UNIVERSITY OF THE WESTERN CAPE

MASTERS' THESIS

Point source simulations and Foreground cleaning techniques for HI intensity mapping

Author:
Sibonelo NGOBESE

Supervisor:
Prof. Mario SANTOS
Co-Supervisor:
Dr. Marta SPINELLI

UNIVERSITY of the
WESTERN CAPE

A thesis submitted in fulfillment of the requirements for the degree of Masters

in the

Astronomy group
Department of Physics and Astronomy

February 28, 2018

Declaration of Authorship

I, Sibonelo NGOBESE, declare that this thesis titled, "*Point source simulations and Foreground cleaning techniques for* **H**I *intensity mapping*" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

UNIVERSITY OF THE WESTERN CAPE

Abstract

FACULTY OF NATURAL SCIENCE
Department of Physics and Astronomy

Masters

Point source simulations and Foreground cleaning techniques for HI intensity mapping

by Sibonelo NGOBESE

We created a full sky point source catalogue at frequencies around 1 GHz, including the intensity, polarisation, rotation measure and spectral index of the point source. These values were matched to data where available and simulations were used otherwise. With this, we were able to generate full sky intensity and polarisation maps from point sources at the frequencies required for intensity mapping. These are crucial for end to end simulations that take into account the telescope beam effects. We showed that the polarisation intensity at these frequencies is reasonably smooth in frequency, making it easier to clean. We proceeded to generate full sky maps for the cosmological **H**I signal and other foregrounds, in particular galactic synchrotron (total and polarised), galactic free-free and extra-galactic free emission. We then analysed how well we should be able to extract the **H**I cosmological signal for a fiducial MeerKAT survey. The signal and foregrounds were combined with the expected MeerKAT noise and convolved with a Gaussian primary beam at all frequencies. The combination of these maps was then used as input in the foreground cleaning simulation taking into account the survey mask which corresponded to about 10% of the sky. The polarised foregrounds were included assuming a 1% leakage. The foreground cleaning techniques used in this analysis were the Principal Component Analysis (PCA), Independent Component Analysis (ICA) and Line Of Sight fitting (LOS). The foreground cleaning relies on a parameter $N_{\rm fg}$, the number of degrees of freedom which represents the modes to be removed. For a case consisting all foregrounds except for polarised synchrotron emission, PCA and ICA recover the **H**I signal with $N_{\rm fg} = 7$ while LOS fails especially at the frequency edges. Considering polarised galactic synchrotron emission as the foreground at 1 % leakage, we show that using $N_{\rm fg} < 60$ does not recover the **H**I signal. Values of $N_{\rm fg} > 83$ do remove the foregrounds but remove signal and noise at the same time. We conclude that polarisation leakage from point sources should not be a problem for these foreground cleaning methods but leakage from polarised synchrotron will be a problem if not calibrated at better than the 1% level. If this is taken care of, then MeerKAT should be able to recover the **H**I intensity mapping signal.

Acknowledgements

Thanks to my family for the ever lasting support and patience. My friends for the advice (Siyanda, Clemence, Siyambonga, Londiwe, Themba and Eliab), assistance and courage. Thanks to Mika and Olorato for the Python tricks and useful modules. To Dr. Marta Spinelli, thanks for all your help and the continuous analysis of my project and discussions. I am very grateful. Thanks to Ridhima Nunhokee for the discussions and input. Thanks to Dr. Jose Fonseca and the UWC Postdoc office for the Coding Cafe and discussions. Thanks to the National Astrophysics and Space Science Programme (NASSP) and the Square Kilometre Array for the financial support throughout the course of my studies. Thanks to the Centre for High Performance Computing (CHPC) for their resources and UKZN for allowing continuous use of "cyclops" all these machines made getting the work done possible. Lastly, to Professor Mario Santos, thank you for giving me an opportunity to work within the top cosmology group (Centre for Radio Cosmology) in South Africa. "*Ngiyabonga Kakhulu*".

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The Lambda Cold Dark Matter model

The standard model of cosmology postulates that the Universe was born in a violent explosion known as the Big Bang. The Universe was in a state of a hot dense plasma seconds after its birth. In the plasma state, there were frequent interactions between photons and baryonic matter. During this phase, there were a variety of elementary particles including the electrons, photons and Dark matter particle and other species of elementary particles. The mean-free path of these photons was very small since they scattered with electrons and were not able to travel a large distance before colliding again. This process is known as Thomson scattering. The evolution of the fluctuations during this phase of the Universe was governed by two opposing forces: gravity and radiation pressure. Gravity was trying to collapse the matter into clumps while radiation pressure was pushing outwards with the photons trying to escape. These left imprints in terms of oscillations which are called Baryon Acoustic Oscillations (BAOs, Eisenstein et al. (2004)). These plasma oscillations are similar to sound waves thus the term acoustic oscillations (traveling at $c_s \sim c/\sqrt{3}$, where $c = 3 \times 10^5$ km/s is the speed of light).

As the Universe expanded and cooled down, the electrons and protons finally combined to give rise to hydrogen atoms, which allowed the photons to travel freely. This then gave rise to the last scattering surface (Hu, 2003), commonly known as the Cosmic Microwave Background (CMB). The CMB is the Big Bang's afterglow which was accidentally detected by Penzias and Wilson (1965). They found a uniform hiss coming from everywhere in the cosmos while working at Bell Labs. The first experiment to detect the CMB fluctuations was the Cosmic Background Explorer (COBE, Smoot (1991, 1992, 1994)) which was a space-based instrument.

With the continuous technological advances over the years, the Wilkinson Microwave Anisotropy Probe (WMAP) satellite was able to deliver much higher resolution CMB images of the young Universe. With a resolution of 13 arcminutes, WMAP was able to resolve intensity fluctuations and polarisation of the CMB (Bennett, 2005). CMB photons from hot and cold spots are incident on an electron via Thomson scattering thus forming linear polarisation (known as $E-$modes). The other type of polarisation is the so-called $B-$modes, which can be a signature of tensor modes created from inflation and thus a "smoking gun" for this event. WMAP combined with supernovae data from Perlmutter et al. (1999) along with BAO data was able to place very strong constraints on cosmology. The CMB fluctuations are an early picture of the large-scale structure (LSS). The CMB allows measuring of cosmological parameters, which are a set of parametric numbers

arising from the ΛCDM model.

The ΛCDM model follows from the theory of general relativity using the solution of Einsteins equations: the Friedmann-Lamaitre-Robertson-Walker (FLRW) metric;

$$ds^2 = -c^2dt^2 + a^2(t)\left[\frac{dR^2}{1-kR^2} + R^2d\Omega^2\right], \tag{1.1}$$

where $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$, $a(t)$ is the scale factor and k gives the geometry of the Universe. The scale factor gives the expansion of the Universe over time t. (R, θ, ϕ) are the spherical coordinates and k = -1, 0, 1 corresponds to a open, flat and closed Universe respectively. For a negatively curved Universe (k=-1) shooting two light beams parallel to each other bend in their path and the light beams diverge in their direction towards infinity. They never cross paths. In the case of a flat Universe (k=0), the light beams remain parallel to each other and never cross. For k= 1 the light beams start off parallel but reach a point where they converge and cross paths. Observations from sub-orbital telescopes (Balloons) found that the Universe has a flat geometry to a good approximation (de Bernardis et al., 2000; Balbi et al., 2002). These experiments showed this by observing the CMB fluctuations on small sky patches and then measured fluctuations on 1° scales. The latest full sky survey providing high precision cosmology was the Planck satellite (Planck Collaboration et al., 2014a,b). Its angular power spectrum is shown in Figure 1.1. The first peak is at multipole moment, $\ell = 200$ which corresponds to 1° in angular scales (as shown in Figure 1.1). The y-axis for this plot is given by $\mathcal{D}_\ell = \ell(\ell+1)C_\ell/2\pi$, where C_ℓ is the angular power spectrum. The red points with error bars are Planck measurements with the solid cyan being the fiducial model. As explained, the first (and highest) peak, tells us about the geometry of the Universe. At about 1°, we know the Universe is flat (e.g. $\Omega_{\rm tot}$=1 - see later). If the peak was at higher ℓ, it would mean the Universe had negative curvature and at lower ℓ (larger than 1°) it would mean positive curvature (closed). However, this interpretation will be confused by the effect of dark energy also. The acoustic peaks tell us about the photon-baryon fluid around the epoch of decoupling (Lahav and Liddle, 2014).

The FLRW cosmological model does well in explaining the expansion of the Universe from a hot Big Bang, to lower redshifts, tied in well with abundant observational support and predictions (Ellis and van Elst, 1999). Assuming a Copernican Universe, every galaxy we see today are receding from us with a velocity H_0D where $H_0 = 100h$ km/s/Mpc is the Hubble constant (where $h = 0.67$ for the latest constraints, Planck Collaboration et al. (2014b)) and D being the distance (Hubble, 1929). Noting that the proper distance to each galaxy can be given by $r(t) = a(t)r_0$ (with r_0 being time independent), the receding velocity is:

$$v(t) = \dot{a}(t)r_0 = \frac{\dot{a}(t)}{a(t)}r(t),$$

where the term $\dot{a}(t)/a(t)$ is the Hubble rate (denoted by $H(t)$). The redshift in terms of the scale factor is given by $1 + z = \frac{a(t_0)}{a(t)}$, where $a(t_0) = 1$ being the scale factor today. This can be written as $1 + z = \frac{\lambda_0}{\lambda_e}$, where λ_0 is the rest wavelength and λ_e is the wavelength for the emitter. This shows the fact that wavelength is stretched by the expansion of space. Using General Relativity combined with assumptions of homogeneity and isotropy (the **cosmological principle**), we can derive a set of

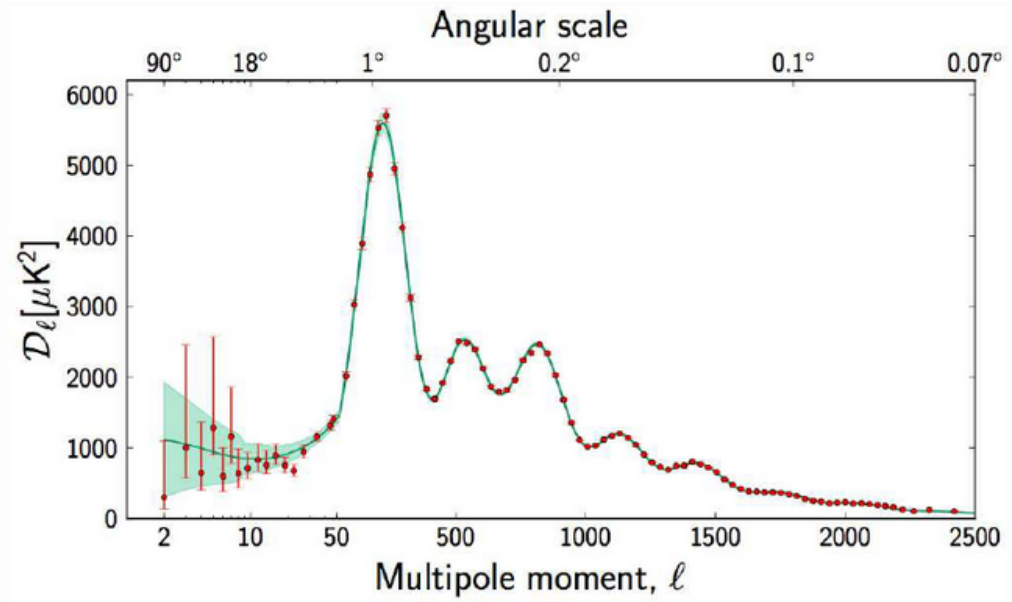

FIGURE 1.1: The CMB angular power spectrum obtained by the Planck observations. The x-axis changes from multipole moment equal to 50 from linear to logarithmic scale. image credit: ESA and the Planck Collaboration (2013)

equations for the expansion of the Universe (Friedmann equations):

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{kc^2}{a^2} - \frac{\Lambda c^2}{3} = \frac{8\pi G\rho}{3}, \tag{1.2}$$

$$2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{kc^2}{a^2} - \Lambda c^2 = -\frac{8\pi Gp}{3}. \tag{1.3}$$

where k is the curvature constant, $G = 6.6408 \times 10^{-11}\ \mathrm{m^3\ kg^{-1}\ s^{-2}}$ the Newtons gravitational constant and Λ is the Einstein's cosmological constant. Simplifying gives:

$$\dot{\rho} = -3\frac{\dot{a}}{a}\left(\rho + \frac{p}{c^2}\right), \tag{1.4}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3}. \tag{1.5}$$

These equations explain the global expansion of the Universe. The first equation can be obtained from the First Law of Thermodynamics, assuming that the expansion of the Universe is an adiabatic process. The second equation explains that the expansion of the Universe is dependent on the energy density and pressure (where $p = p(\rho)$), the cosmological constant is responsible for the Universe accelerated expansion. For a perfect fluid, the equation of state,

$$w = \frac{p}{\rho}, \tag{1.6}$$

where w=1/3, 0 and -1 for radiation, dark matter and a cosmological constant respectively. With a cosmological constant, the term in parenthesis in the equation above will be negative and so $\ddot{a} > 0$ and the Universe will have an accelerated expansion. The true value will depend on the energy content of the Universe. In

order to quantify this, we use the critical density parameters, $\Omega = \rho/\rho_c$, where ρ is the density of the quantity in consideration and ρ_c the critical density of the Universe, $\rho_c = 3H_0^2/8\pi G$. Writing in terms of the density parameters at redshift zero, we get,

$$\frac{H(z)}{H_0} = \sqrt{\Omega_m(1+z)^3 + \Omega_r(1+z)^4 + \Omega_k(1+z)^2 + \Omega_\Lambda}, \tag{1.7}$$

where Ω_m, Ω_r, Ω_k and Ω_Λ represent dark matter plus baryons, radiation, curvature and a cosmological constant, respectively. We can express equation 1.7 as $H(z) = H_0E(z)$ where;

$$E(z) = \sqrt{\Omega_m(1+z)^3 + \Omega_r(1+z)^4 + \Omega_k(1+z)^2 + \Omega_\Lambda}. \tag{1.8}$$

The curvature density $\Omega_k \sim 0$ according to Planck Collaboration et al. (2015) with $w = -1$. In terms of the energy budget in the Universe, there is a lot of missing matter that we still cannot account for. There's a mysterious undetectable component called dark energy and it is a sole contributor to accelerating the expansion of the Universe. The properties of dark energy are poorly understood. Luminous matter in the Universe consists of only 4 % of the matter out there, about 23 % of the non-observable matter is dark matter while dark energy holds 73 % (see Planck Collaboration et al. (2015)). The ΛCDM is the current concordance model. In mathematical terminology, when BAO, CMB data and the assumption that dark energy is the cosmological constant are used to constrain cosmology, *i.e.* $\Omega_{\rm tot} \equiv \sum \Omega_i + \Omega_\Lambda$ where $i = m,\ r,\ k$ yields $\Omega_{\rm tot}$= 1 as mentioned.

The effects of dark matter can only be detected due to its gravitational interaction by measuring the rotation curve of galaxies. Measurements made by Zwicky (1933) showed that there is a lot of mass in galaxies that cannot be accounted for. There are candidates to explain what dark matter could be. Astrophysicists have theorized that dark matter could be a particle that rarely interacts. The Weakly Interacting Massive Particles (WIMPs) (Freese et al., 1988; Ferrer et al., 2006; Fornengo, 2008) was predicted to make up dark matter. An example of the WIMP is the neutrino, an elementary particle that interacts very weakly with standard matter. Massive Compact Halo Objects (MACHOs, Griest (1993)) are other candidates that could be dark matter. Brown dwarfs are failed stars that never made it to the Main sequence and these fall under the MACHOs category. Primordial black holes that formed during the Big Bang, white dwarfs and neutron stars also fall under MACHOs. WIMPs are believed to be non-baryonic while MACHOs are formed from the baryonic material.

Combining CMB data and BAO data along with supernovae data and Hubble constant prior to parameter estimation gives the current known constraints on cosmology. This removes degeneracies in parameter estimation from one CMB dataset thanks to the addition of measurements of the angular diameter distance and the luminosity distance. The angular diameter distance is the ratio of the object's line of sight distance (*i.e.* distance between the observer and the object's location) to its angular size (given in radians). Luminosity distance is another way of calculating the distance to far away galaxies or star systems. This is defined as:

$$d_L^2 = \frac{L_\nu}{4\pi S_\nu}, \tag{1.9}$$

where S_ν is the object's measured the flux and L_ν is the object's luminosity. In terms of the FLRW metric, this formula becomes $d_L = (1+z)r(z)$. The comoving radial distance is:

$$r(z) = \begin{cases} \frac{D_H}{\sqrt{\Omega_k}} \sin\left(\sqrt{-\Omega_k}\frac{\chi(z)}{D_H}\right), & \Omega_k < 0 \\ \chi(z), & \Omega_k = 0 \\ \frac{D_H}{\sqrt{\Omega_k}} \sinh\left(\sqrt{-\Omega_k}\frac{\chi(z)}{D_H}\right), & \Omega_k > 0 \end{cases}$$

where $D_H = c/H_0$ is the cosmological horizon and the radial comoving distance for a flat Universe is given by;

$$\chi(z) = D_H \int_0^z \frac{dz'}{E(z')}, \tag{1.10}$$

where $E(z)$ comes from equation 1.8 with $\Omega_k = 0$. The angular diameter is given by $d_A = \chi(z)/(1+z)$. Supernovae can be used as distance measurement objects. This is done by observing what is classified as type Ia supernovae. When there's a binary system consisting of a white dwarf and companion star (usually a red giant), as the two objects orbit each other, the companion star acretes matter to the white dwarf making it reach a certain mass limit. The mass limit is called the Chandrasekhar mass limit of 1.4 $M_\odot$ and results in an explosion. These explosions can be brighter than the galaxy hosting the binary system. Comparing with the redshift of the host galaxy, this light can thus tell us how much the Universe has expanded and how fast. Just like the BAOs are called "standard rulers", these supernovae are "standard candles" used to measure how much the Universe has accelerated in its expansion (Riess et al., 1998; Perlmutter et al., 1999).

1.2 Radio Astronomy

Karl Jansky started the field of radio-astronomy while working at Bell Laboratories in the 1930s. Radio wavelengths are longer than optical wavelengths and can be received by radio telescopes during cloudy conditions of the sky on Earth. These wavelengths cannot be obscured by dust particles like their optical counterparts. They can traverse the intergalactic medium until they reach our radio telescopes. This implies that the radio telescope does not need to be perfectly shaped like an optical telescope in order to detect the radio emission. Radio astronomy grew enormously in the 1950s when the radio emission from our very own galaxy, the Milky Way was discovered. Radio astronomy started being recognized especially when the CMB was discovered at 3 Kelvin. Other groundbreaking discoveries include:

- Pulsars (Backer et al., 1982; D'Amico et al., 2001).
- Quasars (Porcas et al., 1981).
- Cosmic magnetic fields (Reich, 1990; Furst and Reich, 1990).
- Water Masers (Fiebig and Guesten, 1989; Impellizzeri et al., 2008).
- Gravitational lensing of star-light around super-massive black holes (Bozza and Mancini, 2008).
- The All-sky survey at 408 MHz (Haslam et al., 1982).
- Rotational curves serving as evidence of Dark matter (Carignan et al., 2006).

- Fast radio bursts (Spitler et al., 2014; Petroff, 2017).

Radio astronomy allows probing of gas (mostly diffuse emission) from astronomical objects where other telescopes have difficulties detecting the emission. Radio telescopes will serve as the next "era of precision cosmology" instruments following the steps of CMB. Instead of looking at the CMB redshift, these instruments will probe the intermediate redshifts. These intermediate redshifts provide information about LSS formation from redshift zero to the Epoch of Reionization (EoR) which tells us about the first stars in the Universe (Zaroubi, 2013).

One of the currently proposed ways to probe cosmology with radio telescopes is by observing neutral hydrogen throughout the Universe. Measuring neutral hydrogen (**H**I) faces some challenges since the signal is weak. The **H**I signal is detected by observing the 21 cm emission line using radio telescopes. The 21 cm emission line is emitted when a hydrogen atom undergoes a spin-flip, when an electron changes its spin from being parallel to anti-parallel relative to the proton. The transition results in an emission of a photon traveling in a completely random direction with a frequency equal to 1420 MHz. Recall that the Universe is no longer opaque so these photons are free to move plus they do not interact with dust clouds. Figure 1.2 presents an illustration of the spin-flip transition. Ewen and Purcell (1951) were the first radio astronomers to make the detection of the 21 cm line emission. The transition is highly forbidden. The mean lifetime of the excited state is about 11 million years, but since there is an abundant amount of hydrogen in the Universe we can detect these photons. This is where the idea of "intensity mapping" comes in handy since it will integrate the weak **H**I from each galaxy as a function of frequency. Over large scales, the signal will average over many galaxies and thus, there is no need to worry about individual galaxy properties.

The **H**I emission line can be used to probe flux densities $\sim$1 μJy and above. With the techniques widely used for CMB data analysis, it is hoped that those techniques will come in handy when handling **H**I observations and analysis. Intensity mapping does not require galaxy detection so the only issue that needs to be accounted for is foreground cleaning. However, the instruments must be well calibrated.

1.3 The Square Kilometre Array

The Square Kilometre Array (SKA) will be the largest radio telescope ever built with eventually a square kilometre array of collecting area. It is a radio interferometer, made of many individual elements which can then be correlated to produce observations of the sky. The largest baseline (distance between elements) of such telescope will set its angular resolution. The first observations for this instrument are estimated to be carried out in the 2020s. The SKA will address many different scientific goals:

- Probing the dark ages.
- Galaxy Evolution, Cosmology, and Dark Energy.
- Strong Field Tests of gravity.
- Cosmic Magnetism.

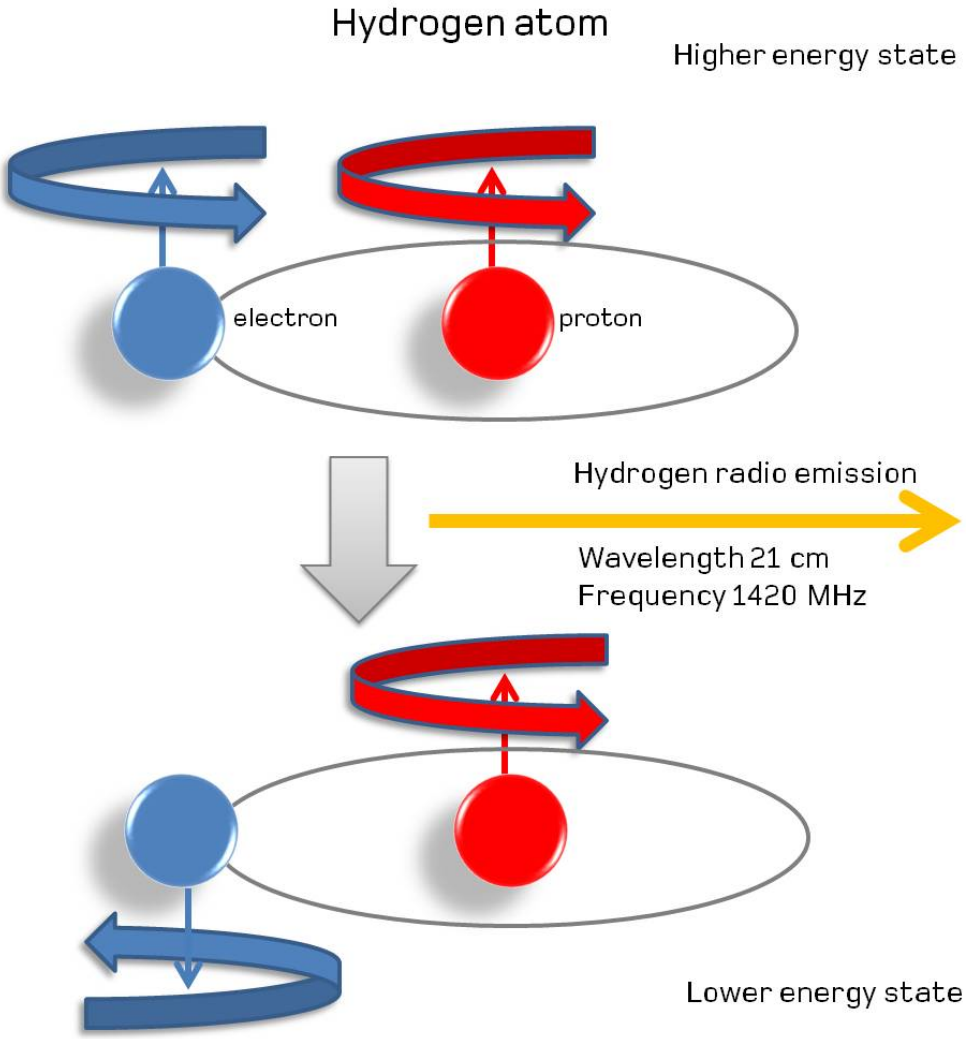

FIGURE 1.2: An artists rendition of what happens in the hydrogen atom. Image credit : SKA Organization (2017a)

- Cradle of Life.
- Exploration of the Unknown.

There are two phases planned for the SKA, SKA1 (Phase 1) and SKA2 (Phase 2). Figure 1.3 shows an artists impression of the radio dish array (SKA-MID) and aperture array (SKA-LOW). SKA-MID will be located in South Africa while SKA-LOW will be built in Australia.

1.3.1 Phase 1

SKA1 is in turn split into 2 instruments: SKA1-LOW, set in Australia and SKA1-MID, set in South Africa.

SKA1-LOW is the low frequency array, operating at frequencies 50-350 MHz and using an aperture array system. It is made of about 130,000 dipole antennas grouped into 476 stations. Its maximum baseline should be 40 km. One of its main objectives is to probe the Epoch of Reionization at high redshifts, when the first stars and galaxies in the Universe formed.

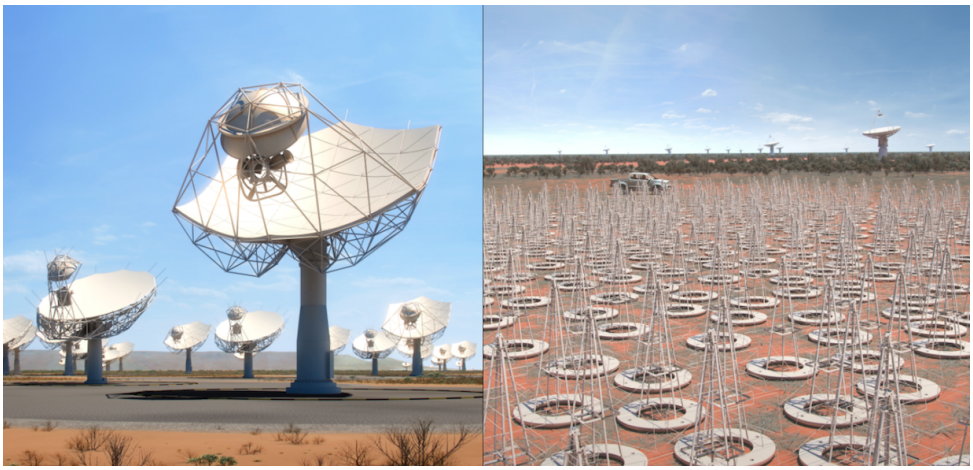

FIGURE 1.3: An artist rendition of the SKA array. *Left*: The SKA-MID instrument that will carry out mid-frequency observations. It will be located in South Africa. *Right*: The SKA-LOW instrument that will carry out low-frequency observations. It will be located in Australia. Image credit: SKA Organization (2017b)

SKA1-MID on the other hand is a an array of dishes operating at frequencies between 350 MHz and 14 GHz. It will be made of 130, 15 m diameter dishes plus 64 MeerKAT dishes (13.5 meters - see below). This telescope will be able to deliver several transformational surveys in cosmology (see the SKA chapters for a review, e.g. Maartens et al. (2015)).

1.3.2 Phase 2

SKA2, the second phase of the SKA projects (aiming to achieve the initial objectives of the SKA), is planned for later, in 2030. It should correspond to a factor of 10x in collecting area and a further increase in the field of view. Although the specifications are not completely defined, it is expected that SKA2 will be again split in two sites, low frequency for Australia and middle frequencies for South Africa. Although the instrument in Australia might follow the SKA1 technology, SKA-MID in South Africa might expand into something completely different made from dense aperture arrays working at a range of frequencies and providing large fields of views and high angular resolution. In terms of Cosmology, SKA2 would provide the ultimate survey. Detecting over one billion galaxies in **H**I over much of the visible sky. Allowing to get immediate spectroscopic redshifts for most of them. These would allow fundamental tests using baryon acoustic oscillations and redshift space distortions. Moreover, it would provide the ultimate weak lensing survey. With galaxy shapes and redshifts measured at the same time.

1.3.3 MeerKAT

MeerKAT, built in South Africa, is considered a SKA1-MID precursor. In fact, its 64 dishes will be absorbed into the SKA system. Each dish is 13.5 m diameter and will be fitted with several single pixel feeds: UHF band (580-1015 MHz), L-band (900-1670 MHz) and S-band (1750 - 3500 MHz). There is also a planned X-band receiver which, if deployed should work at 8 GHz - 14.5 GHz. At the moment, all the 64 dishes have been installed and first operations with all the dishes in L-band is

expected in June 2018.

There are several key science projects already approved for MeerKAT;

- Looking at the Distant Universe with the MeerKAT array (LADUMA, Heywood et al. (2011)).
- MeerKAT Absorption Line Survey (MALS, Gupta et al. (2017)).
- MeerKAT HI Observations of Nearby Galactic Objects: Observing Southern Emitters (MHONGOOSE, de Blok et al. (2017)).
- The Hunt for Dynamic and Explosive Radio Transients with MeerKAT (ThunderKAT).
- MeerKAT International GigaHertz Tiered Extragalactic Exploration survey (MIGHTEE, Jarvis et al. (2017)).
- A MeerKAT HI Survey of the Fornax Cluster.

There is a proposed survey for Cosmology that we will focus on in this thesis: MeerKLASS (MeerKAT Large Area Synoptic Survey, Santos et al. (2017)). It will carry out intensity mapping observations. MeerKLASS will scan an area 4 000 deg^2 in roughly 4 000 hours. It is aimed at the L-band and combines several large sky scientific objectives, although the intensity mapping technique itself can also be applied with success to the UHF band, at higher redshifts. The survey will rely on dish auto-correlation (single-dish mode) as well as cross-correlations (interferometer mode). It will be able to provide **H**I intensity maps (4 000 deg^2), continuum images (~ 5 μJy rms), spectral **H**I lines and polarisation.

1.4 Cosmology with the SKA

Although not the full power of SKA2, we can still achieve competitive and transformational cosmology with SKA1. In Phase 1, we have mentioned that intensity mapping and all-sky continuum surveys will be carried out. It will enhance our understanding of BAOs, dark energy and dark matter. In this section, we cover some of the theory behind on how the radio telescope can be used to study these phenomena.

With a survey of about 10 000 hours over 5000 deg^2 in band 2, SKA1-MID should be able to pick up around 10 million galaxies in **H**I at lower z which can be used for Cosmology (Santos et al., 2015a). Such survey using band 2 will also be optimal for a large Weak lensing survey. Although providing interesting competitive constraints for Cosmology, the signal to noise using **H**I galaxy detections will not be good enough to make transformational measurements. For that, we will need to rely on the **H**I intensity mapping method.

