMERanG data overlaps completely with WMAP data. Furthermore, WMAP observations are more accurate than the BOOMERanG measurements, so a full joint analysis would provide only modest improvements compared to simply replacing the BOOMERanG with WMAP observations. Comparing BOOMERanG and WMAP data is therefore an example of datasets that replace each other as discussed in section 2.3. The separately analysed posteriors of the two experiments are shown in Figure 3.5. It can be seen
Chapter 3. Current Observations and their Information Gains

Table 3.1: Numerical relative entropy estimates in bits for different combinations of CMB data. For the Gaussian approximation, the relative entropy $D$ is split into expected relative entropy $\langle D \rangle$ and Surprise $S = D - \langle D \rangle$. Furthermore, the expected spread $\sigma(D)$ of $D$ around its mean $\langle D \rangle$ and the significance of the Surprise $S/\sigma(D)$ are given. Depending on the analysis strategy, $\langle D \rangle$ and $\sigma(D)$ are given by (2.59) and (2.60) when adding data and by (2.63) and (2.64) when replacing data. For joint analyses, $\langle D \rangle$ and $\sigma(D)$ are calculated as if the data was added independently. The $p$-value is an estimate for the prior probability for observing a Surprise that is greater or equal (less or equal) than $S$ if $S$ is greater (smaller) than zero. The results from the Monte Carlo integration are stated including the estimation uncertainty. The errors of the Gaussian estimates for $D$, $\langle D \rangle$, $S$, and $\sigma(D)$ are of order 0.1.

<table>
<thead>
<tr>
<th>Data combination</th>
<th>Updating scheme</th>
<th>$D$</th>
<th>$\langle D \rangle$</th>
<th>$S$</th>
<th>$S/\sigma(D)$</th>
<th>$p$-value</th>
<th>Monte Carlo estimate of $D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOOMERanG → WMAP 9</td>
<td>replace</td>
<td>22.5</td>
<td>18.4</td>
<td>4.1</td>
<td>1.6</td>
<td>0.07</td>
<td>20.9 ± 0.6</td>
</tr>
<tr>
<td>WMAP 3 → WMAP 5</td>
<td>joint</td>
<td>7.7</td>
<td>5.5</td>
<td>2.2</td>
<td>5.3</td>
<td>0.001</td>
<td>10.5 ± 0.9</td>
</tr>
<tr>
<td>WMAP 5 → WMAP 7</td>
<td>joint</td>
<td>1.4</td>
<td>1.0</td>
<td>0.4</td>
<td>0.6</td>
<td>0.2</td>
<td>1.5 ± 0.7</td>
</tr>
<tr>
<td>WMAP 7 → WMAP 9</td>
<td>joint</td>
<td>1.5</td>
<td>0.3</td>
<td>1.2</td>
<td>0.4</td>
<td>0.3</td>
<td>1.3 ± 0.7</td>
</tr>
<tr>
<td>WMAP 9 → WMAP 9 + SPT</td>
<td>add</td>
<td>4.3</td>
<td>2.1</td>
<td>2.2</td>
<td>2.1</td>
<td>0.04</td>
<td>4.6 ± 0.7</td>
</tr>
</tbody>
</table>

that while $\Omega_b h^2$, $\Omega_c h^2$ and $n_s$ are reasonably well constrained, $H_0$ is almost unconstrained by BOOMERanG data and $A_s$ and $\tau$ are highly degenerate. Nevertheless, the relative entropy estimates from the Gaussian approximation and the Monte Carlo method agree well as can be seen in Table 3.1. The total gain from this replacement is 22.5 bits. This significant update can also be seen in Figure 3.5, which shows that most of this gain can be attributed to a drastic reduction in the volume of the confidence intervals. This is further demonstrated by the decomposition of the relative entropy in Table 3.1, where the dominant contribution (18.4 bits) comes from $\langle D \rangle$, the expected entropy gain. The Surprise from this update is smaller than $2\sigma(D)$. The fact that the WMAP likelihood is not strictly of the Gaussian type used in section 2.6 implies that $\langle D \rangle$ and $\sigma(D)$ are approximations. Nevertheless, the conclusions drawn from Table 3.1 and Figure 3.5 are apparently consistent.

Comparing the individual WMAP releases

The relative entropy gains for the WMAP data releases after collecting three, five, seven, and nine years of data are shown in Table 3.1. As the data of the individual years is correlated, the WMAP team published joint likelihood functions for the overall data, taking those correlations into account. A comparison of the joint analyses of the WMAP data is therefore a mixture of the updating types discussed in section 2.3, but best described by a sequential updating of the constraints from old data with additional new data. For the estimates of $\langle D \rangle$, $S$, and $\sigma(D)$ shown in Table 3.1, it was therefore assumed that the data is complimentary, as a joint analysis cannot be modelled in the framework of section 2.6. The WMAP 3 posterior weakly deviates from Gaussianity, while the other posteriors are well described by a multivariate normal distribution. This deviation from Gaussianity leads to the difference between the relative
3.4. Concordance of CMB Surveys before Planck

Figure 3.5: Marginalised posteriors from BOOMERanG and WMAP 9 data. The contours cover 68% and 95% of the overall posterior volume and the dotted lines show the means.

entropy estimates using Gaussian approximations and the Monte Carlo integration method. The update from WMAP 3 to WMAP 5 shows a significantly larger increase in relative entropy (7.7 bits) as compared to the updates to WMAP 7 and 9 (1.4 and 1.5 bits, respectively). However, on closer inspection it can be seen that the majority of this gain is in the Surprise (5.5 bits) which is at the 5σ(D) level, corresponding to a p-value of 0.001 as the distribution for D is non-Gaussian. This unexpectedly strong change from the three to the five year release can be partly attributed to a change in the likelihood for the low-ℓ temperature power spectrum Dunkley et al. (2009).

Updating WMAP constraints with SPT data

As can be seen in Figure 3.4, SPT and WMAP data have little overlap in the high-ℓ regime where correlations due to cosmic variance are small, and can hence be considered as complementary datasets. As discussed in section 2.3, the constraints from WMAP are therefore compared to the posterior after adding the SPT data to the WMAP constraints. The marginals of prior and posterior when analysing the SPT data with a WMAP 9 prior are shown in Figure 3.6. The estimates for the relative entropy are listed in Table 3.1. Since both distributions are well
approximated by multivariate Gaussians, the Gaussian approximation yields reliable results as can be seen by comparing it to the Monte Carlo integration. Furthermore, the SPT likelihood is a normal distribution in the data. As such, the requirements from section 2.6 are fulfilled to good approximation and splitting $D$ into $\langle D \rangle$ and $S$ according to equations (2.59) and (2.60) is justified. The information gain here is 4.3 bits with 2.1 bits coming from $\langle D \rangle$, which is comparable to the update from WMAP 5 to WMAP 9, and a Surprise at the $2\sigma(D)$ level.

### 3.5 Concordance of WMAP and Planck

The previous section was a study of the agreement between constraints on a flat $\Lambda$CDM cosmology from a historical sequence of CMB experiments. We will now turn our attention to a study of the recent Planck data releases in 2013 (Planck Collaboration, 2014c) and 2015 (Planck Collaboration, 2015c). In particular, we will compare how well the constraints of WMAP agree with Planck in a $\Lambda$CDM universe.

As highlighted in section 2.6, we need mean and covariance of these parameters for each of the

---

**Figure 3.6:** Marginalised distributions of the WMAP 9 constraints with and without SPT data. The contours cover 68% and 95% of the overall posterior volume and the dotted lines show the means.
Table 3.2: Numerical relative entropy estimates in bits for different combinations of CMB data. For the Gaussian approximation, the relative entropy $D$ is split into expected relative entropy $\langle D \rangle$ and Surprise $S = D - \langle D \rangle$. Furthermore, the expected spread $\sigma(D)$ of $D$ around its mean $\langle D \rangle$ and the significance of the Surprise $S/\sigma(D)$ are given. Depending on the analysis strategy, $\langle D \rangle$ and $\sigma(D)$ are given by (2.59) and (2.60) when adding data and by (2.63) and (2.64) when replacing data. For joint analyses, $\langle D \rangle$ and $\sigma(D)$ are calculated as if the data was added independently. The $p$-value is an estimate for the prior probability for observing a Surprise that is greater or equal (less or equal) than $S$ if $S$ is greater (smaller) than zero. We cannot show results from the Monte Carlo integration, since the Planck likelihood depends on a large number of additional nuisance parameters. The errors of the Gaussian estimates for $D$, $\langle D \rangle$, $S$, and $\sigma(D)$ are of order 0.1.

<table>
<thead>
<tr>
<th>Data combination</th>
<th>Updating scheme</th>
<th>Gaussian approximation</th>
</tr>
</thead>
<tbody>
<tr>
<td>WMAP 9 → Planck 13</td>
<td>replace</td>
<td>27.3</td>
</tr>
<tr>
<td>WMAP 9 → Planck 15</td>
<td>replace</td>
<td>7.6</td>
</tr>
<tr>
<td>Planck 13 → Planck 15</td>
<td>joint</td>
<td>68.4</td>
</tr>
</tbody>
</table>

posteriors. We estimate those moments from Monte Carlo Markov chains that were provided by the Planck team for the 2013 and 2015 Planck releases (Planck Collaboration, 2014a, 2015b).

Planck 2013 release

The 2013 data release of Planck (Planck Collaboration, 2014c,b) includes only the temperature power spectrum. In the absence of polarisation data, however, the amplitude of the power spectrum $A_s$ is strongly degenerate with the optical depth to reionisation $\tau$. To break this degeneracy, the Planck team used the large-scale polarisation data from WMAP and we follow the same approach in this section. Whenever we refer to the Planck 2013 constraints in the following, we therefore refer to the constraints from Planck 2013 temperature data when jointly analysed with the large-scale WMAP 9 polarisation data.

Due to cosmic variance correlations between the data from WMAP and Planck, the joint likelihood of both datasets cannot be factorised into the individual likelihoods. For a joint analysis, one would rather have to find the correct joint likelihood function that takes the cosmic variance correlations into account. For simplicity, we hence consider the posteriors from the separately analysed data sets. When estimating expected relative entropy and Surprise, however, we have to ignore the correlations due to cosmic variance and the large scale WMAP polarisation data used in the Planck 2013 analysis to constrain $\tau$. If we were able to take the correlations into account, they would decrease the expected relative entropy and increase the Surprise.

In Table 3.2, we show that there are large gains in relative entropy (27.3 bits) when the WMAP data is replaced by Planck 2013. When studying the decomposition, however, it can be seen
that the contribution from $\langle D \rangle$ to the total entropy gain is dominated by the Surprise part (17.6 bits), though measuring a considerable improvement in precision at 9.6 bits. Furthermore, the Surprise is at levels greater than $4\sigma(D)$ corresponding to a p-value of 0.002. The results shown in Figure 3.7 support these findings and show that though the error contours do decrease considerably with the addition of Planck 2013 constraints another apparent effect is the shift of the confidence intervals. This in fact echoes the results of the Planck Collaboration (2014a) demonstrating shifts of the order of a standard deviation in four parameters when comparing WMAP 9 constraints to the ones from Planck 2013 and WMAP polarisation data. The relative entropy analysis of the posteriors is hence able to detect inconsistencies between posteriors in the case when none of the shifts in the individual parameters is particularly significant on its own while the overall shift in parameter space is nevertheless significantly larger than expected.

To further study the origin of this Surprise contribution, it is illustrative to estimate the relative entropies when replacing large scales ($2 \leq \ell \leq 49$) and small scales ($\ell \geq 50$) of the temperature power spectrum separately. Discrepancies between Planck and WMAP have been found on large scales when taking cosmic variance correlations into account (Planck Collaboration,
3.5. Concordance of WMAP and Planck

However, these discrepancies only have a small effect on the cosmological parameters: The estimated relative entropy of 0.7 bits when replacing large scale data only has a small Surprise contribution of \(-0.1\). Note that the negative Surprise is due to the correlations between Planck and WMAP data in the low-\(\ell\) regime. The overall Surprise is therefore not caused by large scales.

We also found mild tensions between WMAP and SPT in section 3.4. Like SPT, the Planck 2013 data also extends to \(\ell\) scales far beyond \(\ell \approx 1000\). If the tension between WMAP and SPT would be due to a fallacy in the model to explain scales smaller and larger than \(\ell \approx 1000\) at the same time, we would therefore expect that those tensions also show up in the comparison between Planck and WMAP. We tested this by comparing the joint posterior of WMAP and SPT with Planck. Yet, we do not find a better agreement between Planck and the joint constraints.

2015 Planck release

The main results of the previous section is the detection of a significant Surprise when comparing constraints from WMAP 9 data (Bennett et al., 2013; Hinshaw et al., 2013) and the constraints from the 2013 release of Planck data (Planck Collaboration, 2014c,a). In this section, we revisit this comparison in light of the 2015 Planck release and perform a more detailed study of the Surprise between WMAP and Planck. We analyse the Surprise between the posteriors derived from the WMAP 9 data on temperature and polarisation (Bennett et al., 2013; Hinshaw et al., 2013), the Planck 13 temperature data (Planck Collaboration, 2014c,b) together with low-\(\ell\) WMAP polarisation, and finally the Planck 15 data on temperature and polarisation (Planck Collaboration, 2015c).

Table 3.2 shows the numerical values for the observed relative entropy, expected relative entropy, Surprise, and \(p\)-value of the Surprise. As already discussed in the previous section, the large Surprise (\(S = 17.6\)) when going from WMAP 9 to Planck 13 constraints indicates significant tensions between those datasets. In light of the 2015 release of Planck, however, the large Surprise between WMAP 9 and Planck 13 has vanished, implying that the constraints of WMAP 9 and Planck 15 are in good agreement (\(S = -5.1\)). The Surprise between the Planck 13 and Planck 15 releases (\(S = 56.3\)) furthermore shows that the shifts in parameter space are significantly larger than expected a priori (\(p\)-value of \(4 \times 10^{-5}\)).

Looking at the marginalised constraints in Figure 3.7, however, these results are hard to relate to the seemingly mild changes between the Planck 13 and Planck 15 results. To better understand the Surprise between the Planck 13 constraints and the WMAP 9 and Planck 15 results, we argued in section 2.6 that the functional form of the Surprise suggests to standardise the WMAP constraints. Standardisation means that we reparametrise our theory such that the WMAP 9 constraints in the new parameters are uncorrelated with mean 0 and standard deviation 1. The advantage of the standardised parameter is that each component independently contributes to the Surprise (see equation (2.66)), potentially allowing us to identify directions in parameter space that generate the large Surprise.
Chapter 3. Current Observations and their Information Gains

Figure 3.8: Matrix plot of $U$ (defined in equation (3.26)), showing the eigenvectors (columns) of the correlation matrix of the WMAP 9 posterior in the cosmological parameter space together with the inverse eigenvalues given by the diagonal of $P$ (column labels). We plot the absolute value of $U$ to show the strength with which each cosmological parameter contributes to the eigenvectors that define the new parameter space (see colour scale). For the first new parameter (labelled by 120.6), for example, we can see that it gets the strongest contribution from $A_s$, but furthermore points into the $\tau$ and $\Omega_c h^2$ direction.

Standardisation is not unique and we choose to work in the eigenbasis of the correlation matrix for numerical stability. The correlation matrix is related to the covariance matrix via $\text{Corr}(\theta_i, \theta_j) = \text{Cov}(\theta_i, \theta_j) / \sigma(\theta_i) \sigma(\theta_j)$, where $\sigma(\theta_i)$ is the standard deviation of $\theta_i$. Given the eigen-decomposition of the correlation matrix $\text{Corr}$ of the parameters $\Theta$:

$$\text{Corr} = U^T P U$$  \hspace{1cm} (3.26)

and a diagonal matrix $F$ containing the standard deviations of $\Theta$, the following parameters $\Psi$ follow a standard normal WMAP 9 posterior:

$$\Psi = \left(\sqrt{P}\right)^{-1} U F^{-1} (\Theta - \mu(\Theta)).$$  \hspace{1cm} (3.27)

Here, $\sqrt{P}$ is simply containing the square roots of the eigenvalues $P$ and $\mu(\Theta)$ contains the means of the parameters $\Theta$. The matrix $U$ defines how the new parameters $\Psi$ are related to the original cosmological parameters. It is shown in Figure 3.8 together with the inverse
3.5. Concordance of WMAP and Planck

Figure 3.9: Marginalised posteriors of WMAP 9 (black), Planck 13 (blue), and Planck 15 (green) constraints in the parametrisation $\Psi$ given by equation (3.27). The contours cover 68% and 95% of the overall posterior volume and the dotted lines show the means. The labels show the inverse eigenvalues of the correlation matrix given by the diagonal of $P$ defined in equation (3.26). The relation between $\Psi$ and the cosmological parameters is shown in Figure 3.8. By construction, the WMAP 9 constraints on $\Psi$ are uncorrelated, have mean 0 and standard deviation 1. Each deviation from 0 in the means of the Planck constraints independently contributes to the Surprise between WMAP and Planck.