This is a new technique for mapping out the LSS in 3 dimensions by using **H**I emission as a biased tracer of the underlying dark matter density field. This technique does not require to resolve the individual galaxies but integrates their intensity falling within a specific pixel. Foreground subtraction is an important step in making sure we recover the **H**I signal and also polarisation leakage due to the instrument needs to be accounted for. This can be achieved by calibrating the

instrument.

Intensity mapping can be used to constrain Baryonic Acoustic Oscillations (BAOs), modified gravity, redshift space distortions and study spatial curvature and primordial non-Gaussianity (PNG). PNG imply that there existed some level of non-Gaussianity in the statistical distribution of primordial fluctuations predicted by inflationary models. PNG can be probed at ultra-large scales using the power spectrum, see e.g Camera et al. (2015) and references therein for a thorough description of the ultra-large scales mentioned here.

The use of BAOs in cosmology is because they are a statistical "standard ruler", they can be used place stringent constraints on the expansion rate, $H(z)$, and (angular) distance-redshift relation, $d_A(z)$, as functions of redshift, as has been done successfully with recent large galaxy redshift surveys such as BOSS (Dawson et al., 2013) and WiggleZ (Blake et al., 2012). In order to fully probe BAO scales, we need to measure scales with the telescope between 30 arcminutes to 4 degrees and frequency scales from 2 to 35 MHz (around z=1) (Santos et al., 2015c; Maartens et al., 2015).

Unfortunately, SKA1-MID will not have the right baseline distribution in order to probe cosmological scales (such as BAO) using the interferometer. We will then have to rely on the use of the auto-correlations from every single dish in the SKA. Each dish will make a map of the sky and the combination of all the dishes makes for some power detection in terms of signal to noise. It will also allow us to probe very large scales and test cosmology in this new limit (Santos et al., 2015b). The planned **H**I intensity mapping survey for SKA1-MID will cover a huge volume, with areas up to 30 000 deg^2 and redshifts up to z=4.

An example of a telescope that has done single-dish observations of the **H**I signal for intensity mapping is the Green Bank Telescope (GBT, Masui et al. (2010)). Interferometers have usually the advantage of being less sensitive to systematics but cannot probe such large scales (see e.g. CHIME, Bandura et al. (2014)). The intensity mapping survey planned for MeerKAT which we focus on this thesis will also rely on this approach of using single-dish mode. This comes with several calibration challenges that people are now in the process of addressing. With a planned 4 000 hours of observation of 4 000 deg^2 over the L-band, such a survey with MeerKAT has the potential to make the first detections of BAO and redshift space distortions (RSDs) with **H**I intensity mapping. It will also allow providing some of the first constraints on primordial non-Gaussianity using the multi-tracer technique. This will rely on covering the observed area with spectroscopic or photometric data (such as using the DES area). MeerKLASS will also contribute towards other science and in particular, it should be able to detect about 10 million galaxies in the continuum.

1.5 Objectives

One of the objectives of this thesis was to construct a point source catalogue at 1.4 GHz in both intensity and polarisation. The source catalogue was then extrapolated to brightness temperature maps (in intensity and polarisation) in order to use the point source maps as one of the foregrounds in the cleaning techniques. Simulated

HI maps, noise maps that mimic the MeerKLASS survey, foregrounds (intensity and polarisation) were then produced and combined. Several foreground cleaning techniques were then tested in order to quantify how much **HI** signal can be recovered.

Chapter 2 covers the theoretical background of the **HI** emission, foregrounds and theoretical background to power spectrum tools that we will need in analyzing the results from foreground cleaning. Chapter 3 described the construction of the point source catalogue, extrapolating the catalogue to brightness temperature maps and the power spectrum analysis of the point source maps. Chapter 4 goes through the results obtained from applying the foreground cleaning and we draw conclusions from our analysis in Chapter 5.

Chapter 2

HI intensity mapping and Foregrounds Simulations

In this chapter, we theoretically derive the brightness temperature of neutral hydrogen (**H**I) see section 2.1 and we cover the background of intensity mapping in section 2.2. The types of astrophysical contaminants termed "foregrounds" will be explained in section 2.3. The simulation of **H**I and foregrounds and type of survey being mimicked by the simulations are covered in section 2.4 and 2.5 respectively. Lastly, the derivation of the power spectrum is covered in section 2.6.

2.1 HI Brightness temperature

The derivation of the **H**I brightness temperature equation in this section is based on the paper Bull et al. (2015b). The 21 cm intensity can be worked out from the spin temperature $T_{\rm spin}$. The spin temperature describes the **H**I spin in a gas cloud. It is obtained from the Boltzmann equation;

$$\frac{n_1}{n_0} = \frac{g_1}{g_0} e^{\frac{-T_*}{T_{\rm spin}}} = 3e^{\frac{-T_*}{T_{\rm spin}}}, \tag{2.1}$$

where $g_1 = 3$, $g_0 = 1$, n_1 represents the upper state of hyperfine splitting and n_0 represents the lower state hyperfine splitting. The temperature $T_* = h_p\nu_{21}/k_{\rm B} = 0.0682$ K and $k_{\rm B}$ is the Boltzmann constant, ν_{21} is the frequency of emission at 21 cm and h_p is the Planck constant. The linear combination of the two states gives $n_{\rm HI}$ which is the comoving number density of **H**I emission, *i.e.* $n_{\rm HI} = n_1 + n_0$. Thus using $e^x \simeq 1 + x$ for $x \ll 1$, then equation 2.1 becomes $n_1 = 3n_0$ ($T_{\rm spin} \gg T_*$) which can be written as $n_1 = \frac{3}{4} n_{\rm HI}$. In the limit of a small optical depth, if we neglect absorption, we can consider each **H**I atom has an emitter with a luminosity:

$$L_{21} = h_p\nu_{21}A_{10}, \tag{2.2}$$

where $A_{10} = 2.85 \times 10^{-15}\ s^{-1}$ is the spontaneous emission coefficient. An observation with an angular size $d\Omega$ translates into comoving area at redshift z given by $dA = r^2(z)d\Omega$, where $r(z)$ is the comoving line of sight to redshift z and $d\Omega$ is the angular size of the clump. A frequency resolution, $d\nu_{\rm obs}$, translates into comoving distance,

$$dr = \frac{\lambda_{21}(1+z)^2 d\nu_{\rm obs}}{H(z)}, \tag{2.3}$$

where λ_{21} is the 21 cm wavelength. Relating the observed flux to the luminosity (equation 2.2) using the luminosity distance and knowing that $dS_\nu = I_\nu d\Omega d\nu_{\rm obs}$, we

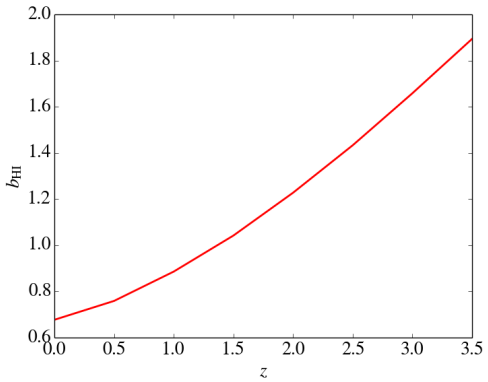

FIGURE 2.1: The HI bias as a function of redshift.

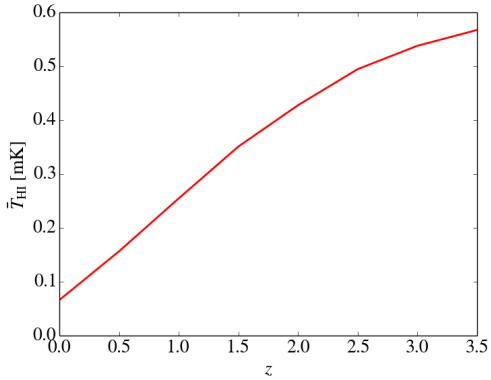

FIGURE 2.2: The average HI brightness temperature as a function of redshift.

solve for intensity:

$$I_\nu = \frac{3h_p c}{16\pi}\frac{A_{10} n_{\rm HI}}{H(z)}. \tag{2.4}$$

Using Rayleigh-Jeans approximation, $I_\nu = 2k_{\rm B}T_{\rm HI}\nu^2/c^2$ we solve for $T_{\rm HI}$. The brightness temperature:

$$T_{\rm HI} = \frac{3h_p c^3}{32\pi}\frac{A_{10}}{k_{\rm B}\nu_{21}^2}\frac{(1+z)^2}{H(z)} n_{\rm HI}. \tag{2.5}$$

The comoving number density can be written as $n_{\rm HI} = \Omega_{\rm HI}\frac{\rho_c}{m_{\rm p}}(1+\delta_{\rm HI})$ where $\Omega_{\rm HI} = \frac{\rho_{\rm HI}(1+z)^{-3}}{\rho_c}$, the comoving HI fraction where $\rho_{\rm HI}$ is the HI density, $m_{\rm p}$ is the proton mass and $\delta_{\rm HI}$ is the HI density contrast. The HI density fluctuation, $\delta_{\rm HI}$ is expected to trace the underlying dark matter field, $\delta_{\rm cdm}$ with an homogeneous bias on linear scales:

$$\delta_{\rm HI} = b_{\rm HI}\delta_{\rm cdm}, \tag{2.6}$$

If the model for $\Omega_{\rm HI}$ is known it will be possible to work out the HI bias and average brightness temperature, $\bar{T}_{\rm HI}$. Figure 2.1 and 2.2 shows the evolution of the HI bias and average brightness temperature as a function of redshift respectively. Once the HI bias can be quantified, we can then calculate the HI power spectrum:

$$P_{\rm HI}(k,z) = \left(b_{\rm HI} + f\mu^2\right)^2 D^2(z)\bar{T}_{\rm HI}^2 P_{\rm cdm}(k). \tag{2.7}$$

where f is the logarithmic derivative of the growth factor (D) by scale factor ($f \equiv d\ln D/d\ln a$) (Kaiser, 1987; Hamilton, 1998), μ is cosine of the angle to the line-of-sight, k is the wavenumber and $P_{\rm cdm}$ is the dark matter power spectrum generated from CAMB (Lewis et al., 2000). Figure 2.3 shows the estimated HI power spectrum with changing redshift.

2.2 HI intensity mapping

Intensity mapping is a novel technique proposed to do LSS cosmology in three dimensions by using HI as a biased tracer of the dark matter density field (Seo and Eisenstein, 2007; Chang et al., 2008; Loeb and Wyithe, 2008; Peterson et al., 2009;

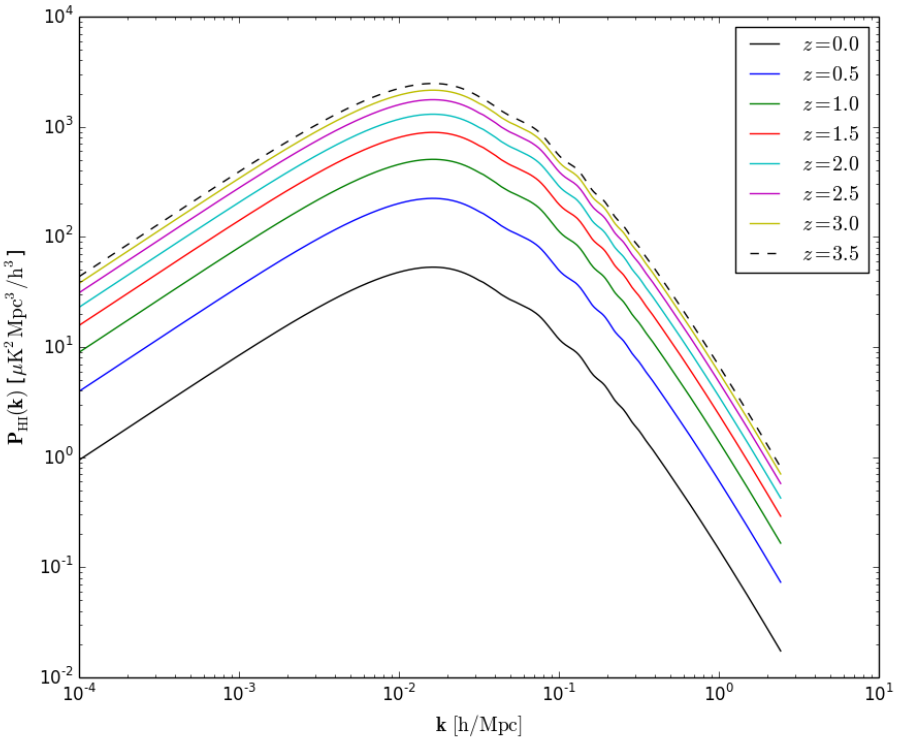

FIGURE 2.3: HI power spectrum in terms of the primordial power spectra at different redshifts. The overall amplitude of the power spectrum grows with redshift.

Camera et al., 2013; Battye et al., 2013; Santos et al., 2015c; Bull et al., 2015b,a; Smoot and Debono, 2017). What is measured is the intensity of the red-shifted 21 cm line over the sky in a range of redshifts without the requirement to resolve individual galaxies. The SKA will be dedicated to making 3D maps of equation 2.5 and then using those maps to constraint cosmology using equation 2.7. The first observational indirect detection of $\Omega_{\mathbf{HI}}$ came from the GBT in 2012. Masui et al. (2013) placed constraints on this quantity by cross correlating with WiggleZ observations. The experiment measured

$$\Omega_{\mathbf{HI}} b_{\mathbf{HI}} r = [0.43 \pm 0.07(stat.) \pm 0.04(sys.)] \times 10^{-3},$$

where r is the galaxy-hydrogen correlation coefficient. This measurement was made from data spanning the redshift range $0.6 < z < 1$ over two fields that equal 41 square degrees in 190 hours of integration time.

2.3 Foreground Emission

Nearly all emission from normal galaxies comes from synchrotron radiation due to relativistic electrons and free-free radiation from HII regions. At higher frequencies (200 GHz), dust absorption dominates over this type of radio emission. The radio emission independent of the temperature of the source is called non-thermal radiation (e.g synchrotron emission). Thermal radiation is instead electromagnetic emission that is influenced by temperature (or heat, e.g free-free emission).

In this section, we explain in detail these types of radio emission that act as contaminants to the cosmological signal. We will use the term foregrounds in this project to refer to astrophysical contaminants to the **H**I signal. Fortunately, most of the astrophysical foregrounds have smooth spectra, meaning they can be modeled by a power-law dependent on frequency thus making it possible to subtract them. Leakage of polarised foreground emission is also a problem but this can be solved by careful calibration of the telescope. Noise contribution due to the telescope, spillover from the Earth's ground, and the Radio Frequency Interference (RFI) which observations also need to be accounted for. In this thesis, we will only account for foregrounds and noise.

2.3.1 Galactic synchrotron

Galactic synchrotron is expected to be the dominant contaminant to 21 cm observations with an amplitude of 4-5 orders of magnitude greater than the cosmological signal. This arises when highly energetic cosmic ray electrons move around the magnetic field from our galactic plane resulting in synchrotron emission (Rybicki and Lightman, 1979). This type of emission is also present in other galaxies because they have magnetic fields as well. The synchrotron emission relies on the electron density n_e and a perpendicular magnetic field $B_\perp$ to the line of sight. Using the energy distribution of the cosmic rays, one can determine the spectral index of the synchrotron emission. In the radio-mm frequency range, the electron density have a power-law of index p ($n_e(E) \propto E^{-p}$). The index p depends on the cosmic ray properties. The frequency dependence of synchrotron emission is also a power law;

$$I_{\rm sync}(\nu) \propto \nu^{\beta_s+2}, \tag{2.8}$$

where $\beta_s = -(p+3)/2$. This works out to be -3 for a value of p = 3. Using the Rayleigh-Jeans approximation, ($T(\nu) = \frac{c^2}{2k_B\nu^2} I(\nu)$), we get that $T_{\rm sync} \propto \nu^{\beta_s}$. p varies with direction on the sky and with frequency. Past observations showed that from 408 MHz to 10 GHz β_s varies from -2.8 to -3.2 (Reich and Reich, 1988; Davies et al., 1996; Bennett et al., 2003). There is also a general steepening of the spectrum as one probe high galactic latitudes. This emission is also highly polarised. There is an associated polarisation fraction for synchrotron emission which goes as $\Pi = \frac{3p+3}{3p+7}$ (Pacholczyk, 1970). The polarised emission changes its polarisation angle as it traverses the intergalactic medium due to Faraday rotation. The rotation measure for these particles is given by $RM = \int n_e B_\parallel dl$, where dl represents integration over the line of sight and $B_\parallel$ is the magnetic field in our line of sight (see section 3.2.3). The polarisation component of this signal is very problematic for foreground cleaning techniques. This is due to limited measurements of the emission available. There are although simulations for it, one example is the Hammurabi code (Waelkens et al., 2009). Generating models of the polarised component is a complex subject beyond the scope of this project.

2.3.2 Free-Free emission

This type of emission is radiated when free electrons are accelerated by ions. There is free-free emission coming from our very own Milky Way, called galactic free-free and from outside our galaxy, the extra-galactic free-free. The emission traces the warm ionized medium. This emission is usually called Bremsstrahlung radiation. It is in general fainter than the synchrotron or dust emission (thermal), except in

active star-forming regions in the galactic plane. In terms of frequency, these foregrounds can also be modeled as power laws since they are spectrally smooth. These astrophysical sources can be found in accretion disks (Sancisi et al., 2008), **H**II regions, galactic hot corona and intergalactic gaseous medium in clusters or groups (Alves et al., 2012). In brightness temperature, this kind of emission is several orders of magnitude greater than the **H**I signal. Spectral index for this type of emission is dependent on the electronic temperature T_e and varies slowly with frequency (Bennett et al., 2003; Dickinson et al., 2003). On high frequency intervals, 10 GHz to 100 GHz, the spectral index of this emission varies from 2.12 to 2.20 for 4 000 < T_e < 10 000 K, see the Planck Sky Model (PSM, Delabrouille et al. (2013)). PSM generates the free-free emission spatial variations in three ways: by using **H**α as a proxy for mapping free-free emission at any frequencies or a free-free emission map from WMAP called Maximum Entropy Method (Bennett et al., 2003) or a map obtained by Miville-Deschênes et al. (2007). Other observations of this emission have been dedicated to mapping our galaxy. This type of emission can be found in the intergalactic medium as well as the extra-galactic because of the presence of ionized regions exists in every galaxy.

2.3.3 Extra-galactic point sources

Extragalactic point sources can be regarded as radio emission from objects beyond our Milky Way galaxy. Types of point sources are Active Galactic Nuclei (AGN) and star-forming galaxies which can be classified into star-burst galaxies and quiescent galaxies. AGNs are quite different from normal galaxies since the light they emit is not the total sum of the starlight in the galaxy. AGNs are powered by a supermassive black hole in the center and are observed to have jets. The jets are comprised of plasma of high energetic relativistic synchrotron particles (Begelman et al., 1984) extending from the center of the AGN. AGNs are believed to form when gas accretes onto a disk (Fabian, 2012). Examples of AGNs are Quasars, blazars and Radio galaxies (see Padovani et al. (2017) for a recent review). A quasar is an AGN viewed at a tilted angle, a blazar is an AGN viewed face on and a radio galaxy is viewed edge-on (Blandford and Rees, 1974; Blandford and Königl, 1979; Smidt et al., 2017). Figure 2.4 shows an example of the AGN[1] powered by a central source, a supermassive black hole. The name quasar means quasi-stellar or star-like meaning these are galaxies that look like stars since they are point-like in the optical. The most studied AGN is Cygnus A (Perley et al., 1984; Djorgovski et al., 1991). The frequency range of point sources overlaps with the range covered by the **H**I cosmological signal making the point sources one of the foreground candidates to the **H**I cosmological signal.

2.4 Simulation software

In this section, we cover what simulation was used and how it works. The Cosmological Realizations for Intensity Mapping Experiments (**CRIME**,Alonso et al. (2014)) comes with a suite of three routines that allow us to generate the cosmological signal maps alongside desired foregrounds. The three executables making up the simulation are Get**H**I, ForGet and JoinT and these are written in C. This simulation can be downloaded from Github[2]. The code comes with OpenMP

[1] https://imagine.gsfc.nasa.gov/science/objects/active_galaxies1.html
[2] https://github.com/damonge/CRIME

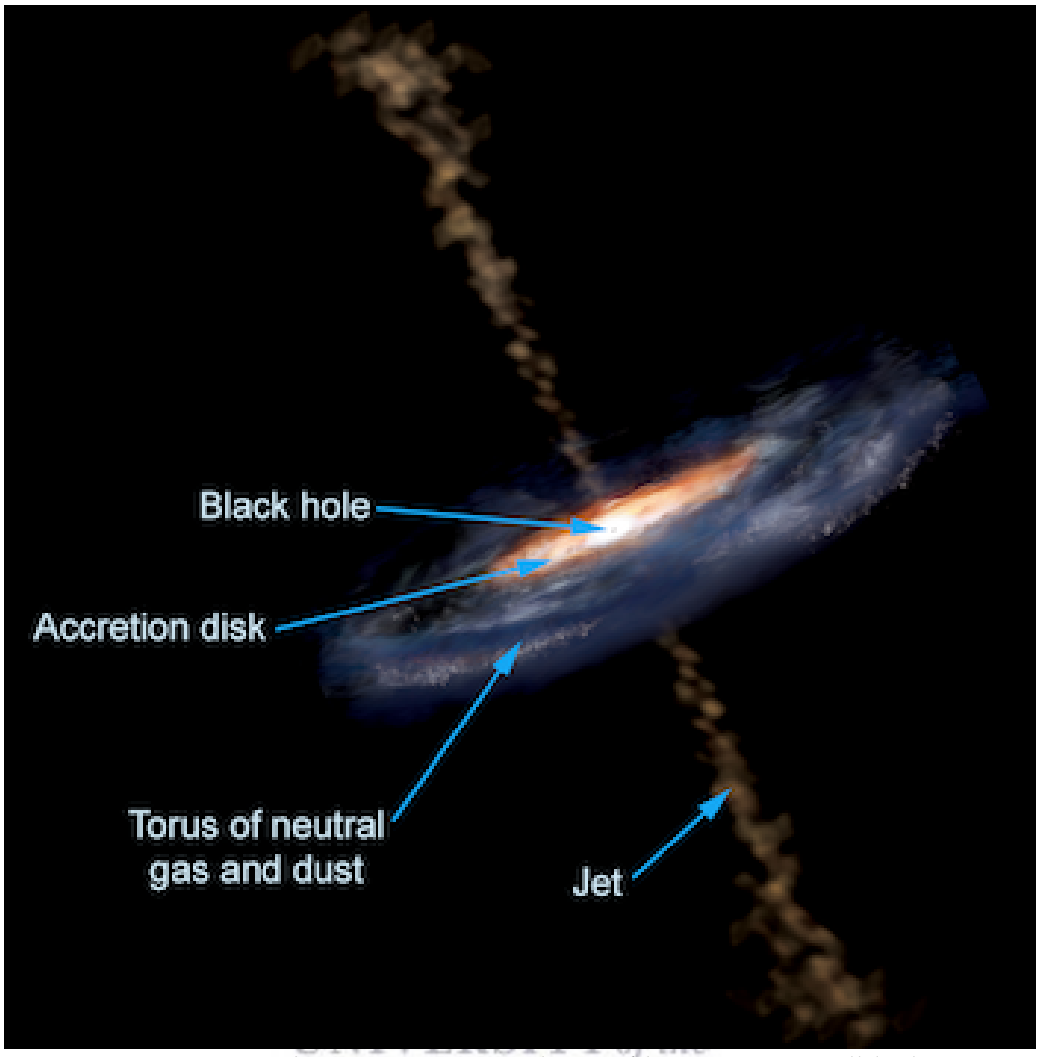

FIGURE 2.4: This is an artist's rendition of how an AGN will look like. In this type of galaxy, the HI halo is a toroidal cloud around the accretion disk. Image credit: National Aeronautics and Space Administration (2016)

and MPI-parallelization capabilities. JoinT allows generating the noise realization maps based on the type of instrumental parameters chosen. The basis for most of the CRIME simulations is a linear dark matter field. To do this, lognormal distributions of the late-time density field were used. Lognormal distribution is based on a random field $x_{\rm LN}$ given by,

$$\delta_{\rm LN} = \exp(\delta_G), \tag{2.9}$$

where the term δ_G is the Gaussian random field. This means the lognormal distribution is never negative. To make sure that the lognormal overdensity field has zero mean, equation 2.9 is modified to have the form,

$$1 + \delta_{\rm LN} = \exp\left(\delta_G - \frac{\sigma_G^2}{2}\right). \tag{2.10}$$

The σ_G term guarantees that $<\delta_{\rm NL}> = 0$. These lognormal transformation fields were used in Cole et al. (2005) and Wilman et al. (2008). These fields cannot be expected to

TABLE 2.1: Foreground $C_\ell(\nu_1, \nu_2)$ model parameters for the pivot values $\ell_{\rm ref} = 1000$ and $\nu_{\rm ref} = 130$ MHz.

Foreground	A [mK^2]	β	α	ξ
Galactic synchrotron	700	2.4	2.8	4
Point sources	57	1.1	2.07	1
Galactic free-free	0.088	3	2.15	35
Extragalactic free-free	0.014	1	2.1	35

able to describe the galaxy density on small scales ($\lesssim 5\, h^{-1}$Mpc, Kitaura et al. (2010)) meaning this is applicable for scales well in the linear regime only.

2.4.1 HI Simulation

Get**H**I is the code that allows one to make **H**I maps. Get**H**I requires a parameters file containing the cosmological parameters, a matter power spectrum generated from the same cosmological parameters (Lewis et al., 2000), pixel resolution, number of points in a grid and the desired frequency range. The resolution parameter (called $N_{\rm side}$) and the number of points in a grid (called $N_{\rm grid}$) are the only parameters used by the code to determine the total run time and the required disk space to store the generated **H**I maps. The typical CPU-run time on a shared-memory machine is $\sim$ 30 minutes (with 64-80 cores) while on a Distributed memory machine ($i.e$ cluster) it only takes a few minutes. This computational output holds for $N_{\rm side} = 512$, $N_{\rm grid} = 2048$ and if the frequency resolution is 1 MHz. A typical disk space of 3.5 gigabytes is required to store 150 channels of the **H**I maps covering redshifts $z \in (0.42, 0.67)$.

2.4.2 Simulating foregrounds

The foregrounds are generated by the ForGet routine in **CRIME**. ForGet only requires a file (a frequency table) containing the frequencies covered by the **H**I maps. Note the frequencies should match with the ones used to generate **H**I maps in order to be able to add these simulated maps together and do accurate foreground cleaning. The different foregrounds generated are galactic synchrotron (unpolarized and polarised), galactic and extra-galactic free-free emission and point sources. Following Santos et al. (2005), foregrounds are generated as Gaussian realizations of;

$$C_\ell(\nu_1, \nu_2) = A \left(\frac{\ell_{\rm ref}}{\ell}\right)^\beta \left(\frac{\nu_{\rm ref}^2}{\nu_1 \nu_2}\right)^\alpha \exp\left(-\frac{\log^2(\nu_1/\nu_2)}{2\xi^2}\right). \tag{2.11}$$

The parameter α describes the foregrounds mean frequency dependence and ξ is the frequency-space correlation length that parameterizes the characteristic frequency scale over which foregrounds are correlated. Foregrounds with a correlation length that is small will be difficult to subtract because they will be less smooth in frequency. A denotes the amplitude and β parametrize the foregrounds distribution on angular scales. The parameter values can be found in Table 2.1.

Galactic synchrotron

The galactic synchrotron simulated in this project uses the Haslam map at 408 MHz extrapolated to other frequencies (Haslam et al., 1982). Since the Haslam maps have

poor resolution in small angular scales, we use equation 2.11 for small fluctuations in the synchrotron emission. The Gaussian realizations of $C_\ell(\nu_1, \nu_2)$ use the parameters given in Table 2.1. Synchrotron emission is also polarised. If the spectral index (p) of the cosmic ray electrons, that have energy density $n(E) \propto E^{-p}$, is $p = 2.5$, then the polarisation fraction $\Pi_0 = \frac{3p+3}{3p+7}$ is equal to 0.7. The galactic synchrotron emission is more significant towards the galactic plane (low b) because of the relevance increase of the effects of Faraday rotation. In turn this has been used in the **CRIME** simulation as a way to constrain the radial power spectrum of Stokes Q, of which **CRIME** can reproduce the effect if the Faraday-correlation length (ξ_ψ) is 0.5 $\mathrm{rad/m^2}$ (see section 5.2.1 in Alonso et al. (2014)).

Free-free emission

Simulating free-free emission also follows Gaussian realizations of $C_\ell(\nu_1, \nu_2)$ given by equation 2.11 using the parameters given in Table 2.1. This emission can be divided into two parts as suggested by Table 2.1 and in these simulations, this type of foreground has no polarisation information.

Point sources

Simulating point sources can also be done using the ForGet routine but in this project, we gave focus to generating our own point source maps using existing datasets like NRAO VLA Sky Survey (NVSS, Condon (1986)) and Sydney University Molonglo Sky Survey (SUMSS, Mauch et al. (2003)) and SKADS Simulated Skies (S^3, Wilman et al. (2008)) at 1.4 GHz. This will be covered in more details in Chapter 3. The point sources from S^3 were Poisson sampled to fill in the missing flux density intervals not covered by NVSS and SUMSS (see Chapter 3). This is not the case for point sources generated from **CRIME** which are Gaussian realizations of equation 2.11.

2.5 Assumed Survey

In simulating the noise component of the total intensity we use the JoinT routine from **CRIME**. The simulations presented in this project mimic the single dish experiment. The parameters used are similar to the ones of MeerKAT. We considered only $\lesssim$9% of the sky, the patch of the sky we chose overlaps with DES. Table 2.2 summarizes the instrumental specifications for the noise rms used in the simulations. The noise rms is shown in equation 2.12,

$$\sigma_N = \frac{T_{\mathrm{sys}}}{\sqrt{2t_{\mathrm{pix}}\Delta\nu}}, \tag{2.12}$$

where,

$$t_{\mathrm{pix}} = t_{\mathrm{obs}} N_{\mathrm{dish}} \frac{\Omega_{\mathrm{pix}}}{\Omega_{\mathrm{sur}}}, \tag{2.13}$$

and $\Omega_{\mathrm{sur}} = 4\pi f_{\mathrm{sky}}$ represents the survey area, Ω_{pix} represents the pixel area, t_{obs} represents the integration time and N_ν denotes the number of frequency channels (Santos et al., 2015b; Bigot-Sazy et al., 2015; Olivari et al., 2016). In the previous version of the code, $\Omega_{\mathrm{beam}} \sim \lambda/D_{\mathrm{dish}}$ (where λ is the wavelength of the emission) was used instead of Ω_{pix}, which was incorrect because it is the number of pointings that set the time per pointing and this is in turn set by the way we decide to

TABLE 2.2: Instrumental parameters used in the simulations.