The marginalised posteriors in the new parameter space $\Psi$ are shown in Figure 3.9. Note that Figure 3.9 shows the same constraints as Figure 3.7 in a different parametrisation of the $\Lambda$CDM model. Note also that the new parametrisation is entirely derived from the WMAP 9 constraints without any knowledge of the Planck data. By construction, the WMAP 9 constraints have mean 0, standard deviation 1, and show no correlations between the individual $\psi_i$ components of $\Psi$. The striking feature of Figure 3.9 is the deviation of the Planck 13 constraints from the WMAP 9 mean by more than 5$\sigma$ in the direction of the parameter space with the dominant eigenvalue. The Planck 15 constraints, on the other hand, agree with the WMAP 9 constraints in this direction to great accuracy, thereby explaining the vanishing Surprise.
Chapter 3. Current Observations and their Information Gains

Looking at the eigenvector that corresponds to the largest eigenvalue (see first column of Figure 3.8), we see that the direction that is responsible for the tension between Planck 13 and the WMAP 9 and Planck 15 constraints is dominated by $A_s$ with additional contributions from $\tau$ and $\Omega_c h^2$. Larson et al. (2015) come to a very similar conclusion based on a simulation-based approach to comparing WMAP and Planck data. They point out that there is a mismatch in amplitude between Planck 13 data and WMAP 9 for small $\ell$ which dominate the constraints on the cosmological parameters and also conclude that the shifts in the $\Omega_c h^2-A_s$ direction are larger than expected when taking cosmic variance correlations into account. $A_s$ and $\tau$ dominantly affect the overall amplitude of the temperature power spectrum. The Planck team indeed changed their calibration scheme from the 2013 to the 2015 release (Planck Collaboration, 2015a,d), and the strong change in the Surprise is most likely a direct consequence of this change in the data.

It is not easy to interpret the actual magnitude of the negative Surprise between WMAP 9 and Planck 15. In principle, a negative Surprise indicates that the agreement between the constraints is better than expected from statistical fluctuations. The $p$-value of 0.07 for the Surprise being $-5.1$ or smaller is estimated assuming independence of Planck and WMAP power spectra. However, the measurements are correlated due to cosmic variance as they both measure the same CMB. We would therefore expect the measurements to be more consistent than predicted from the WMAP 9 posterior. The confidence of the over-consistency between WMAP 9 and Planck 15 is small anyway and taking the correlations into account would most likely diminish it.

3.6 Cosmological Probes at Low Redshifts

We now proceed to an additional set of cosmological probes beyond the CMB. As discussed in section 1.4, the CMB mainly probes the physical properties of the universe at the time of the decoupling of radiation and charged particles during recombination. Dark energy, however, started to dominate the energy density of the universe only recently. Deviations from a cosmological constant would hence show up in inconsistencies between cosmological probes at low redshifts and the CMB. There are many different ways of probing cosmology with observations at low redshifts, and we touched upon a few of them in sections 1.1 and 1.3. We will briefly summarise the datasets used in our updates before reporting the results on the agreement of the low redshift probes with WMAP and Planck.

Distance ladder observations

As a measurement of the Hubble constant with the distance ladder, we used the recent measurement based on the NGC 4258 maser distance conducted by Efstathiou (2014). The likelihood for $H_0$ is then given by a Gaussian with mean $H_0 = 70.6$ km/s/Mpc and standard deviation $\sigma = 3.3$ km/s/Mpc.
Supernovae type Ia

As supernova dataset, we use a compilation of data from the supernova legacy survey, Sloan digital sky survey (SDSS) and the Hubble space telescope analysed by Conley et al. (2011). We use the publicly available data and covariance matrix in a likelihood that follows code used by Conley et al. (2011). Additional nuisance parameters for the absolute magnitude, stretch, and colour are marginalised at the posterior level.

Baryon acoustic oscillations

We follow the selection of the Planck Collaboration (2015b) for our compilation of BAO data from the 6dF Galaxy Survey (Beutler et al., 2011), the Baryon Oscillation Spectroscopic Survey (Anderson et al., 2014), the WiggleZ Dark Energy Survey (Kazin et al., 2014), and a measurement from the SDSS data release 7 by Ross et al. (2015). We assume a Gaussian likelihood using the provided covariances. For simplicity, we ignore the possible correlations arising from the overlap between the SDSS III field and the WiggleZ field.

Weak lensing

As a final probe, we include the weak lensing shear correlation function measured from Canada-France-Hawaii-Telescope Lensing Survey (CFHTLenS) data by Kilbinger et al. (2013). We assume a Gaussian likelihood and use data of the 2d comic shear correlation functions, ignoring the cosmological parameter dependence of the covariance matrix. We apply no cuts at small angles, but consider the full shear correlation data published by CFHTLenS. The theory prediction for the correlation function uses the Eisenstein & Hu transfer function (Eisenstein and Hu, 1999) with non-linear corrections from halofit (Smith et al., 2003; Takahashi et al., 2012) as well as the Limber approximation (Limber, 1953).

3.7 Concordance at Low Redshifts

In this section, we will use current data from the distance ladder, supernovae type Ia, Baryonic Acoustic Oscillations (BAOs), and weak gravitational lensing to update the constraints on $\Lambda$CDM from both WMAP 9 and Planck 15. With the exception of the Planck 15 constraints, all posteriors are estimated with MCMC chains from CosmoHammer. The Planck 15 is assumed to be a Gaussian with mean and covariance estimated from the official Planck 15 chains.

WMAP prior

The results of updating WMAP 9 constraints with low-redshift data are given in Table 3.3. We can see that all datasets achieve relatively small information gains: weak lensing contributes the most at 1.25 bits, closely followed by the BAO data (1.10 bits). All datasets show a negative
Chapter 3. Current Observations and their Information Gains

Table 3.3: Relative entropy estimates in bits for updates of the WMAP constraints using low-redshift data. All values are calculated using the Gaussian approximation. Since we are adding data, \( \langle D \rangle \) and \( \sigma(D) \) are given by (2.59). The \( p \)-value is an estimate for the prior probability for observing a Surprise that is greater or equal (less or equal) than \( S \) if \( S \) is greater (smaller) than zero. The errors of the estimates for \( D \), \( \langle D \rangle \), \( S \), and \( \sigma(D) \) are of order 0.04.

| Data combination          | \( D \) | \( \langle D \rangle \) | \( S \) | \( |S/\sigma(D)|\) | \( p \)-value |
|---------------------------|--------|------------------------|-------|-----------------|-------------|
| WMAP 9 + \( H_0 \)        | 0.04   | 0.26                   | -0.22 | 0.7             | 0.03        |
| WMAP 9 + SNe              | 0.10   | 0.16                   | -0.06 | 0.33            | 0.50        |
| WMAP 9 + BAOs             | 1.10   | 1.68                   | -0.58 | 0.63            | 0.25        |
| WMAP 9 + weak lensing     | 1.25   | 1.34                   | -0.09 | 0.10            | 0.37        |

Table 3.4: Relative entropy estimates in bits for updates of the Planck constraints using low-redshift data. All values are calculated using the Gaussian approximation. Since we are adding data, \( \langle D \rangle \) and \( \sigma(D) \) are given by (2.59). The \( p \)-value is an estimate for the prior probability for observing a Surprise that is greater or equal (less or equal) than \( S \) if \( S \) is greater (smaller) than zero. The errors of the estimates for \( D \), \( \langle D \rangle \), \( S \), and \( \sigma(D) \) are of order 0.04.

| Data combination          | \( D \) | \( \langle D \rangle \) | \( S \) | \( |S/\sigma(D)|\) | \( p \)-value |
|---------------------------|--------|------------------------|-------|-----------------|-------------|
| Planck 15 + \( H_0 \)     | 0.04   | 0.07                   | -0.03 | 0.29            | 0.44        |
| Planck 15 + SNe           | 0.04   | 0.05                   | -0.01 | 0.09            | 0.35        |
| Planck 15 + BAOs          | 0.25   | 0.47                   | -0.22 | 0.43            | 0.47        |
| Planck 15 + weak lensing  | 4.84   | 0.45                   | 4.38  | 8.70            | 2.3 \times 10^{-4} |

Surprise, highlighting that the observations are consistent with the WMAP constraints within \( \Lambda \)CDM.

It is worth mentioning here, that the impact of the other probes is suppressed by the choice of the model. A flat \( \Lambda \)CDM cosmology can be constrained with CMB data alone. As CMB observations are more mature than the other measurements, adding the low-redshift probes is affecting the constraints only mildly. For other models, however, adding probes to the CMB is crucial for breaking degeneracies in the extended parameter space. Analysing the constraints in extended model spaces goes beyond the scope of this thesis, but we have shown in Grandis et al. (2016b) that the information gains for \( H_0 \), supernova, and BAO observations are greatly amplified in models with a free curvature component \( \Omega_K \) or a dark energy component with a free equation of state parameter \( w \).

**Planck prior**

In Table 3.4, we show the updates of the Planck 15 constraints with observations at low-redshifts. We see that the impact of distance ladder, supernovae, and BAO observations on the constraints has been further diminished by the even stronger Planck constraints. The gain from adding the weak lensing results from CFHTLenS (Kilbinger et al., 2013), however, is
greatly enhanced. This is mostly due to a highly significant Surprise contribution that shows that the Planck and CFHTLenS constraints are not consistent within $\Lambda$CDM.

The weak lensing results of CFHTLenS have recently received much attention and revised results which are in better agreement with Planck have been published (Köhlinger et al., 2016; Kitching et al., 2016; Asgari et al., 2016; Liu et al., 2016; Joudaki et al., 2016). It will be interesting to see how the consistency between weak lensing and CMB evolves in the future, with upcoming results both from Planck and weak lensing surveys such as the Dark Energy Survey.

3.8 Discussion

In this chapter, we have applied the concepts of information gain and Surprise introduced in chapter 2 to a range of cosmological datasets. We started with CMB experiments ranging from BOOMERanG, WMAP, and SPT to the most recent results from the Planck team. We furthermore considered updates to the constraints from the CMB using cosmological probes at low redshifts such as $H_0$ measurements from the distance ladder, and supernovae type Ia, BAOs, and weak lensing observations.

Since the parameters of a flat $\Lambda$CDM cosmology are well constrained by the CMB, we base our analyses of the information gains on the analytic expression for Gaussian distributions derived in chapter 2. In this limit, information gain and Surprise depend only on the mean and covariance of the posteriors. We discussed how those moments can be robustly estimated from samples generated with our MCMC framework CosmoHammer.

The relative entropy gains in units of bits from BOOMERanG, WMAP, SPT, and Planck surveys range from about 1 to 30. In general, the numbers are driven by the contributions from the expected relative entropy, $\langle D \rangle$, but in three cases the Surprise is found to dominate the results. In terms of expected relative entropy, the step from BOOMERanG to WMAP 9 is the biggest ($\langle D \rangle \sim 18$ bits), followed by the update of WMAP by Planck 15 data ($\langle D \rangle \sim 13$ bits). The addition of SPT data to the WMAP constraints leads to an expected relative entropy gain of $\langle D \rangle \sim 2$ bits.

While most CMB datasets turn out to be consistent within $\Lambda$CDM, some combinations show a significant Surprise. The update from WMAP 3 to WMAP 5, for example, shows a large Surprise of $S = 5.5$ which is 5.3 standard deviations away from expectations. This result might be caused by both the deviations of the WMAP 3 posterior from a normal distribution and the adjustments of the likelihood function for the low-$\ell$ temperature power spectrum by the WMAP team.

The biggest Surprise values, however, are generated by the Planck 13 data release. Comparing the Planck 13 constraints to WMAP 9 ($S = 17.6$, significant at a confidence of $p = 2 \times 10^{-5}$) and Planck 15 results ($S = 56.3$, significant at a confidence of $p = 4 \times 10^{-5}$), we find that it strongly disagrees with both. We also find that Planck 2015 is in good agreement with the
WMAP 9 constraints as indicated by a negative Surprise ($S = -5.1$ bits). By analysing the posteriors in the principal components of the WMAP 9 constraints, we were able to detect the direction in parameter space that is the primary cause for the Surprise in the update from WMAP 9 to Planck 13. It is mainly composed by $A_s$, $\tau$, and $\Omega_c h^2$ and shifts along this direction lead to a change in the amplitude of the predicted temperature power spectrum. The inconsistency between Planck 13 and WMAP 9 as detected by the Surprise is hence most likely the consequence of a systematic in the calibration of the Planck data which was resolved in the 2015 release of the Planck team (Planck Collaboration, 2015a,d; Larson et al., 2015).

Including the low-redshift probes in the analysis of the flat $\Lambda$CDM constraints from CMB data, we find that the largest information gains are achieved by BAO and weak lensing measurements at about 1 bit. The contribution of the $H_0$ measurement from distance ladder observations and luminosity distance from supernova type Ia data were smaller by about an order of magnitude. These small gains are explained by the restricted parameter space of a flat $\Lambda$CDM cosmology that is already well constrained by the CMB alone. Measuring the gains in extended spaces significantly enhances the contributions from other probes (Grandis et al., 2016b). The only tension within flat $\Lambda$CDM in this analysis was detected in the update of the Planck 15 constraints with the weak lensing results from CFHTLenS by Kilbinger et al. (2013). In light of recent results based on revised analyses of the CFHTLenS data (Köhlinger et al., 2016; Kitching et al., 2016; Asgari et al., 2016; Liu et al., 2016; Joudaki et al., 2016), however, this discrepancy seems to be resolved.

Our study shows that the Surprise is a reliable measure of agreement between cosmological constraints. The analysis of state-of-the-art datasets from a large variety of cosmological probes showed only very few tensions that are most likely related to systematics in the data. As more cosmological probes are able to put tight constraints on $\Lambda$CDM parameters, estimating the Surprise between the constraints can help to detect systematic issues in data or model in the future.
4 Future Observations and HI Intensity Mapping

As we have seen in the previous chapter, our main sources of cosmological information are the CMB and observations of astronomical objects out to redshifts of $z \approx 2$. As the CMB mainly probes the state of the universe at the time of decoupling, i.e. at a redshift of approximately $z \approx 1100$, large volumes of the universe are left unobserved to date. And while upcoming optical surveys like DES\textsuperscript{1}, DESI\textsuperscript{2}, LSST\textsuperscript{3}, Euclid\textsuperscript{4}, and others will move the boundary at low redshifts to about $z \approx 3$, detecting large galaxy samples at even higher redshifts becomes increasingly complex.

In this context, intensity mapping of atomic and molecular line transitions is emerging as a promising probe of cosmological evolution as well as of the connection between galaxy evolution and the growth of large-scale structure in the universe (Santos et al., 2015; Croft et al., 2016; Chang et al., 2010; Masui et al., 2013; Switzer et al., 2013; Comaschi and Ferrara, 2016; Silva et al., 2015; Gong et al., 2011; Li et al., 2016; Mashian et al., 2015). The ability to detect a line over wide areas of the sky implies that one has accurate redshift information over large cosmological volumes and forecasts indicate great potential for the recovery of cosmological information from intensity mapping surveys (Wyithe et al., 2007; Loeb and Wyithe, 2008; Visbal et al., 2009; Bull et al., 2015; Chang et al., 2015). The 21 cm line of neutral hydrogen (HI) is a particularly promising future probe, being the target of several dedicated efforts using radio telescopes such as BAOBAB (Pober et al., 2013), BAORadio (Ansari et al., 2012b), BINGO (Battye et al., 2013), and CHIME (Bandura et al., 2014). A successful HI intensity mapping experiment will map the integrated emission of HI within the beam of the instrument over large parts of the sky. Due to the high frequency resolution of radio telescopes, such experiments would produce maps of the large-scale structure of HI in the universe within thin redshift slices. The angular and redshift resolution in these surveys will correspond to comparable physical scales.