$D_{\rm dish}$	13.5 m
$t_{\rm obs}$	4 000 h
$\Delta\nu$	1 MHz
N_ν	150
$N_{\rm dish}$	64
$T_{\rm sys}$	20 K
$(\nu_{\rm min}, \nu_{\rm max})$	(850, 1000) MHz
$(z_{\rm min}, z_{\rm max})$	(0.42, 0.67)
$f_{\rm sky}$	0.0886
sky area	3 655 deg^2

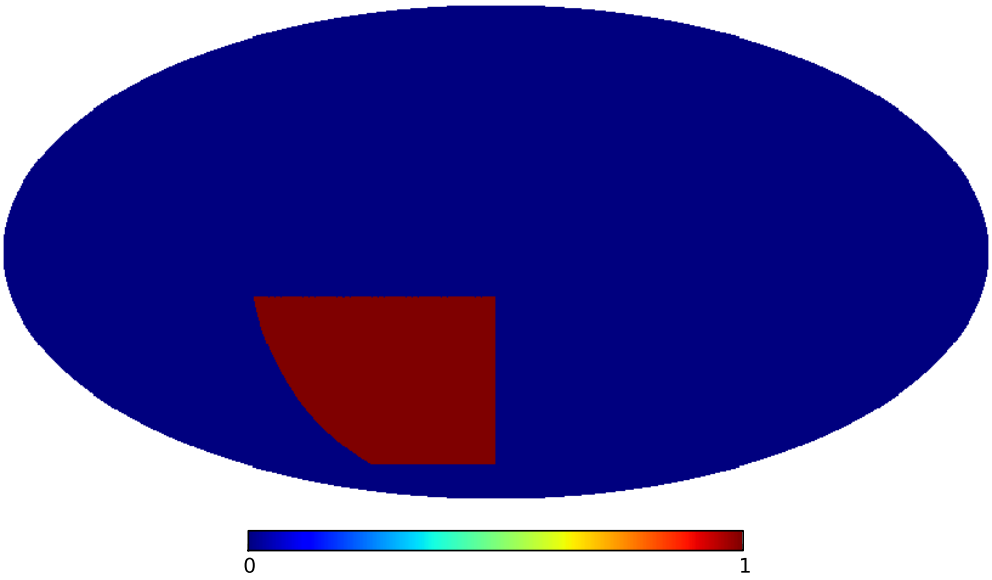

FIGURE 2.5: The assumed survey overlapping with DES region. Red represents the simulated region while blue is where our analysis is blind. The map projection shown is Equatorial coordinates.

"mosaic" the sky, not the telescope beam. Typically we use a pixelisation that is smaller than the beam size. This also makes the actual noise rms constant in frequency since $\Omega_{\rm pix}$ does not scale scale with λ as $\Omega_{\rm beam}$. This was corrected in the current version of the simulation. In our simulations, we consider a pixel resolution, $\theta_{\rm pix} = 0.11$ deg (with pixel area given by $\Omega_{\rm pix} = \theta^2_{\rm pix}$). This gives a noise rms value of 0.245 mK for the instrumental parameters in Table 2.2. Figure 2.5 shows the area we have chosen for these simulations.

The **CRIME** routine JoinT creates the Gaussian realization of the noise rms corresponding to the survey chosen. For example, if the survey area is full sky then full sky Gaussian realizations of the noise rms will be created by JoinT. We then manually join the noise, **HI** and foreground maps using Python and save in **HEALPix** fits files. These maps will be used as input in the foreground cleaning code. Figure 2.6 shows the noise rms simulated from Table 2.2. As it can be seen that the rms fluctuates around the value σ_N, which is expected. Figure 2.7 shows

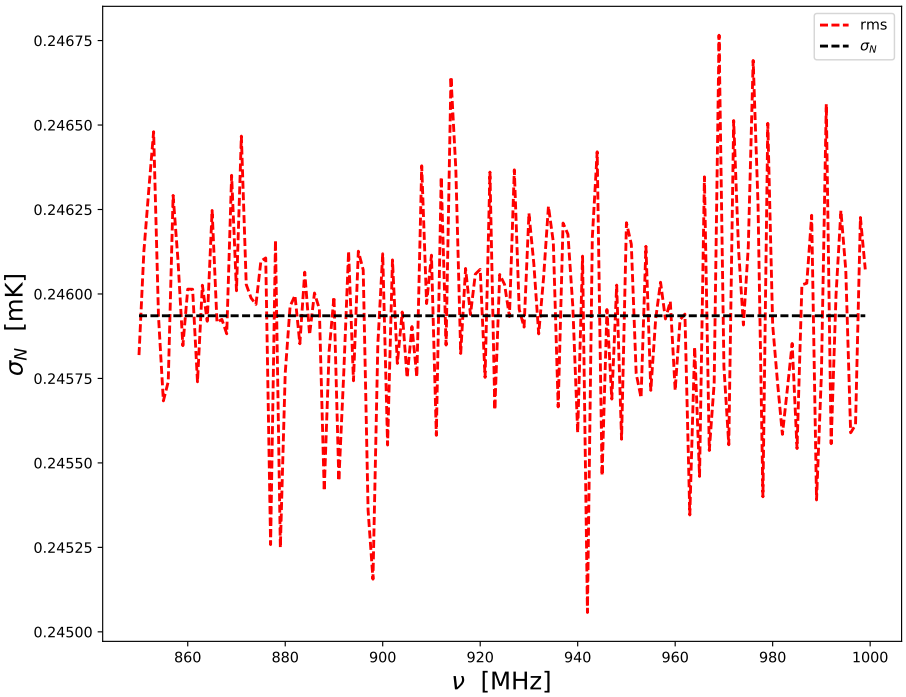

FIGURE 2.6: The noise rms for the region simulated. The black line is the theoretical rms of the noise (with a value of ~ 0.245 mK for the instrumental parameters in Table 2.2) maps given by equation 2.12 with the red dashed line representing the Gaussian realization of σ_N at each frequency.

all the foreground contaminants simulated at 950 MHz. The map viewing convention used here is the galactic coordinates. For our analysis, the noise realization was simulated for the patch shown in Figure 2.5. As it has been known in literature the synchrotron emission becomes more depolarized as one moves to high galactic latitudes (Strom, 1973; Alonso et al., 2014). It is convenient to look at the types of foregrounds in terms of lines of sight and we do this by picking out random galactic latitudes within our survey area. Figure 2.8 and 2.9 show four different lines of sight of the non-polarised and polarised foregrounds as a function of frequency. Since the simulated sky is about 9% of the sky we do not see a significant change in galactic synchrotron emission as a function of galactic latitude b. Stokes Q and U are calculated from $|P|\cos(2\phi)$ and $|P|\sin(2\phi)$ respectively, this implies that when Q (U) is near a zero and thus changing sign (resulting in a cusp in $|T_b|$), U (Q) is in anti-phase and thus at its maximum.

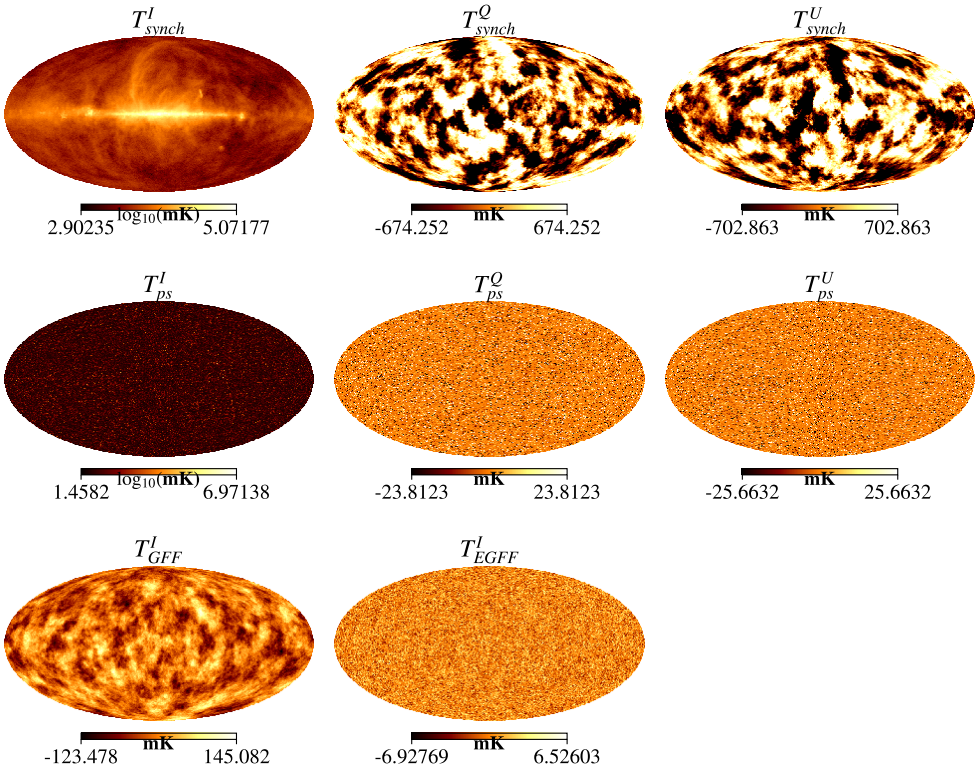

FIGURE 2.7: Different maps of the foreground contaminants at frequency 950 MHz used as input. *Top row*: The galactic synchrotron (subscript *synch*) shown in total intensity and polarisation (Q and U). *Middle row*: The point source (subscript *ps*) maps obtained from our catalogue for both total intensity and polarisation (see Chapter 3). *Bottom row*: The galactic and extra-galactic free-free emission with subscript *GFF* and *EGFF* respectively. The range of the polarised foreground components has been scaled for clear viewing by using their rms.

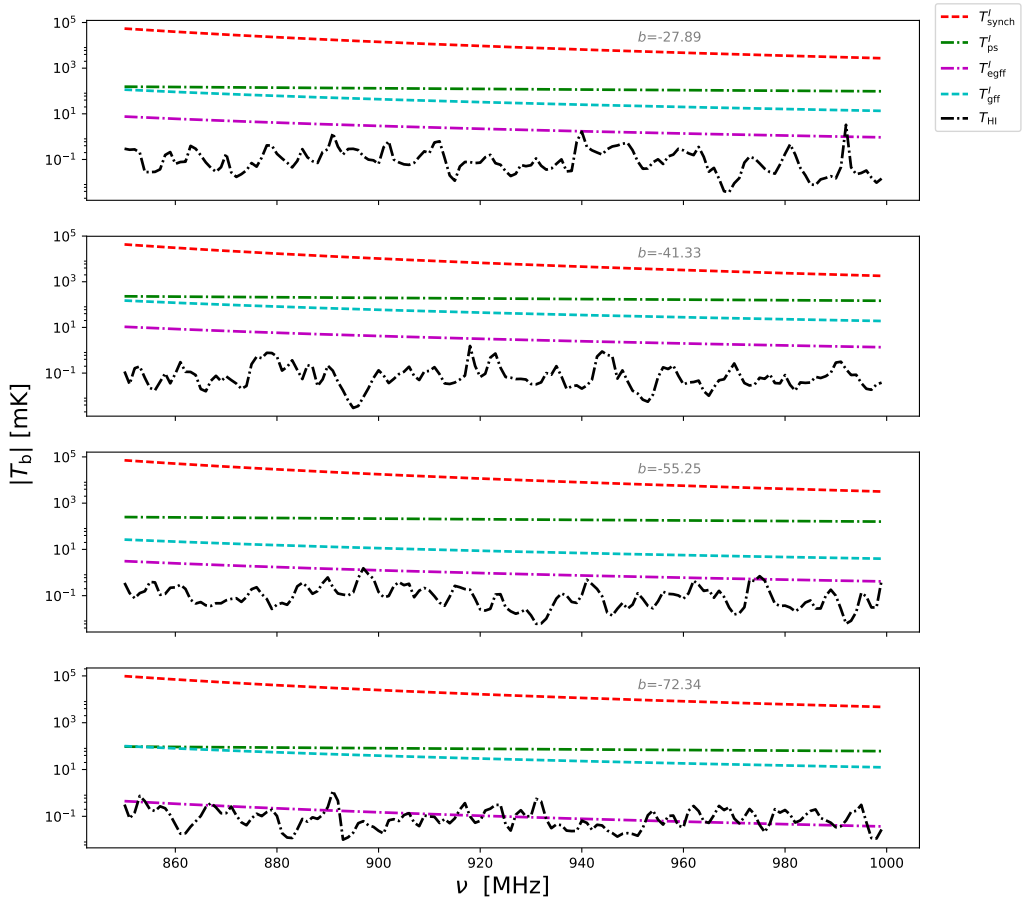

FIGURE 2.8: Brightness temperature of the foregrounds intensity as a function of frequency. Galactic synchrotron (in red), the point sources (in green) are from our catalogue and galactic (in cyan) and extra-galactic (in magenta) emission. The galactic synchrotron is 4-5 orders of magnitude greater than the cosmological HI signal (in black). The lines of sight are chosen in terms of the galactic latitude parameter, b and these lines of sight fall within our simulated region.

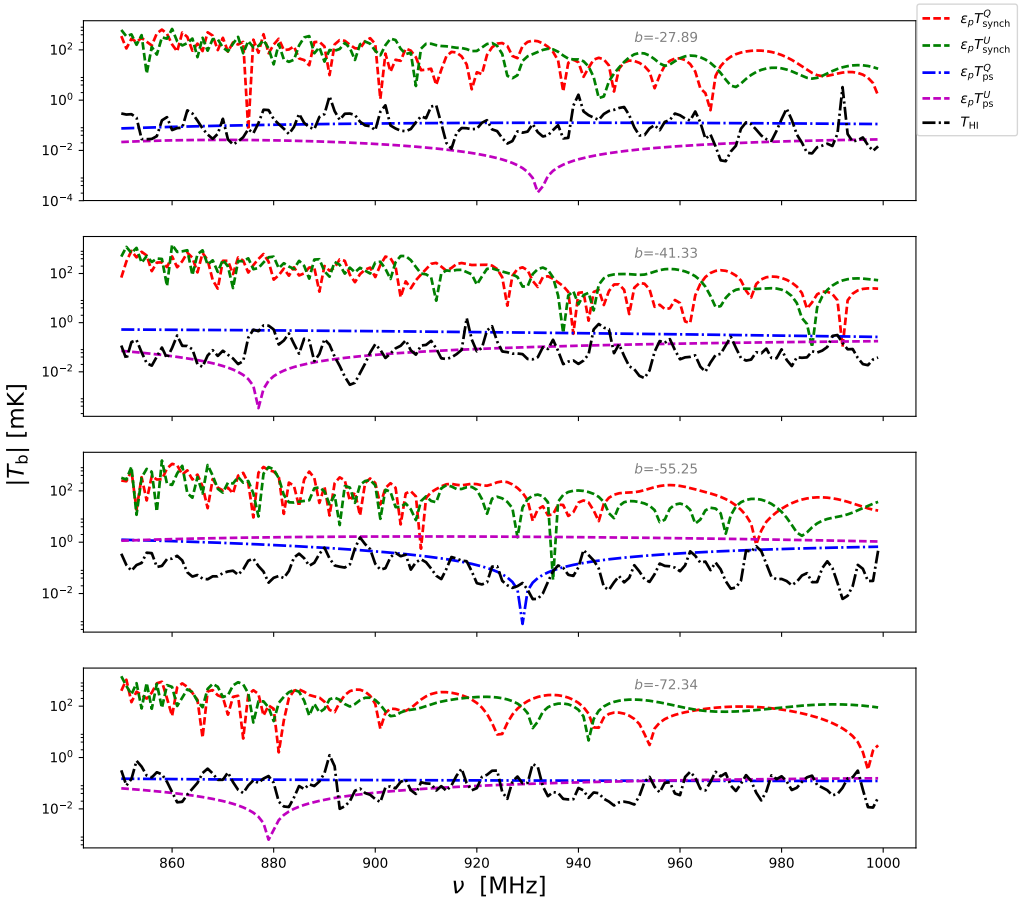

FIGURE 2.9: The absolute brightness temperature of each foreground polarised intensity falling within a pixel as a function of frequency. The galactic synchrotron Stokes Q (in red), galactic synchrotron Stokes U (in green), point source Stokes Q (in blue), point source Stokes U (in magenta) and the input HI (in black). Only 1 % of the Stokes Q and U emission was considered for these foregrounds. The lines of sight are chosen in terms of the galactic latitude parameter, b and these lines of sight fall within our simulated region. The galactic synchrotron emission does not change with b since the simulated area is only 9% of the sky.

2.6 Power spectrum tools

To get the angular and radial power spectrum **fg_rm** was used to output these quantities for our frequency range.The angular power spectrum is determined using a **HEALPix** function that performs a spherical transform of a map. Its starts by calculating the spherical harmonics coefficients of the map formally called the $a_{\ell m}$'s:

$$a_{\ell m}(\nu) = \int T(\nu, \hat{\mathbf{n}}) Y^*_{\ell m}(\hat{\mathbf{n}}) d\Omega_{\hat{\mathbf{n}}}, \tag{2.14}$$

where $Y^*_{\ell m}(\hat{\mathbf{n}})$ are the spherical harmonics and $T(\nu, \hat{\mathbf{n}})$ are the brightness temperature fluctuations at line of sight $\hat{\mathbf{n}}$. Taking the Fourier Transform of the 2-point correlation

of brightness temperature yields the angular power spectrum;

$$\hat{C}_\ell = \frac{1}{2\ell+1}\sum_{\mathrm{m}=-\ell}^{m=\ell}|a_{\ell m}|^2. \tag{2.15}$$

The radial power spectrum is used instead of the full 3D power spectrum because the BAO in the transverse direction are smeared out because of low angular resolution (as already mentioned we would need angular resolutions of at least 30 arcminutes in order to resolve the BAO in the transverse direction). The radial power spectrum is defined as:

$$P_\parallel(k_\parallel) = \frac{\Delta\chi}{2\pi N_\theta}\sum_{i=1}^{N_\theta}\widetilde{\Delta T}(k_\parallel,\hat{\mathbf{n}}_i)\widetilde{\Delta T^*}(k'_\parallel,\hat{\mathbf{n}}_i); \tag{2.16}$$

where $\widetilde{\Delta T}$ is the Fast Fourier Transform (FFT) of the brightness temperature along the line of sight, $\Delta\chi = \chi(z_{\max}) - \chi(z_{\min})$ is the comoving width of the redshift shell and N_θ is the number of lines of sight. We use frequency resolution of 1 MHz. This means the FFT is determined at constant intervals of the frequency wavenumber δk_ν = $2\pi/\Delta\nu$. The radial wavenumber is given by,

$$k_\parallel = \frac{\nu_{21}H(z_{\mathrm{eff}})}{(1+z_{\mathrm{eff}})^2}k_\nu, \tag{2.17}$$

where for each bin a constant effective redshift, z_{eff}, has been assumed and k_ν is the wavenumber of the transverse direction.

Chapter 3

Point source simulations

In this chapter, we explain the procedure used to obtain a realistic full sky point source catalogue in total intensity and polarisation. Such realistic point source maps are important because they will contaminate the signal at some level depending on the effects of the primary beam. Therefore, we need to take them into account when developing realistic simulations of the telescope observations. Firstly, we explain the surveys and simulations used to create our catalogue at 1.4 GHz and the conversion of the catalogue to brightness temperature maps. The full description of how we constructed our catalogue can be found in section 3.2. We then cover all the mathematical details and assumptions made before gridding the brightness temperature to **HEALPix** maps (section 3.4). Finally concluding with the obtained maps and power spectrum for the point sources.

3.1 Surveys

In this section, we go through the surveys used in providing point source information at 1.4 GHz. To create our point source catalogue, we start from already existing catalogues *i.e.* NVSS (section 3.1.1), SUMSS (section 3.1.2) and the simulated S^3 (section 3.1.3).

3.1.1 NRAO VLA Sky Survey

The Very Large Array (VLA) carried out the observational experiment called National Radio Astronomy Observatory VLA Sky Survey (NVSS, Condon (1986)) . This radio observatory is based in San Agustin, New Mexico with 27 antennas. The location was chosen to be far from radio frequency interference. The antennas are aluminum based with a dish diameter of 25 m. With a resolution range of 0.2-0.4 arcseconds, the frequencies probed are 1-50 GHz. Each dish is alt-az mounted $i.e$ it tilts up and down and rotates by 360°. The total dishes move on train rails making a "Y" shape (see Figure 3.1). This moving allows optimizing the baselines for sensitivity at the required scales. NVSS[1] covers the Northern Sky from declination $\delta \geqslant$-40° (82 % of the sky). The radio observations were done at 1.4 GHz with a resolution of 45 arcseconds. NVSS has 2326 $4° \times 4°$ continuum patches with three planes containing Stokes I, Q and U images. Total intensity and polarisation measurements are reported in units of mJy. The NVSS has 2 $\times 10^6$ sources with a flux density limit of $\sim$ 2.5 mJy with the types of sources being radio galaxies, quasars, low-luminosity AGNs, normal galaxies and starburst galaxies. The root mean square (rms) brightness fluctuations are $\sigma \approx$ 0.45 mJy/beam (equivalent to 0.14 K) for Stokes I and are $\sigma \approx$ 0.29 mJy/beam (0.09 K) for Stokes Q and U. The spectral energy distribution of these sources was not measured accurately but the

[1] www.heasarc.gsfc.nasa.gov/W3Browse/radio-catalog/nvss.html

FIGURE 3.1: The VLA looking South. Image Credit: NRAO (2008)

data indicates that these sources have steep spectra. The description of the NVSS catalogue can be found in Condon et al. (1998).

3.1.2 Sydney University Molonglo Sky Survey

The Sydney University Molonglo Sky Survey (SUMSS, Mauch et al. (2003)) surveyed the southern radio sky ($\delta \leqslant -30°$) at 843 MHz with similar resolution and sensitivity to NVSS. It was carried out using the Molonglo Observatory Synthesis Telescope (MOST) and covers about 43% of the sky. SUMSS[2] catalogue has 107 765 sources over a total area of 3500 deg^2. In the region where the survey overlaps with the NVSS, about 90% of the sources within SUMSS can be found in NVSS. SUMSS has a mask covering the regions within the galactic latitudes $|b| < 10°$. The sources observed with NVSS can be observable with SUMSS given that they are situated in the part of sky visible to both surveys. The rms for the mosaics is 1.2-1.3 mJy/beam with the mode being 1.25 mJy/beam. Below $\delta \leqslant -50°$ the SUMSS catalogue has sources that peak around 8 mJy while at $\delta \geqslant -50°$ the sources have flux densities around 18 mJy. SUMSS does not have polarised flux density measurements. The calculated median spectral index, α=-0.83 between 1.4 GHz (NVSS flux densities) and 843 MHz (Mauch et al., 2003).

3.1.3 S^3: SKA Simulated Skies

SKA Simulated Skies (S^3) is a suite of simulations developed to mimic the upcoming SKA observations of the radio sky (Wilman et al., 2008). One of the simulations of the radio sky used in this project is the S^3 Source Extractor (S^3-SEX)[3].

[2] http://www.astrop.physics.usyd.edu.au/sumsscat/
[3] http://s-cubed.physics.ox.ac.uk/

S^3-SEX is a semi-empirical simulation that traces the underlying dark matter density field using extrapolated luminosity functions. The simulation models the distribution of large-scale astrophysical sources, not their internal properties. It covers a 20×20 deg^2 sky area with roughly 320 million sources. It reaches a redshift of 20 and flux density limits of 10 nJy at 151 MHz, 610 MHz, 1.4 GHz, 4.86 GHz and 18 GHz. The types of sources simulated are Radio-quiet AGNs, Radio-loud AGN of the Fanaroff-Riley types (FR I and FR II) (Fanaroff and Riley, 1974) and star-forming galaxies (classified into star-burst and quiescent galaxies). The FR I and FR II fluxes are obtained from partitioning the flux between the core and extended lobes and hot spots. The spectral index distribution peaks around α ∼-0.7 to -0.8 due to the extended radio-loud AGN and the radio-quiet AGN. The observed frequencies satisfying this spectral distribution are 610 MHz and 1.4 GHz. The S^3 simulations ignores polarisation so we had to add it (see section 3.2.2).

3.2 Catalogue description

In this section, we explain in more detail, how the full sky catalogue of the radio point sources was obtained. We explain briefly how we obtained polarisation from NVSS data and estimated the full catalogue polarisation information.

3.2.1 Total intensity catalogue

The aim is to construct a mock catalogue of radio sources at 1.4 GHz in both intensity and polarisation and then turn it into a **HEALPix** map at 1.4 GHz (again both in intensity and polarisation). We start by querying the S^3-SEX online database to retrieve a 20×20 deg^2 simulated radio sources at 1.4 GHz. These simulations have total flux density information only, statistically extrapolated from luminosity functions. Using Poisson statistics, we populated the full sky using a **HEALPix** pixelisation scheme and turn it to a full sky catalogue. To make the catalogue more realistic, we add the NVSS (at 1.4 GHz) and SUMSS catalogues (extrapolated from 0.843 GHz to 1.4 GHz). The polarisation catalogue is drawn statistically from NVSS polarisation properties. The procedure is explained in more detail in this section.

We use the S^3-SEX to retrieve extra-galactic radio source data that mimics what the SKA will see in future observations (see section 3.1.3). We obtained the point source data for fluxes covering $S \geq 10$ μJy at 1.4 GHz by querying the online database. All the information has a corresponding right ascension and declination associated with it. We use the full area (20×20 deg^2) to infer the number of galaxies (treated as point sources, whether they are AGN, star-forming or star-burst was neglected) that will populate the full sky. We assume that the number of galaxies is proportional to the area covered. In order to translate the data retrieved from the query to our very own source catalogue at 1.4 GHz we follow this procedure:

- We pixelise the full sky using pixels of equal area, with resolution $\theta_{\rm pix} = \sqrt{\Omega_{\rm pix}} = 0.11$ deg, where $\Omega_{\rm pix}$ is the pixel area (see section 2.4).
- We group the S^3-SEX data into logarithmic bins of flux density. Every flux bin i has $N^i_{\rm src/bin}$ sources.
- We calculate $\bar{N}^i_{\rm src/pix} = N^i_{\rm src/bin}\left(\frac{A_{\rm pix}}{A_{\rm gal}}\right)$, the mean number of sources per pixel assuming that the average number of galaxies of a certain flux, scales with area considered. $A_{\rm gal}$ is the S^3-SEX area (20×20 deg^2).

- The actual number of galaxies per pixel in flux bin i is obtained by Poisson sampling from a distribution with mean $\bar{N}^i_{\mathrm{src/pix}}$.
- We give each resulting galaxy a spectral index.

The spectral indexes were generated from a Gaussian distribution around an average spectral index, α=-0.8, with a width of 0.2 (for $S \propto \nu^{\alpha}$). This is in agreement with the S^3 results (see section 3.1.3). The above process allows us to extend the S^3-SEX simulated sources to full sky. In the following we will refer to these as S^3-sources. This catalogue can be improved with realistic flux density information. This will allow us to have a more realistic simulation once we take into account the region that is going to be observed by the telescope. In particular, strong point source on the sky can have a relevant effect on the observations. We considered the flux density data from the NVSS catalogue (see section 3.1.1) at 1.4 GHz and the SUMSS catalogue (see section 3.1.2) at 843 MHz. The NVSS catalogue covers the Northern sky from a declination of -40° ($i.e.$ 82% of the full sky) while SUMSS is a Southern sky survey from the declination of -30° and below. The following explains in detail the steps taken to add observational data from the two surveys:

- The NVSS catalogue has rms = 0.45 mJy/beam. Since we want to consider only clearly detected sources, we select the sources that are above 10σ. In practice, we fix our flux cut at 5 mJy.
- For sources in the SUMSS catalogue, the flux density is extrapolated from 843 MHz to 1.4 GHz using $\alpha = -0.83$ (see section 3.1.2). The same cut is applied to retain only sources with fluxes that are greater than or equal to 5 mJy.
- Each retained NVSS and SUMSS sources are assigned a spectral index from the Gaussian distribution, $\mathcal{N}$(-0.8, 0.2).
- For the full sky we retain S^3-sources with flux densities below 5 mJy to account for the underlying sources that are too far away or faint to be detected by NVSS or SUMSS.
- Since the SUMSS catalogue the galactic is masked, we refilled the region $|b|$ <10° with S^3-sources that have flux densities above the flux cut.

The hitmap in Figure 3.2 shows the survey areas for NVSS and SUMSS.
In order to make sure that the catalogue is valid, we reconstruct the differential source counts diagram at 1.4 GHz. This is shown in Figure 3.3. The scatter at high flux density is due to Poisson sampling a flux density bin that has few sources close to zero sources per bin. The differential source count plot reconstructed is in agreement with previous work done in literature (Condon, 1984; Padovani et al., 2017).

3.2.2 Polarised flux density catalogue

We want to construct the polarisation component of our source catalogue. We recall that the complex polarised flux density can be written as

$$P = Q + iU = |P|\exp(i2\phi), \tag{3.1}$$

where $|P| = \sqrt{Q^2 + U^2}$ is the polarised flux density and ϕ is the polarisation angle. Since the S^3 simulation, SUMSS and other radio observations do not have

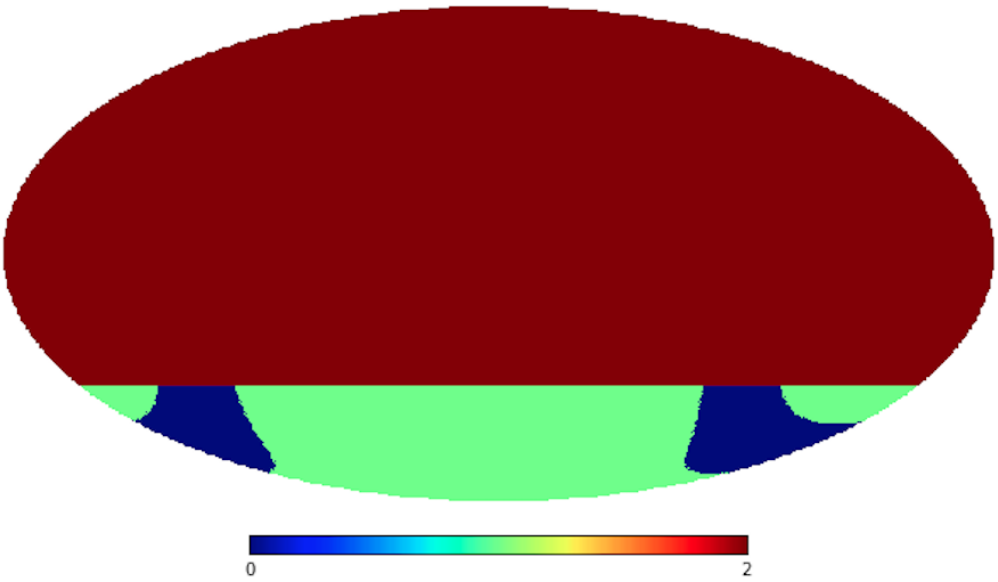

FIGURE 3.2: Survey coverage for NVSS and SUMSS. The red colour is the NVSS coverage and SUMSS is in green. The blue is the SUMSS masked area *i.e.* $|b|$ <10° which was filled with S^3-sources.

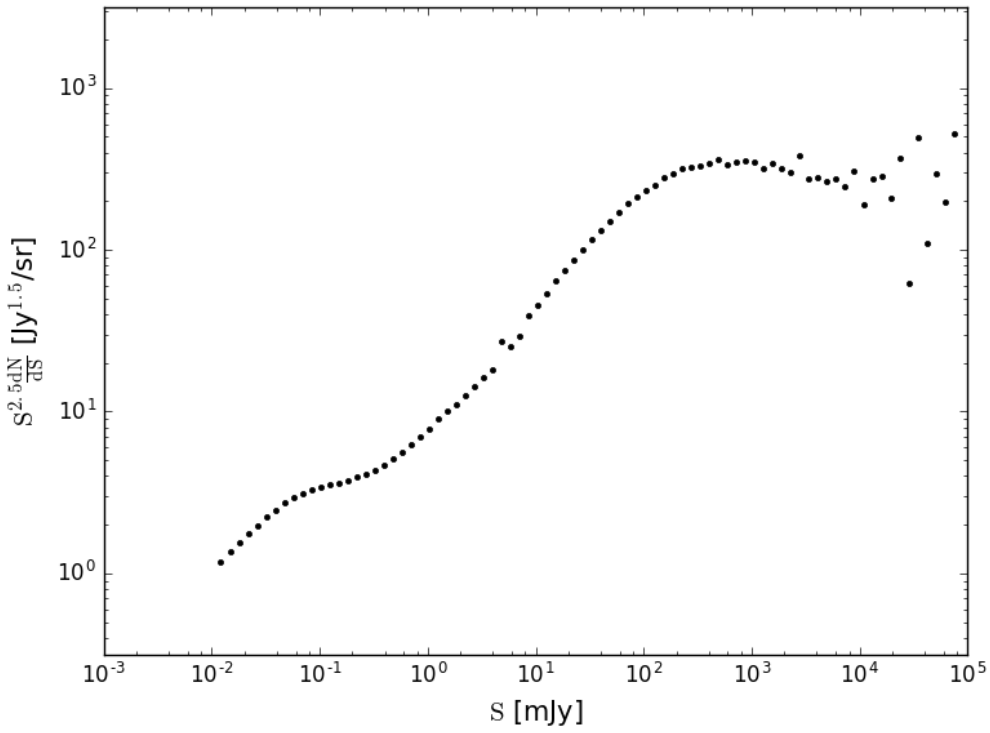

FIGURE 3.3: The normalized intrinsic source counts plot of the catalogue. Below 5 mJy are the sources from S^3 simulation while above are NVSS sources and SUMSS at 1.4 GHz.

polarisation, we use the NVSS to infer polarisation fraction of all point sources in our catalogue. To achieve this:

- We bin the NVSS intrinsic flux density into 5 large logarithmic bins.

- At each interval, we calculate the polarisation fraction using the relation, $\Pi_{\rm bin}=\frac{|P|}{S}$, where $|P|$ is the polarised flux density and S is the flux density.
- For each bin, we determine the sample mean and standard deviation, $\bar{\Pi}_{\rm bin}$ and $\sigma_{\bar{\Pi}_{\rm bin}}$, of the polarisation fraction.

The type of distribution obtained from the NVSS catalogue is shown in Figure 3.4. The error bars in polarization fraction extend to the negatives because of the bias correction done on the NVSS data set which introduced negative values of $|P|$ to the catalogue. For statistical purposes the non-physical values were not ignored when polarisation fraction was determined per total flux density bin. The 5 bins were made large to ensure a Gaussian distribution for all polarisation flux density for all bins. Once we have $\bar{\Pi}_{\rm bin}$ and $\sigma_{\bar{\Pi}_{\rm bin}}$ using the NVSS sources, we then bin the full point sources catalogue in terms of flux density and assign Gaussian distributed polarisation fractions. For example, all the sources within our catalogue that fall within i-th bin will be given a polarisation fraction $\Pi_{\rm Gauss}$ drawn from $\mathcal{N}(\bar{\Pi}_i\ ,\sigma_{\bar{\Pi}_i})$. From this, we infer what is the polarised flux density ($|P| = \Pi_{\rm Gauss}S$) for every point source in our catalogue. If the drawn value is below zero we consider it non-physical and set it to zero. The point sources of our catalogue that lie outside the first (last) NVSS bin were given a polarisation fraction of the first (last) bin. Figure 3.5 shows the differential source count plot for the polarisation component of our catalogue. The source counts for polarization flux density are in agreement than what is obtained in literature (Tucci et al., 2004).

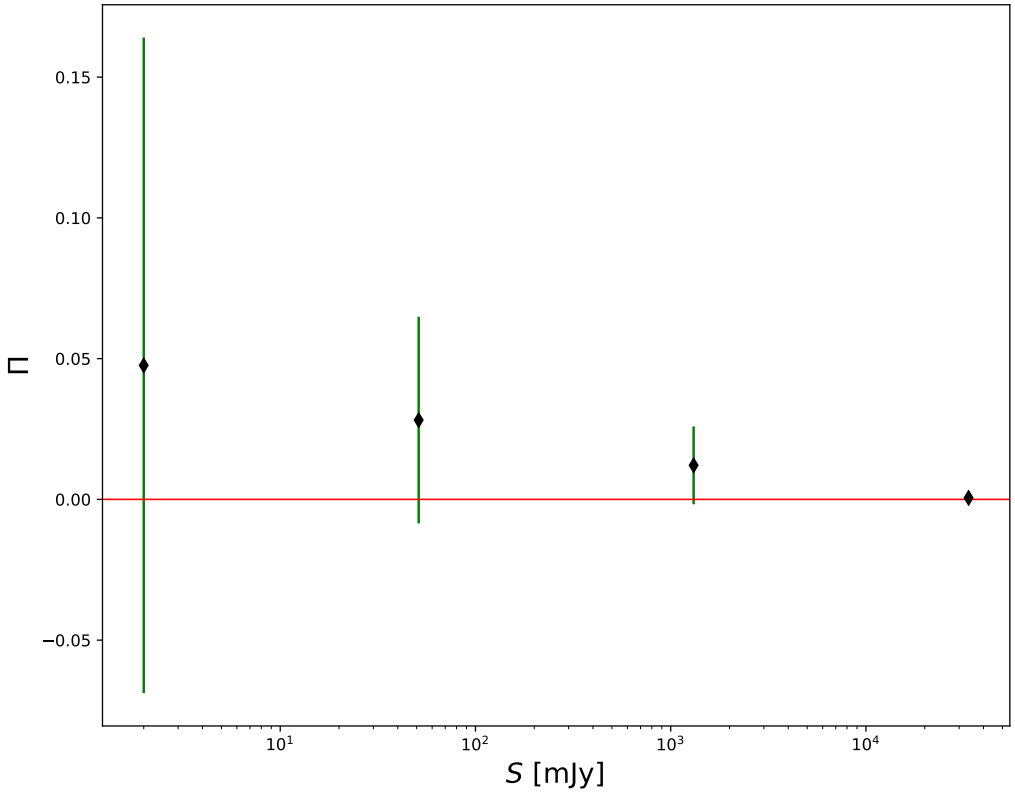

FIGURE 3.4: The mean value $\bar{\Pi}_i$ and 1 sigma $\sigma_{\bar{\Pi}_i}$ error of the polarisation of the NVSS sources as a function of the total flux bins.

3.2.3 Polarisation angle

A linear polarisation signal experiences Faraday rotation when in presence of a magnetic field. Since the interstellar medium is full of the magnetic fields we expect the intrinsic polarisation angle of our sources to be Faraday rotated.

$$\phi = \phi_0 + RM\lambda^2, \tag{3.2}$$

where ϕ_0 is the intrinsic polarisation angle and the Faraday depth is

$$RM = \frac{e^3}{2\pi(m_e c^2)^2} \int_0^{\mathrm{s}} n_e(s', \hat{\mathbf{n}}) B_{\parallel}(s', \hat{\mathbf{n}}) ds', \tag{3.3}$$

where ds is the line of sight element traversed by the light as the electrons (given by the electron density, n_e) gyrate around parallel magnetic fields ($B_{\parallel}$) along our line of sight. Note that the rotation is frequency dependent.

To assign an intrinsic polarisation angle to every source in our catalogue we draw a value from a Uniform distribution, $\mathcal{U}$ (0°, 360°). For assigning a value of Faraday depth RM we use a Gaussian distribution $\mathcal{N}$(5.6, 20) rad/m^2. The rotation measure distribution is given in Figure 3.6 and it was extrapolated by Nunhokee et al. (2017) using the paper by Taylor et al. (2009). The final polarisation angle for every source at a given frequency can be calculated using equation 3.2.

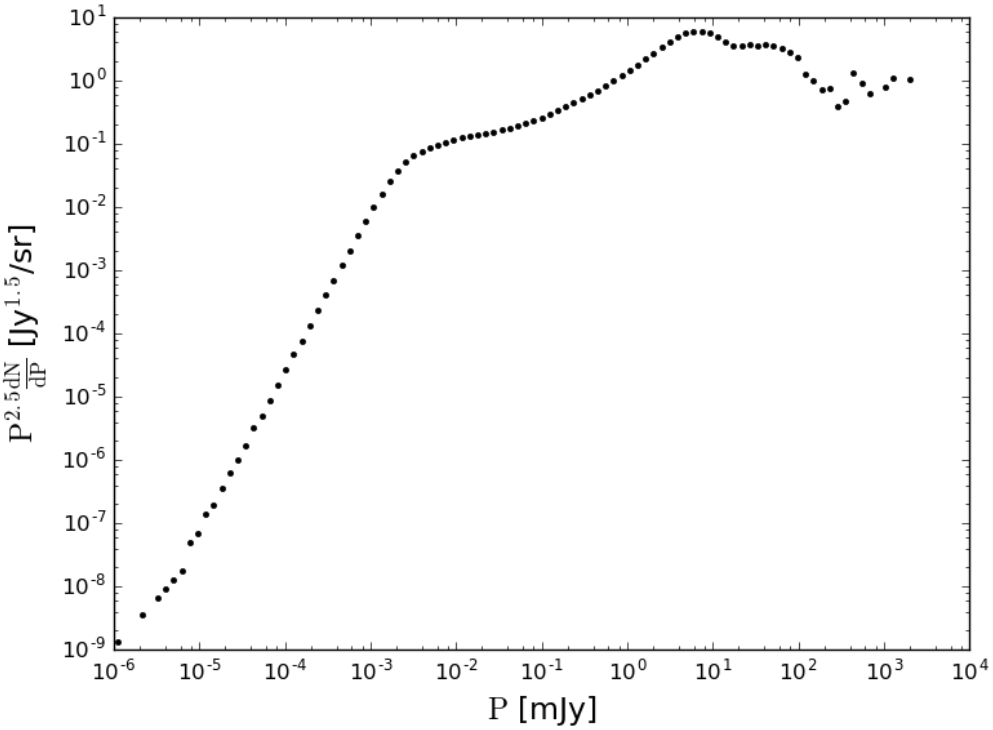

FIGURE 3.5: The polarisation component of the differential source counts plot of the catalogue. This is the differential source counts plot reconstructed from the total polarisation $|P|$.

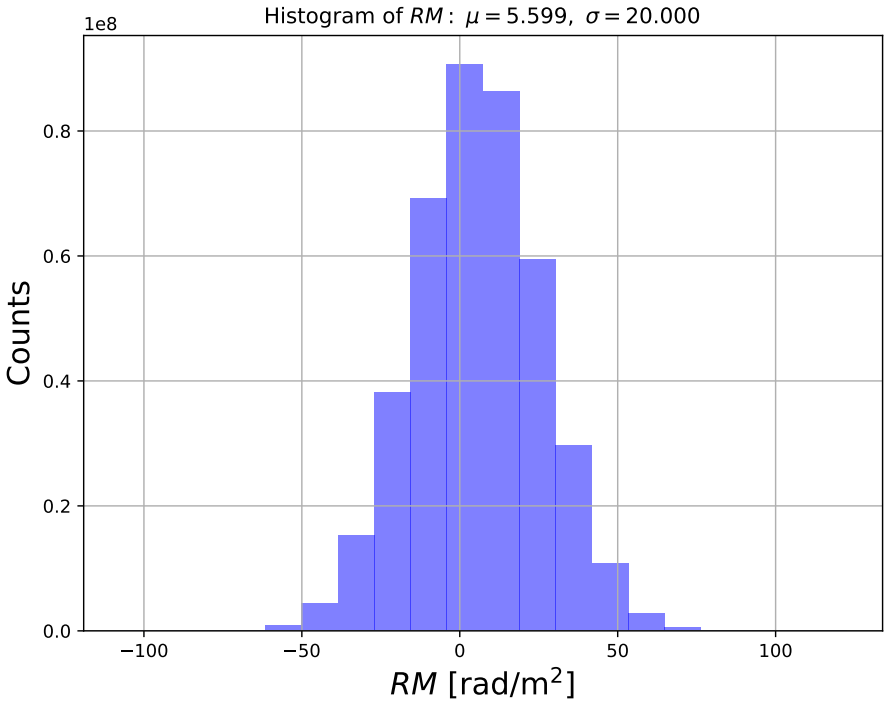

FIGURE 3.6: A Gaussian distribution of rotation measure values.

TABLE 3.1: Table of the full catalogue at 1.4 GHz. The total number of the radio sources in this catalogue is ≃408.8 million.

RA	DEC	α	$S_{1.4\ \mathrm{GHz}}$	$\lvert P\rvert$	ϕ_0	RM
0.00067	-4.045	-0.6721	106.8	9.0973	181.07031	-11.11024
0.00104	-20.080	-0.9359	7.60	2.8678	277.30053	19.82430
0.00163	0.683	-0.7882	149.50	7.7268	199.85905	3.11680
0.00387	8.132	-0.8392	8.80	0.3296	20.55178	-10.954840
.						
.						
.						
139.100	-47.945	-0.8564	2.63683e+06	2210.3860	48.16041	17.95413
147.811	-51.062	-1.0231	2.63683e+06	1302.3953	228.8445	-2.65341
196.842	-57.589	-0.7582	2.63683e+06	1192.8902	43.90959	40.79391

3.3 Catalogue format

In this section, we show what contents are in the catalogue. and we show a snippet of the catalogue. As aforementioned, the catalogue is full sky pushed down to a minimum flux density of 10 μJy. In total, the number of sources is 408 807 228. The catalogue is organized in this way:

- Column 1: The source's right ascension coordinate given in degrees.
- Column 2: The source's declination coordinate given in degrees.
- Column 3: The spectral index.
- Column 4: The source's flux density at 1.4 GHz given in mJy.
- Column 5: The source's polarised flux density at 1.4 GHz given in mJy.
- Column 6: The intrinsic polarisation angle in degrees.
- Column 7: The rotation measure given in rad/m^2.

An extract of the catalogue is given in Table 3.1.

3.4 Point source maps

In this section, we explain how we convert flux density of our catalogue to brightness temperature (section 3.4.1). We then grid the results to **HEALPix** maps and represent them in terms of the angular and radial power spectrum (section 3.4.2).

3.4.1 From catalogue to maps

The next step is to convert from the flux density information in our catalogue to brightness temperature maps at different frequencies. We first extrapolate our flux density both in intrinsic and polarised intensity to a generic frequency using the following equations:

$$S(\nu) = S_{1.4\ \mathrm{GHz}} \left(\frac{\nu}{1.4\ \mathrm{GHz}}\right)^{\alpha}, \tag{3.4}$$

and

$$P(\nu) = |P| \exp\left(2i(\phi_0 + RM\frac{c^2}{\nu^2})\right), \tag{3.5}$$

where $S_{1.4\ \mathrm{GHz}}$ is the flux density from the catalogue and $|P|$ is the polarised flux density at 1.4 GHz. We choose the frequency range of interest to be (850, 1000) MHz. We convert to units of temperature using the Rayleigh Jeans approximation (Bigot-Sazy et al., 2015),

$$T_{\mathrm{ps}}(\nu, p) = \frac{c^2}{2k_{\mathrm{B}}\nu^2}\frac{1}{\Omega_{\mathrm{pix}}}\sum_{i=1}^{N_p} S_i(\nu). \tag{3.6}$$

Here k_{B} is the Boltzmann constant, S_i is the flux of point source i falling in pixel p and Ω_{pix} is the pixel area.

We choose the **HEALPix** pixelization scheme (Górski et al., 2002). We choose the resolution parameter N_{side} to be 512 which corresponds to a pixel resolution of 0.11 degree as required. We used equation 3.6 to convert Stokes Q, U and $|P|$ from flux density to brightness temperature maps. The bright sources in Figure 3.7 arise from the NVSS catalogue.

Figure 3.8 and 3.9 are the Stokes Q and U of our catalogue at 1400 MHz respectively. Polarisation fraction is given by $\Pi = \left(\sqrt{(T^Q_{\mathrm{ps}})^2 + (T^U_{\mathrm{ps}})^2}\right)/T^I_{\mathrm{ps}}$ and is shown in Figure 3.10. The average polarisation fraction of our extra-galactic source map at 1.4 GHz is 4.5% which is in agreement with Tucci et al. (2004).

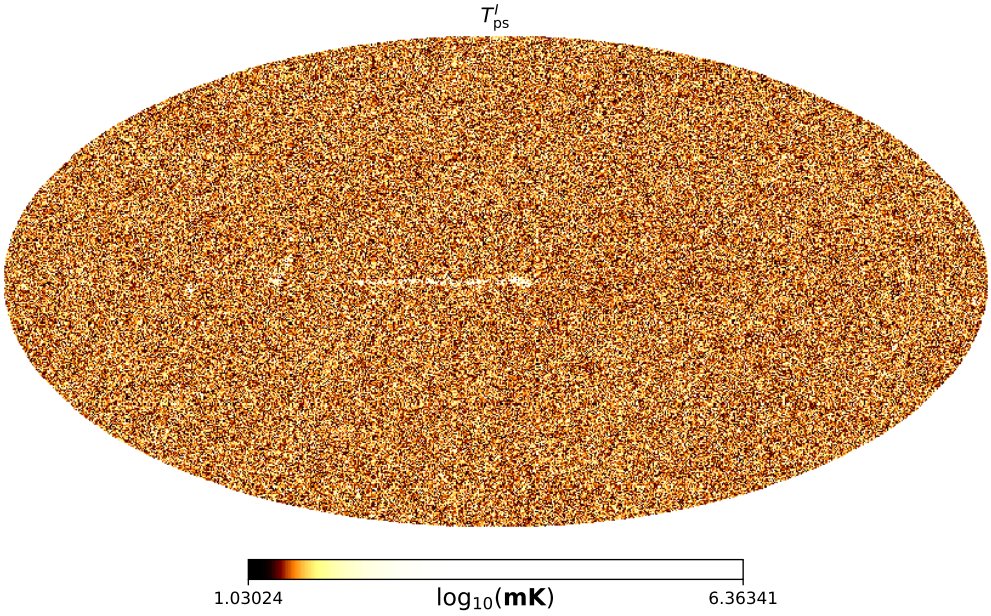

FIGURE 3.7: Point source total intensity map at 1400 MHz. There are bright sources recovered from the NVSS catalogue. The map convention used here is rotated Equatorial-to-galactic coordinates.

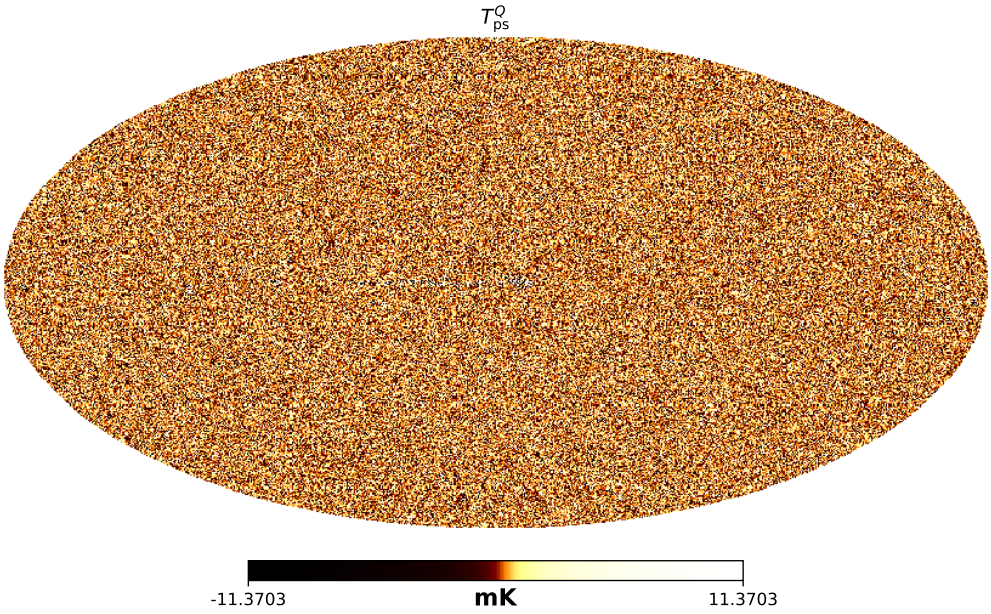

FIGURE 3.8: Point source Stokes Q map at 1400 MHz. The map convention used here is rotated Equatorial-to-galactic coordinates.

3.4.2 Point source analysis

In this section we look at using statistical ways to study the fluctuations of point sources over all scales. We begin by representing point source fluctuations of randomly chosen lines of sight for Stokes Q and U. This is shown in Figure 3.11 and the four random lines of sight suggest that the point source polarisation emission is smooth in frequency. This might be a bit counter-intuitive as one usually expects to see fluctuations in polarisation across frequency due to Faraday rotation. The

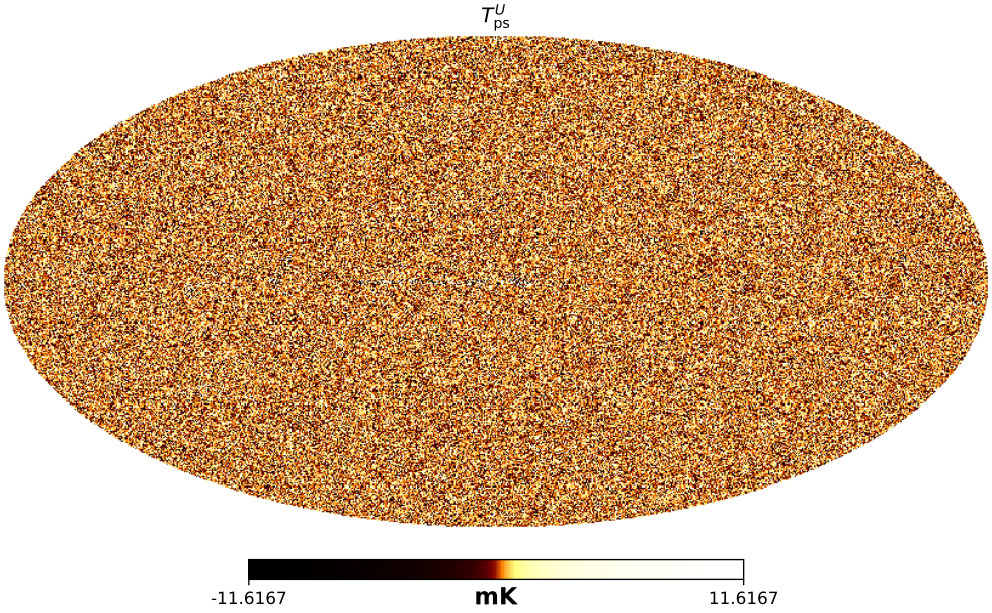

FIGURE 3.9: Point source Stokes U map at 1400 MHz. The map convention used here is rotated Equatorial-to-galactic coordinates.

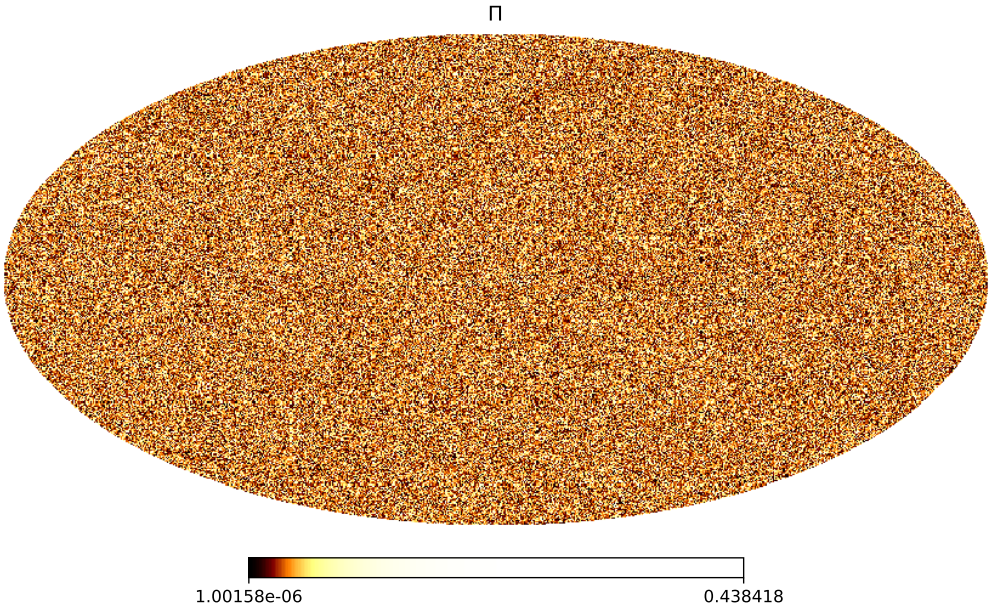

FIGURE 3.10: Polarisation fraction of the point sources at 1400 MHz. The map convention used here is rotated Equatorial-to-galactic coordinates.

reason we do not see that in this case is because the period of the wave associated with the Faraday rotation is typically much larger than the interval of wavelengths we are considering. For instance, the values for point sources Q due to Faraday rotation should follow: $|P|\cos(2\phi_0 + 2RM\lambda^2)$. The period is then $\tau = \pi/RM$. Even the largest RM values in our catalogue (50 rad/m^2) will have τ= 0.063 m^2. The wavelength range we are considering gives $\Delta\lambda^2$ = 0.034 m^2 that is a factor two smaller than the smallest period we can get. Therefore we only see a smooth curve

for all cases except for those lines of sight that contain point sources with very large RM, where a change in sign in the curve is possible. The **H**I signal (denoted by $T_{\mathbf{HI}}$) was included for further emphasis on how these point sources behave as a function of frequency. This is shown in Figure 3.12 and 3.13 for Stokes Q and U respectively.

We construct angular and radial power spectrum of the point source maps (using the methods described in section 2.6). In this exercise, there's a **HEALPix** function that gives us the spherical harmonic of the maps and calculates the angular power spectrum (equation 2.15). Figure 3.14 shows the angular power spectrum for the total intensity at 1.4 GHz. Figure 3.15 shows the angular power spectrum of the point source types of emission at 1.4 GHz. Figure 3.16 shows the angular power spectrum for point source Stokes parameters at 950 MHz compared to the **H**I signal. The power spectrum shown in this section is a full sky point source catalogue. The point source power spectra are reasonably flat because it is essentially dominated by the Poisson distribution. The source of clustering, in particular for small ℓ comes from the NVSS and SUMSS sources. In any case, for the propose of testing the foreground contamination, this approach should be enough (and includes the stronger sources on the sky), since it is the structure in frequency that matters the most.

The overall amplitude of the angular power spectrum seems to be higher in figure 3.16 (10^{-3} to 10^{-2} mK^2) as compared to the values at 1.4 GHz (10^{-4} to 10^{-3} mK^2) shown in Figure 3.15. This is expected because the point source emission is higher at a lower frequency. **H**I evolves with angular scales compared to Stokes Q and U that have flatter spectrum (see Figure 3.16) which is due to our Poisson distribution construction as explained. Figure 3.17 shows how the angular power spectrum for point sources compares to other foregrounds.

Figure 3.18 shows the radial power spectrum for Stokes Q and U for the range (860, 990) MHz. For this analysis, we left out the first and last 10 channels on the edges of our frequency range ($i.e.$ 10 MHz subtraction on each edge) to avoid any boundary effects. The explanation of how the 3D radial power spectrum is calculated can be found in section 2.6. The spectrally smooth component pile up at low $k_{\parallel}$ this is why the amplitude of the point sources is high at low $k_{\parallel}$ and a gradually decreases with increasing $k_{\parallel}$ as seen in Figure 3.18. The rotation measure introduces structure to the point sources radial power spectrum. The gradual decrease overall scales of $k_{\parallel}$ also prove that point source polarisation is a smooth function of frequency. This should make it easy for the foreground cleaning methods to eliminate this signal from the total intensity maps.

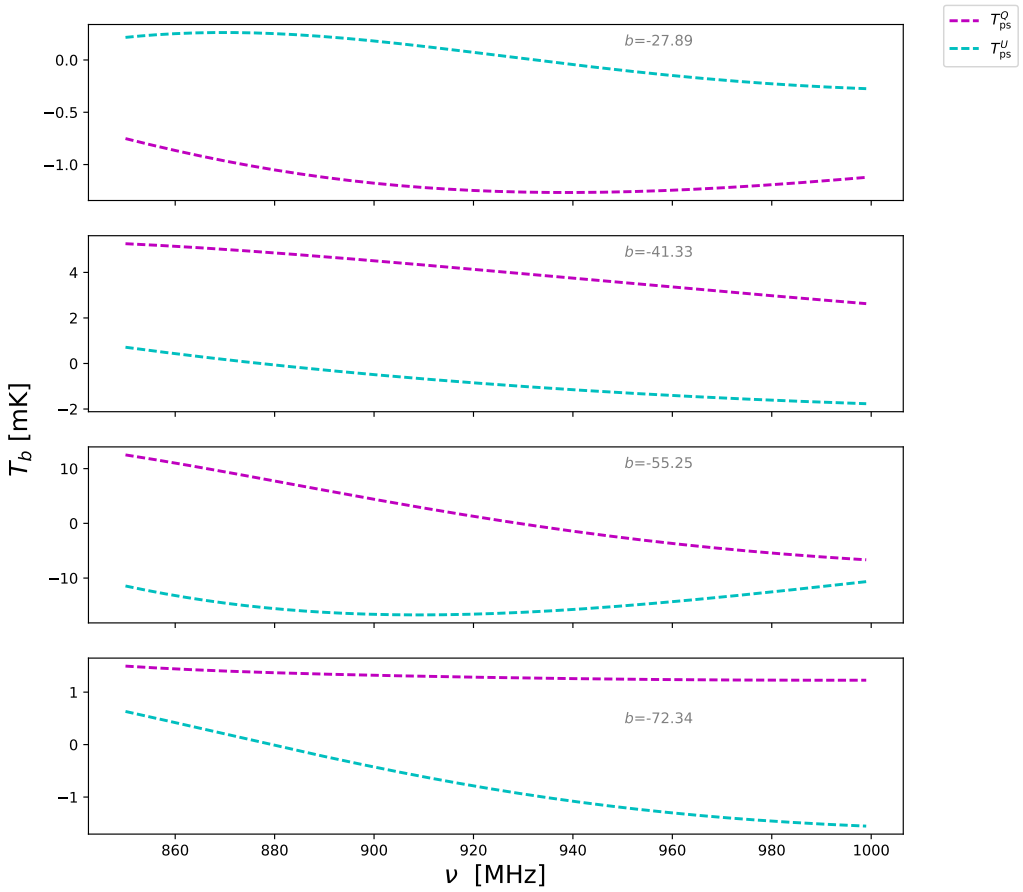

FIGURE 3.11: The randomly chosen lines of sight for pixels falling in our patch. Stokes Q (in magenta) and U (in cyan) for our frequency range. The lines of sight are chosen in terms of the galactic latitude parameter, b and these lines of sight fall within our simulated region.