\textsuperscript{1}http://www.darkenergysurvey.org/
\textsuperscript{2}http://desi.lbl.gov/
\textsuperscript{3}http://www.lsst.org/
\textsuperscript{4}http://sci.esa.int/euclid/
Chapter 4. Future Observations and HI Intensity Mapping

In the following, we will briefly review the principles of post-reionisation HI intensity mapping. We will then discuss a possible method for simulating the expected signal. In sections 4.4 and 4.5, we use these simulations to study the clustering of HI as measured by the angular power spectrum. We end this chapter with a discussion of challenges that we faced during our observations with a single-dish radio telescope in Bleien, Switzerland.

This chapter is based on Seehars et al. (2016). In section 4.6, we further summarise results from manuscripts that are currently in preparation (Monstein et al., 2016; Akeret et al., 2016; Chang et al., 2016).

### 4.1 HI Intensity Mapping

We start with a brief review of the principles of post-reionisation HI intensity mapping. We discuss the observational principles for the example of a single dish experiment before giving a summary of the theoretical expectations for the clustering of the large-scale HI distribution. For more details see e.g. Pritchard and Loeb (2012) for a review.

#### Cosmological 21 cm line emission

Mapping the matter distribution in our universe is a central part of observational cosmology. Traditionally, this is done by studying the distribution and the shapes of galaxies in wide field surveys. In HI intensity mapping, the idea is to study the distribution of HI by mapping the distribution of redshifted flux from its 21 cm line emission. After reionisation, the bulk of HI is expected to be clumpy and associated with galaxies. E.g., at redshifts $z \lesssim 1.5$, damped Lyman-$\alpha$ (DLA) systems have been found associated with MgII and FeII absorption systems (Rao and Turnshek, 2000; Rao et al., 2006). HI emission line stacking techniques applied to star forming galaxies at low redshifts $z \sim 0.2-0.3$ have yielded measurements of $\Omega_{\text{HI}} \sim 10^{-3}$ (Lah et al., 2007, 2009). And finally, HI emission studies in the local universe ($z < 0.03$) have yielded individual detections of a large number of HI-rich galaxies, with $\Omega_{\text{HI}} \sim \text{few } \times 10^{-4}$ (Martin et al., 2010; Zwaan et al., 2005). See also Padmanabhan et al. (2015) for a compilation of HI observations. In this work, we will distribute HI in dark matter halos using a phenomenological prescription proposed in Padmanabhan et al. (2016) that is consistent with these observations of $\Omega_{\text{HI}}$ and the hypothesis that HI is typically associated with star forming galaxies\(^5\).

When mapping the flux from individual clouds of neutral hydrogen for example with the large beam of a single dish radio telescope, the flux of discrete but non-resolved sources on the sky is integrated within a physical volume defined by the size of the beam and a range in frequency. Very large single dish instruments like the Green Bank Telescope with a diameter of 100 m have a resolution of order 10$'$ at $z \approx 1$. In frequency, however, radio telescopes can

---

\(^5\)We do note that recent measurements of the clustering of DLA systems at redshifts $2 \lesssim z \lesssim 3$ (Font-Ribera et al., 2012) are discrepant with direct HI observations on which our model is built (Padmanabhan et al., 2016). The resolution of this discrepancy is unclear at present, and goes beyond the scope of this chapter.
4.1. HI Intensity Mapping

have channels with a bandwidth of \( \sim 1 \text{ MHz} \) and smaller, corresponding to redshift bins of width \( \Delta z \approx 0.003 \) at redshift \( z \approx 1 \). In co-moving distances, the intensity maps from such an instrument would hence have comparable angular and radial resolution of order \( \sim 10 h^{-1} \text{ Mpc} \) at this redshift. In the more distant future, however, intensity mapping surveys with long baseline interferometers like the SKA will be able to improve the angular resolution by many orders of magnitude.

Given an average density of neutral hydrogen \( \rho_{\text{HI}}(z) \) at redshift \( z \) or equivalently its ratio to the critical density today \( \Omega_{\text{HI}}(z) \), the average brightness temperature of 21 cm flux is given by (Battye et al., 2013):

\[
\bar{T}(z) \approx 44 \mu K \left( \frac{\Omega_{\text{HI}}(z) h}{2.45 \times 10^{-4}} \right) (1 + z)^2 E(z) \tag{4.1}
\]

with \( E(z) = H(z)/H_0 \) being the normalised Hubble parameter. At redshift \( z \approx 1 \), this corresponds to brightness temperatures of approximately 100 \( \mu K \) and is therefore roughly four orders of magnitude sub-dominant to the brightness of our own galaxy at these frequencies which is at the \( \sim 1 \text{ K} \) level even at high galactic latitudes. One of the key challenges of this technique is hence the ability to separate the signal from the Galaxy and other extra-Galactic radio sources (Ansari et al., 2012a; Wolz et al., 2014; Alonso et al., 2015; Bigot-Sazy et al., 2015).

Large-scale structure of HI

The fluctuations of 21 cm brightness temperatures are expected to be a biased version of the fluctuations in the matter density field. Hence, the Fourier transformed temperature fluctuations in the map at large scales can be written as

\[
\delta T(k, z) = \bar{T}(z) b_{\text{HI}}(k, z) \delta(k, z), \tag{4.2}
\]

where \( \delta \) is the dark matter overdensity and \( b_{\text{HI}}(k, z) \) is the bias of HI relative to \( \delta \).

On even larger scales, the power spectrum of the matter density fluctuations is well described by linear theory predictions and the halo bias is scale independent. Assuming a simple relation \( m_{\text{HI}}(m) \) between HI mass \( m_{\text{HI}} \) and halo mass \( m \), the large scale bias of HI \( b_{\text{HI}} \) can be predicted from the halo mass function using:

\[
b_{\text{HI}}(z) = \int dm n(m, z) \frac{m_{\text{HI}}(m)}{\bar{\rho}_{\text{HI}}(z)} b(m, z), \tag{4.3}
\]

where \( \bar{\rho}_{\text{HI}}(z) = \int dm n(m, z) m_{\text{HI}}(m) \) is the mean density of HI, \( n(m, z) \) is the unconditional differential number density of halos and \( b(m, z) \) is the corresponding halo bias (we use the forms from Sheth and Tormen, 2002). The large scale power spectrum of HI is hence expected to be well described by

\[
P_{\text{HI}}(k, z) \approx \left[ \bar{T}(z) b_{\text{HI}}(z) D(z) \right]^2 P(k), \tag{4.4}
\]
where \( P(k) \) is the linear theory power spectrum of the matter overdensities at \( z = 0 \) and \( D(z) \) is the linear growth factor.

As wide-field surveys probe a spherical region rather than a cartesian box, different methods for measuring the two-point correlations have been proposed and compared (see e.g. Asorey et al., 2012; Nicola et al., 2014; Lanusse et al., 2015). One approach is to perform a tomographic analysis based on the angular correlations (or equivalently the power spectrum) of the field within bins of redshift. This approach is convenient for intensity mapping as techniques for its estimation from maps are well established from the CMB (Tegmark and de Oliveira-Costa, 2001; Bond et al., 1998; Szapudi et al., 2001b,a; Hivon et al., 2002; Efstathiou, 2004; Chon et al., 2004), there is no need for assuming a cosmology during the analysis (Bonvin and Durrer, 2011; Challinor and Lewis, 2011), and cross-correlating different surveys is straightforward (Eriksen and Gaztanaga, 2015). Furthermore, intensity mapping surveys will produce maps of brightness temperature fluctuations within bins in frequency and thus a tomographic analysis within these bins does not erase any information. The angular power spectrum \( C_\ell(z,z') \) of the temperature fluctuations is related to \( P(k) \) via

\[
C_\ell(z,z') = \frac{2}{\pi} \int dz W(z) \hat{T}(z) D(z) b_{HI}(z) \int dz' W'(z') \hat{T}(z') D(z') b_{HI}(z') \times \int k^2 dk P(k) j_\ell(k R(z)) j_\ell(k R(z')) ,
\]

(4.5)

where \( W, W' \) are the redshift window functions for the two tomographic bins around \( z, z' \) and \( R(z) \) is the co-moving distance to redshift \( z \).

An additional contribution to the angular power spectrum arises from the shot noise of the discrete sources. Given a population of unresolved, Poisson distributed point sources with neutral hydrogen mass \( m_{HI} \) distributed over the full sky, the power spectrum of the resulting intensity map is given by \( C^{sn}_\ell = \left( \frac{\bar{T}}{m_{HI}} \right)^2 \int dm_{HI} n_{HI}(m_{HI}) m_{HI}^2 n_{HI}(m_{HI}) \) with \( n_{HI}(m_{HI}) \) being the differential source count per steradian and \( \bar{m}_{HI}(z) \) being the mean hydrogen mass of the population (Tegmark and Efstathiou, 1996). For our maps, this contribution will however turn out to be negligible (\(< 5 \times 10^{-3} \mu K^2 \) at the lowest redshift) due to the large number of low mass halos.

**Estimation of angular power spectra from intensity maps**

Given a map with brightness temperatures \( T_i \) drawn from a field with underlying angular power spectrum \( C_\ell \) and its coefficients \( a_{\ell m} \) of the map’s expansion in spherical harmonics, the angular power spectrum can be estimated as

\[
\hat{C}_\ell = \frac{1}{2\ell + 1} \sum_{m=-\ell}^{\ell} |a_{\ell m}|^2 .
\]

(4.6)
4.2 Simulating HI Intensity Maps

When only parts of the sky are available, recovering the spherical harmonics from the masked field is more complicated. The pseudo $C_\ell$ estimator (Peebles, 1973) is a standard approach to this problem: In this approximation, the masked pixels are set to the mean value of the unmasked part before the full sphere is decomposed in spherical harmonics. The power spectrum is then estimated via

$$\hat{C}_\ell^\text{pseudo} = \frac{1}{(2\ell + 1)f_{\text{sky}}} \sum_{m=-\ell}^{\ell} |a_{\ell m}|^2$$

(4.7)

with $f_{\text{sky}}$ being the fraction of unmasked sky (see also Tegmark and de Oliveira-Costa, 2001; Bond et al., 1998; Szapudi et al., 2001b; Hivon et al., 2002; Efstathiou, 2004, for more sophisticated approaches).

In addition to the pseudo $C_\ell$ estimator, we also consider the publicly available PolSpice package (Szapudi et al., 2001a; Chon et al., 2004) in our analyses. PolSpice first estimates the correlation function of the masked fluctuations, corrects for the mask, and then calculates the demasked angular power spectrum from the corrected correlation function. To avoid artefacts from the decomposition of an incomplete correlation function of the masked field, the correlation function has to be smoothed with an apodisation function. We used Gaussian random fields with a mask corresponding to those applied to the simulated intensity maps to fix the parameters of a Gaussian apodisation (width of $\sigma = 90^\circ$ and a maximum angle of the correlation function $\beta_{\text{max}} = 120^\circ$).

4.2 Simulating HI Intensity Maps

Predicting the constraining power of 21 cm experiments requires accurate knowledge of the expected signal. Several approaches have been developed for this purpose. Analytic methods based on perturbation theory and the halo model can be used to predict some of the statistical properties of the signal, such as the power spectrum (Wyithe et al., 2007; Battye et al., 2013; Pontzen, 2014; Bull et al., 2015; Padmanabhan et al., 2016). For a fast production of wide field intensity maps, simulations have been developed based on log-normal random fields (Alonso et al., 2014). A number of different numerical approaches have also been used to understand the effect of non-linear and baryonic processes on the signal. They range from hydrodynamical simulations (Duffy et al., 2012; Dave et al., 2013) that model both collisionless dark matter and the hydrodynamics of baryons to approaches where the HI is simply assigned to resolved halos from dark matter only simulations (Bagla et al., 2010; Khandai et al., 2011).

In this section, we follow an alternative and complementary approach by simulating low-redshift HI intensity maps ($z < 1$) using a combination of dark matter fields from N-body simulations and a halo model prescription for assigning HI to sub-resolution dark matter halos. The advantage of this approach is that it produces maps that have both realistic clustering from N-body simulations and large volumes, as necessary for future HI intensity mapping experiments. For this purpose, we use N-body simulations of a $2.6 \, h^{-1} \, \text{Gpc}$ box (Wechsler et al.,...
2016) with a mass resolution of $1.6 \times 10^{11} h^{-1} M_{\odot}$. Below this mass resolution, we model the halo distribution using a conditional halo mass function (Sheth and Tormen, 2002). We then generate HI intensity maps by assigning HI mass to the halos following a phenomenological prescription (Padmanabhan et al., 2016).

**Matter density fields**

We use 10 N-body simulations run with the L-Gadget2 code (based on Gadget2 (Springel et al., 2001; Springel, 2005)). These simulations are also being used to simulate galaxy catalogues for the Dark Energy Survey (Wechsler et al., 2016; Busha et al., 2013; Erickson et al., 2013), and a subset of these catalogues have been used previously in several Dark Energy Survey studies (Vikram et al., 2015; Bonnett et al., 2015; Leistedt et al., 2015; Becker et al., 2015; Park et al., 2015). In this work, we use only the dark matter distribution from these simulations. Each of the 10 simulations contains $2048^3$ particles of mass $1.6 \times 10^{11} h^{-1} M_{\odot}$ in a $2.6 h^{-1} \text{Gpc}$ box and produces lightcones on the fly by writing all particles that cross the lightcone surface at each time step of the simulation. The simulations use a flat $\Lambda$CDM cosmology with matter density relative to critical $\Omega_m = 0.286$, baryon density relative to critical $\Omega_b = 0.047$, Hubble parameter $h = 0.72$, root mean squared density fluctuations at $8 h^{-1} \text{Mpc}$ given by $\sigma_8 = 0.82$ and spectral index of the primordial power spectrum $n_s = 0.96$. We use the lightcone output to create HEALPix maps of the projected matter distribution with an angular resolution of about $7'$ ($n_{\text{side}} = 512$) by counting the number of particles per pixel, where a pixel is the cosmological volume defined by a HEALPix cell and a redshift bin. We choose the redshift bins such that they correspond to the redshift of the 21 cm line in equally spaced bins in frequency between 750 MHz and 1050 MHz with a bandwidth of 25 MHz per channel. This results in 12 maps of the radially averaged density field between redshifts $z \approx 0.35$ and $z \approx 0.89$ with negligible shot noise and redshift bin width from $\Delta z \approx 0.033$ at the lowest to $\Delta z \approx 0.061$ at the highest redshift. In co-moving scales, our pixels have an angular and radial size of $2 h^{-1} \text{Mpc}$ and $82 h^{-1} \text{Mpc}$ at $z \approx 0.369$ and $4 h^{-1} \text{Mpc}$ and $110 h^{-1} \text{Mpc}$ at $z \approx 0.863$, respectively. As the co-moving distance to $z = 0.89$ $(2.1 h^{-1} \text{Gpc})$ is bigger than half the size of the box $(1.3 h^{-1} \text{Gpc})$, a full-sky lightcone would overlap in the simulated box. To balance overlap and sky coverage, we analyse four quadrants of the lightcone separately. This procedure hence yields 40 realisations of the matter density field, each covering a quarter sky. The volume of one quadrant corresponds to roughly 53% of the total box volume. 11% of the quadrant volume is overlapping in the box, but most of this overlapping volume appears at different redshifts. The largest overlap within a single redshift bin is 3% in the highest redshift bin around $z = 0.86$. 

4.2. Simulating HI Intensity Maps

Halo assignment

As mentioned in section 4.1, the dominant part of HI is expected to be found within galaxies. The challenge of simulating a wide field HI intensity mapping survey is hence that the signal is expected to come from relatively low mass halos that cannot be easily resolved by N-body simulations with the box sizes needed for the sky coverage corresponding to wide field surveys.

Taking a cue from the coarseness of the instrument resolution for HI intensity mapping experiments which erases information about the angular positions of individual sources, one possible prescription for turning density fields into intensity maps is to neglect the discrete origin of the 21 cm intensity and to assume that overdensity $\delta$ and brightness temperature fluctuations $\delta T$ are linearly related:

$$\delta T(z, \Theta) \approx b_{\text{HI}}(z) \bar{T}(z) \delta(z, \Theta).$$  \hfill (4.8)

In this case, the intensity map would be a rescaled version of the density field. In Alonso et al. (2014), for example, the authors followed this approach in their work on creating fast intensity mapping simulations using random log-normal fields. As we expect the HI bias to be closely related to the halo bias, we know that a scale independent, linear bias prescription is only valid at large, linear scales. A possible improvement over equation (4.8) would be a non-linear bias as in de la Torre and Peacock (2013) which however still ignores possible scale dependencies and the stochastic relation between halo and density field.

We therefore choose to go one step further in exploiting the low resolution of the intensity maps by combining a simulated large scale density field with the conditional mass function of dark matter halos as derived from the halo model. The coarseness of the angular resolution of intensity mapping experiments indeed means that we are not interested in simulating the exact locations of the relevant dark matter halos, but only their total number in each pixel. The conditional mass function $N(m|M)$ calibrated by Sheth and Tormen (2002), gives the differential number of halos in the mass range $(m, m + dm)$ within a Lagrangian volume containing mass $M$,

$$N(m|M)dm = \frac{M |T(s|S)|}{m \sqrt{s - S}} \exp\left(-\frac{(B(s) - \delta_{\text{lin}}(M))^2}{2(s - S)}\right) \frac{ds}{s - S}$$ \hfill (4.9)

where $s$ is the variance of the linear field when smoothed on the Lagrangian scale of the halo mass $m$, $S$ is the corresponding variance at scale $M$, $\delta_{\text{lin}}(M)$ is the linearised overdensity corresponding to the mass $M$, and $B(s)$ is the moving barrier shape associated with ellipsoidal collapse:

$$B(s, z) = \sqrt{a} \delta_{sc}(z) \left(1 + \beta \left(\frac{\delta_{sc}^2(z)}{s}\right)^{-\alpha}\right).$$ \hfill (4.10)

where $\delta_{sc}(z)$ is the critical overdensity for collapse at redshift $z$ and the fitting parameters are $a = 0.707$, $\alpha = 0.615$, and $\beta = 0.485$. Finally, $T(s|S)$ is the 5th order Taylor series of $B(s) - \delta_{\text{lin}}(M)$
What we have available from the pixelated density fields is the mass $M$ or equivalently the (non-linear) overdensity $\delta$ contained in each pixel. Ideally we would like to use a mass function conditioned on this $\delta$. In order to apply equation (4.9) to the density field, however, we need to relate $\delta$ to its linearised overdensity $\delta_{\text{lin}}(M)$. In order to maintain a physical link between $\delta$ and $\delta_{\text{lin}}$, we do this by assuming that the density in each pixel approximately follows spherical evolution, in which case we can write (Bernardeau, 1994):

$$1 + \delta = \left(1 - \frac{\delta_{\text{lin}}}{\delta_{\text{sc}}}\right)^{-\delta_{\text{sc}}},$$

(4.11)

where we have suppressed the mass and redshift dependence. We have tested the approximation by checking that the distribution of $\delta_{\text{lin}}$ returned by this procedure is approximately Gaussian for our density fields (skewness $\sim -0.14$ and excess kurtosis $\sim -0.10$ at the lowest redshift). Going to maps with even smaller redshift bins might be desirable in the future, as radio telescopes can have frequency resolutions that are better than the 25 MHz chosen for our simulations. This would however either require better modelling of $\delta_{\text{lin}}$, for example by taking the initial conditions of the simulation into account, or fitting the conditional mass function directly to the non-linear density field.

Furthermore, the relation between mass and variance requires us to choose a smoothing filter. The derivation of the mass function from excursion set theory assumes a top-hat filter in Fourier space (Bond et al., 1991), while halos are typically defined as localised objects in real space. As the results only depend on the variance of the filtered field, standard practice is to use a spherical top-hat filter in real space to relate the halo mass $m$ to the variance $s$. In Sheth and Tormen (2002), also the mass $M$ of the local overdensity is related to a variance $S$ with a spherical top-hat filter. For the pixelated density field, however, a spherical filter is a poor approximation for relating $M$ and $S$, since our pixels have complicated elongated shapes. In order to be consistent, we therefore set $S$ to be the actual variance of the linearised pixelated density field across each map. We have checked that the procedure described above leads to an average mass function that is within 10% of the unconditional mass function obtained by setting $S \to 0$ and $\delta_{\text{lin}} \to 0$ in equation (4.9) (see Figure 4.1).

When populating the density field with halos, we first define a minimum and maximum halo mass. We then Poisson sample the number of halos in each pixel around a mean of

$$N(M) = \int_{m_{\text{min}}}^{m_{\text{max}}} dm \frac{N(m|M)}{N(M)},$$

(4.12)

where $M$ is the mass in the pixel. Finally, we sample a mass $m$ for each of the halos from the distribution $N(m|M)/N(M)$. We end up with a sample of halo masses that is drawn from the conditional mass function for each individual pixel on the map.

---

7In principle, one might worry that this procedure does not conserve mass (Sheth and Lemson, 1999); in our case, however, we have $m_{\text{max}} \ll M$ and we have checked that mass conservation is not a problem in practice.
4.2. Simulating HI Intensity Maps

Figure 4.1: Number of sampled halos within a given mass bin and averaged over all pixels in the lowest (blue) and highest (green) redshift bin. The dashed black lines show the predictions of the unconditional mass function from Sheth and Tormen (2002). The distribution of the sampled halos agrees with the prediction to better than 10%.

When following the described procedure, we however ignore the effect of peculiar velocities on the simulated maps. Peculiar velocities of the halos can shift the 21 cm line of the HI within some of the halos into neighbouring frequency bins, thereby distorting the line-of-sight boundaries between the pixels. As the prospects of redshift space distortion (RSD) measurements with the 21 cm line are promising (Raccanelli et al., 2015; Bull et al., 2015), we propose a way for including them in the simulations at the end of this section.

HI intensity maps

To simulate the sky as seen by a HI intensity mapping survey, we need to turn the distribution of halos and their masses into a distribution of HI mass. For a summary of different approaches to this problem see Padmanabhan et al. (2015) and Padmanabhan et al. (2016). Our approach closely follows work in Bagla et al. (2010): For halos that satisfy a constraint on the circular velocity of the halo, we assign a fraction of the halo mass as HI mass. As proposed in Padmanabhan et al. (2016), we use a halo mass to HI mass ratio \( \alpha = 0.15 \) together with an
Algorithm 1 Algorithm for simulating HI intensity maps.

1: create mass map of matter distribution from N-body lightcone output
2: for mass $M$ in pixel $i$ of the mass map do
3:   turn $M$ into an overdensity $\delta$
4:   turn $\delta$ into a linearised overdensity $\delta_{lin}$ using eq. (4.11)
5:   calculate $N(m|\delta)\, dm$ using eq. (4.9)
6:   sample the number of halos $n$ from a Poisson distribution centered on $N(M)$, eq. (4.12)
7:   sample $n$ halo masses $\{m_j\}_{j=1, \cdots, n}$ from $N(m|\delta)/N(M)$ using inversion sampling
8:   turn $\{m_j\}$ into HI masses $\{m_{HI, j}\}$ using eq. (4.13)
9:   set pixel $i$ of HI mass map to $\sum_j m_{HI, j}$
10: end for
11: turn HI mass map into 21 cm intensity map using eq. (4.1)

exponential cut-off at circular velocities $v_{c,0} = 30 \text{ km/s}$ and $v_{c,1} = 200 \text{ km/s}$:

$$m_{HI}(m) = \alpha (1 - Y_p) \frac{\Omega_b}{\Omega_m} m \exp\left(-\left(\frac{v_{c,0}}{v_{c}(m)}\right)^3\exp\left(-\left(\frac{v_{c,0}}{v_{c,1}}\right)^3\right)\right), \quad (4.13)$$

where $m$ is again the mass of the halo. $\Omega_b$ and $\Omega_m$ are set by the simulations and we use a Helium fraction of $Y_p = 0.24$. Using equation (4.13) to assign a HI mass to all the halos in a given pixel, the total HI mass in the pixel is simply given by the sum over all halos. We can finally rescale the resulting HI mass map into an intensity map of 21 cm brightness temperature by using equation (4.1).

Algorithm 1 gives a summary of the proposed procedure for generating HI intensity maps. For the dark matter halos, the main assumptions of our procedure are that the linearised field is well approximated by equation (4.11) (line 4) and that the conditional mass function sampling within our pixels (line 5 and 6) is a good description of the halo distribution. In section 4.3 we study those steps in more detail. Our main arguments for the validity of our approach are the good agreement of the resulting unconditional halo mass function (see Figure 4.1) and large scale bias (see Figures 4.4 and 4.9) with the analytical results from Sheth and Tormen (2002). The analytic expressions have been shown to be in reasonable agreement with N-body simulations (Reed et al., 2007; Crocce et al., 2010; Watson et al., 2013; Khandai et al., 2015; Hellwing et al., 2016), and the fact that we are reproducing the analytical results at the mass function and bias level is encouraging. We leave more detailed investigations of the accuracy of our procedure with high-resolution simulations of smaller volumes for future work.

Figure 4.2 shows the simulated HI intensity maps at the lowest and highest redshift ($z \approx 0.369, 0.863$, respectively). In Figure 4.3, we show the relation between dark matter overdensity $\delta$ and brightness temperature fluctuations $\delta T$ induced by the conditional halo mass function sampling. Similar to findings in studies with high resolution N-body simulations (see e.g. Manera and Gaztañaga, 2011), we find that the relation between halo overdensity and matter overdensity deviates from a linear behaviour for large overdensities. Figure 4.3 also shows the individual distributions of $\delta$ and $\delta T$ together with a log-normal distribution of the same mean
4.2. Simulating HI Intensity Maps

Figure 4.2: Simulated HI intensity maps in units of µK brightness temperature at \( z \approx 0.86 \) (top) and \( z \approx 0.369 \) (bottom). The zoom regions have side lengths of 15°.

and variance.

The left panel of Figure 4.4 shows a comparison of \( \Omega_{\text{HI}}(z) \) resulting from our prescriptions along with the data points of the compilation in Padmanabhan et al. (2015) consisting of damped Lyman-\( \alpha \) and intensity mapping observations. We can see that equation (4.13) yields a somewhat high \( \Omega_{\text{HI}} \) compared to the data compilation, as the prescription was fitted to another data compilation and for a cosmology with a lower \( \sigma_8 \). \( \Omega_{\text{HI}} \) could be matched more precisely by adjusting the parameter \( \alpha \) in equation (4.13) which would simply result in an overall rescaling of all maps. As all our results are independent of the overall amplitude, we decided to simply adopt the value of \( \alpha \) proposed in Padmanabhan et al. (2016). The mapping between halo mass and HI mass will not be one-to-one in reality, and more work has to go into the question on how much coarse graining is allowed when relating halo and HI mass.
Chapter 4. Future Observations and HI Intensity Mapping

Figure 4.3: Relation between matter overdensity $\delta$ and temperature fluctuations $\delta T/\bar{T}$ in the simulated maps at redshift $z \approx 0.863$ (right) and $z \approx 0.369$ (left). The one-dimensional histograms show the individual distributions of $\delta$ and $\delta T/\bar{T}$ along with a log-normal distribution with the same mean and variance (black line). The two-dimensional histograms show the non-linear, stochastic relation between $\delta$ and $\delta T/\bar{T}$ as compared to a simple linear bias model (black line). Both the colour-scaling of the two-dimensional histograms and the y-axis of the one-dimensional histograms are log-scaled.

as more data becomes available. A more detailed modelling of e.g. the scatter around this relation is however beyond the scope of this work.

Including peculiar velocities

When assigning the halos to pixels in the maps, one has to take into account that, depending on the line-of-sight velocities of the halos and the position of the halos within the radial extension of the bin, the 21 cm line in some of the halos will shift into the neighbouring frequency bins. Until now, we neglected this effect called redshift space distortions (RSDs) and we will now discuss a possible way of including them in our algorithm.

The bin-width of our maps corresponds to a frequency bandwidth of $\Delta f = 25$ MHz. For a halo at the centre of the bin to escape the bin, it would thus require velocities larger than $3 \times 10^3$ km s$^{-1}$ which are rarely achieved. At the boundaries of the bin, however, the RSDs due to peculiar velocities are distorting the shape of this boundary.

To model the effect of RSDs on our maps, we start, as before, with dark matter maps for which we assign each particle of the N-body simulation within the lightcone to the bin which
4.2. Simulating HI Intensity Maps

![Graph showing comparison of density and large scale biasing of HI as a function of redshift.]

Figure 4.4: Comparison of density (left panel) and large scale biasing of HI (right panel) as a function of redshift as predicted by the unconditional mass function from Sheth and Tormen (2002) (green) for the prescription of eq. (4.13), as measured on our simulated intensity maps (red), and from the compilation in Padmanabhan et al. (2015) (blue). The values from the unconditional halo model agree well with the simulations by construction, with small deviations coming from the imperfect linearisation (equation (4.11)). The $\Omega_{\text{HI}}$ values from our prescription are slightly high with respect to the compilation in Padmanabhan et al. (2015). Once better data is available, a more significant mismatch in amplitude could be easily fixed by adjusting $\alpha$ in equation (4.13).

corresponds to its position in redshift according to its co-moving distance. In addition to the number $n_{ij}$ of particles within pixel $i$ of redshift bin $j$, we now additionally count the number $\tilde{n}_{ij}(k)$ of particles within this volume that fall into redshift bin $k$ when furthermore taking their line of sight velocity into account. The number of particles $\tilde{n}_{ik}$ within pixel $i$ of redshift bin $k$ of the distorted dark matter maps (including RSDs) is then given by $\tilde{n}_{ik} = \sum_j \tilde{n}_{ij}(k)$. Only a small fraction of the particles gets redistributed by this procedure, but due to the coherence of the effect on large scales, even these mild changes enhance the large scale power spectrum of the field (Kaiser, 1987).

Our approach for assigning halos to pixels is based on the insight that the number of halos within each pixel can be modelled using the conditional mass function of Sheth and Tormen (2002). The procedure however does not model the positions of the halos within the pixels or their velocities. Assuming that the halo velocities are unbiased with respect to the dark matter flow (see e.g. Desjacques and Sheth, 2010, for a discussion), we can however use the information about the distorted dark matter particle distribution $\tilde{n}_{ij}(k)$ in order to imprint the RSDs on the HI intensity maps. Starting from the unperturbed dark matter density field, we follow the procedure described in section 4.2 to assign an HI mass $m_{ij}^{\text{HI}}$ to pixel $i$ of redshift bin $j$. Instead of adding the complete HI mass $m_{ij}^{\text{HI}}$ to this pixel, we however distribute the mass over all redshifts according to the fraction $\tilde{n}_{ij}(k)/n_{ij}$ of dark matter particles that got redistributed due to their line of sight velocity. In pixel $i$ of redshift bin $k$, the distorted HI
Chapter 4. Future Observations and HI Intensity Mapping

maps consequently contain a mass of

\[ \tilde{m}_{ij}^{\text{HI}} = \sum_j m_{ij}^{\text{HI}} \tilde{n}_{ij}(k)/n_{ij}. \] (4.14)

In section 4.4 we will show that those modifications indeed yield distorted maps that exhibit angular power spectra consistent with linear theory predictions for RSDs.

4.3 Testing the Sub-Grid Halo Model

In this section, we compare our prescription for sampling halos from the density field with the resolved halos in the simulation. Our sub-grid model involves the linearisation of the non-linear density field (Bernardeau, 1994) and the sampling of halos from the conditional mass function from Sheth and Tormen (2002). In the following, we first analyse the linearisation before proceeding to a comparison between the resolved and sub-sampled halo populations.

Linearisation

Figure 4.5 shows the distribution of overdensities in the dark matter maps before and after linearisation for the lowest and highest redshift bin. We expect the linearised overdensities to follow a Gaussian distribution of mean zero, but the variance of the Gaussian depends on the filtering of the field by our pixels and is hence hard to predict from the linear power spectrum. We therefore also plot a Gaussian with the same mean and variance as the distribution of linearised overdensities. We can see in Figure 4.5 that the agreement is reasonable, but the distribution of overdensities is shifted (means of \(-0.27\) and \(-0.09\) for the low and high redshift bin) and skewed (skewness of \(-0.14\) and \(-0.10\) for the low and high redshift bin) towards negative overdensities in particular for the lowest redshift bins. This means that equation (4.11) fails at correctly reproducing the high overdensity tail of the linearised field for our pixels.