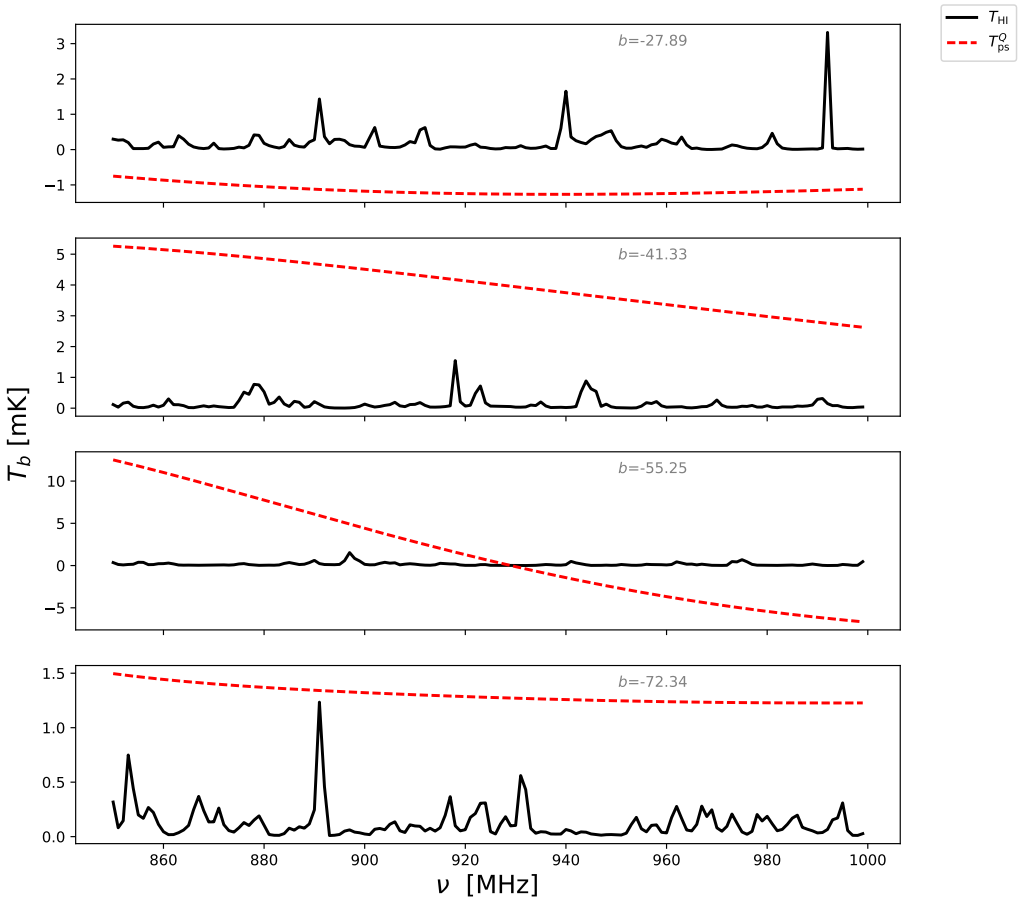

FIGURE 3.12: Line of sight are given in terms of galactic latitude parameter, b for Stokes Q (in red) compared to **H**I (in black).

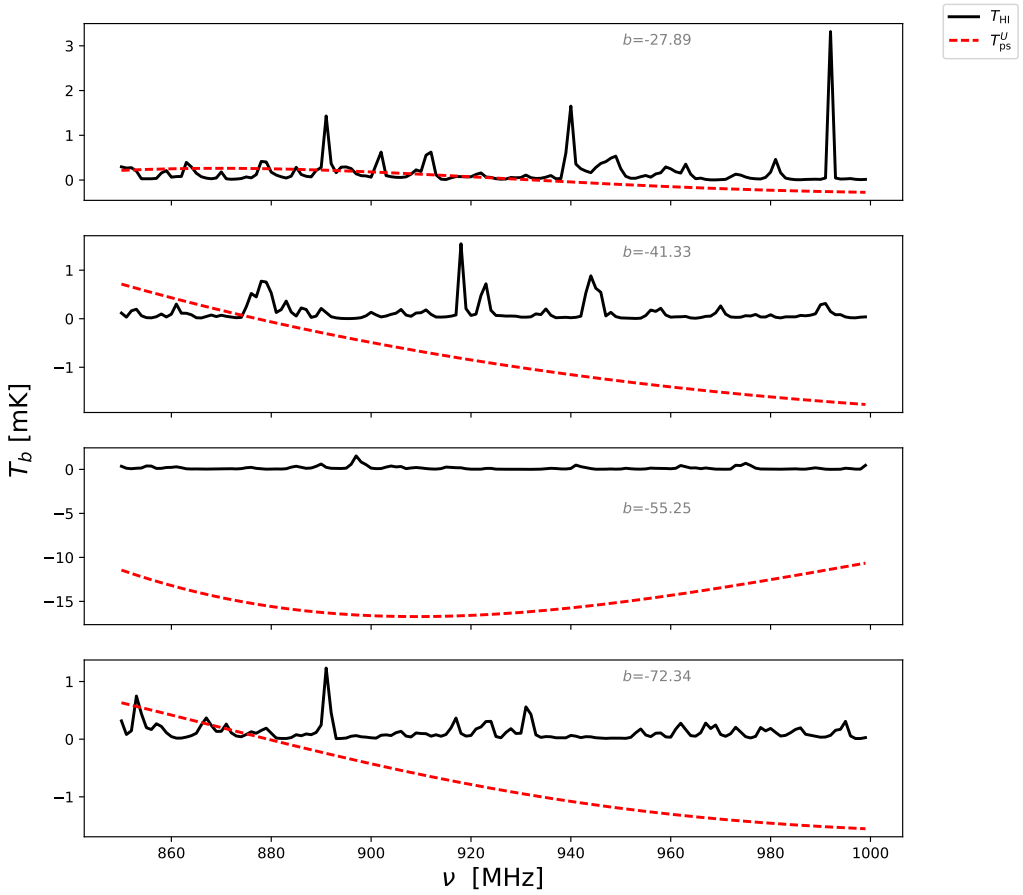

FIGURE 3.13: Line of sight are given in terms of galactic latitude parameter, b for Stokes U (in red) compared to **H**I (in black) chosen from our simulated region.

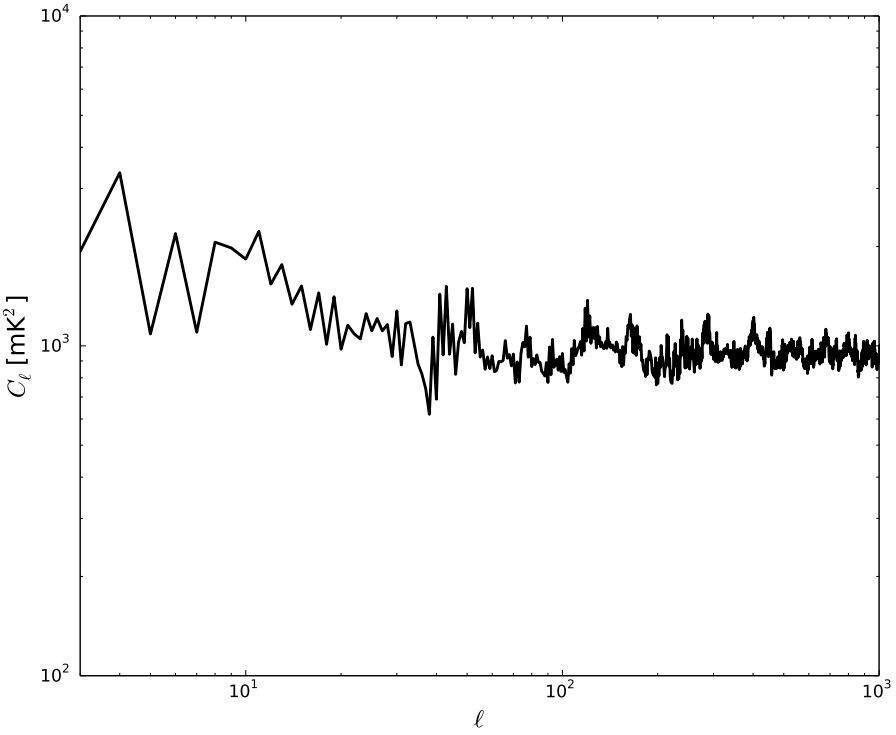

FIGURE 3.14: The angular power spectrum of point source total intensity at 1400 MHz. Note that the only source of clustering in our simulation is coming from the NVSS and SUMSS sources, since for the rest we assumed a Poisson distribution

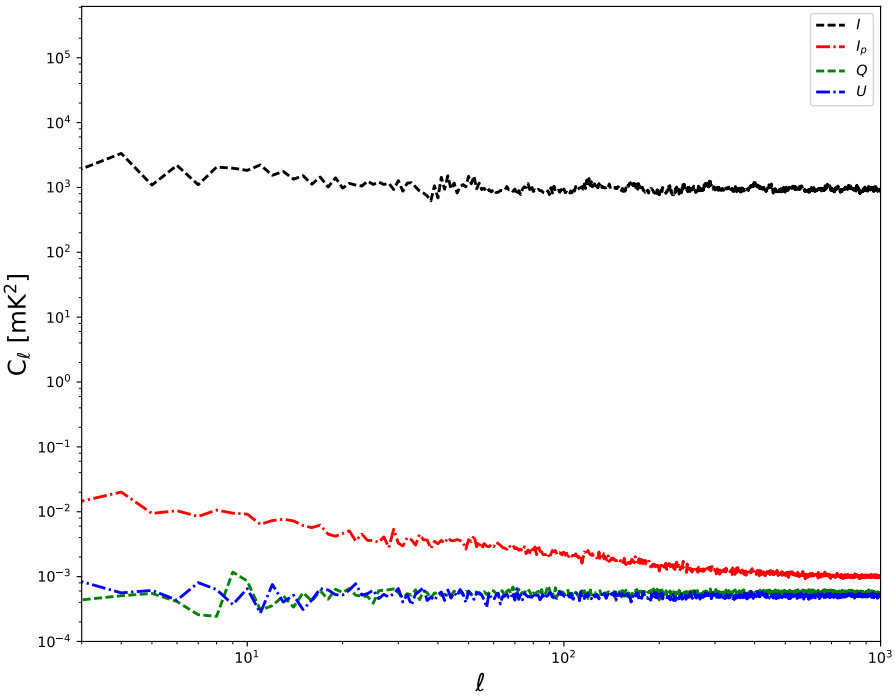

FIGURE 3.15: The angular power spectrum of the total intensity, I (in black), I_p (in red), Stokes Q (in green) and U (in blue) for the point sources at 1400 MHz.

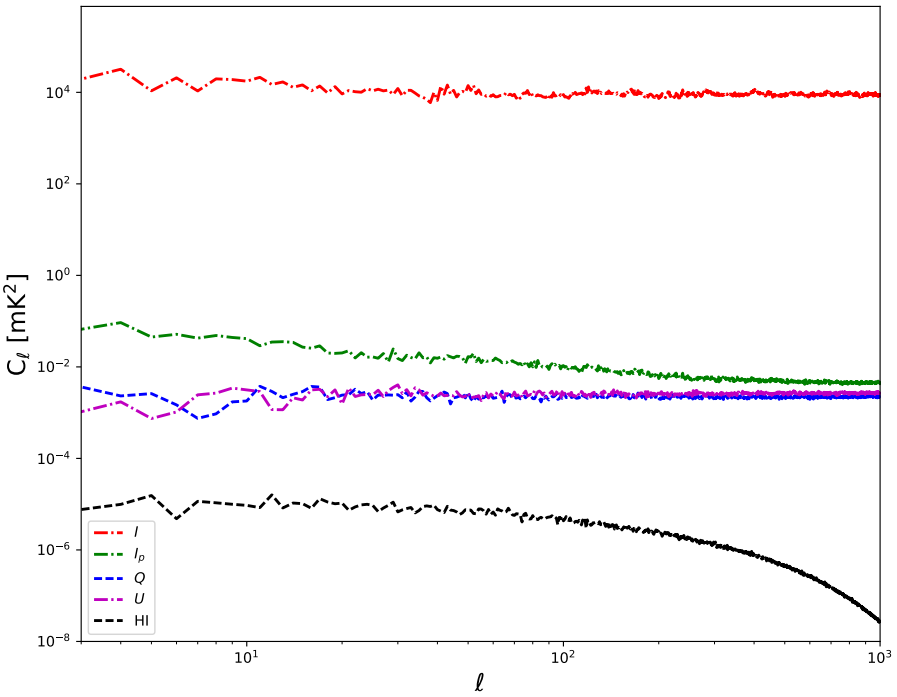

FIGURE 3.16: Angular power spectrum of the total intensity I (in red), Stokes I_p (in green), Q (in blue), U (in magenta) and **H**I (in black) at 950 MHz.

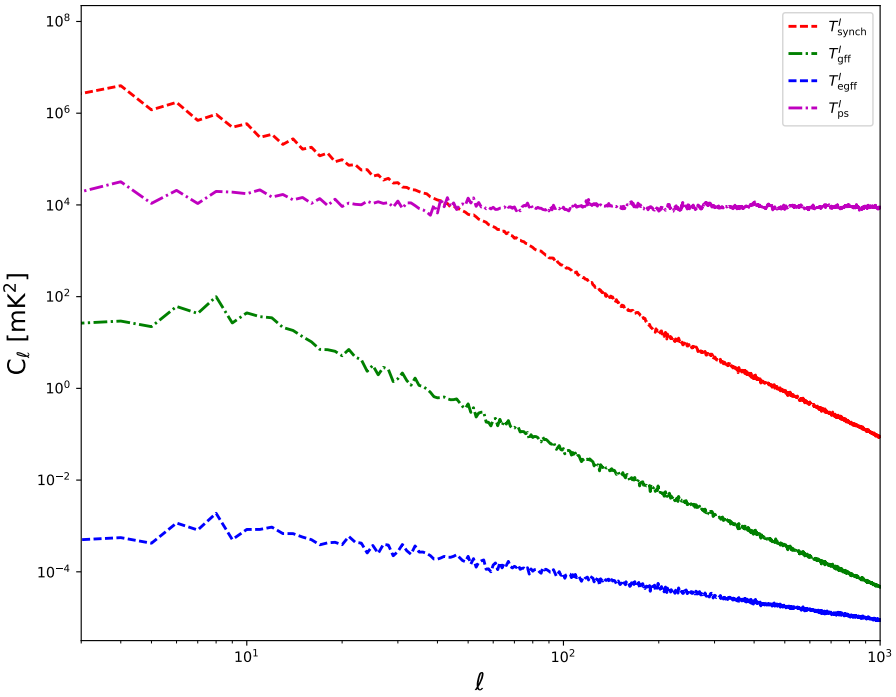

FIGURE 3.17: Angular power spectrum of the foregrounds simulated at 950 MHz. The galactic synchrotron (*red*) dominates over large scales, $\ell \lesssim 50$ compared to the galactic free-free (*green*), extra-galactic free-free (*blue*) and point source (*magenta*) emission. The point sources dominate on small scales $\ell \gtrsim 50$ for this frequency slice shown.

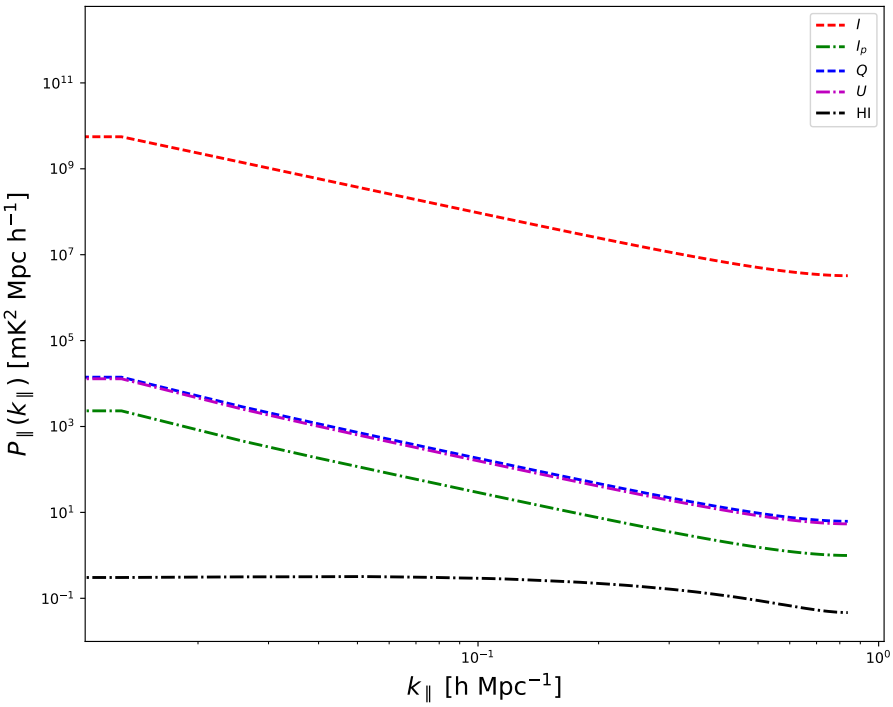

FIGURE 3.18: Radial power spectrum of Stokes I (in red), I_p (in green), Q (in blue), U (in magenta).The radial power of the cosmological signal is shown in black.

Chapter 4

Foreground cleaning

In this chapter, we go through the foreground separation techniques in detail (in section 4.1). We start by explaining the polynomial fitting (section 4.1.1), followed by the Principal Component Analysis (section 4.1.2) and lastly Independent Component Analysis (section 4.1.3). We apply the foreground cleaning methods to two different cases of simulated sky maps, case 1: the total intensity has all foregrounds except polarised galactic synchrotron and case 2: has the polarised galactic synchrotron as a foreground contributing to the total intensity. We include our point source maps, described in Chapter 3, in one of the case studied (section 4.2). We consider a MeerKAT-like experiment as described in Chapter 2 (in section 2.5). We then present the foreground cleaned results in terms of maps, the line of sight for brightness temperature as a function of frequency and power spectrum (2D and 3D).

4.1 Foreground cleaning techniques

In this section, we look at the mathematical background of the methods implemented in the companion code **fg_rm** of the **CRIME** simulation. The methods presented here are useful to a wide spectrum of science from pattern recognition, face recognition, and image compression etc. These methods are called blind and non-blind cleaning methods. The cleaning methods explained in this section were extracted from CMB experiments (Tegmark, 1997; Dupac and Giard, 2002; Crill et al., 2003; Keihänen et al., 2005; Patanchon et al., 2008; Alonso et al., 2015). The three models tested for foreground removal are Polynomial fitting (also known as Line of sight fitting, LOS), Independent Component Analysis (ICA) and the Principal Component Analysis (PCA). The basic mathematical assumption behind these foreground cleaning methods is that the total signal can be described by the following equation,

$$T(\nu, \hat{\mathbf{n}}) = \sum_{k=1}^{N_{\rm fg}} f_k(\nu) S_k(\hat{\mathbf{n}}) + T_{\rm cosmo}(\nu, \hat{\mathbf{n}}) + T_{\rm noise}(\nu, \hat{\mathbf{n}}), \tag{4.1}$$

where the total measured intensity is represented by $T(\nu, \hat{\mathbf{n}})$. The components describing the foregrounds are given by $S_k(\hat{\mathbf{n}})$. $f_k(\nu)$ represents the set of smooth functions dependent on frequency and lastly, $T_{\rm cosmo}(\nu, \hat{\mathbf{n}})$ and $T_{\rm noise}(\nu, \hat{\mathbf{n}})$ represent the cosmological signal and noise respectively. In other words the terms under a summation represent the foreground maps. For a set of N_ν channels using a line of sight $\hat{\mathbf{n}}$, equation 4.1 can be written in matrix form,

$$\mathbf{x} = \hat{\mathbf{A}} \cdot \mathbf{s} + \mathbf{r}, \tag{4.2}$$

where $\mathbf{r} = T_{\rm cosmo}(\nu, \hat{\mathbf{n}}) + T_{\rm noise}(\nu, \hat{\mathbf{n}})$, $\mathbf{x} = T(\nu, \hat{\mathbf{n}})$, $\hat{\mathbf{A}} = f_k(\nu)$ and $\mathbf{s} = S_k(\hat{\mathbf{n}})$. The term $\mathbf{x}$, represents the total radiation received from a direction $\hat{\mathbf{n}}$ and at frequency ν. The main goal behind the blind separation problem is to determine the component mixing matrix $\hat{\mathbf{A}}$ so that the term, $\mathbf{r} = \mathbf{x} - \hat{\mathbf{A}} \cdot \mathbf{s}$ is recovered as accurately as possible. The next following subsections look at how the three models determine the mixing matrix and recover $\mathbf{r}$. For these foreground methods there is an associated degrees of freedom parameter, $N_{\rm fg}$, which gives the number of modes to be removed.

4.1.1 Line Of Sight fitting

This technique is the most intuitive because it chooses a basis of smooth functions (a set of polynomials) that can describe the foregrounds. Using least squares fitting on equation 4.2 we minimize the χ^2 with respect to s, *i.e.* $\frac{\partial \chi^2}{\partial s}$=0 where;

$$\chi^2 = (\mathbf{x} - \hat{\mathbf{A}} \cdot \mathbf{s})^T \hat{\mathbf{N}}^{-1} (\mathbf{x} - \hat{\mathbf{A}} \cdot \mathbf{s}), \tag{4.3}$$

where $\hat{\mathbf{N}}$ is the covariance matrix of the cosmological signal plus the noise. The solution to minimizing equation 4.3:

$$\hat{\mathbf{s}} = (\hat{\mathbf{A}}^T \hat{\mathbf{N}}^{-1} \hat{\mathbf{A}})^{-1} \hat{\mathbf{A}}^T \hat{\mathbf{N}}^{-1} \mathbf{x}. \tag{4.4}$$

We assume that the noise is uncorrelated in frequency thus N_{ij} = diag(σ_i^2), where σ_i contains both contribution from the noise and cosmological signal. Note that the derivation of this latter is not trivial since it is linked to the solution we are looking for. Nevertheless it can be approximately known through simulation or multi-steps foreground cleaning, starting with only noise in the covariance matrix. In the analysis section of this thesis, we denote the recovered HI signal from using this method as "LOS".

4.1.2 Principal Component Analysis

PCA was invented by Karl Pearson (see Pearson (1901) and Hotelling (1933)). The modern description of PCA can be found in Jolliffe (2002). The aim of this technique is to reduce dimensionality in multivariate data while making sure the information is preserved. In the transformed basis, the principal components contain almost all the variation present in the data. This is useful in foreground cleaning since the foregrounds and the signal are correlated at different levels in frequency. PCA requires no specific prior information on the spectra of the foreground and the noise. In general, PCA is a way of identifying patterns in data and highlights the differences and similarities. It is mostly used in pattern recognition, image processing and in Machine Learning algorithms. To understand the PCA algorithm, we start with basic linear algebra. An eigenvector of a linear transformation $\mathbf{A}$ is a non-zero vector that satisfies

$$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}, \tag{4.5}$$

where λ is the eigenvalue corresponding to the eigenvector $\mathbf{v}$. To calculate the eigenvalue, we need to compute the characteristic polynomial. The eigenvalues of $\mathbf{C}$ will be the values of λ that satisfy the equation $|\mathbf{C} - \lambda \mathbf{I}| = 0$. We can define a matrix $\mathbf{U}$ that contains the linear independent eigenvectors of $\mathbf{C}$. This $\mathbf{U}$ matrix will diagonalize $\mathbf{C}$ *i.e.* $\mathbf{U}^T \mathbf{C} \mathbf{U} = \mathbf{\Lambda}$ where $\Lambda_{ij} = \lambda_i \delta_{ij}$.

Let's imagine that $\mathbf{C}$ is a correlation matrix $N_\nu \times N_p$ where N_p is the number of pixels. If it describes a system very correlated its entries will be all be close to 1. In the extreme case of complete correlation every entry will be 1. If we compute the eigenvalues we will find one eigenvalue which is $\lambda_1 = N_\nu$ and all the other eigenvalues will be zero. This means that information about the correlation is contained in the first eigenvector $\mathbf{v_1}$ corresponding to λ_1. In our case of foreground subtraction, we can try to subtract the foregrounds using the fact that the information about them will be contained in the eigenvectors corresponding to the largest eigenvalues. Indeed this is because the foregrounds are strongly correlated in frequency while the **H**I signal is mostly uncorrelated.

Let's assume we have N_ν frequencies. We can compute the frequency covariance matrix of our problem as,

$$\Sigma_{ij} = \frac{1}{N_p}\mathbf{x}\mathbf{x}^T = \frac{1}{N_p}\sum_{p=1}^{N_p} T(\nu_i, \hat{\mathbf{n}}_p)T(\nu_j, \hat{\mathbf{n}}_p), \tag{4.6}$$

where $T(\nu_i, \hat{\mathbf{n}}_p)$ is the brightness temperature for a frequency channel ν_i with the $i = 1, \cdots, N_\nu$ and line of sight $\hat{\mathbf{n}}_p$ where $p = 1, \cdots, N_p$. The correlation coefficient matrix can be obtained by

$$C_{ij} = \frac{\Sigma_{ij}}{\Sigma_{ii}^{1/2}\Sigma_{jj}^{1/2}}. \tag{4.7}$$

We can diagonalize the matrix $\hat{\mathbf{C}}$ by

$$\mathbf{U}^T\hat{\mathbf{C}}\mathbf{U} = \hat{\mathbf{\Lambda}} \equiv \text{diag}(\lambda_1, \cdots, \lambda_{N_\nu}), \tag{4.8}$$

where we have ordered the eigenvalues from large to small *i.e.* $\lambda_i > \lambda_{i+1}\ \forall i$. We need to choose the number of eigenvalues that are large and so corresponds to the foregrounds. Once we have guessed the number of foregrounds, we can build the matrix $\mathbf{U}_{\text{fg}}$ from the columns of $\mathbf{U}$ corresponding to these eigenvalues. We then reconstruct the brightness temperature model as

$$\mathbf{x} = \mathbf{U}_{\text{fg}}\mathbf{s} + \mathbf{r}, \tag{4.9}$$

which has the same form as equation 4.2. To get the foreground maps we project $\mathbf{x}$ on the eigenvector basis of $\hat{\mathbf{C}}$,

$$\mathbf{s} = \mathbf{U}_{\text{fg}}^T\mathbf{x}. \tag{4.10}$$

To include the covariance of the noise $\mathbf{N}^{-1}$, requires us to apply weights to the brightness temperature in equation 4.6, *i.e.* $T(\nu_i, \hat{\mathbf{n}}_p)/\sigma_i$ where the standard deviation $\sigma_i = \sqrt{\sigma^2_{\text{noise},i} + \sigma^2_{\text{cosmo},i}}$ is combination cosmological signal and noise. This means before projecting $\mathbf{x}$, it is multiplied by σ_i which introduces $\mathbf{N}^{-1}$ thus making equation 4.10 to be equivalent to equation 4.4. The last step is to compare the reconstructed foreground maps to the original input maps which will leave behind noisier cosmological maps.

4.1.3 Independent Component Analysis

We present a second blind separation technique that does not require any prior knowledge about the foregrounds and the mixing matrix being solved. The

following derivation follows from Hyvärinen and Oja (2000) and Alonso et al. (2015). If we ignore the noise+signal term (**n**), equation 4.2 can be written as

$$\mathbf{x} = \mathbf{A}\mathbf{s}. \tag{4.11}$$

As aforementioned we have no prior information on **A** and **s**. Fortunately, it turns out if we assume a statistical property on **s** then it makes it possible to solve the problem more easily. ICA assumes the independence of **s**, in equation 4.2 *i.e.* in terms of probability theory P(X=**s**) = $\Pi_i P(X_i = s_i)$. X_i represents our random variable. We recall that **s** represents the foreground components. Equation 4.11 can be formally inverted,

$$\mathbf{s} = \hat{\mathbf{W}}\mathbf{x}. \tag{4.12}$$

Recall, we want to make sure that **s** is independent thus we are maximizing the non-Gaussianity of the matrix $\hat{\mathbf{W}}$. Using y to represent one component of equation 4.12, we have,

$$y = w_{11}x_1 + w_{12}x_2 + w_{13}x_3 + \cdots + w_{1k}x_k. \tag{4.13}$$

In other words y is just a linear combination of some s_i. If we maximize the non-Gaussianity, we are imposing statistical independence, this is the key assumption of ICA. The Central Limit Theorem assures that the sum of linearly independent variable will be more Gaussian than the single distribution for one or two variables. This means that if we add more terms like y, the system will tend to be more Gaussian *i.e.* if more components of **s** are mixed. Then we need to make sure there exists at least one component that is non-Gaussian otherwise this technique will not work. Now, the non-Gaussianity of y (equation 4.13) can be measured in two ways, firstly using Kurtosis or Negentropy. The statistical technique of kurtosis can be used as a measure of non-Gaussianity of a random variable. This means finding the fourth moment of that particular random variable. Kurtosis is defined as,

$$\text{kurt}(y) = E(y^4) - 3(E(y^2))^2, \tag{4.14}$$

where E represents the expectation (or average) of y. Kurtosis is not used much because of the fourth moment in the first expectation of equation 4.14, it is computational expensive for ICA experiments. The second measure of non-Gaussianity as aforementioned is given by negentropy. Negentropy is given by $J(y) = H(y_{\text{Gauss}}) - H(y)$ where $H(y)$ is the entropy of our variable y and y_{Gauss} is a random Gaussian variable with unit variance..

$$H(Y) = -\sum_i P(Y = a_i)\log P(Y = a_i). \tag{4.15}$$

The entropy is maximal for all Gaussian variables with the same variance making the negentropy, $J(y)$, always positive for non-Gaussian variables. Since the information on the probability distribution is usually not available, some algorithm like FastICA (Hyvärinen, 1999) use an approximation to J given by,

$$J(y) \sim \sum_i k_i \left[\langle G_i(y)\rangle_\theta - \langle G_i(y_{\text{Gauss}})\rangle_\theta\right], \tag{4.16}$$

where k_i are positive constants, $\langle . \rangle_\theta$ denotes averaging over all pixels and G_i is a set of non-quadratic equations functions. The main functions that FastICA uses are;

$$G(y) = \exp(-y^2/2), \quad G(y) = \frac{1}{a}\text{logcosh}(ay), 1 \leq a \leq 2. \tag{4.17}$$

4.2 Foreground cleaning results

In this section, we apply the cleaning techniques described in section 4.1 to the maps generated as in section 3.4.1. The foreground cleaning provides two sets of results, the recovered maps, and the recovered power spectrum. We consider two cases. Case 1 comprises of all foregrounds except the polarised synchrotron and case 2 has only the polarised synchrotron emission as the astrophysical foreground. In case 1 the total intensity used as input to the foreground cleaning code is given by,

$$T = T_{\mathbf{HI}} + T^{I}_{\rm ps} + \epsilon_p T^{Q}_{\rm ps} + T^{I}_{\rm synch} + T^{I}_{\rm Gff} + T^{I}_{\rm EGff} + T_{\rm noise}, \tag{4.18}$$

where $T^{I}_{\rm ps}, T^{I}_{\rm synch}, T^{I}_{\rm Gff}, T^{I}_{\rm EGff}$ denote the point sources from our catalogue, the galactic synchrotron, galactic and extra-galactic free-free emission respectively. $T^{Q}_{\rm ps}$ denotes the Stokes Q emission for point sources of our catalogue and ϵ_p is the leakage fraction term. We use a polarisation leakage of 1% which is in agreement with Jelić et al. (2010) and Alonso et al. (2014). The cosmological signal in the equation is denoted by $T_{\mathbf{HI}}$. The noise, denoted by $T_{\rm noise}$ in equation 4.18, was added to account for the instrumental effects thus making the simulations more realistic. As described in Chapter 2, we consider a MeerKAT-like experiment, meaning the results shown in this section are for the sky patch chosen. In case 2 we only considered the polarised synchrotron emission as aforementioned. The total intensity is thus given by,

$$T = T_{\mathbf{HI}} + \epsilon_p T^{Q}_{\rm synch} + T_{\rm noise}, \tag{4.19}$$

where $T^{Q}_{\rm synch}$ denotes Stokes Q emission for the galactic synchrotron. Our way of taking into account polarisation leakage is simplified since the Mueller matrix approach was not followed (Jelić et al., 2010). The polarisation considered for both cases is from Stokes Q. For our catalogue polarisation from Stokes, U behaves similarly to Stokes, Q. The cleaning routine **fg_rm** provides the three cleaning methods explained in section 4.1: LOS, PCA, and ICA. The parameter that we can tune in the foreground cleaning is the number of degrees of freedom to be removed $N_{\rm fg}$.