Using the angular power spectrum estimators briefly introduced in section 4.1, we can also study the clustering of the linearised field. Figure 4.6 shows the angular power spectrum of the non-linear and linearised field as compared to the respective predictions from linear theory and non-linear corrections. The theory prediction (black) for the matter power spectrum are calculated with CLASS (Blas et al., 2011; Di Dio et al., 2013) and halofit (Smith et al., 2003; Takahashi et al., 2012). We can see that, overall, the correlations of the linearised field follow the linear theory prediction. The bottom panel shows, however, that the linearised field has deviations of around 15% when compared to its prediction while the non-linear field agrees with the non-linear prediction to better than 5%.
4.3. Testing the Sub-Grid Halo Model

Figure 4.5: Histogram of non-linear (blue) and linearised (green) matter overdensities in our maps at the lowest \((z \approx 0.369, \text{ left})\) and highest \((z \approx 0.863, \text{ right})\) redshift. The non-linear overdensities have a tail towards high overdensities and are, by definition, cut off at an underdensity of -1. The distribution of the linearised overdensities is overplotted with a Gaussian distribution of the same mean and variance (black line). We can see that the linearised overdensities approximately follow the Gaussian distribution, but are shifted and skewed to negative values in particular for the lowest redshift bin.

Comparison of halo populations

In this section, we compare the resolved halos in the simulation to the halos sampled from the density field of the same simulation with our procedure. We used rockstar (Behroozi et al., 2013) for finding halos in the lightcones of the simulation and use the mass \(M_{200,c}\) within the region where the density is 200 times higher than the critical density as halo mass.

Figure 4.7 shows the agreement between the number of halos within a given mass range as observed from the population of resolved and sampled halos as well as the distribution predicted from the unconditional mass function from Sheth and Tormen (2002). The simulations have a mass resolution of \(1.6 \times 10^{11} \, h^{-1} M_\odot\) and the distribution of resolved halos is therefore not complete for masses below approximately \(10^{13} h^{-1} M_\odot\). The exponential cut-off scale in the HI assignment, however, is at roughly \(10^{12} h^{-1} M_\odot\) so in our model the resolved halos would
Chapter 4. Future Observations and HI Intensity Mapping

Figure 4.6: Angular power spectrum of non-linear (blue) and linearised (green) matter overdensities. The spectrum of the linearised field and is supposed to follow the linear theory power spectrum (black dashed line). The theory prediction (black) for the matter power spectrum are calculated with CLASS (Blas et al., 2011; Di Dio et al., 2013) and halofit (Smith et al., 2003; Takahashi et al., 2012). We can see that, in particular for the lowest redshift bin, the estimated and predicted linear power spectra deviate by around 15%.

not contribute significantly to the overall HI distribution. Close to $10^{13} \, h^{-1} \, M_\odot$, the resolved and sampled halo populations indeed agree to about 10%. The high mass tail of the sampled halo distribution at low redshifts, however, diverges from the resolved population. This is most likely due to the shortcomings of the linearisation process at the high overdensities discussed in the previous section. As these halos do not contain significant amounts of HI in our prescription, this mismatch is however not critical for our application. At the highest redshift bin our method seems to extend even to this high mass end.

In Figure 4.8 we show the distribution of pixel masses in the matter density and the halo field at the highest redshift. The halos are either the resolved halos from the simulations (filled contours) or the sampled halos (lines). The plot shows that both the distribution and the scatter in the halo mass distribution of the sampled halos is very similar to the distribution of the resolved halos. There are mild differences in the scatter that are most likely related to the simple Poisson sampling scheme. For the same reason, some of the pixels slightly violate mass conservation. None of these shortcomings are expected to affect the halos within the mass range relevant for our intensity maps.
4.4 The Angular Power Spectrum of HI Intensity Maps

Using the maps described in the previous section, we study the wide field clustering properties of the HI intensity fluctuations, focusing in particular on the impact of non-linearities. HI intensity maps are characterised by having both the wide field and the relatively low angular resolution of Cosmic Microwave Background (CMB) maps and the three dimensional, non-Gaussian large-scale structure of galaxy surveys. We therefore measure the angular power spectrum $C_\ell$ of the HI intensity maps using the pseudo $C_\ell$ estimator (Peebles, 1973) and the publicly available Po1SpicE estimator (Szapudi et al., 2001a; Chon et al., 2004) that were briefly introduced in section 4.1. These estimators have been developed for the CMB and have also been applied to galaxy surveys (see e.g. Thomas, S A and Abdalla, F B and Lahav, O, 2011; Becker et al., 2015). We assess their relative performance in this new regime and compare the
Chapter 4. Future Observations and HI Intensity Mapping

Figure 4.8: Comparison of mass distributions in pixels for the density field and the halos in the highest redshift bin. We only consider halos with mass greater than $10^{13} h^{-1} M_\odot$. The filled contours show the distribution of the resolved halos and the line contours show the distributions of the sampled halos. Note that the plot is not showing the pixels which contain no halos.

results to expectations from analytic models. The non-linearities are strongest at low redshifts, and we will in the following focus on the auto-correlation spectrum $C_\ell(z)$ in the lowest redshift bin. We contrast the results with the auto-correlation spectrum in the highest redshift bin to compare the performance for different levels of non-Gaussianity.

Results from simulations

In Figure 4.9, we show the angular power spectrum as estimated from the matter density and HI intensity maps using PolSpice, averaged over all 40 realisations. We also show the theory predictions for the angular power spectrum of the matter density field from equation (4.5) as calculated with CLASS (Blas et al., 2011; Di Dio et al., 2013), using halofit (Smith et al., 2003; Takahashi et al., 2012) to model the non-linearities. The bottom panel shows the ratio of the estimated to the predicted power spectra.
4.4. The Angular Power Spectrum of HI Intensity Maps

Figure 4.9: Angular power spectrum as estimated from the matter density (green) and HI intensity maps (blue) using PolSpice, averaged over all realisations of the simulations. We show the dimensionless spectrum for the intensity maps, scaled by the mean brightness temperature. The theory prediction (black) for the matter power spectrum are calculated with CLASS (Blas et al., 2011; Di Dio et al., 2013) and halofit (Smith et al., 2003; Takahashi et al., 2012). The thick lines show the spectra when ignoring RSDs, while the thin lines are showing the results for the distorted maps. The bottom panel shows the ratio between the predicted matter spectrum and the estimated spectra for the matter and intensity maps without RSDs. As expected, the spectrum of the matter field is consistent with the prediction and their ratio is close to one. Deviations after $\ell \approx 700$ are due to pixel effects. The ratio of intensity mapping spectra and theory prediction for the matter field should be compared to the large scale HI bias from equation (4.3) (red). We see that the HI bias matches the ratio at large scales ($\ell \lesssim 50$) and that there is a scale dependence at smaller scales.

As expected, the matter power spectrum is consistent with the theory prediction over most scales with a deviation at small scales ($\ell > 700$) due to pixel effects. At the power spectrum level, the effect of non-linearities in the density field is hence well modelled by the halofit corrections to the linear power spectrum. The agreement also confirms that, to the precision we are interested in, the overlap of a few percent per mass map within the simulated box (see also section 4.2) is not affecting the angular power spectra.

The ratio between the intensity mapping spectra and the predicted matter density is seen to be close to $b_{\text{HI}}^2$ on large scales ($\ell \lesssim 50$). Even though the agreement with the analytic prediction for the large scale bias is expected, it is a non-trivial consequence of the conditioning of the mass function on the overdensities within each pixel. Figure 4.9 hence shows that our prescription for assigning halos to the density field reproduces the halo statistics at the two-point level.
Chapter 4. Future Observations and HI Intensity Mapping

on large scales remarkably well. On smaller scales, our modelling of the halo distribution with the conditional mass function yields a mildly scale dependent bias in particular at the lowest redshift, where the ratio falls by $\sim 10\%$ of its large scale value at $\ell \approx 200$. The right panel of Figure 4.4 shows a comparison of $b_{HI}$ from equation (4.3) with the large-scale bias in the simulated intensity maps, estimated as the square root of the average ratio of the intensity mapping angular power spectra and the theory predictions on scales $10 < \ell < 50$. Figure 4.4 also shows good agreement with a compilation of bias values in Padmanabhan et al. (2015), derived from a selection of theoretical prescriptions from the literature.

Effect of redshift space distortions

An important effect that changes the apparent clustering of observed intensity maps are redshift space distortions. We described a possible way of including this effect in our simulations in section 4.2. By estimating the angular power spectra of those distorted maps, we can study the validity of this approach as well as the effect of redshift space distortions on the performance of the angular power spectrum estimators.

Following the procedure outlined in section 4.2, we created distorted dark matter and HI intensity maps from four independent simulations. Figure 4.9 shows a comparison of the angular power spectra as estimated from the dark matter maps with (thin green line) and without (thick green line) RSDs. Figure 4.9 also shows the angular power spectra of the intensity maps when including redshift space distortions (thin blue lines). In linear theory, the redshift space distortions lead to a modification of the power spectrum (Kaiser, 1987) that results in an enhancement of the angular power spectrum at large scales but does not affect scales much smaller than the size of the redshift bin (see e.g. Padmanabhan et al., 2007; Bonvin and Durrer, 2011, for more details).

The top panel of Figure 4.10 additionally shows that the enhancement of the large scale correlations in the distorted maps is as expected from the theoretical prediction by CLASS (see Di Dio et al. (2013) for details on RSDs in CLASS). In the middle panel of Figure 4.10 it can be seen that the power spectrum estimated from the distorted maps follows the theoretical prediction for a biased tracer with unbiased velocities. The bottom panel finally shows the agreement of the large scale bias in the distorted maps with the analytical predictions.

4.5 Covariance of the Angular Power Spectrum

In this section, we will study the covariance of the angular power spectrum estimators which is of particular importance for interpreting upcoming HI surveys and for assessing their constraining power. As for the power spectrum, we apply covariance estimators used for the CMB and galaxy surveys. We estimate the covariance of the aforementioned $C_\ell$ estimators using jackknife and analytical estimates based on Gaussian statistics and compare the results to the covariance estimated from a suite of simulated HI intensity maps.
4.5. Covariance of the Angular Power Spectrum

We first describe our covariance estimation procedure in more detail. To get some intuition for the expected performance of the different estimators, we then study the simplified case of stochastic field simulations. These stochastic field simulations provide an ideal testbed since independent realisations can be generated quickly. We next discuss the results from the simulated intensity maps described in section 4.2, focusing in particular on the effect of non-linearities on the covariance of the angular power spectrum. Finally, we briefly show the effect of redshift space distortions on the estimated covariances.

**Estimation**

We describe and compare three standard approaches to the problem of covariance estimation: estimation from multiple simulations, estimators based on Gaussian random fields and a jackknife approach.
Multiple simulations As already mentioned in section 4.2, we use the density field from 10 independent N-body simulations as the basis for our intensity maps. We use each of the boxes to create 4 simulations which cover a quarter of the sky. If we focus on the first quadrant of each simulation, then we estimate the covariance in this case as:

$$\text{Cov}(C_\ell, C_{\ell'}) = \frac{1}{n-1} \sum_{i=0}^{n} (C_i^\ell - \bar{C}_\ell)(C_i^\ell' - \bar{C}_{\ell'}),$$

(4.15)

where $n = 10$, $C_i^\ell$ is the power spectrum estimated from the $i^{th}$ realisation (this could be either the pseudo $C_\ell$ or PolSpice estimator), $\bar{C}_\ell = (1/n) \sum_{i=0}^{n} C_i^\ell$ is the mean over all realisations and the $n - 1$ denominator is chosen to yield an unbiased estimator of the underlying covariance. We do this for each quadrant and take the final covariance matrix estimator to be the arithmetic mean of Cov($C_\ell, C_{\ell'}$) over all four quadrants. This way we avoid that correlations between the four quadrants introduced through the finite size of the simulated box affect the estimated covariance.

Gaussian approximation In the case of an isotropic Gaussian random field on a full sphere, the coefficients $a_{\ell m}$ of the spherical harmonic decomposition are independent Gaussian random variables and identically distributed for each $\ell$. Consequently, the estimator $\hat{C}_\ell$ (4.6) is chi-squared distributed with mean $C_\ell$ and variance (Knox, 1995)

$$\text{Var}(\hat{C}_\ell) = \frac{2}{2\ell + 1} C_\ell^2.$$  

(4.16)

The covariance Cov($C_\ell, C_{\ell'}$) is zero for $\ell \neq \ell'$ in this case.

As for the estimation of the $C_\ell$s, these considerations get more involved in the case of partial sky coverage (see e.g. Tegmark and de Oliveira-Costa, 2001; Bond et al., 1998; Szapudi et al., 2001b; Hivon et al., 2002; Efstathiou, 2004; Cabre et al., 2007). A standard approximation for the variance of the pseudo $C_\ell$ estimator is given by Hivon et al. (2002):

$$\text{Var}(\hat{C}_\ell^{\text{pseudo}}) \approx \frac{2}{(2\ell + 1)f_{\text{sky}}} C_\ell^2,$$  

(4.17)

where as before $f_{\text{sky}}$ is the fraction of unmasked sky and Cov($\hat{C}_\ell^{\text{pseudo}}, \hat{C}_{\ell'}^{\text{pseudo}}$) is approximated to be zero for $\ell \neq \ell'$. Since this approximation neglects correlations between individual $\ell$ scales as introduced by the mask, it is a good approximation only for bandpowers

$$B_\ell = \frac{1}{2\Delta \ell + 1} \sum_{\ell = \ell - \Delta \ell}^{\ell + \Delta \ell} C_\ell,$$  

(4.18)

where $\Delta \ell$ is chosen such that the correlations between the bandpowers is small. Following the
4.5. Covariance of the Angular Power Spectrum

definition of $B_\ell$, the variance of the bandpowers relates to the $C_\ell$ variance as follows:

$$\text{Var}(B_\ell) = \frac{1}{(2\Delta\ell + 1)^2} \sum_{\ell'=\ell-\Delta\ell}^{\ell+\Delta\ell} \sum_{\ell''=\ell-\Delta\ell}^{\ell+\Delta\ell} \text{Cov}(C_{\ell'}, C_{\ell''}).$$  (4.19)

Besides the pseudo $C_\ell$ approach, we will also use the PolSpice estimator for the covariance (Efstathiou, 2004) which is based on Gaussian field statistics and takes the mask into account. We calculate the Gaussian field estimators for each simulation individually.

**Jackknife estimation**  Resampling techniques aim at estimating the covariance of an estimator by manipulating the sample on which it is based. The advantage of resampling techniques is that they do not assume a field with specific statistical properties. The disadvantage is that they typically assume independence of different parts of the sample which is not satisfied by correlated fields. In a jackknife approach, one studies the behaviour of an estimator when parts of the data are ignored in the estimation process. In the context of our maps this implies that we study the distribution of angular power spectra as derived from maps where a subset of the simulated brightness temperature pixels is masked. We split the simulated part of the sky into $n_{\text{jack}}$ parts and estimate all power spectra $C_{\ell,i}^{\text{jack}}$ (using either the pseudo $C_\ell$ or PolSpice estimators) when the $i$th of the $n_{\text{jack}}$ parts is masked out. The covariance of the angular power spectrum is then estimated as

$$\text{Cov}_{\text{jack}}(C_\ell, C_{\ell'}) = \frac{n_{\text{jack}} - 1}{n_{\text{jack}}} \sum_{i=1}^{n_{\text{jack}}} (C_{\ell,i}^{\text{jack}} - \bar{C}_\ell^{\text{jack}})(C_{\ell',i}^{\text{jack}} - \bar{C}_{\ell'}^{\text{jack}}).$$  (4.20)

with $\bar{C}_\ell^{\text{jack}} = (1/n_{\text{jack}}) \sum_{i=1}^{n_{\text{jack}}} C_{\ell,i}^{\text{jack}}$. In our analysis, we define the masked-out patches as pixels in coarser HEALPix maps. The size of the patches was determined after an analysis of stochastic fields. As for the Gaussian field estimate, we calculate a jackknife estimate individually for each simulation.

**Presentation of estimates**  To present our results on the estimated covariances, we decompose the covariance into its diagonal part (the variance) and its off-diagonal part (the correlation matrix) as given by:

$$\sigma^2(C_\ell) = \text{Cov}(C_\ell, C_\ell),$$  (4.21)

$$\text{Corr}(C_\ell, C_{\ell'}) = \frac{\text{Cov}(C_\ell, C_{\ell'})}{\sigma(C_\ell)\sigma(C_{\ell'})}. $$  (4.22)

To reduce the statistical noise of the estimated correlation matrix, we also show the correlation $\text{Corr}(C_\ell, C_{\ell+\delta\ell})$ as a function of $\delta\ell$ when averaged over neighbouring $\ell$ values (we decide to
average over \( n = 100 \ell \) values):

\[
\text{Corr}(C_\ell, C_{\ell+\delta\ell}) = \frac{1}{n} \sum_{i=-n/2}^{n/2} \text{Corr}(C_{\ell+i}, C_{\ell+i+\delta\ell}).
\]

(4.23)

For simplicity, we will call the results regarding the covariance estimated from the multiple, independent realisations the \textit{simulation estimator} in the following.

### Stochastic field models

In this section we analyse the properties of pseudo \( C_\ell \) and PolSpice estimators when applied to masked random fields. We will compare the results from many realisations of maps with the same power spectrum to the results from both Gaussian field predictions and resampling methods. The mask is chosen such that it matches the mask which we also impose on our simulated intensity maps. We will use two types of random fields: Gaussian fields and log-normal fields.

**Gaussian fields** As input power spectrum, we use the angular power spectrum for redshift \( z \approx 0.863 \) with a bin width of \( \Delta z \approx 0.06 \) as calculated from linear theory with halofit corrections using CLASS (Blas et al., 2011; Di Dio et al., 2013). The highest redshift map is closest to a Gaussian random field and closest to the linear theory prediction, so this analysis is used as a mock for the high redshift behaviour of the covariance. We rescale each density field to an intensity map using equation (4.8). To estimate the variance and covariance of the angular power spectrum estimator, we apply jackknife and the predictions for Gaussian fields to 100 random realisations of the field. We compare these results to the values inferred from \( 10^4 \) random realisations from the HEALPix routine \texttt{synfast}.

Considering the diagonal of the covariance matrix \( \sigma^2(C_\ell) = \text{Cov}(C_\ell, C_\ell) \), the top panel of Figure 4.11 shows a comparison of \( \sigma(C_\ell) \) estimates from Gaussian fields together with the results from the \( 10^4 \) realisations of a Gaussian random field with the same input power spectrum. The pseudo \( C_\ell \) estimator overestimates \( \sigma(C_\ell) \) by a factor of \( \sim 2 \), while the prediction from PolSpice agrees well with the result from the \( 10^4 \) realisations. The pseudo \( C_\ell \) estimate for the error is getting closer to the correct value if we consider the variance of bandpowers that are averaged over bins in \( \ell \). This however comes at the cost of loosing more and more information on the spectrum itself due to the averaging.

The bottom panel of Figure 4.11 shows the results for the jackknife variance estimates. We use the PolSpice estimator for estimating the angular power spectrum of the maps, but the results are similar when using pseudo \( C_\ell \) values. To create the jackknife sample, we mask patches that correspond to pixels in coarser HEALPix maps. Each colour in the bottom panel of Figure 4.11 refers to a different patch size. The number of jackknives \( n_{\text{jack}} \) given in the legend is inversely proportional to the size of the patch. We can see that the error estimate
4.5. Covariance of the Angular Power Spectrum

Figure 4.11: Standard deviation $\sigma(C_\ell)$ of the angular power spectrum from multiple realisations of the masked Gaussian field (black) as compared to the pseudo $C_\ell$ (blue) and PolSpice (green) estimators for Gaussian fields (top panel) and to different jackknife configurations (bottom panel). While the PolSpice estimate is consistent with the result from $10^4$ realisations, the pseudo $C_\ell$ estimate overestimates $\sigma(C_\ell)$ by a factor of $\sim 2$ to compensate for the lack of correlations. As we will later demonstrate for the simulated intensity maps, this mismatch can be reduced by considering bandpowers that are binned in $\ell$. The jackknife error estimates become more and more biased to higher values as $n_{\text{jack}}$ increases, while the variance of the error estimate decreases.
Figure 4.12: Standard deviation $\sigma(C_{\ell})$ of the angular power spectrum from multiple realisations of the masked log-normal field as compared to predictions from Gaussian field statistics for pseudo $C_{\ell}$ and PolSpice estimators. While the pseudo $C_{\ell}$ estimator is again biased to higher $\sigma(C_{\ell})$ by a factor of 2, the PolSpice estimator only shows a mild deviation of a view percent at small scales (inset in the top right corner).

with the smallest bias is coming from the jackknife with largest patch size. The higher we go in $n_{\text{jack}}$—i.e. the lower in patch size—the smaller is the noise on the estimate but the larger is the disagreement between the jackknife estimate and the repeated simulations. To compromise between variance and bias of the jackknife estimate, we chose $n_{\text{jack}} = 48$ as our jackknife configuration for the remaining analysis. For a jackknife with $n_{\text{jack}} = 48$, $\sigma(C_{\ell})$ is biased to higher values by about 40%.

Log-normal fields We want to use a log-normal field with correlations that are close to the prediction from the intensity map of the lowest redshift slice ($z \approx 0.366 \pm 0.026$) in order to approximate its non-Gaussian nature. As we cannot easily simulate a log-normal field with specified power spectrum, we start from a Gaussian field created from an input power spectrum given by the linear theory prediction at this redshift. When exponentiating a Gaussian field $X$ with mean $\mu = 0$ and variance $\sigma^2$, the resulting field is log-normally distributed with mean $\exp(\sigma^2/2)$ and variance $(\exp(\sigma^2) - 1) \exp(\sigma^2)$. The following field $Y$ can hence be shown
4.5. Covariance of the Angular Power Spectrum

Figure 4.13: Correlations $\text{Corr}(C_\ell, C_\ell')$ of the angular power spectrum from multiple realisations of the masked log-normal field as compared to predictions from Gaussian field statistics for pseudo $C_\ell$ and PolSpice estimators. The correlations are shown for pseudo $C_\ell$ (left) and PolSpice (right) and split into large scales ($\ell = 25$, top) and small scales ($\ell = 650$, bottom).

We use the map as given by $Y$ as an approximation for the desired intensity map, but need to be aware that its angular power spectrum is not necessarily consistent with the input to the initial Gaussian field.

As shown in Figure 4.12, the variances as predicted using Gaussian field statistics for both pseudo $C_\ell$ and PolSpice are affected by the change in the underlying distribution only very mildly at small scales. As expected, the jackknife results are independent of the underlying distribution and are equivalent to the ones for Gaussian fields shown in the previous section.

We now turn our attention to the correlation matrix $\text{Corr}(C_\ell, C_\ell') = \frac{\text{Cov}(C_\ell, C_\ell')}{\sigma(C_\ell)\sigma(C_\ell')}$. Figure 4.13 shows the correlations between individual $\ell$ values at large scales around $\ell = 25$ and small scales around $\ell = 650$.

To be log-normally distributed with variance $\sigma_T^2$ and mean equal to $T$:

$$Y = T \exp \left( \frac{\sqrt{\log((\sigma_T/T)^2 + 1)}}{\sigma} X - \frac{1}{2} \log((\sigma_T/T)^2 + 1) \right). \quad (4.24)$$

We use the map as given by $Y$ as an approximation for the desired intensity map, but need to be aware that its angular power spectrum is not necessarily consistent with the input to the initial Gaussian field.
Chapter 4. Future Observations and HI Intensity Mapping

Figure 4.14: Mean correlation $\langle \text{Corr}(C_\ell, C_{\ell+\delta\ell}) \rangle$ for the PolSpice angular power spectrum of the masked log-normal field, averaged over 100 $\ell$s. We show estimates from the multiple realisations, the jackknife, and the predictions for Gaussian fields. Correlations between large scales ($0 \leq \ell < 100$) are shown on the left, small scales ($600 \leq \ell < 700$) on the right.

scales around $\ell = 650$. As expected, the large scale part of the correlation matrix is not affected by the non-Gaussianity as the perturbations on large scales are still approximately Gaussian. Looking at the correlations between individual $\ell$ values on small scales, however, we find small amounts of extra correlations introduced by the log-normal field for scales separated by $\delta\ell > 5$. Due to the assumption of an isotropic Gaussian field, the extra correlations are not estimated correctly by PolSpice. The jackknife estimator, however, is able to pick these correlations up for $\delta\ell \gtrsim 15$ while it is biased to negative correlations for $5 \lesssim \delta\ell \lesssim 15$. The variance of the jackknife estimator (shown by the shaded area) is too large to resolve the extra correlations when estimated from a single realisation of the field. Figure 4.14 hence shows the correlation $\text{Corr}(C_\ell, C_{\ell+\delta\ell})$ as a function of $\delta\ell$ when averaged over 100 neighbouring $\ell$ values. Again, there are additional correlations at small scales ($600 \leq \ell < 700$, right panel), but this time the noise in the jackknife estimate is just about small enough to detect those correlations (for $\delta\ell \gtrsim 15$) even from a single realisation.

Results from simulations

After studying the simplified case of stochastic random field simulations in the previous section, we now analyse the covariance of the angular power spectra from the simulated HI intensity maps described in section 4.2. We first discuss the variance and correlations of the undistorted maps before considering the effect of redshift space distortions on the estimated covariance.
4.5. Covariance of the Angular Power Spectrum

Figure 4.15: Standard deviation $\sigma(C_\ell)$ of pseudo $C_\ell$ (top) and Po1Spice (bottom) estimators as estimated from the 10 independent simulations (black), the jackknife (blue), and the Gaussian assumption (green) for the lowest ($z \approx 0.369$, left panel) and highest ($z \approx 0.863$, right panel) redshift bin of the simulations. The bottom panel of each plot shows the ratio of the jackknife and Gaussian estimator to the result from the 10 realisations. The shaded areas in the bottom panels show the standard deviation of the ratio, estimated from the distribution of the jackknife and Gaussian estimator over the 10 realisations. For the pseudo $C_\ell$ estimator, $\sigma(C_\ell)$ is overpredicted by the Gaussian estimator by a factor of $\sim 2$. Within the noise, the Po1Spice estimate is consistent with the result from the independent simulations. The jackknife estimate overestimates $\sigma(C_\ell)$ by about 40%, independently of scale, redshift, and estimator.
Figure 4.16: Standard deviation $\sigma(B_\ell)$ of pseudo $C_\ell$ bandpowers defined in equation (4.18) as estimated from the 10 independent simulations, the jackknife, and the Gaussian assumption for the lowest ($z \approx 0.369$, left panel) and highest ($z \approx 0.863$, right panel) redshift bin of our simulations. The bottom panel shows the ratio of the jackknife and Gaussian estimator to the result from the 10 realisations. The shaded areas in the bottom panels show the standard deviation of the ratio, estimated from the distribution of the jackknife and Gaussian estimator over the 10 realisations. The bandpowers are averaged over a bandwidth of $\Delta \ell = 5$. The bias of the pseudo $C_\ell$ estimate for $\sigma(B_\ell)$ is dropping from 90% in the case of the unbinned spectrum (Figure 4.15) to below 10% for the bandpowers.

Variance Figure 4.15 shows the different estimates of $\sigma(C_\ell)$ for the lowest ($z \approx 0.369$) and highest ($z \approx 0.863$) redshift bin of the intensity map simulations presented in section 4.2. For all redshifts and estimators, the jackknife estimator (4.20) is consistently biased\textsuperscript{8} to higher $\sigma(C_\ell)$ by approximately $\sim 40\%$, an effect that is also seen for the stochastic field models in the previous section. The \texttt{PolSpice} estimator is consistent with the simulation estimator within the statistical noise. We find that the pseudo $C_\ell$ estimate for $\sigma(C_\ell)$ given in equation (4.17) predicts a diagonal error on the angular power spectrum that is a factor of $\sim 2$ higher than the simulation estimator, compensating for the lack of correlations (see Figure 4.15). To account for this effect, we also show the results for the variance of pseudo $C_\ell$ bandpowers $B_\ell$ for a

\textsuperscript{8}Whenever we talk about bias in this section, we refer to the bias of an estimator in the statistical sense, i.e. as a mismatch between expected value of the estimator and the true, underlying quantity. This is not to be confused with the concept of halo and HI bias discussed in sections 4.1 and 4.4.
4.5. Covariance of the Angular Power Spectrum

Figure 4.17: Correlations $\text{Corr}(C_\ell, C_{\ell'})$ for the PolSpice angular power spectrum as estimated from the 10 independent simulations, the jackknife, and the Gaussian assumption for the lowest ($z \approx 0.369$, top panels) and highest ($z \approx 0.863$, bottom panels) redshift bin of the simulations. Correlations between large scales (around $\ell = 25$) are shown on the left, small scales (around $\ell = 650$) are shown on the right. The shaded areas show the standard deviation of the jackknife estimator from the 10 independent simulations. The estimate from the 10 realisations is consistent with the PolSpice result within the noise.

Overall, the variance is hence in good agreement with predictions for Gaussian fields, even though the temperature fluctuations follow the complex distribution imprinted by gravitational clustering. Being more biased while having a similar variance, the results from the jackknife approach are clearly outperformed by the Gaussian estimate. The results from this section are consistent with our analysis of stochastic fields in the previous section.

**Correlations**  For a full-sky Gaussian field that is statistically homogeneous and isotropic, we do not expect any correlations between different $\ell$ scales. Non-zero off-diagonal elements in
the correlation matrix can however be introduced by the finite mask of the map and through the non-Gaussian nature of the matter density field. Since the PolSpice estimate of the angular power spectrum and its covariance already accounts for the mask, it is particularly interesting to compare the latter with correlation estimates from the independent simulations and the jackknife, thereby enabling us to distinguish masking effects from non-Gaussianity.

Figure 4.17 shows Corr($C_{\ell}, C_{\ell'}$) as a function of $\ell'$ for the lowest (top row) and the highest (bottom row) redshift bin and at small ($\ell = 650$, left column) and large scales ($\ell = 25$, right column). Overall, the PolSpice estimate appears to be in good agreement with the simulation estimator within the noise. The jackknife estimate is biased to negative correlations for scales separated by $5 \lesssim \delta \ell \lesssim 15$ and is in good agreement with the simulation and PolSpice estimate otherwise.

To reduce the statistical noise of the simulation estimator for the correlations, we also show
the correlation $\text{Corr}(\hat{C}_\ell, \hat{C}_{\ell+\delta \ell})$ as a function of $\delta \ell$ when averaged over neighbouring $\ell$ values (we decide to average over 100 $\ell$s). The result is shown in Figure 4.18. Again, we find good agreement between the PolSpice estimate and the simulation results. Yet, for small scales and low redshifts (top right of Figure 4.18), there are $\sim 5\%$ extra correlations for $\delta \ell > 15$ in both jackknife and simulation estimators. The bias to negative correlations of the jackknife estimate for $5 \lesssim \delta \ell \lesssim 15$ can be seen in all four panels of Figure 4.18.

The pseudo $C_\ell$ estimator for the covariance neglects correlations, hence its correlation matrix is simply the identity matrix. As expected, both the simulation and the jackknife estimator show that the pseudo $C_\ell$ covariance is not diagonal in practice. Its off-diagonal contributions from mask and non-Gaussianity are very similar to those reported for the PolSpice covariance.