4.2.1 Case 1

In this case, we are considering all the foregrounds except the polarised galactic synchrotron (see equation 4.18) and all three foreground subtraction techniques were applied to this case with the number of degrees of freedom to be removed, $N_{\rm fg}$ of 5, 6 and 7. Figure 4.1 shows the recovered **HI** maps for ICA, PCA and LOS three different values $N_{\rm fg}$. ICA and PCA do a good job when recovering **HI** signal even for $N_{\rm fg} = 5$. The LOS method does badly in recovering the cosmological signal especially for lower $N_{\rm fg}$ values but start to converge once 7 modes are removed. To investigate further the behaviour of the foreground cleaning algorithm we show in Figure 4.2 four random lines of sight. The **CRIME** simulation does include the **HI** average $\bar{T}_{\mathbf{HI}}$. Since the **HI** average is smooth in frequency it will be removed by the foreground cleaning methods. The recovered signal contains the **HI** fluctuations and the noise. This allows us to compare $T^{\rm rec}_{\mathbf{HI}} - T_{\rm noise}$ to $T_{\mathbf{HI}} - \bar{T}_{\mathbf{HI}}$ where $T^{\rm rec}_{\mathbf{HI}}$ is the recovered signal from either LOS, PCA or ICA. The polynomial fitting method does badly on the frequency edges while PCA and ICA recover some of the cosmological signals on most of the frequency range. For case 1 PCA is the best method to use. Figure 4.3 shows this in more detail. Figure 4.4 is similar to 4.3 but we have

included the input noise realization along each chosen line of sight to infer how much the noise level can affect the cleaning method.

To investigate further we show the angular power spectrum for case 1 in Figure 4.5. Even in this case, the LOS method is clearly behaving poorly in recovering the **H**I signal for low $N_{\rm fg}$ values. On small scales, there is a discrepancy between the input **H**I and the recovered **H**I signal that can be accounted for by noise. To see this better we plot in Figure 4.6 the angular power spectrum with the noise angular power spectrum subtracted. For PCA and ICA, the angular power spectrum now overlaps at high ℓ with the input **H**I signal. For the LOS method on small scales, the reconstruction gets better for higher number of modes subtracted but for $N_{\rm fg} = 5$ is ineffective.

We can also investigate the radial power spectrum of the recovered **H**I signal. We can see that the same conclusion results for the radial power spectrum. This can be seen in Figure 4.7 and 4.8. Note that we cut 10 MHz from our frequency edges meaning the calculation of $P_{\parallel}(k_{\parallel})$ was done for (860, 990) MHz to avoid boundary effects that introduced by the foreground cleaning. In Figure 4.8 we clearly see that at large scales (low $k_{\parallel}$) the smooth components have been eliminated since there is no power on those scales. Some radial power from the **H**I was depleted as well. Again we conclude that at large scales ($i.e.$ small $k_{\parallel}$) the recovered **H**I fails to match the input **H**I signal while on small scales ($i.e.$ large $k_{\parallel}$) the noise dominates over the input **H**I.

In conclusion, when including all the expected foregrounds, except for the polarised galactic synchrotron leakage, the foreground cleaning methods work quite well. This is valid even when we include the polarised leakage from point sources (even strong ones). As explained before, this is mainly due to the fact that such polarisation is reasonably smooth in frequency at the frequencies we are considering here.

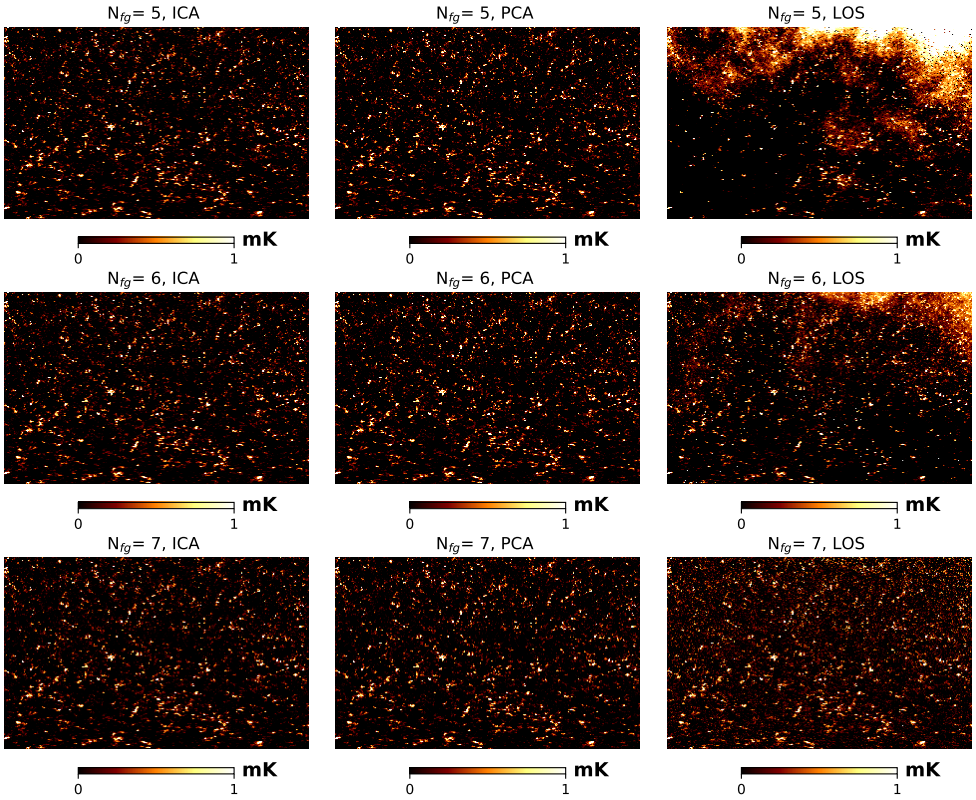

FIGURE 4.1: The recovered **H**I signal maps at 950 MHz for different foreground cleaning methods ICA, PCA, and LOS. Different rows correspond to $N_{\rm fg}$= 5, 6 and 7. In these maps the noise contribution has been subtracted.

4.2.2 Case 2

In this case, we only considered polarised galactic synchrotron (Stokes Q) as a foreground contaminant to the total intensity (see equation 4.19). In the light of the result of the previous case, we decide to use PCA since was the best behaving foreground cleaning technique. We start by using $N_{\rm fg} = 7$ for case 2 but it does not succeed in recovering the input **H**I signal. This can be seen in Figure 4.9 where we show four random lines of sight comparing the recovered **H**I signal to the input **H**I signal. Indeed there is a lot of contamination from foreground emission left in the recovered signal. Even using the angular power spectrum tool it is clear that removing 7 number of modes fails in reconstructing the **H**I signal (see Figure 4.10). Indeed case 2 requires a larger number of modes to be subtracted rather than stopping at $N_{\rm fg}$=7 because of the complex frequency behaviour of the polarised galactic synchrotron. The subtraction is so difficult that the reconstructed **H**I signal looks like the input polarised galactic synchrotron leakage.

Values of $N_{\rm fg}$ below 60 do poorly in recovering the **H**I signal for this case. With this conclusion, we try to strongly increase the number of degrees of freedom to be more aggressive in the removal of foregrounds. We show in Figure 4.11 the recovered **H**I signal for $N_{\rm fg}$=7, 63, 77, 84, 91, 100. In map space, there is clearly structure left behind in the recovered map from removing $N_{\rm fg}$=7 while at other $N_{\rm fg}$ values the maps look different compared to the case $N_{\rm fg}$=7 while we expect random **H**I fluctuations as in Figure 4.1. We show this in Figure 4.12, 4.13, 4.14 and

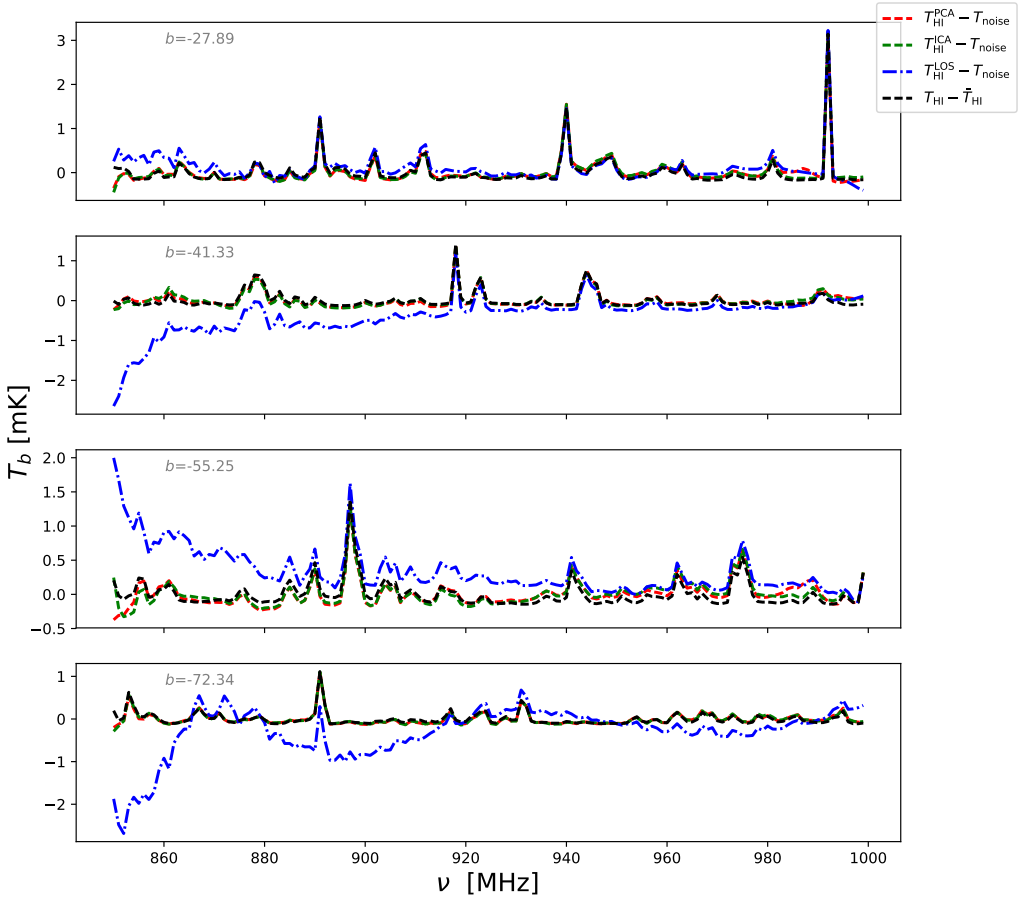

FIGURE 4.2: Four different lines of sight of the recovered **H**I signal for case 1 with $N_{fg} = 7$. PCA is shown in red dash, ICA in dashed green and LOS in dashed blue while the input signal is in dashed black. The lines of sight are chosen in terms of the galactic latitude parameter, b and these lines of sight fall within our simulated region.

Figure 4.15 in terms of lines of sight plots for N_{fg}=63, 77, 84, 91, 100 respectively. Unfortunately removing a high number of foreground modes leads to depletion of the **H**I signal and removing less leaves the bothersome foregrounds behind. Using lines of sight as a Figure of Merit does not give robustness in explaining the recovered **H**I signal. We then turn to using statistical methods to measure the overall structure of the **H**I signal. This is why the angular and radial power spectrum was used for further investigation. These tools can tell us about the overall distribution and amplitude of the signal in terms of frequency.

We consider the power spectrum for the recovered signal from case 2. Removing the noise power spectrum (both angular and radial) is not ideal because the recovered signal suffered a lot of depletion due to over-cleaning. The angular power spectrum for case 2 is shown in Figure 4.17 and 4.18. The latter has noise angular power spectrum subtracted.

The radial power spectrum of the recovered **H**I signal from the set of N_{fg} values in Figure 4.19. The smooth components (at large scales) along with **H**I signal were greatly depleted at low $k_{\parallel}$ as we increase the number of modes to be removed as shown in Figure 4.19. On small scales (high $k_{\parallel}$) the noise power spectrum

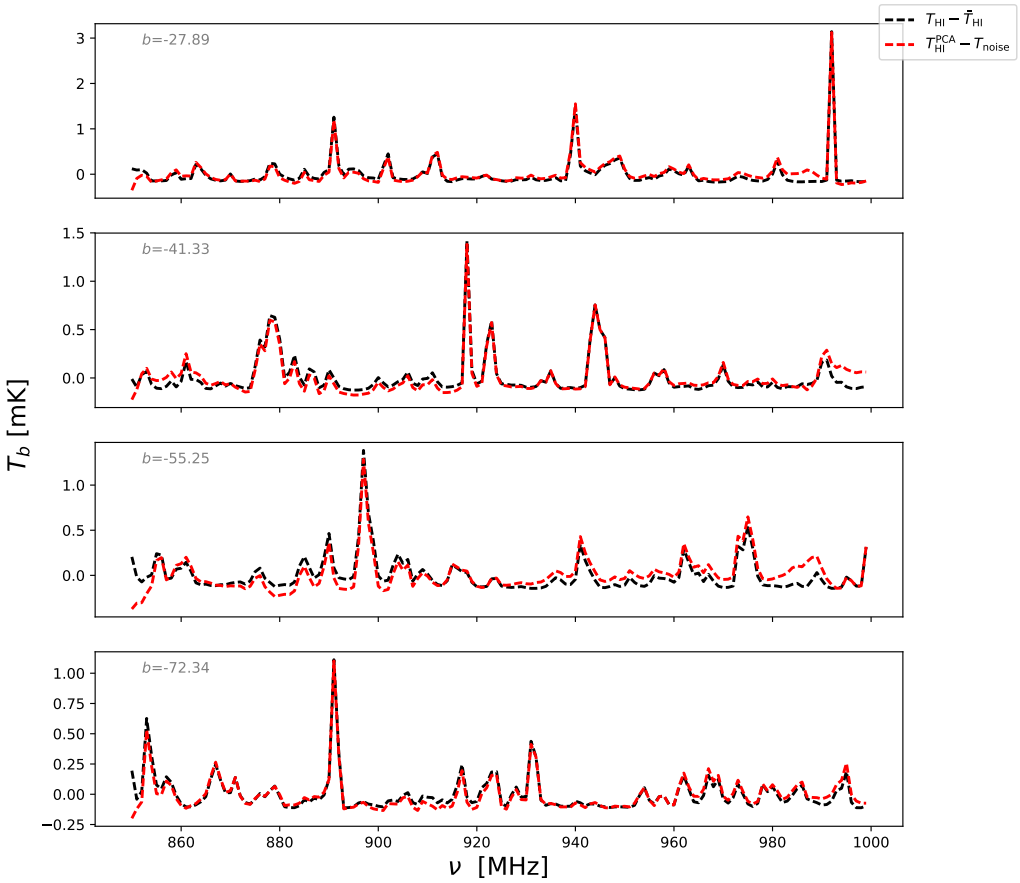

FIGURE 4.3: The recovered **H**I signal (red-dashed) from the PCA method compared to input **H**I (black-dashed). As aforementioned the number of modes that were removed is 7. The lines of sight are chosen in terms of the galactic latitude parameter, b and these lines of sight fall within our simulated region.

dominates the recovered **H**I signal. Some of the input noise power spectrum is also removed since it should be present on small scales (large $k_{\parallel}$). This is due to over-cleaning. What we can conclude from case 2 is that an uncalibrated leakage at 1 % cannot be cleaned using blind foreground cleaning methods. If galactic Stokes Q or U were known then we could subtract this contamination. This is an example of template fitting because this would imply removing these foregrounds from the signal before imposing the foreground separation methods.

We can see that at the 1% level, the polarisation leakage is well above the **H**I signal. Moreover, it seems current methods such as PCA are unable to properly clean this contamination. The solution would be to make sure the contamination is below the **H**I signal, but this will require leakage terms of the order of 0.01% according to our simulations (see figures 4.20, 4.21, 4.22 and 4.23). This will be a crucial problem for **H**I intensity mapping and other cleaning methods should be explored in order to see if the polarisation contamination can be removed more efficiently.

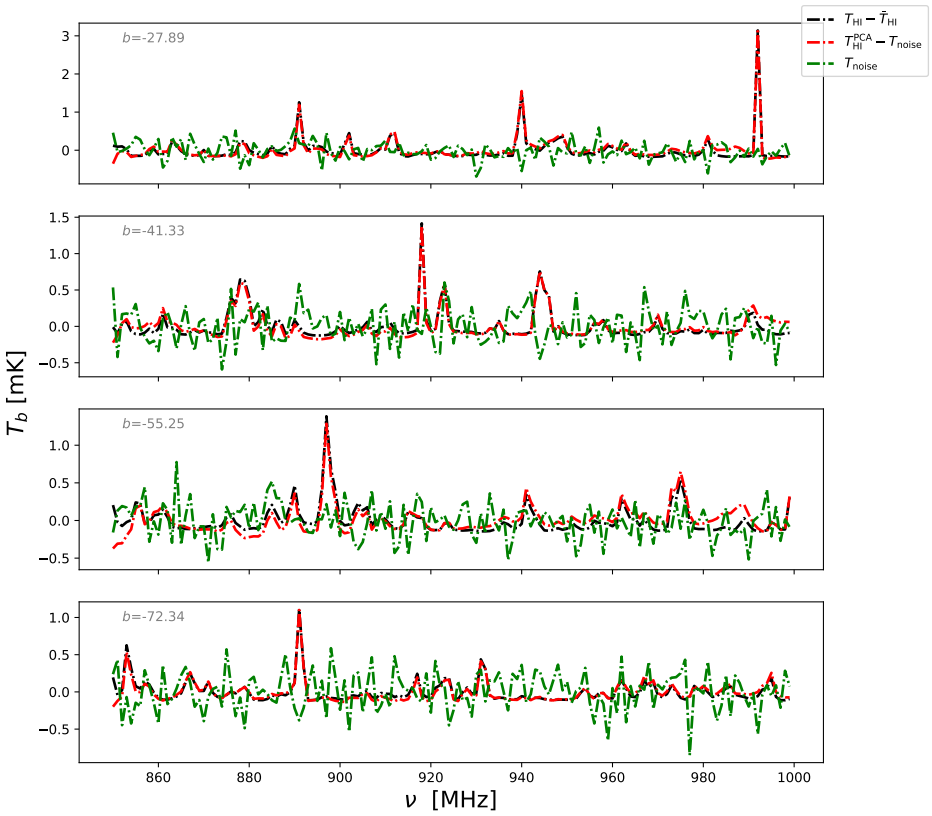

FIGURE 4.4: The recovered HI signal (red) from the PCA method compared to input HI (black). Including the input noise (green-dashed) along each line of sight. The lines of sight are chosen in terms of the galactic latitude parameter, b and these lines of sight fall within our simulated region.

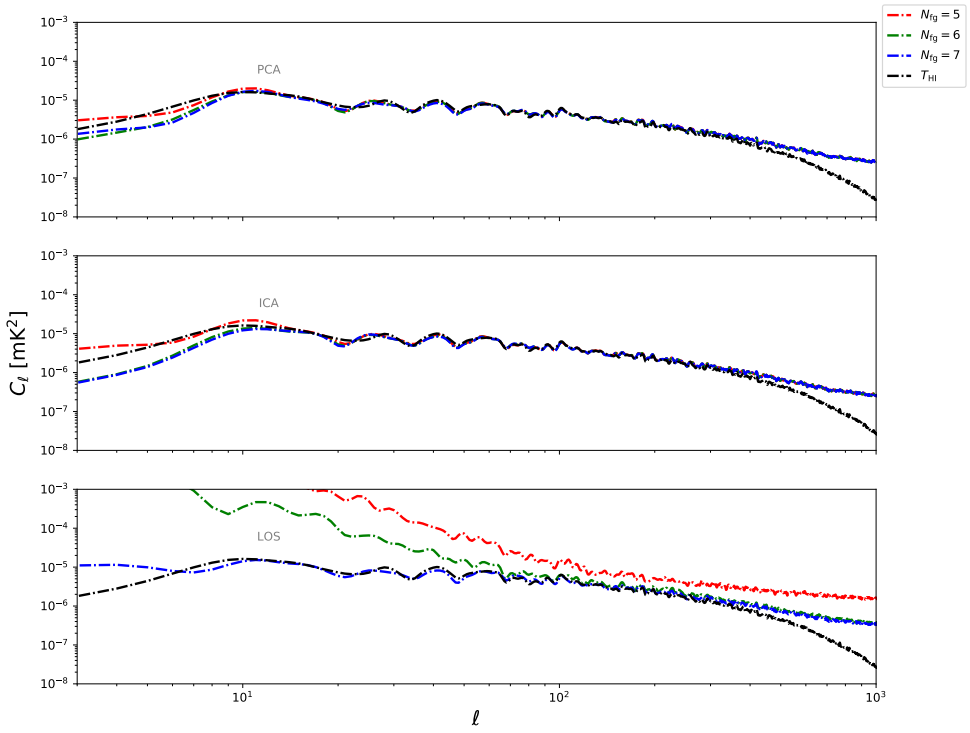

FIGURE 4.5: The recovered **H**I signal at 950 MHz. For this plot, the noise has not been subtracted off from the recovered **H**I signal for all three techniques.

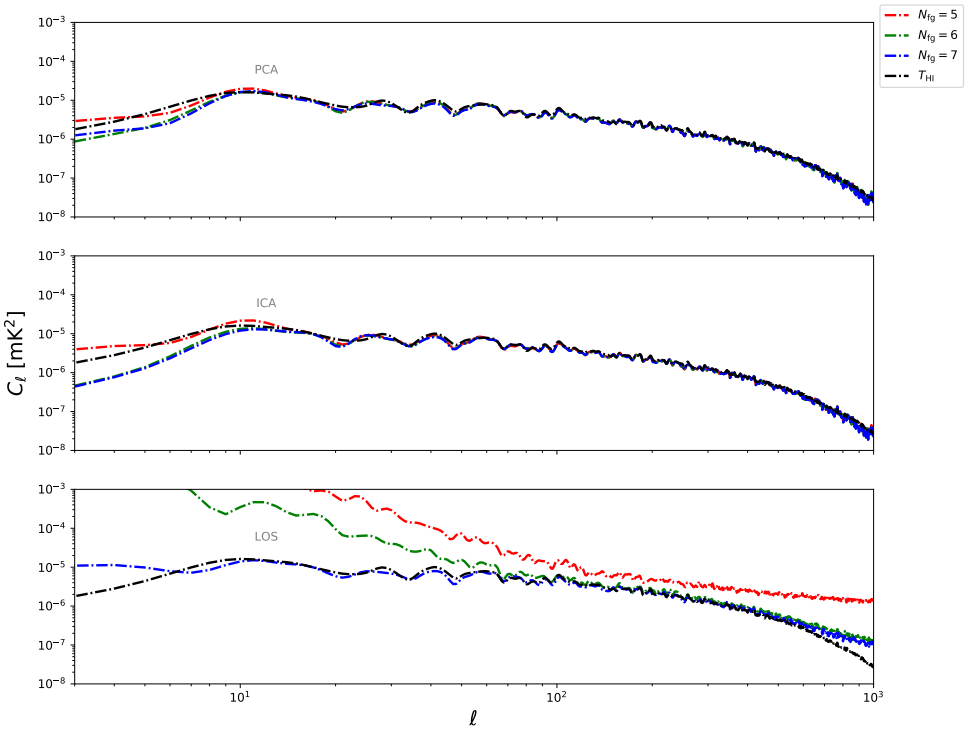

FIGURE 4.6: The angular power spectrum at 950 MHz for case 1 from ICA, PCA, and LOS respectively. The angular power spectrum for the noise was subtracted from the recovered HI for the three different modes.

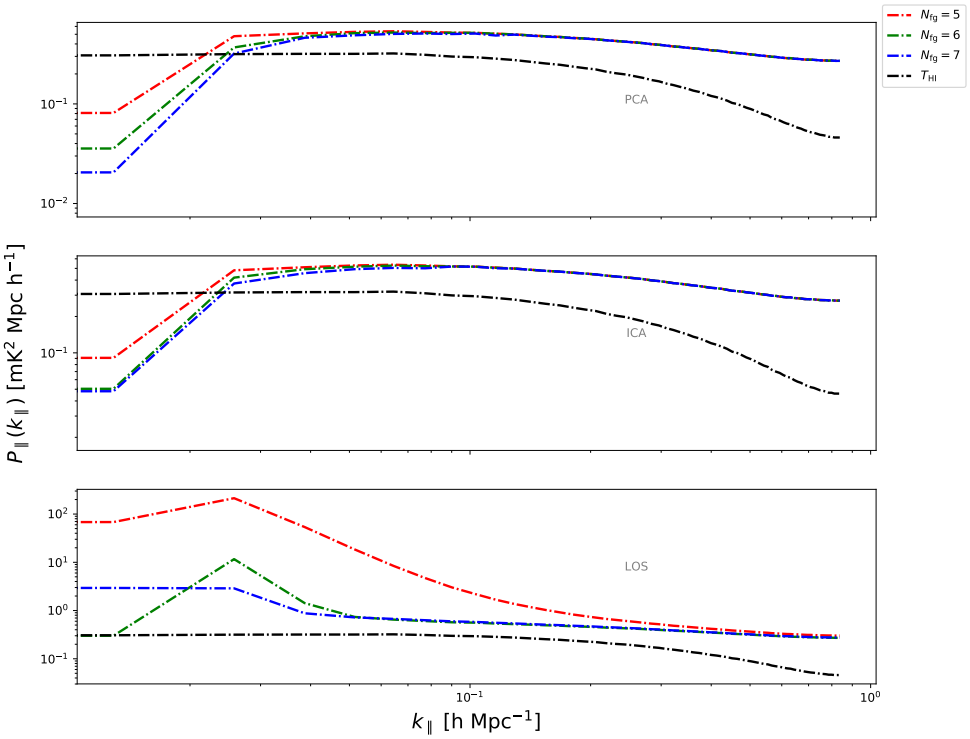

FIGURE 4.7: The radial power spectrum of case 1 without subtracting the noise power spectrum. 10 MHz was subtracted from the frequency edges before determining the power spectrum which helps in avoiding boundary effects that arise due to foreground cleaning.

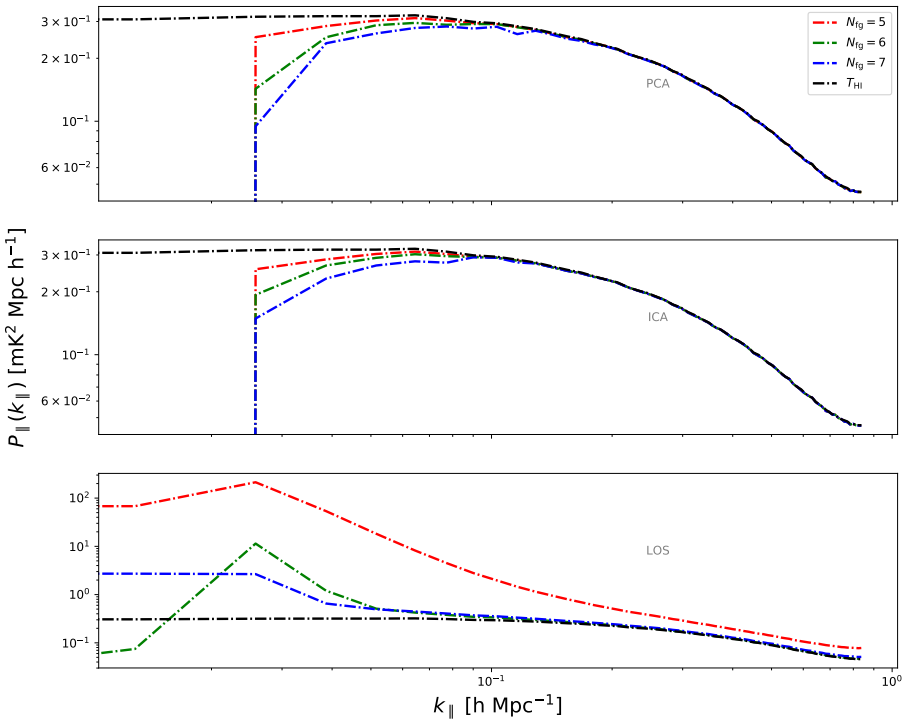

FIGURE 4.8: The recovered HI signal in terms of the power spectrum with the noise power spectrum subtracted. On small scales, the recovered signal lies on top of the input HI while on large scales the foreground cleaning fails to recover the HI signal. This is more evident for the polynomial fitting method.