The overall agreement between the correlations predicted for a masked Gaussian field, the simulation estimator, and the jackknife is remarkably good. While the correlations between
Chapter 4. Future Observations and HI Intensity Mapping

\[ z = 0.369, \ 600 \leq \ell < 700 \]

\[ \langle \text{Corr}(C_{\ell}, C_{\ell+\delta\ell}) \rangle \]

Simulations w/o RSDs
Simulations w/ RSDs
PolSpice

Figure 4.20: Correlations of the angular power spectra of the HI intensity maps with (green) and without (blue) RSDs as estimated from our analysis of four realisations of the simulations. The corresponding estimates for Gaussian random fields from PolSpice are shown in black. To reduce the noise in the correlation matrix estimate, we show the correlations between neighbouring \( \ell \) values averaged over \( 600 \leq \ell < 700 \).

individual \( \ell \) values estimated from the simulations are too noisy for a detailed comparison, the averaged results in Figure 4.18 indicate deviations of approximately 8% for small scales at the lowest redshift \( z \approx 0.369 \) which are not detectable at the highest redshift \( z \approx 0.863 \). As for the variance, the results on the correlation matrix are in line with our findings from the analysis of stochastic fields in the previous section. While the correlations introduced through the mask are present at all scales, the correlations from the non-Gaussianity induced by non-linear gravitational clustering only affect small scales. Instrumental noise will also mostly affect small scales, suppressing the contributions from non-Gaussianity for realistic surveys.

Effect of redshift space distortions Using the intensity maps from four realisations of the N-body simulation, we furthermore studied if the RSDs affect the agreement between the Gaussian estimate and the estimate from the independent realisations for variance and covariance of the HI angular power spectra. Figure 4.19 shows that, within the noise of the estimators,
the variance follows the Gaussian prediction from PolSpice for both the simulations with and without RSDs. We verified that the same is true for the correlation matrix at high redshifts or large scales. Figure 4.20 shows the results for the averaged correlations \( \langle \text{Corr}(\hat{C}_\ell, \hat{C}_{\ell+\delta\ell}) \rangle \) between small scales \((600 \leq \ell < 700)\). As for the simulations without RSDs, we find an excess of correlations between neighbouring \( \ell \) scales for high \( \ell \) values \((600 \leq \ell < 700)\). Figure 4.20 however shows that, again within the noise of the estimate, the magnitude of the excess is broadly consistent with the estimate from the maps without RSDs.

### 4.6 The Bleien Galactic Survey

As already mentioned in section 4.1, one of the key observational challenges in HI intensity mapping is the weakness of the signal: Not only is it several orders of magnitude fainter than the galactic foregrounds, but the expected average brightness temperature fluctuations of order 100 mK also put challenging requirements on the stability of radio telescopes.

The Bleien Galactic Survey (BGS) is a survey in the frequency range 990 MHz–1260 MHz with a receiver design comparable to concepts for HI intensity mapping surveys, but on a small (dish-size of 7 m) radio telescope in Bleien, Switzerland. The motivation of BGS is to learn about the instrument and data analysis challenges that will also arise in future HI intensity mapping surveys. The main scientific product will be maps of the foregrounds to the HI intensity mapping signal in the observed frequency range with a resolution of about 1 MHz. In the following, we will briefly describe the instrument, the simulation and data analysis pipelines, and the expected data product of BGS.

This section is a summary of a series of papers about BGS that are currently in preparation (Monstein et al., 2016; Akeret et al., 2016; Chang et al., 2016).

#### Instrument description

We collect our data with a 7 m-class radio telescope in Bleien, Switzerland (geographic latitude 47°20’23” north, longitude 8°6’42” east, and altitude 469 m). The site is a radio protected area within a radius of 1.5 km, but the data collection is nevertheless significantly affected by radio-frequency interference (RFI) from mobile phone transmitters, airplane communication, and other sources. The telescope is equipped with a dual-polarisation, corrugated horn antenna that under-illuminates the dish by a factor of 0.5 in order to reduce side-lobes and spill-over from the ground.

We use a pseudo correlation receiver design (Mennella et al., 2003) similar to the one proposed for the BINGO (BAO from Integrated Neutral Gas Observations) survey (Battye et al., 2013). The main difference is that the signal from the sky is correlated with an electronic reference source rather than a separate antenna observing a reference position on the sky. In addition, the phase of the pseudo correlation receiver is switched after each observation, such that
the sky signal is given by the difference between two consecutive observations. Overall, the receiver is designed to minimise correlated noise in the data stream which would require frequent recalibration. The receiver system is stable to at least tens of hours (Monstein et al., 2016).

The receiver is coupled to a fast Fourier transform (FFT) spectrometer with a frequency resolution of 48.8 kHz. The data is integrated for three seconds for each phase, resulting in a total integration time of six seconds per spectrum. The overall noise temperature of the system is measured at approximately 80 K (Monstein et al., 2016). As of right now, only one of the two linear polarisations of the instrument is recorded such that the telescope effectively acts as a bolometer and is not sensitive to polarisation.

**Survey description**

In a further effort to increase the stability of the system over time, we use a drift scan observation strategy similar to what is proposed for BINGO (Battye et al., 2013). In a drift scan, the telescope is pointing to a fixed (azimuth, elevation) position. As time goes by, the astronomical sky slowly drifts through the beam of the telescope. Within 24 h, we hence complete a full 360° scan of the equatorial declination corresponding to the date and the horizontal (azimuth, elevation) coordinate of the pointing. To fill the largest possible region of the sky, the telescope elevation is moved everyday at 24:00 universal time. The survey strategy in equatorial coordinates is therefore given by consecutive 360° azimuth stripes at different declinations.

In addition to the drift scan on survey days, we collect additional calibration data at the beginning and end of the survey as well as every 8 days during the survey. The calibration data consists of transit observations of known calibration sources such as Cassiopeia A, Cygnus A, and Taurus A.

In the first season of observations, we collected data in one polarisation over approximately four months from December 17th, 2015 to April 15th, 2016. Each day, the telescope generates 14′000 spectra corresponding to 9′000 float numbers in double precision. The data volume per day is hence about 1 GB. Currently, the survey is replicated to study the stability of our results and increase the sensitivity of the observations.

**Hide and Seek**

To analyse the generated data, we decided to simultaneously develop two pipelines: The HI Data Emulator (HIDE) for simulating the data stream from the telescope and the Signal Extraction and Emission Kartographer (SEEK) for processing the data. The advantage of the simultaneous development is that each analysis step in SEEK can be cross-validated against HIDE simply by applying the analysis pipeline to simulated data.
Both HIDE and SEEK are implemented in Python and based on the workflow engine Ivy\(^9\). Ivy is a generic framework for chaining a collection of modules which are executed independently while all the interaction between the modules is collected in a container object (called context) that is passed along the chain. A workflow is specified via a configuration file that is a Python module containing a list of modules and parameters for those modules. Parallel execution of modules can be straightforwardly implemented through a map-reduce step: The parallelisable workload is first distributed among the available processes by splitting the contexts in the map step. After the execution of the workload, the contexts of the processes are merged into a single context in the reduce step that can then be passed further along the workflow.

In the following, we will describe the HIDE and SEEK configurations that we used for simulating and analysing BGS. HIDE and SEEK are however more general and the pipelines can be easily modified in order to model and analyse other instruments. The main data structures in HIDE and SEEK are \texttt{HEALPix} maps for the representation of the sky and time-ordered data (TOD), i.e. a sequence of spectra, for representing the data stream from the telescope.

The simulation pipeline HIDE for the BGS survey consists of the following modules:

- **Load beam profile**: We model the beam as an airy disk with a frequency dependent width inferred from sun transits.

- **Define survey strategy**: The survey strategy is generated from a schedule file that specifies time and coordinates of each pointing. The same schedule file was also used for operating the real telescope. The resulting pointings are stored in the same format as the pointing information we get from the telescope.

- **Load astronomical signal**: We use the Global Sky Model (GSM) (de Oliveira-Costa et al., 2008) as input signal for our simulations. The GSM is a compilation of available observations from 10 MHz to 100 GHz which is interpolated to predict the sky at unobserved frequencies. It is given as a \texttt{HEALPix} map of brightness temperature in units of Kelvin. To correctly simulate the calibration days, we need to furthermore modify the GSM such that the point sources used for calibration exhibit the observed spectrum.

- **Convolve signal with beam**: We perform a numerical convolution in real space by querying the beam response at the \texttt{HEALPix} pixels in the neighbourhood of the pointing. Beam response and signal are then integrated to find the flux in Kelvin.

- **Apply gain**: We transform the TOD from units of Kelvin into internal instrument units (ADU) by using a frequency dependent gain that corresponds to the observed values.

- **Add baseline**: We add a frequency dependent baseline to the TOD which is independent of time, but a function of elevation. This baseline is inferred from the real data and models contributions to the signal from ground pick-up and other instrumental effects that depend on the pointing.

\(^9\)http://github.com/cosmo-ethz/ivy
Chapter 4. Future Observations and HI Intensity Mapping

- **Add noise**: We add white noise (in ADU) to the data, where the standard deviation of the noise is directly inferred from the real TOD from the instrument.

- **Add RFI**: We finally add RFI to the TOD by randomly placing bursts such that the overall RFI contamination mimics the properties of the real data. This means that the fraction of masked pixels per frequencies is chosen such that it fits the fraction observed from real data. We furthermore model a variation in RFI contamination with time, since more RFI is observed during the day.

- **Write TOD**: The time-ordered data is written to disk in the same format as the real data to ensure that the simulated data can be directly processed by SEEK.

Using parallelisation and just-in-time compilation with HOPE (Akeret et al., 2015) for the most demanding tasks in the pipeline, we are able to simulate the full survey in about two hours. Short runtimes like this make it possible to perform a great amount of tests for the analysis pipeline.

The analysis pipeline SEEK for the BGS survey consists of the following modules:

- **Find files**: We traverse the file system to find all files that correspond to the pattern defined in the configuration file. We automatically split the files into survey data and calibration data.

- **Calculate gain**: We load the calibration data and estimate the gain of the telescope from the transits of known point sources. The resulting gain relates the internal units (ADU) to Kelvin.

- **Load data**: We load the TOD corresponding to the survey data from the file system.

- **Apply gain**: We apply the gain file to the TOD to transform the units from ADU to Kelvin.

- **Process coordinates**: We process the coordinate files containing the pointing information of the telescope in horizontal coordinates to attach an equatorial coordinate to each spectrum in the TOD.

- **Mask objects**: We mask known objects such as the sun and the moon whenever they are close to the pointing of the telescope.

- **Mask RFI**: We apply the SumThreshold algorithm (Offringa et al., 2010) to the TOD which detects RFI contamination through an iterative procedure of smoothing and thresholding of the data.

- **Remove baseline**: We remove the baseline of the telescope from the data by subtracting the median flux over the day.

- **Create map**: We create the map by taking the mean over all data points within each HEALPix pixel after another outlier-rejection step.
4.6. The Bleien Galactic Survey

Analysing the full dataset with HIDE in the outlined configuration takes less than one hour. A full loop of simulation and analysis of BGS data is therefore feasible within a few hours.

The radio sky at 990–1260 MHz

We are currently still in the process of analysing the BGS data. We plan to analyse frequencies from 990 MHz to 1260 MHz at a resolution of approximately 1 MHz, resulting in 276 maps of the astronomical sky at those frequencies. Our scanning strategy yields a sky coverage of approximately 70% corresponding to an area of 29'000 square degrees.

Due to the varying level of RFI contamination, however, the quality of the maps will be heterogeneous. From our simulations, we expect the sky coverage to vary from 0% to 70% with an average sky coverage of approximately 40% over all frequencies. Furthermore, the average number of observations per pixel in our simulations drops from approximately 31 without RFI masking to about 15. Efficiently cleaning the RFI without losing too much data is hence one of the key challenges of our observations at Bleien. While the RFI situation will be better for many of the proposed sites of HI intensity mapping surveys, RFI contamination will remain a serious issue due to the high brightness of many RFI sources. Sources far away from the beam can easily introduce RFI contamination which is well below the noise in the TOD, but larger than the targeted sensitivity in the map.

At an $n_{\text{side}}$ of 64, corresponding to a resolution of approximately $0.9^\circ$, we expect from our simulations that the maps have a statistical error of approximately 0.2 K per pixel on average (0.09 K for the best frequencies). The largest systematic error in our current analysis strategy is introduced by the baseline removal. Estimating the baseline from the median of the TOD over each day is effectively removing elevation dependent changes in the baseline from ground pick up and other day to day variations in the instrument. We however also subtract the median of each declination scan of the astronomical sky. This systematic over-subtraction yields a declination and frequency dependent bias of approximately 0.5 K which is significantly higher than our statistical error. An additional systematic offset of approximately 0.05 K is caused by undetected RFI in the maps.

We are currently working on improvements in the baseline removal which will hopefully lead to a systematic error that is smaller than the statistical contribution. We expect gain variations between calibration days to be the next leading contributor to the systematic error budget. As the sensitivity of future HI intensity mapping surveys is at least one order of magnitude smaller than the BGS errors, many more systematic uncertainties will need precise modelling in order to be able to extract cosmological information.
4.7 Discussion

Intensity mapping of the redshifted 21 cm emission line of neutral hydrogen (HI) is a promising technique for studying the large-scale distribution of matter in the universe and will potentially yield complementary information to more traditional galaxy redshift surveys. The majority of the HI content of the low redshift \( z \lesssim 1 \) universe is believed to reside in dense clumps inside galaxies. The low angular resolution of radio telescopes combined with high frequency resolution then means that 21 cm intensity mapping surveys will map an unresolved (in angular position) population of galaxies containing HI in fine redshift bins. In this chapter we have simulated wide field HI intensity maps and used them to study the statistical properties of the expected brightness temperature signal, in particular the angular power spectrum \( C_\ell \) and its covariance. We have also described observations of the Bleien Galactic Survey (BGS) which intends to create maps of the foregrounds to HI intensity mapping in the frequency range 990 MHz to 1260 MHz.

We have simulated HI intensity maps at redshifts \( 0.35 \lesssim z \lesssim 0.9 \), based on a suite of 10 N-body simulations of a 2.6 h\(^{-1}\) Gpc box (Wechsler et al., 2016) and a mass resolution of \( 1.6 \times 10^{11} \) h\(^{-1}\) \( M_\odot \). We created HEALPix (Górski et al., 2005) maps of the dark matter density field from the N-body simulations in equally spaced bins in frequency. Since our simulations do not resolve all the low mass halos that are expected to host HI, we have used the coarse density field in each pixel to statistically sample these halos using a halo mass function conditioned on the overdensity of the pixel following Sheth and Tormen (2002). The low angular resolution of the intensity map means that we do not need to assign locations to individual halos within each pixel. Instead, each of these halos is assigned an amount of HI determined by its mass following a phenomenological prescription based on Padmanabhan et al. (2016). The resulting distribution of HI has then been used to generate 12 maps of redshifted 21 cm brightness temperature with an angular resolution of 7\( \prime \) between \( z \approx 0.35 \) and \( z \approx 0.89 \) and redshift bin widths from \( \Delta z \approx 0.033 \) at the lowest to \( \Delta z \approx 0.061 \) at the highest redshift.

The main assumptions of our prescription are that the halo distribution within each pixel is determined by its overdensity and well described by the conditional mass function (Sheth and Tormen, 2002), the linearised overdensity in each pixel is well approximated by the value derived from spherical evolution (Bernardeau, 1994), and that the HI content of the halos is determined by the mass of the halo (Padmanabhan et al., 2016). As a (non-trivial) validity check for the halo subsampling, we confirmed that our procedure reproduces the unconditional mass function and the large scale bias to better than 10%. With high-resolution simulations of smaller volumes, the accuracy of our procedure could be assessed in more detail and we leave such a study to future work. Assessing the validity of the HI assignment is more complicated. The model from Padmanabhan et al. (2016) is able to reproduce data of sky-averaged values such as the HI density \( \Omega_{HI} \) or the bias \( b_{HI} \). Due to a lack of data at these redshifts, a more detailed study of the relation between halo and HI on an object by object basis would have to resort to hydrodynamic simulations (see e.g. Villaescusa-Navarro et al., 2016). We furthermore
showed a way for including redshift space distortions in our simulations and verified that they do not affect the main conclusions drawn from the analysis of the undistorted maps in the following.

Our brightness temperature maps of the cosmic signal are sufficiently realistic to study the impact of non-linear effects on the large-scale clustering properties of HI intensity maps. In our analysis of the maps, we have focused on the angular clustering of the brightness temperature fluctuations as measured by the angular power spectrum. We have used two estimators for the power spectrum: the pseudo $C_\ell$ approach (Peebles, 1973) and the publicly available PolSpice package (Szapudi et al., 2001a; Chon et al., 2004). We have verified that the estimated angular power spectra of the simulated intensity maps agree well with predictions from linear theory and the halo model on large scales. We have found that the bias of HI relative to dark matter, as estimated from the ratio of angular power spectra, is close to unity and mildly scale dependent. For the lowest redshift of our simulations ($z \approx 0.366$), the bias falls below its large scale value by $\sim 10\%$ at $\ell \approx 200$.