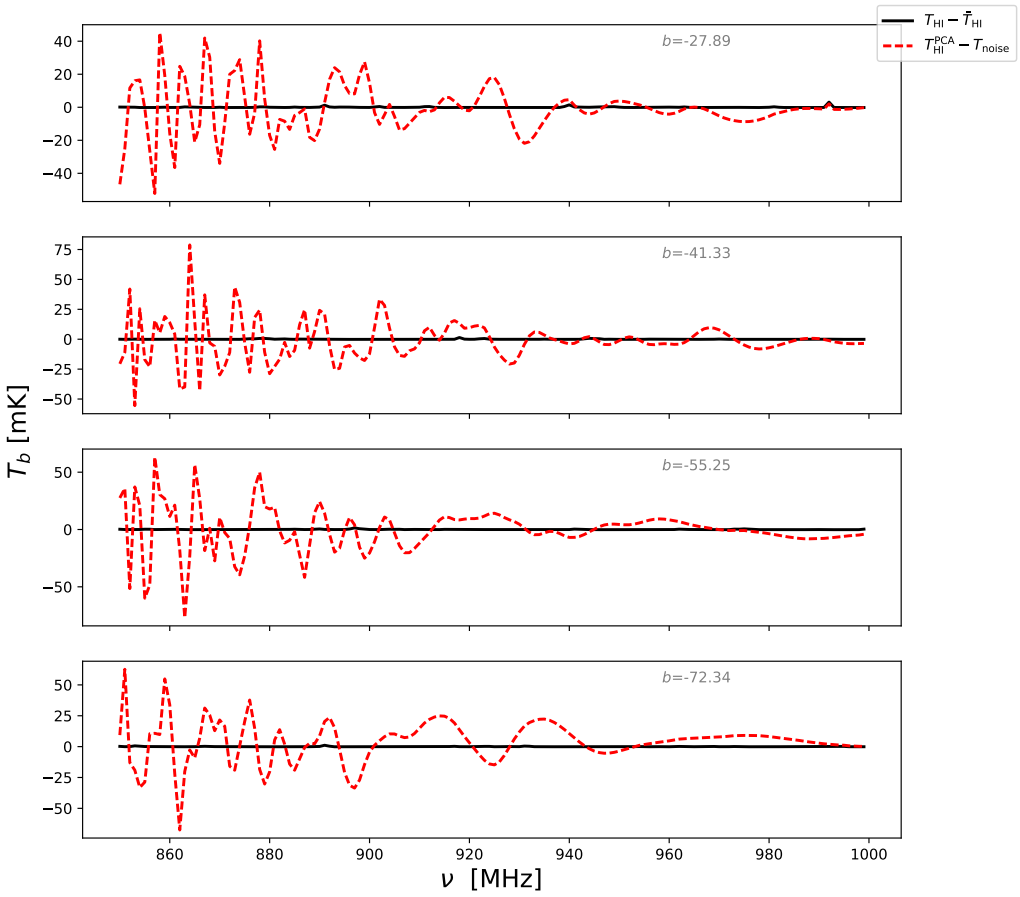

FIGURE 4.9: The recovered **H**I signal from the PCA method with $N_{\rm fg}$=7. The PCA does a bad job in recovering the **H**I signal for this case. There is a lot of contamination from foreground emission left in the recovered signal. This motivates the idea of increasing the number of modes to be removed. The lines of sight are chosen in terms of the galactic latitude parameter, b and these lines of sight fall within our simulated region.

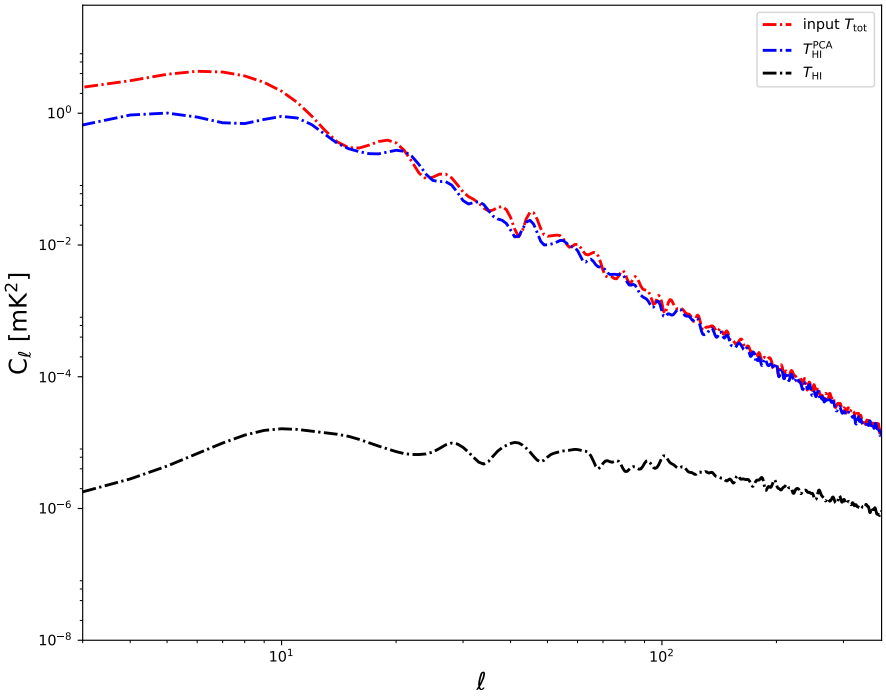

FIGURE 4.10: Recovered HI signal (blue) after 7 modes have been removed compared to the HI (black) signal. The input total signal power spectrum (simulated at 950 MHz) which is a combination of the HI signal, noise and the polarised galactic synchrotron (Stokes Q) is shown in red.

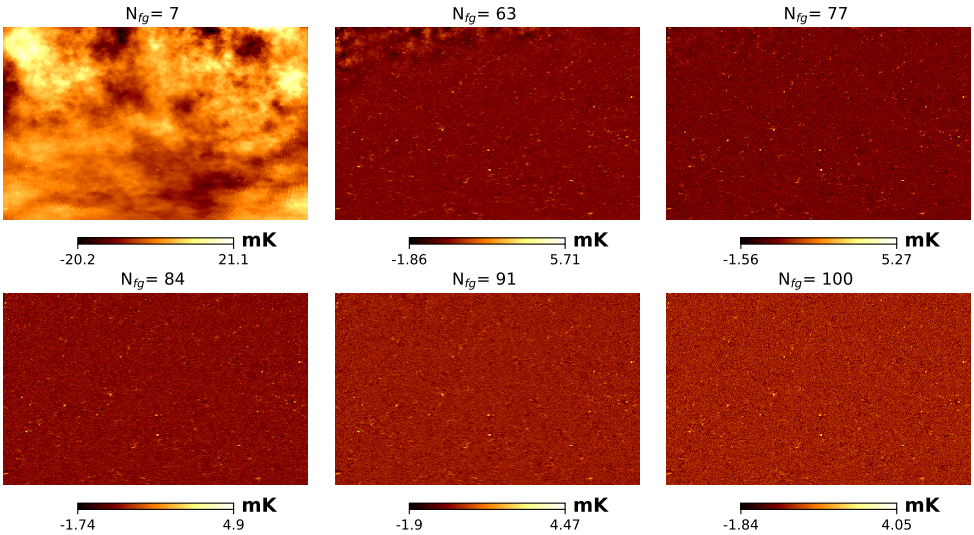

FIGURE 4.11: The recovered HI signal maps at 950 MHz for case 2 using PCA for different $N_{\rm fg}$ values. The noise contribution was not subtracted for these maps.

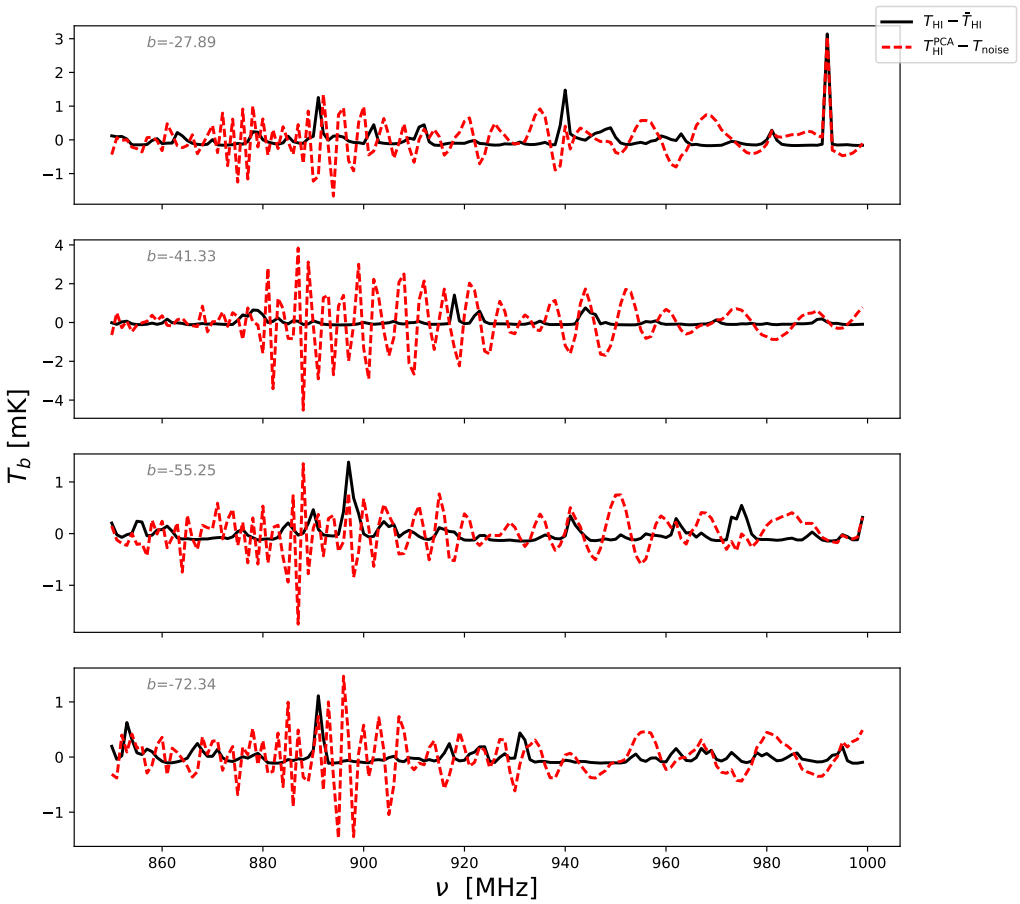

FIGURE 4.12: Recovered **H**I signal (red) compared to the input **H**I (black) after 63 modes have been removed. PCA fails in recovering some of the peaks and actual randomness in the signal. The lines of sight are chosen in terms of the galactic latitude parameter, b and these lines of sight fall within our simulated region.

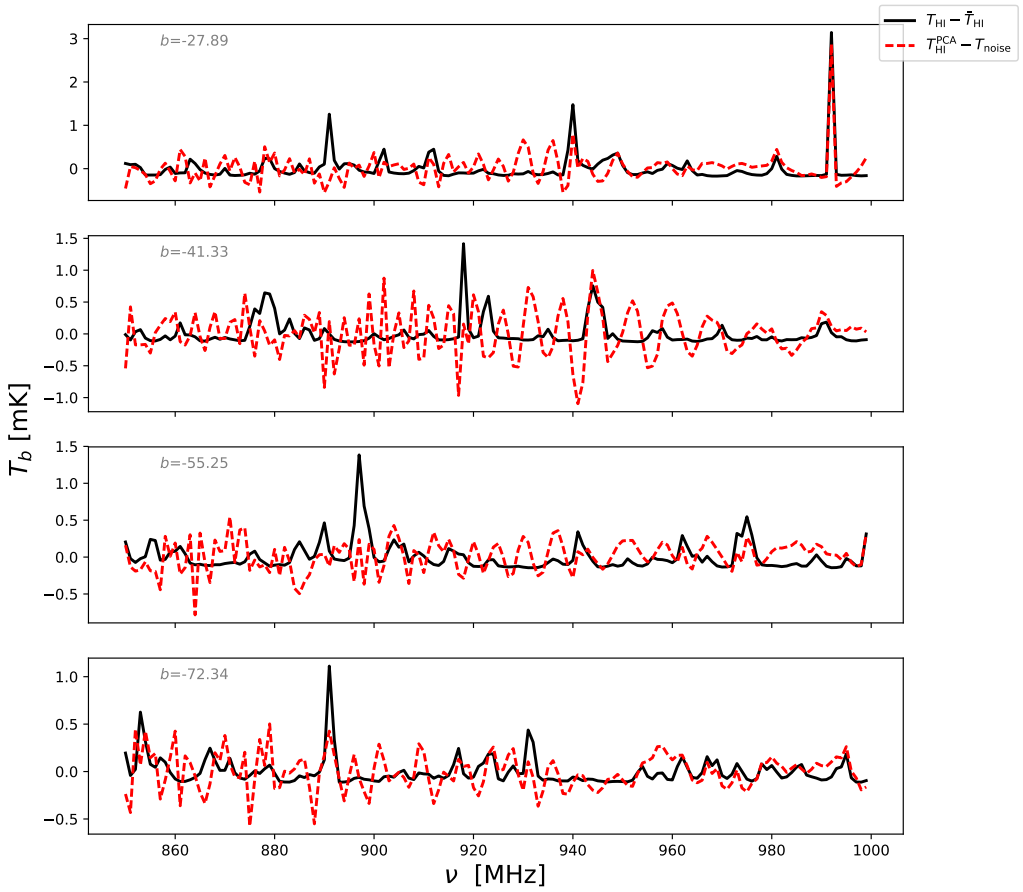

FIGURE 4.13: Recovered **H**I signal (red) compared to the input **H**I (black) after 77 modes have been removed. The lines of sight are chosen in terms of the galactic latitude parameter, b and these lines of sight fall within our simulated region.

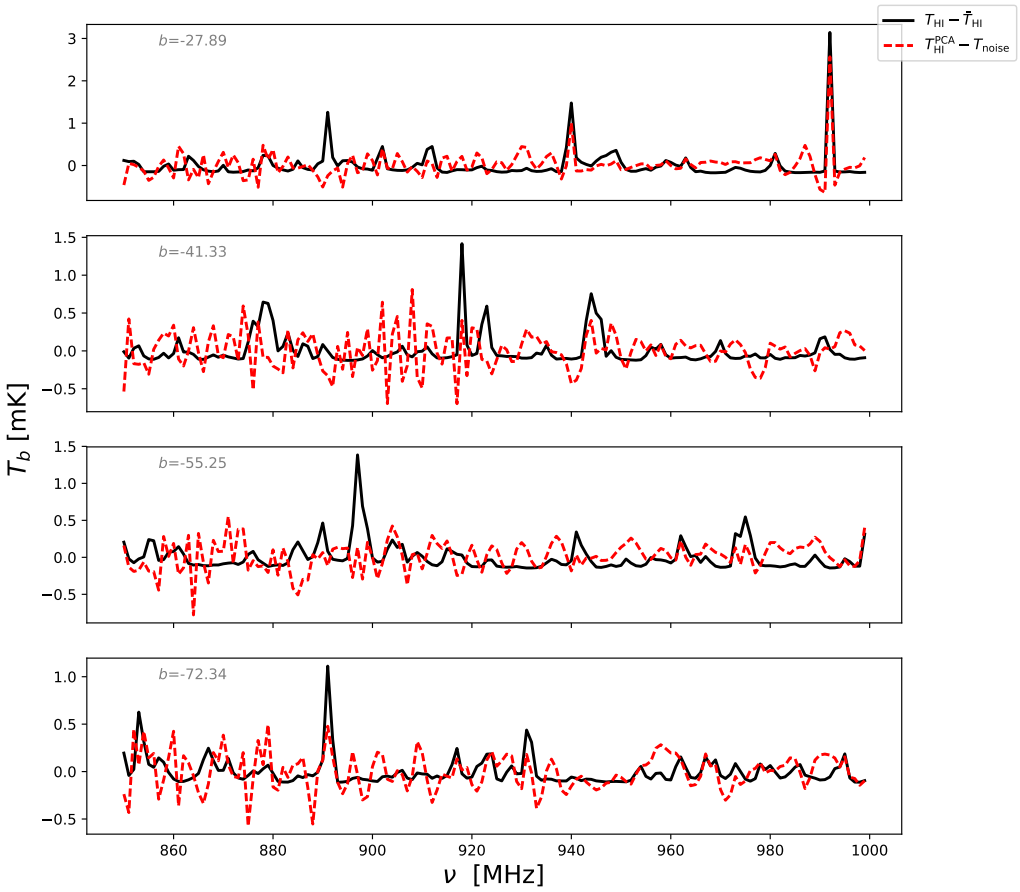

FIGURE 4.14: Recovered **H**I signal (red) compared to the input **H**I (black) after 84 modes have been removed. The lines of sight are chosen in terms of the galactic latitude parameter, b and these lines of sight fall within our simulated region.

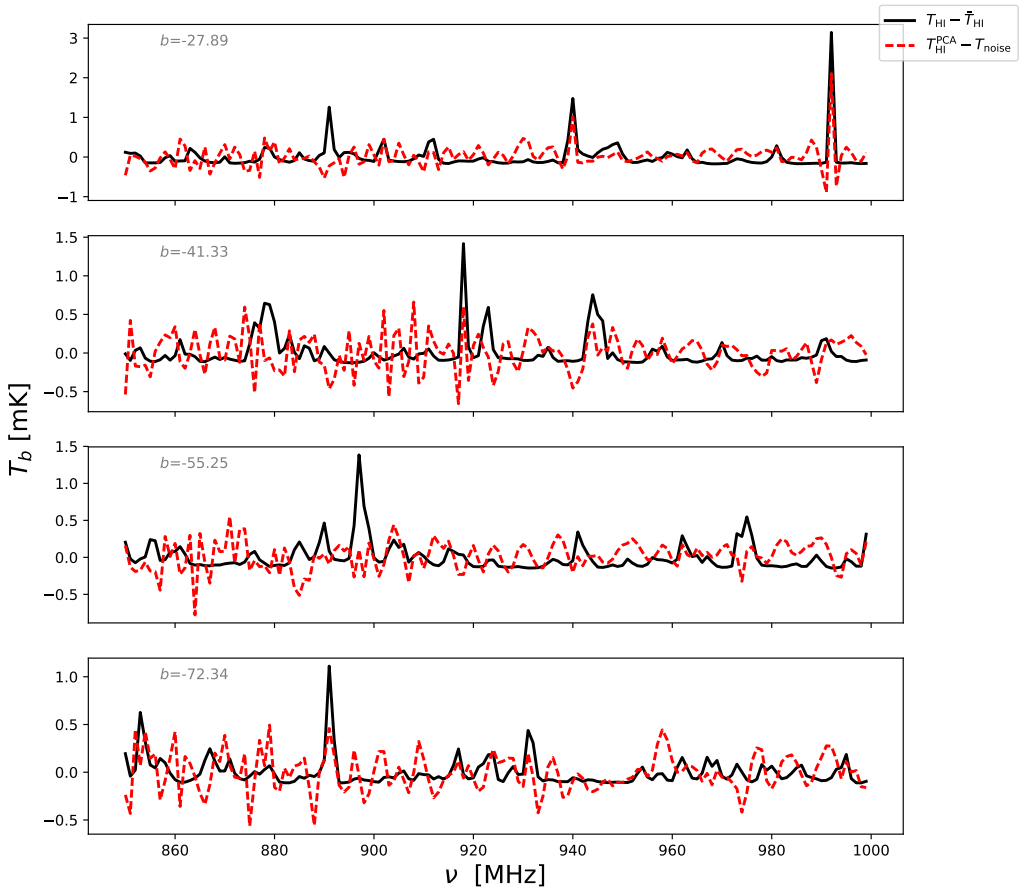

FIGURE 4.15: Recovered **H**I signal (red) compared to the input **H**I (black) after 91 modes have been removed. The lines of sight are chosen in terms of the galactic latitude parameter, b and these lines of sight fall within our simulated region.

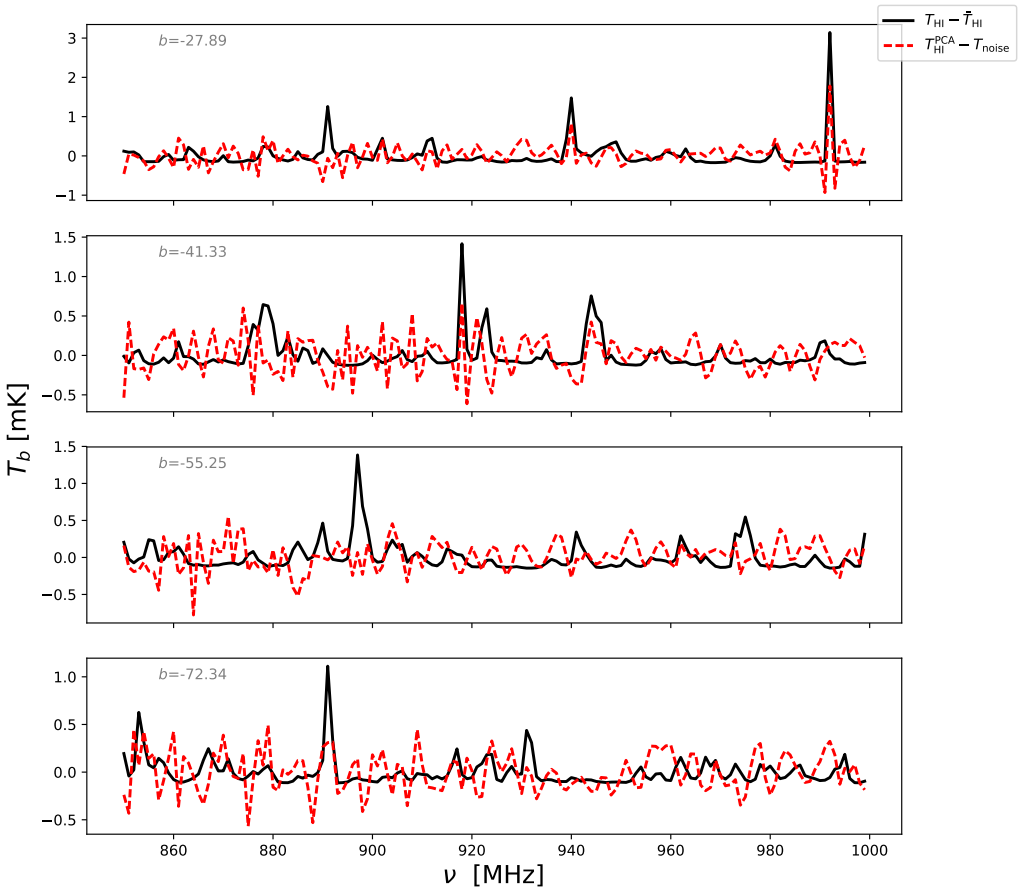

FIGURE 4.16: Recovered **H**I signal (red) compared to the input **H**I (black) after 100 modes have been removed. PCA fails in recovering some of the peaks and actual randomness in the signal. The lines of sight are chosen in terms of the galactic latitude parameter, b and these lines of sight fall within our simulated region.

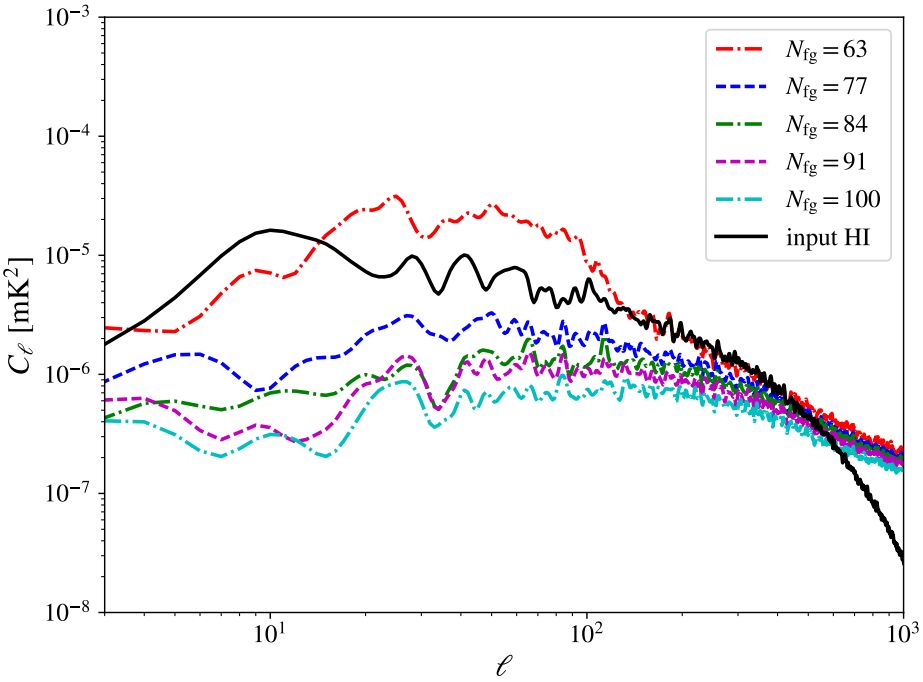

FIGURE 4.17: The recovered **H**I signal, from the PCA method at different numbers of modes removed ($N_{\rm fg}$). As expected at small scales the noise dominates. $N_{\rm fg}$ equal to 63 is the only value that recovers an **H**I signal power that is above the expected input signal on some scales while the other values deplete the **H**I signal. The angular power spectrum shown is at 950 MHz.

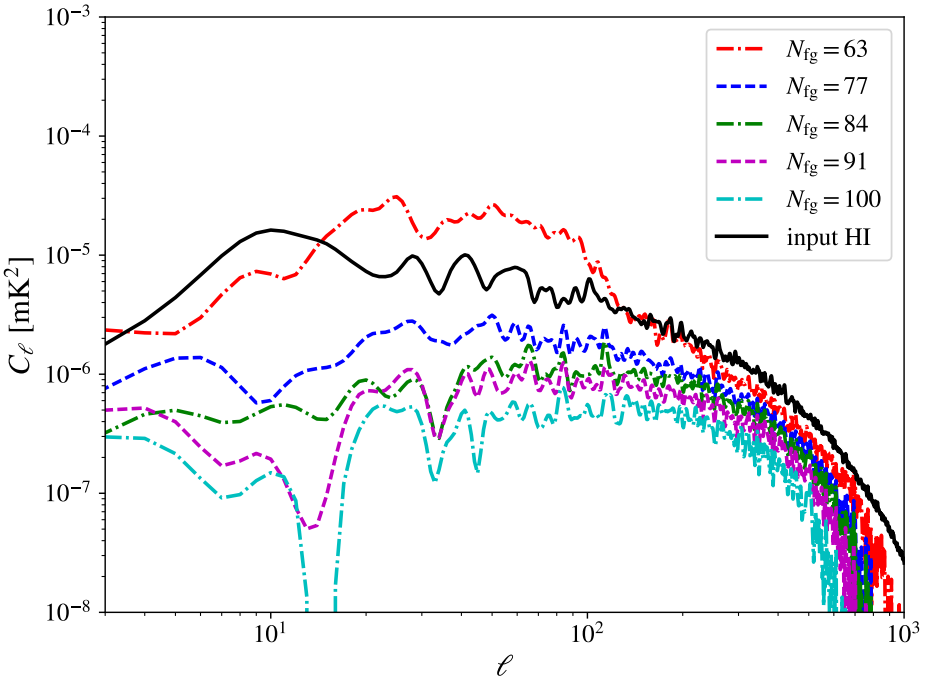

FIGURE 4.18: The recovered **H**I signal when the noise has been subtracted. High numbers of $N_{\rm fg}$ are subjected to signal over-cleaning. The angular power spectrum shown is at 950 MHz.

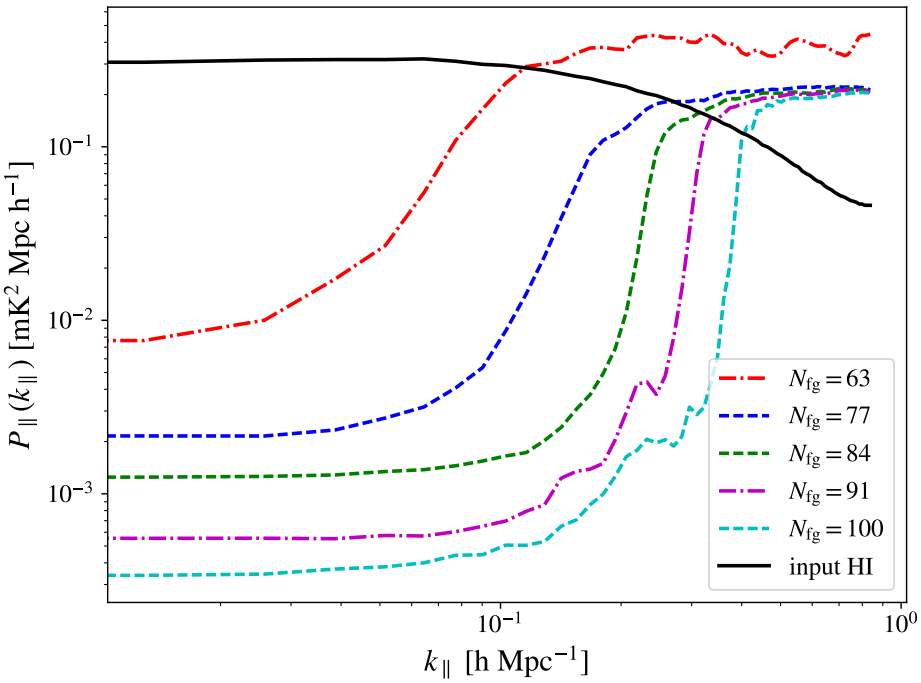

FIGURE 4.19: The recovered HI signal, from the PCA method at different numbers of modes removed ($N_{\rm fg}$). On large scales (small values of $k_{\parallel}$), the power is close to zero while on small scales it is above the expected cosmological HI signal since it is dominated by noise.

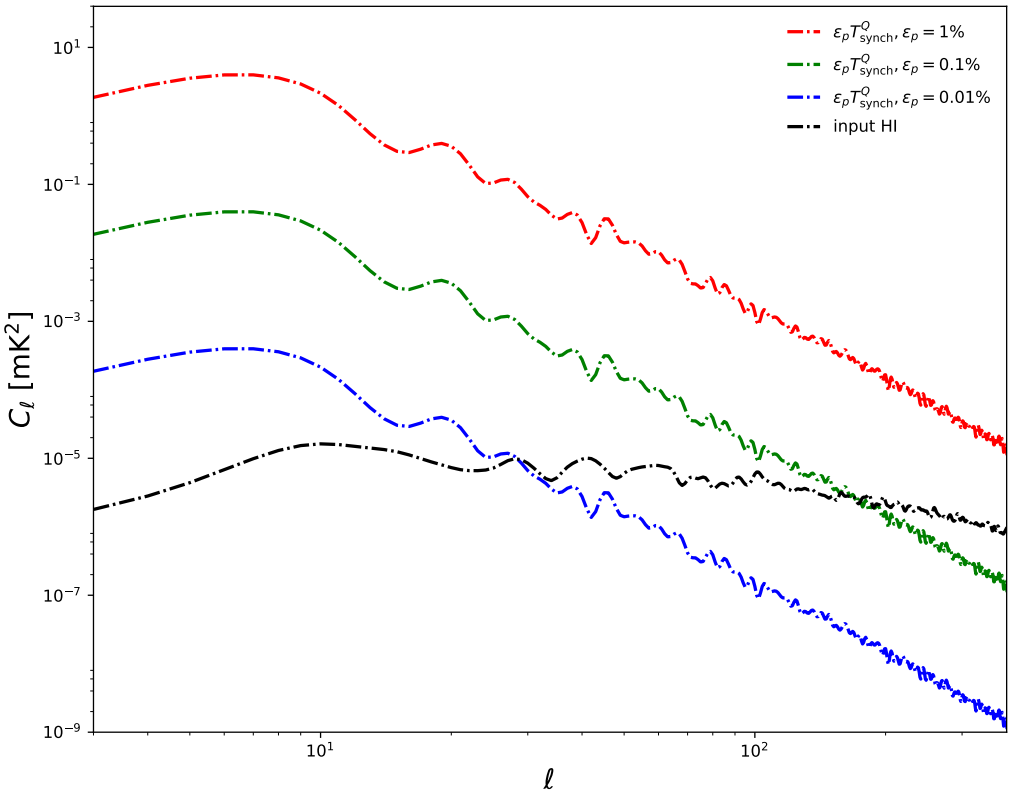

FIGURE 4.20: The angular power spectrum for polarised galactic synchrotron emission at different polarisation leakage fractions compared to the HI signal. The angular power is scaled by ϵ_p^2 for the leakage terms 1% (*red*), 0.1% (*green*) and 0.01% (*blue*) respectively.