Using the multiple N-body realisations, we have estimated the covariance of both $C_\ell$ estimators and compared the results to estimators from jackknife resampling and analytic predictions based on Gaussian statistics. Treating the covariance from multiple simulations as a noisy estimate of the true covariance, we have found good agreement with the PolSpice covariance estimate based on Gaussian statistics. This shows that even for our simple survey geometry (a contiguous quarter of the sky), most off-diagonal correlations are introduced by the mask. Only for small scales and low redshifts, the results indicate an excess of $\sim 8\%$ off-diagonal non-Gaussian correlations that are not captured by the PolSpice estimate. The jackknife estimator overpredicted the error on the angular power spectrum by approximately $\sim 40\%$, independently of scale, redshift and $C_\ell$ estimator. It was however able to trace the extra correlations introduced by the non-Gaussian nature of the brightness temperature fluctuations. The pseudo $C_\ell$ estimator for the covariance has no off-diagonal correlations by construction and overestimated the variance by a factor of two. Looking at the variance of pseudo $C_\ell$ bandpowers, i.e. binned $C_\ell$ values that are approximately uncorrelated, the lack of correlations however balanced the excess in variance and the estimate converged to the results from the multiple realisations. It is worth noting that the HI prescription from Padmanabhan et al. (2016) leads to a distribution of brightness temperature fluctuations that is less skewed to high contrasts than the matter density field and hence also more Gaussian. The lack of HI in high overdensities is supported by physical (Bagla et al., 2010) and observational (Haynes et al., 2011; Papastergis et al., 2013) arguments, but changes in the HI prescription could nevertheless influence some of the conclusions on the covariance of the angular power spectrum drawn from our maps.

Our analysis showed that our approach to simulating HI intensity maps can be used to simulate and study upcoming surveys. It also allows for the development of data analysis techniques that improve the extraction of cosmological information and the separation of the cosmological signal and contaminating components such as astrophysical foregrounds or human-made radio frequency interference (RFI). As galaxy surveys cover greater parts of
the sky to increasing depth, cross-correlations between intensity mapping and galaxy surveys are of particular interest as they are less sensitive to systematic effects. In future work, such simulations can thus be used to additionally model the connection between halos and galaxies in order to study the cross-correlation signal between HI and galaxy surveys.

Apart from the described theoretical challenges in modelling and interpreting the HI intensity mapping signal, there are many observational problems associated with this new technique. Our observations for the BGS are a first step towards understanding some of the issues that arise in single-dish intensity mapping. We have shown that the combination of a pseudo-correlation receiver, an under-illuminated dish, and a drift scan survey strategy as proposed for surveys such as BINGO (Battye et al., 2013) indeed lead to observations that are stable over hours. Elevation dependent fluctuations in the baseline, however, lead to systematic effects in the final maps that are hard to calibrate and need to be modelled. Furthermore, the data loss due to RFI contamination as well as the detection of RFI close to the noise level in the TOD are challenging problems in particular at our site in Switzerland.

The data analysis is not yet completed, but the main data products will be 276 maps of the astronomical sky at frequencies from 990 MHz to 1260 MHz at a resolution of approximately 1 MHz. Using two simultaneously developed packages for the simulation (HIDE) and the analysis (SEEK) of the data stream from the telescope, we can simulate and analyse the full survey over approximately four months within hours. This allows us to conduct extensive tests of the performance of the analysis pipeline. From the analysis of the simulated data, we expect a sky coverage up to 70% with an average statistical error per channel of approximately 0.2 K at a HEALPix resolution of $n_{\text{side}} = 64$ corresponding to an angular resolution of approximately 0.9°. The systematic error in the current analysis, however, is at a level of approximately 1 K and dominates over the statistical contribution.

A future upgrade of the system at Bleien to two identical receiver chains for both linear polarisations would allow us to additionally distinguish polarised from unpolarised components of the foregrounds. Even though the HI intensity mapping signal is expected to be unpolarised, a thorough understanding of the polarised foregrounds is necessary in the future since instrumental effects allow the polarised component to leak into the unpolarised signal. As it is well known that the polarised emission from the galaxy is less smooth in frequency than the unpolarised emission, polarisation leakage is a major issue for future HI intensity mapping surveys.

To conclude, we have seen that there are many obstacles for HI intensity mapping to fulfil its potential. The situation for observations at radio frequencies on earth is becoming increasingly difficult due to more and more contamination from man-made emitters, but many of those technological problems could be mitigated by remote sites or even a space mission in the future. On the theoretical side, future work needs to focus on the understanding of the relation between matter and neutral hydrogen on large-scales. While there is hope that the complicated physical processes on small scales can be coarse-grained to a robust effective
model on cosmological scales—similar to our approach for the simulated intensity maps—
those models have to be tested against more rigorous data. HI intensity mapping will hence
generate exciting problems in both observational and theoretical aspects as more instruments
come online in the next years. Once these challenges are overcome, adding the cosmological
information in the yet unobserved volumes probed by HI intensity mapping to the probes
discussed in chapter 3 will provide a powerful test for the cosmological concordance model
\( \Lambda \)CDM.
The current standard model of cosmology (ΛCDM) describes the present universe as dominated by a dark matter component and a cosmological constant. While its phenomenological nature and loose connection to microscopic physics leaves hope for exciting new physics to be discovered through cosmological observations, ΛCDM remains to be in good agreement with a large range of individual cosmological probes. To further improve our understanding of the evolution of our universe, we can aim at improving the current observations, developing new observational techniques, or advancing our ability to combine and compare the constraints from the currently available probes. In this thesis, we focus on the latter two aspects.

To this end, we developed a framework for comparing datasets based on the relative entropy between two posterior distributions. Within a given model, the relative entropy quantifies the information that is gained by adding complementary data to existing constraints or replacing old data with more precise measurements. We use the information gain in two ways: First, it allows us to rank different datasets by their contribution to our knowledge on the parameters of ΛCDM. Second, we use it as a measure of difference between multivariate posterior distributions. Given such a measure of difference—and a model which correctly describes both probes—we can use the constraints inferred from one dataset to predict the amount of variation in the constraints that we expect from a comparison with another dataset. We consequently proposed the difference between the observed relative entropy and the expected relative entropy as a measure of agreement between datasets and called it the Surprise. A large Surprise indicates that the deviations between the constraints is larger than expected, implying that the model is not able to describe both observations within the noise of the measurement. It could hence be triggered either by shortcomings in the model or systematics in the data.

We applied the information gain and the Surprise to a wide range of cosmological probes. Starting from a historical sequence of CMB experiments, we focussed on the comparison of observations from the WMAP and Planck satellite missions, the two most powerful cosmologi-
Chapter 5. Conclusions

cal datasets currently available. We further complemented the constraints from WMAP and Planck with data from distance ladder, supernova type Ia, BAO, and weak lensing observations. Among all comparisons, the 3-year WMAP constraints (Spergel et al., 2007), the 2013 Planck constraints (Planck Collaboration, 2014b), and the 2013 analysis of CFHTLenS weak lensing data (Kilbinger et al., 2013) generated large Surprise values, indicating tensions in data or model. Re-analyses of those datasets (Hinshaw et al., 2013; Planck Collaboration, 2015b; Köhlinger et al., 2016), however, alleviated most of the tensions, thereby showing that the Surprise presumably detected systematics in the data rather than deviations from $\Lambda$CDM.

Even though the Surprise proved to be a very sensitive measure of tensions, the current observations considered in this thesis remain in agreement with $\Lambda$CDM according to our measure. Our analysis also showed that CMB observations dominate our current knowledge on the parameters of $\Lambda$CDM, achieving information gains which are an order of magnitude larger than the values from any other probe. Future large-scale structure observations have the potential to close this gap and will hence put $\Lambda$CDM to a much more stringent test.

Particularly promising for our focus on information gains within $\Lambda$CDM is the technique of HI intensity mapping. It is unique in its capability to map the large-scale structure of neutral hydrogen (HI) in volumes between us and the CMB which are currently unobserved. As such, it can measure the perturbations to the isotropic and homogeneous background in larger volumes than any other cosmological probe. Yet, extracting the cosmological information from the three-dimensional, spherical volumes probed by HI intensity maps is challenging. With an expected brightness temperature of approximately 100 $\mu$K, the signal is very faint and sub-dominant to foregrounds at the same wavelength by several orders of magnitude. A thorough understanding of the expected HI signal as well as its foregrounds is hence crucial for the success of this technique.

As a step towards a better understanding of the HI signal, we proposed a scheme for simulating the large-scale structure of HI. It is based on a combination of dark matter fields from N-body simulations and a phenomenological prescription for assigning HI to sub-resolution dark matter halos that is consistent with current observations. We showed that our scheme is sufficiently accurate to study the impact of non-linear effects on the large-scale clustering properties of HI intensity maps. We further reported progress of ongoing efforts in the Bleien Galactic Survey (BGS) which aims at mapping the sky at frequencies between 990 MHz to 1260 MHz with a 7 m radio telescope in Bleien, Switzerland. BGS is designed to test strategies for single-dish drift scan surveys such as BINGO (Battye et al., 2013) while improving our understanding of the foregrounds to HI intensity mapping at these frequencies by producing maps that will cover 70% of the sky at a frequency resolution of 1 MHz. Preliminary analyses showed that even though our system is stable, controlling the systematics due to radio frequency interference and pick up from other terrestrial emitters is challenging already at a sensitivity of about 100 mK.

Irrespective of the observational technique, cosmology will face exciting developments while
processing, analysing, and interpreting the massive amounts of data that will be generated by the next generation of cosmological surveys. Aiming for a better understanding of our universe, this thesis highlighted that analysing the consistency between the individual probes can provide a powerful test for $\Lambda$CDM. Although we showed that there are many challenges for future large-scale structure surveys such as HI intensity mapping, we think that it would be instructive to look ahead: the full potential of future surveys might be unlocked by maximising their joint prospects for testing $\Lambda$CDM.
Bibliography


Bibliography


Bibliography


List of Symbols

\( a \) \hspace{1cm} \text{Scale factor.}
\( A_s \) \hspace{1cm} \text{Amplitude of the primordial power spectrum.}

\( b_{\text{HI}} \) \hspace{1cm} \text{HI bias.}

\( C_\ell \) \hspace{1cm} \text{Angular power spectrum.}
\( \chi(a) \) \hspace{1cm} \text{Co-moving distance to scale factor } a.

\( \mathcal{D} \) \hspace{1cm} \text{Data vector.}
\( D(p_2||p_1) \) \hspace{1cm} \text{Relative entropy of probability densities } p_2 \text{ and } p_1.
\( \langle D \rangle \) \hspace{1cm} \text{Expected value of the relative entropy } D(p_2||p_1) \text{ as predicted from the posterior predictive of } \mathcal{D}_1.

\( h \) \hspace{1cm} \text{Dimensionless Hubble constant.}
\( H_0 \) \hspace{1cm} \text{Hubble constant, i.e. } H(a) \text{ evaluated at } a = 1.

\( m_{\text{HI}}(m) \) \hspace{1cm} \text{Relation between HI mass and halo mass.}

\( N(m|M) \) \hspace{1cm} \text{Differential number of halos in the mass range } (m, m + dm) \text{ within a larger volume containing mass } M \text{ (conditional mass function).}

\( n(m,z) \) \hspace{1cm} \text{Differential number density of halos in the mass range } (m, m + dm) \text{ at redshift } z \text{ (halo mass function).}

\( \mathcal{N}(\Theta; \bar{\Theta}, \Sigma) \) \hspace{1cm} \text{Multivariate Gaussian distribution in } \Theta \text{ with mean } \bar{\Theta} \text{ and covariance } \Sigma.$
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_s$</td>
<td>Spectral index of the primordial power spectrum.</td>
</tr>
<tr>
<td>$\Omega$</td>
<td>Density parameter $\rho/\rho_c$.</td>
</tr>
<tr>
<td>$\Omega_X$</td>
<td>Density parameter of species $X$ today.</td>
</tr>
<tr>
<td>$p(\mathcal{D})$</td>
<td>Evidence for the data $\mathcal{D}$.</td>
</tr>
<tr>
<td>$p(\mathcal{D}_2</td>
<td>\mathcal{D}_1)$</td>
</tr>
<tr>
<td>$p(\mathcal{D}</td>
<td>\Theta)$</td>
</tr>
<tr>
<td>$p(\Theta)$</td>
<td>Prior distribution for the parameters $\Theta$.</td>
</tr>
<tr>
<td>$p(\Theta</td>
<td>\mathcal{D})$</td>
</tr>
<tr>
<td>$p_X(\Theta)$</td>
<td>Shorthand notation for the posterior of data $\mathcal{D}_X$: $p_X(\Theta) \equiv p(\Theta</td>
</tr>
<tr>
<td>$\rho$</td>
<td>Energy density.</td>
</tr>
<tr>
<td>$S$</td>
<td>Surprise component of the relative entropy.</td>
</tr>
<tr>
<td>$\Theta$</td>
<td>Multivariate parameter vector.</td>
</tr>
<tr>
<td>$\bar{\Theta}$</td>
<td>Mean of parameters $\Theta$ under some distribution.</td>
</tr>
<tr>
<td>${\Theta_t}_{t=0,\ldots,N}$</td>
<td>Sample from a probability density of size $N + 1$.</td>
</tr>
<tr>
<td>$w$</td>
<td>Equation of state parameter.</td>
</tr>
<tr>
<td>$z$</td>
<td>Redshift.</td>
</tr>
</tbody>
</table>
Index

$p$-value, 48
21 cm line, see Neutral hydrogen

Angular power spectrum, 14, 51, 70, 83 estimation, 70
Autocorrelation function, 47
Autocorrelation time, 47

Baryon acoustic oscillations, 63
Bayes’ theorem, 20
Bleien Galactic Survey, 101
Boltzmann equation, 7
perturbed, 10
BOOMERanG, 51

Co-moving distance, 5
Co-moving horizon, 5
Conformal time, 5
Cosmic Microwave Background, 12, 14, 50
Cosmic variance, 51
CosmoHammer, 45
Cosmological constant, 4
Cosmological probes, 14
Covariance, 21
of the angular power spectrum, 86

Density parameter, 4
Distance ladder, 62

Einstein tensor, 3
Einstein's equations, 3
perturbed, 6, 9
Einstein-Boltzmann equations, 9
equation of state, 4
Evidence, 20

Expected relative entropy, 27
for Gaussian distributions, 30, 32
Fisher matrix, 23
Friedmann equations, 3
Gaussian distribution
multivariate, 21
univariate, 21
Goodman-Weare algorithm, 44
Growth function, 10
Halo mass function, 69, 81
conditional mass function, 73
HI bias, 69, 85
HI intensity mapping, 67
HIDE, 103
Hubble parameter, 4
Inflation, 3
Information gain, 24

Kullback-Leibler divergence, see relative entropy
Likelihood, 20
Linear model, 21
Linear power spectrum, 69
Linearised overdensity, 73, 80
Markov chain Monte Carlo, 42
Mean, 21
Metropolis-Hastings algorithm, 42
Monte Carlo estimation, 47
N-body simulations, 72
Neutral hydrogen, 68, 75

Peculiar velocities, see Redshift space distortions
Planck, 53
PolSpice, 71
Posterior, 20
Posterior predictive, 27
for Gaussian distributions, 29
Power spectrum, 11
matter, 11
primordial, 11
Prior, 20
Pseudo $C_{\ell}$, 71
Python, 45, 48, 103
Redshift, 5
Redshift space distortions, 78, 86, 100
Reionisation, 13
Relative entropy, 19, 24
for Gaussian distributions, 26
SEEK, 104
Shot noise, 70
Sound horizon, 13
South Pole Telescope, 52
Stress-energy tensor, 3
Supernova type Ia, 15, 63
Surprise, 27
for Gaussian distributions, 31, 32, 37
Transfer function, 10
Weak lensing, 63
Wilkinson Microwave Anisotropy Probe, 52