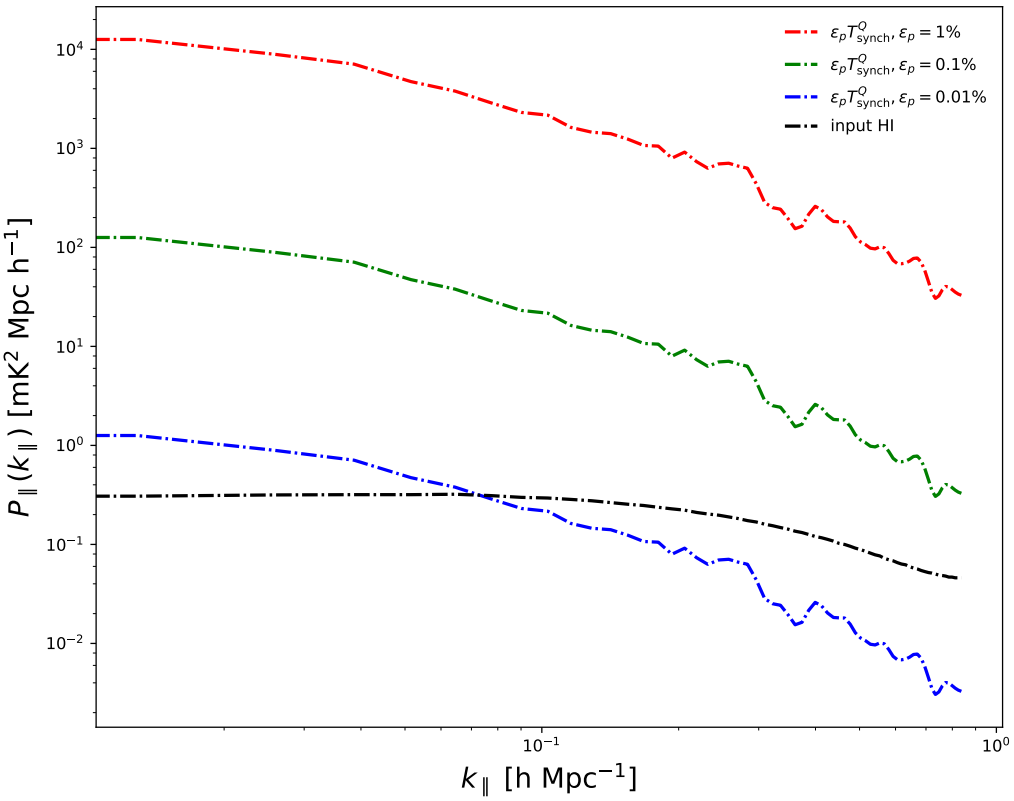

FIGURE 4.21: The radial power spectrum for polarised galactic synchrotron emission at different polarisation leakage fractions compared to the HI signal. The radial power is also scaled by ϵ_p^2 for the leakage terms 1% (*red*), 0.1% (*green*) and 0.01% (*blue*) respectively.

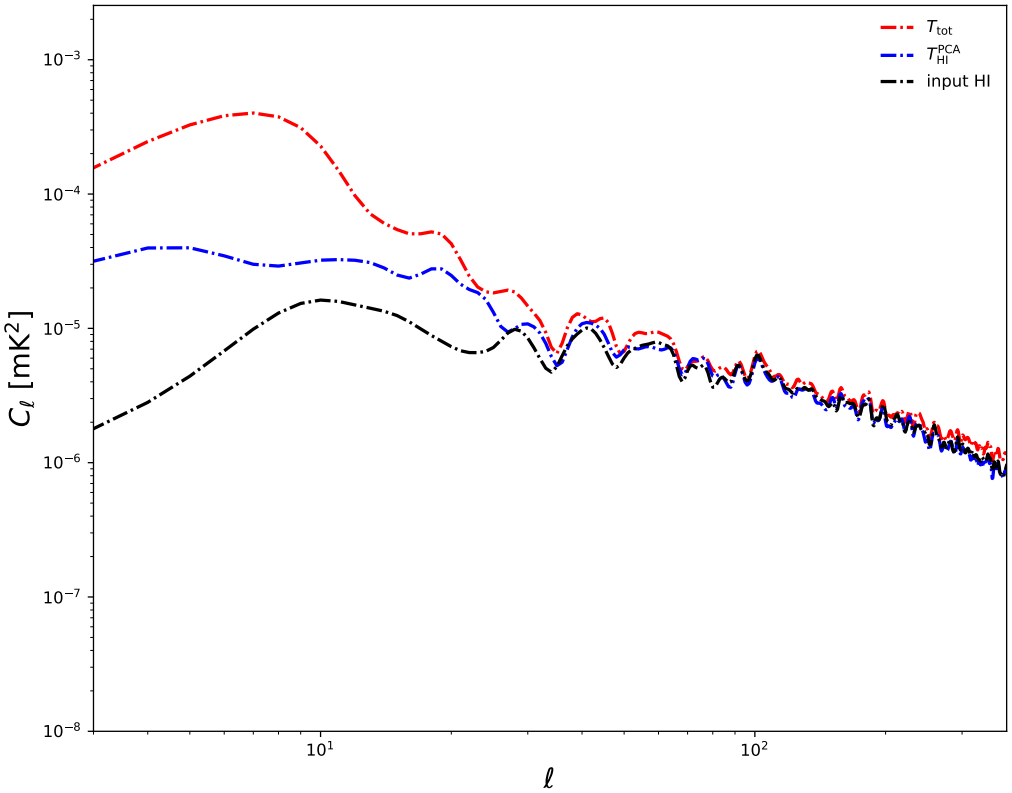

FIGURE 4.22: The angular power spectrum of the total intensity (red) given by equation 4.19 where the polarisation leakage is 0.01%, the recovered **H**I signal ($blue$) from PCA (where $N_{\rm fg}$=7) and the input **H**I signal ($black$) respectively. Note on small scales, PCA was able to recover the **H**I signal while on large scales some contamination from polarised galactic synchrotron emission is still present.

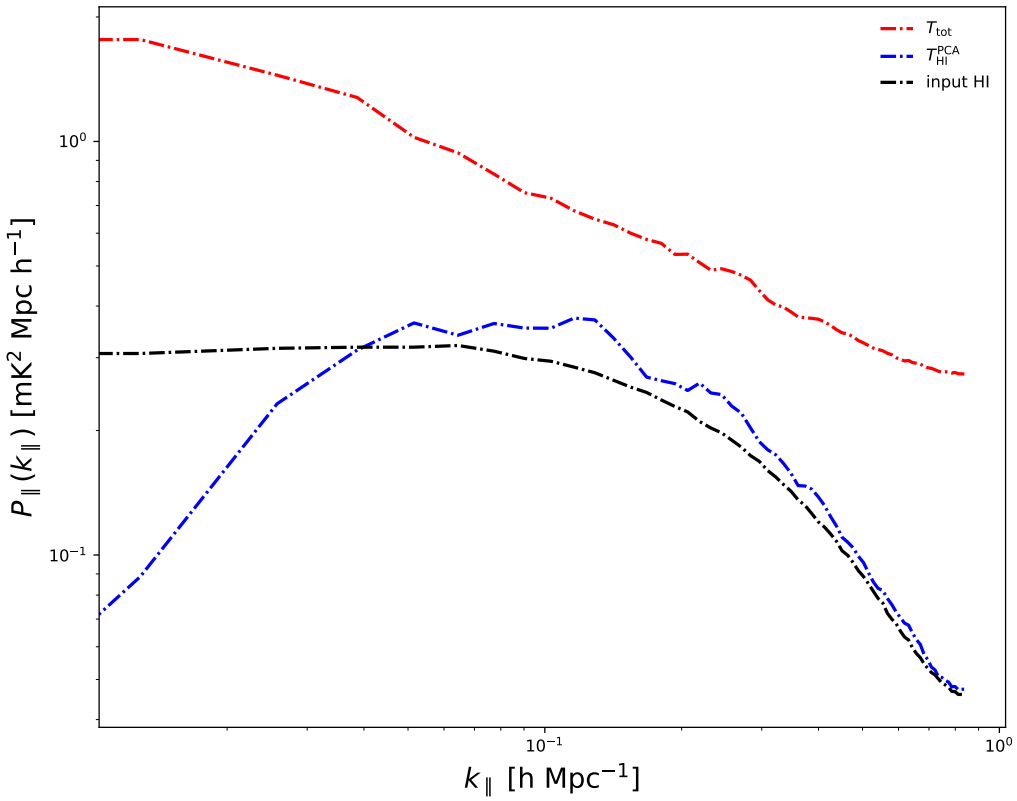

FIGURE 4.23: The radial power spectrum of the total intensity (*red*) given by equation 4.19 where the polarisation leakage is 0.01%, the recovered **H**I signal (*blue*) from PCA (where $N_{\rm fg}$=7) and the input **H**I signal (*black*) respectively. At polarisation leakage < 0.01%, the cleaning methods, PCA and ICA should be able to recover the **H**I signal.

Chapter 5

Conclusion

In this thesis, I have analysed the cosmological **H**I intensity mapping signal (at $z<$ 1) and the foregrounds that are expected to contaminate these observations. Typically one assumes that the much stronger foregrounds can be removed due to their smooth frequency structure as compared to the **H**I signal. However, possible foreground frequency fluctuations or the convolution of such strong foregrounds with instrumental fluctuations will make such cleaning more challenging. This is true for instance with the primary beam effects, such as frequency dependence and polarisation leakage. I summarise the main results:

- Full sky, point source catalogues were generated down to low fluxes both in intensity and polarisation. The differential source counts for both intensity and polarisation are compatible with previous studies as shown in Figure 3.3 and 3.5. The catalogue includes NVSS and SUMSS point sources where available which allows for a more realistic simulation of the observations. **HEALPix** temperature maps in intensity and polarisation were generated from this at different frequencies.

- Usually one expects the polarised synchrotron emission to show fluctuations in frequency. For point sources, one should see a sinusoidal behaviour due to the Faraday rotation. However, the polarised point source maps from our simulations show a smooth dependence as a function of frequency at all galactic latitudes (especially at high b). This is basically due to the fact that the frequency range we consider (in $\Delta\lambda^2$ units) is much smaller than the typical periods of the waves, which is set by π/RM. Even large rotation measures will give larger periods than the "length" we are considering in the analysis.

- We note that all the simulated foregrounds are well above the 21cm signal. The point source intensity dominates over all other foregrounds on small scales while on large scales ($\ell \lesssim 50$, see Figure 3.17) the galactic synchrotron dominates. For the polarised emission, the galactic synchrotron dominates over all scales in our simulation.

- The foreground cleaning methods used in this thesis are Line-of-sight fitting (section 4.1.1), PCA (section 4.1.2) and ICA (section 4.1.3).

- We considered a cosmology survey with MeerKAT (called MeerKLASS) covering about 4 000 (9 % of the sky) over 4 000 hours. We simulated the noise rms realizations accordingly and performed the foreground subtraction on the fraction of the sky chosen. We considered two cases for the foreground contamination. Case 1 was a combination of the **H**I signal with all foregrounds and noise except polarised galactic synchrotron (see section 4.2). We used the point sources from our catalogue as mentioned above. Including our point sources, further tests the robustness of the foreground cleaning

methods. We used the number of degrees of freedom $N_{\rm fg}$ (DOF) to quantify the number of modes to be removed.

- The value of $N_{\rm fg}$ relates to the number of foregrounds considered. In our analysis, we obtained that using $N_{\rm fg} = 7$ is the optimal number for case 1. Increasing to higher values yields no gain but ends up over-cleaning (see chapter 4). At 1% of Stokes Q for point sources and other foregrounds, the PCA and ICA were able to remove the foregrounds and recover the **H**I signal (see section 4.2.1). While polynomial fitting failed to remove the foregrounds and suffered severely on the frequency edges. This holds for angular power spectrum on the edge frequencies and also the line of sight of the recovered **H**I signal. Contrary to initial expectations, cleaning of the leakage signal from polarised point sources seems relatively easy. This is due to the smooth nature of the point source polarisation over the frequencies considered (see figure 3.11).

- Considering the polarised synchrotron emission in case 2, we saw from our analysis that removing $N_{\rm fg}$ less than 60 does not recover the **H**I signal (see section 4.2.2). Aggressively using $N_{\rm fg}$ values beyond 80 does recover part of the **H**I signal. However, most of the signal and noise itself is removed with such an aggressive cleaning. The power spectrum (both radial and angular) in case 2 shows that the PCA method cannot recover the **H**I signal except partially on small scales when a large $N_{\rm fg}$ is used ($\ell > 100$ for the angular power spectrum). Power is completely removed from the line of sight power spectrum for $k_{\parallel} < 0.2\ h\ \mathrm{Mpc}^{-1}$.

- The analysis in case 2, considers polarisation leakage at the 1% level which is shown to severely contaminate the signal. The options will be to either find a different cleaning technique that can deal with this leakage or correct for the leakage in the calibration process to make sure it is well below the 1% level (which is the usually the standard case for telescopes). In our simulations, a leakage of at most 0.01% is required in order for the cleaning methods to work (this will basically put the foreground at the same level as the HI signal - see figures 4.20 and 4.21). The simulations produced here are available online and can be used for a range of tests.

Another avenue that is now being explored is to consider a more detailed model for the MeerKAT beam in the simulations, in particular, the frequency dependence that can generate extra structure once convolved with our point source maps. The ultimate goal will be to have an accurate simulation pipeline that can test the extraction of the cosmological signal in different settings and, for instance, show possible biases on the recovery of the baryon acoustic oscillations.

Bibliography

Alonso, D., Bull, P., Ferreira, P. G., and Santos, M. G. (2015). Blind foreground subtraction for intensity mapping experiments. MNRAS, 447:400–416.

Alonso, D., Ferreira, P. G., and Santos, M. G. (2014). Fast simulations for intensity mapping experiments. MNRAS, 444:3183–3197.

Alves, M. I. R., Davies, R. D., Dickinson, C., Calabretta, M., Davis, R., and Staveley-Smith, L. (2012). A derivation of the free-free emission on the Galactic plane between ℓ= 20deg and 44deg. MNRAS, 422:2429–2443.

Backer, D., Kulkarni, S., Heiles, C., Davis, M., and Goss, M. (1982). Millisecond Pulsar in 4C 21.53. IAU Circ., 3743.

Balbi, A., Ade, P., Bock, J., Borrill, J., Boscaleri, A., de Bernardis, P., Ferreira, P. G., Hanany, S., Hristov, V., Jaffe, A. H., Lee, A. T., Oh, S., Pascale, E., Rabii, B., Richards, P. L., Smoot, G. F., Stompor, R., Winant, C. D., and Wu, J. H. P. (2002). Constraints on Cosmological Parameters from MAXIMA-1. In Gurzadyan, V. G., Jantzen, R. T., and Ruffini, R., editors, *The Ninth Marcel Grossmann Meeting*, pages 2195–2196.

Bandura, K., Addison, G. E., Amiri, M., Bond, J. R., Campbell-Wilson, D., Connor, L., Cliche, J.-F., Davis, G., Deng, M., Denman, N., Dobbs, M., Fandino, M., Gibbs, K., Gilbert, A., Halpern, M., Hanna, D., Hincks, A. D., Hinshaw, G., Höfer, C., Klages, P., Landecker, T. L., Masui, K., Mena Parra, J., Newburgh, L. B., Pen, U.-l., Peterson, J. B., Recnik, A., Shaw, J. R., Sigurdson, K., Sitwell, M., Smecher, G., Smegal, R., Vanderlinde, K., and Wiebe, D. (2014). Canadian Hydrogen Intensity Mapping Experiment (CHIME) pathfinder. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 9145 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 22.

Battye, R. A., Browne, I. W. A., Dickinson, C., Heron, G., Maffei, B., and Pourtsidou, A. (2013). H I intensity mapping: a single dish approach. MNRAS, 434:1239–1256.

Begelman, M. C., Blandford, R. D., and Rees, M. J. (1984). Theory of extragalactic radio sources. *Reviews of Modern Physics*, 56:255–351.

Bennett, C. L. (2005). First Year WMAP Observations. In Colless, M., Staveley-Smith, L., and Stathakis, R. A., editors, *Maps of the Cosmos*, volume 216 of *IAU Symposium*, page 18.

Bennett, C. L., Hill, R. S., Hinshaw, G., Nolta, M. R., Odegard, N., Page, L., Spergel, D. N., Weiland, J. L., Wright, E. L., Halpern, M., Jarosik, N., Kogut, A., Limon, M., Meyer, S. S., Tucker, G. S., and Wollack, E. (2003). First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Foreground Emission. ApJS, 148:97–117.

Bigot-Sazy, M.-A., Dickinson, C., Battye, R. A., Browne, I. W. A., Ma, Y.-Z., Maffei, B., Noviello, F., Remazeilles, M., and Wilkinson, P. N. (2015). Simulations for single-dish intensity mapping experiments. *ArXiv e-prints*.

Blake, C., Brough, S., Colless, M., Contreras, C., Couch, W., Croom, S., Croton, D., Davis, T. M., Drinkwater, M. J., Forster, K., Gilbank, D., Gladders, M., Glazebrook, K., Jelliffe, B., Jurek, R. J., Li, I.-h., Madore, B., Martin, D. C., Pimbblet, K., Poole, G. B., Pracy, M., Sharp, R., Wisnioski, E., Woods, D., Wyder, T. K., and Yee, H. K. C. (2012). The WiggleZ Dark Energy Survey: joint measurements of the expansion and growth history at z < 1. MNRAS, 425:405–414.

Blandford, R. D. and Königl, A. (1979). Relativistic jets as compact radio sources. ApJ, 232:34–48.

Blandford, R. D. and Rees, M. J. (1974). A 'twin-exhaust' model for double radio sources. MNRAS, 169:395–415.

Bozza, V. and Mancini, L. (2008). Gravitational Lensing of Stars Surrounding Supermassive Black Holes. In Kleinert, H., Jantzen, R. T., and Ruffini, R., editors, *The Eleventh Marcel Grossmann Meeting On Recent Developments in Theoretical and Experimental General Relativity, Gravitation and Relativistic Field Theories*, pages 1710–1712.

Bull, P., Camera, S., Raccanelli, A., Blake, C., Ferreira, P., Santos, M., and Schwarz, D. J. (2015a). Measuring baryon acoustic oscillations with future SKA surveys. *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, page 24.

Bull, P., Ferreira, P. G., Patel, P., and Santos, M. G. (2015b). Late-time Cosmology with 21 cm Intensity Mapping Experiments. ApJ, 803:21.

Camera, S., Raccanelli, A., Bull, P., Bertacca, D., Chen, X., Ferreira, P., Kunz, M., Maartens, R., Mao, Y., Santos, M., Shapiro, P. R., Viel, M., and Xu, Y. (2015). Cosmology on the Largest Scales with the SKA. *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, page 25.

Camera, S., Santos, M. G., Ferreira, P. G., and Ferramacho, L. (2013). Cosmology on Ultralarge Scales with Intensity Mapping of the Neutral Hydrogen 21 cm Emission: Limits on Primordial Non-Gaussianity. *Physical Review Letters*, 111(17):171302.

Carignan, C., Chemin, L., Huchtmeier, W. K., and Lockman, F. J. (2006). The Extended H I Rotation Curve and Mass Distribution of M31. ApJ, 641:L109–L112.

Chang, T.-C., Pen, U.-L., Peterson, J. B., and McDonald, P. (2008). Baryon Acoustic Oscillation Intensity Mapping of Dark Energy. *Physical Review Letters*, 100(9):091303.

Cole, S., Percival, W. J., Peacock, J. A., Norberg, P., Baugh, C. M., Frenk, C. S., Baldry, I., Bland-Hawthorn, J., Bridges, T., Cannon, R., Colless, M., Collins, C., Couch, W., Cross, N. J. G., Dalton, G., Eke, V. R., De Propris, R., Driver, S. P., Efstathiou, G., Ellis, R. S., Glazebrook, K., Jackson, C., Jenkins, A., Lahav, O., Lewis, I., Lumsden, S., Maddox, S., Madgwick, D., Peterson, B. A., Sutherland, W., and Taylor, K. (2005). The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications. MNRAS, 362:505–534.

Condon, J. J. (1984). Cosmological evolution of radio sources. ApJ, 287:461–474.

Condon, J. J. (1986). Radio characteristics of galactic nuclei. PASP, 98:152–158.

Condon, J. J., Cotton, W. D., Greisen, E. W., Yin, Q. F., Perley, R. A., Taylor, G. B., and Broderick, J. J. (1998). The NRAO VLA Sky Survey. AJ, 115:1693–1716.

Crill, B. P., Ade, P. A. R., Artusa, D. R., Bhatia, R. S., Bock, J. J., Boscaleri, A., Cardoni, P., Church, S. E., Coble, K., de Bernardis, P., de Troia, G., Farese, P., Ganga, K. M., Giacometti, M., Haynes, C. V., Hivon, E., Hristov, V. V., Iacoangeli, A., Jones, W. C., Lange, A. E., Martinis, L., Masi, S., Mason, P. V., Mauskopf, P. D., Miglio, L., Montroy, T., Netterfield, C. B., Paine, C. G., Pascale, E., Piacentini, F., Polenta, G., Pongetti, F., Romeo, G., Ruhl, J. E., Scaramuzzi, F., Sforna, D., and Turner, A. D. (2003). BOOMERANG: A Balloon-borne Millimeter-Wave Telescope and Total Power Receiver for Mapping Anisotropy in the Cosmic Microwave Background. ApJS, 148:527–541.

D'Amico, N., Lyne, A. G., Manchester, R. N., Camilo, F. M., Kaspi, V. M., Bell, J., Stairs, I. H., Crawford, F., Morris, D., and Possenti, A. (2001). The Parkes multibeam pulsar survey: Preliminary results. *X-ray Astronomy: Stellar Endpoints, AGN, and the Diffuse X-ray Background*, 599:598–601.

Davies, R. D., Watson, R. A., and Gutierrez, C. M. (1996). Galactic synchrotron emission at high frequencies. MNRAS, 278:925–939.

Dawson, K. S., Schlegel, D. J., Ahn, C. P., Anderson, S. F., Éric Aubourg, Bailey, S., Barkhouser, R. H., Bautista, J. E., Beifiori, A., Berlind, A. A., Bhardwaj, V., Bizyaev, D., Blake, C. H., Blanton, M. R., Blomqvist, M., Bolton, A. S., Borde, A., Bovy, J., Brandt, W. N., Brewington, H., Brinkmann, J., Brown, P. J., Brownstein, J. R., Bundy, K., Busca, N. G., Carithers, W., Carnero, A. R., Carr, M. A., Chen, Y., Comparat, J., Connolly, N., Cope, F., Croft, R. A. C., Cuesta, A. J., da Costa, L. N., Davenport, J. R. A., Delubac, T., de Putter, R., Dhital, S., Ealet, A., Ebelke, G. L., Eisenstein, D. J., Escoffier, S., Fan, X., Ak, N. F., Finley, H., Font-Ribera, A., Génova-Santos, R., Gunn, J. E., Guo, H., Haggard, D., Hall, P. B., Hamilton, J.-C., Harris, B., Harris, D. W., Ho, S., Hogg, D. W., Holder, D., Honscheid, K., Huehnerhoff, J., Jordan, B., Jordan, W. P., Kauffmann, G., Kazin, E. A., Kirkby, D., Klaene, M. A., Kneib, J.-P., Goff, J.-M. L., Lee, K.-G., Long, D. C., Loomis, C. P., Lundgren, B., Lupton, R. H., Maia, M. A. G., Makler, M., Malanushenko, E., Malanushenko, V., Mandelbaum, R., Manera, M., Maraston, C., Margala, D., Masters, K. L., McBride, C. K., McDonald, P., McGreer, I. D., McMahon, R. G., Mena, O., Miralda-Escudé, J., Montero-Dorta, A. D., Montesano, F., Muna, D., Myers, A. D., Naugle, T., Nichol, R. C., Noterdaeme, P., Nuza, S. E., Olmstead, M. D., Oravetz, A., Oravetz, D. J., Owen, R., Padmanabhan, N., Palanque-Delabrouille, N., Pan, K., Parejko, J. K., Pâris, I., Percival, W. J., Pérez-Fournon, I., Pérez-Ràfols, I., Petitjean, P., Pfaffenberger, R., Pforr, J., Pieri, M. M., Prada, F., Price-Whelan, A. M., Raddick, M. J., Rebolo, R., Rich, J., Richards, G. T., Rockosi, C. M., Roe, N. A., Ross, A. J., Ross, N. P., Rossi, G., Rubiño-Martin, J. A., Samushia, L., Sánchez, A. G., Sayres, C., Schmidt, S. J., Schneider, D. P., Scóccola, C. G., Seo, H.-J., Shelden, A., Sheldon, E., Shen, Y., Shu, Y., Slosar, A., Smee, S. A., Snedden, S. A., Stauffer, F., Steele, O., Strauss, M. A., Streblyanska, A., Suzuki, N., Swanson, M. E. C., Tal, T., Tanaka, M., Thomas, D., Tinker, J. L., Tojeiro, R., Tremonti, C. A., Magaña, M. V., Verde, L., Viel, M., Wake, D. A., Watson, M., Weaver, B. A., Weinberg, D. H., Weiner, B. J., West, A. A., White, M., Wood-Vasey, W. M., Yeche, C., Zehavi, I., Zhao, G.-B.,

and Zheng, Z. (2013). The baryon oscillation spectroscopic survey of sdss-iii. *The Astronomical Journal*, 145(1):10.

de Bernardis, P., Ade, P. A. R., Bock, J. J., Bond, J. R., Borrill, J., Boscaleri, A., Coble, K., Crill, B. P., De Gasperis, G., De Troia, G., Farese, P. C., Ferreira, P. G., Ganga, K., Giacometti, M., Hivon, E., Hristov, V. V., Iacoangeli, A., Jaffe, A. H., Lange, A. E., Martinis, L., Masi, S., Mason, P., Mauskopf, P. D., Melchiorri, A., Miglio, L., Montroy, T., Netterfield, C. B., Pascale, E., Piacentini, F., Pogosyan, D., Pongetti, F., Prunet, S., Rao, S., Romeo, G., Ruhl, J. E., Scaramuzzi, F., Sforna, D., and Vittorio, N. (2000). Detection of anisotropy in the Cosmic Microwave Background at horizon and sub-horizon scales with the BOOMERanG experiment. *ArXiv Astrophysics e-prints*.

de Blok, W. J. G., Adams, E. A. K., Amram, P., Athanassoula, E., Bagetakos, I., Balkowski, C., Bershady, M. A., Beswick, R., Bigiel, F., Blyth, S.-L., Bosma, A., Booth, R. S., Bouchard, A., Brinks, E., Carignan, C., Chemin, L., Combes, F., Conway, J., Elson, E. C., English, J., Epinat, B., Frank, B. S., Fiege, J., Fraternali, F., Gallagher, J. S., Gibson, B. K., Heald, G., Henning, P. A., Holwerda, B. W., Jarrett, T. H., Jerjen, H., Józsa, G. I., Kapala, M., Klöckner, H.-R., Koribalski, B. S., Kraan-Korteweg, R. C., Leon, S., Leroy, A., Loubser, S. I., Lucero, D. M., McGaugh, S. S., Meurer, G. R., Meyer, M., Mogotsi, M., Namumba, B., Oh, S., Oosterloo, T. A., Pisano, D. J., Popping, A., Ratcliffe, S., Sellwood, J. A., Schinnerer, E., Schröder, A. C., Sheth, K., Smith, M. W. L., Sorgho, A., Spekkens, K., Stanimirovic, S., van der Heyden, K., van Driel, W., Verdes-Montenegro, L., Walter, F., Westmeier, T., Wilcots, E., Williams, T., Wong, O. I., Woudt, P. A., and Zijlstra, A. (2017). An Overview of the MHONGOOSE Survey: Observing Nearby Galaxies with MeerKAT. *ArXiv e-prints*.

Delabrouille, J., Betoule, M., Melin, J.-B., Miville-Deschênes, M.-A., Gonzalez-Nuevo, J., Le Jeune, M., Castex, G., de Zotti, G., Basak, S., Ashdown, M., Aumont, J., Baccigalupi, C., Banday, A. J., Bernard, J.-P., Bouchet, F. R., Clements, D. L., da Silva, A., Dickinson, C., Dodu, F., Dolag, K., Elsner, F., Fauvet, L., Faÿ, G., Giardino, G., Leach, S., Lesgourgues, J., Liguori, M., Macías-Pérez, J. F., Massardi, M., Matarrese, S., Mazzotta, P., Montier, L., Mottet, S., Paladini, R., Partridge, B., Piffaretti, R., Prezeau, G., Prunet, S., Ricciardi, S., Roman, M., Schaefer, B., and Toffolatti, L. (2013). The pre-launch Planck Sky Model: a model of sky emission at submillimetre to centimetre wavelengths. A&A, 553:A96.

Dickinson, C., Davies, R. D., and Davis, R. J. (2003). Towards a free-free template for CMB foregrounds. MNRAS, 341:369–384.

Djorgovski, S., Weir, N., Matthews, K., and Graham, J. R. (1991). Discovery of an infrared nucleus in Cygnus A - an obscured quasar revealed? ApJ, 372:L67–L70.

Dupac, X. and Giard, M. (2002). Map-making methods for cosmic microwave background experiments. MNRAS, 330:497–505.

Eisenstein, Zehavi, and Hogg (2004). Baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies. *Astrophys. J.*, 129(3):450–467.

Ellis, G. F. R. and van Elst, H. (1999). Cosmological Models (Cargèse lectures 1998). In Lachièze-Rey, M., editor, *NATO Advanced Science Institutes (ASI) Series C*, volume 541 of *NATO Advanced Science Institutes (ASI) Series C*, pages 1–116.

ESA and the Planck Collaboration (2013). Planck's power spectrum of temperature fluctuations in the cosmic microwave background. http://sci.esa.int/jump.cfm?oid=51555. Accessed: 2017-10-01.

Ewen, H. and Purcell, E. (1951). Observation of a Line in the Galactic Radio Spectrum: Radiation from Galactic Hydrogen at 1,420 Mc./sec. *Nature*, 168(4270):356–356.

Fabian, A. C. (2012). Observational Evidence of Active Galactic Nuclei Feedback. ARA&A, 50:455–489.

Fanaroff, B. L. and Riley, J. M. (1974). The morphology of extragalactic radio sources of high and low luminosity. MNRAS, 167:31P–36P.

Ferrer, F., Krauss, L. M., and Profumo, S. (2006). Indirect detection of light neutralino dark matter in the next-to-minimal supersymmetric standard model. Phys. Rev. D, 74(11):115007.

Fiebig, D. and Guesten, R. (1989). Strong magnetic fields in interstellar H2O maser clumps. A&A, 214:333–338.

Fornengo, N. (2008). Status and perspectives of indirect and direct dark matter searches. *Advances in Space Research*, 41:2010–2018.

Freese, K., Frieman, J., and Gould, A. (1988). Signal modulation in cold-dark-matter detection. Phys. Rev. D, 37:3388–3405.

Furst, E. and Reich, W. (1990). Probing the Interstellar Magnetic Field by Polarization Observations of Supernova Remnants. In Beck, R., Wielebinski, R., and Kronberg, P. P., editors, *Galactic and Intergalactic Magnetic Fields*, volume 140 of *IAU Symposium*, page 73.

Górski, K. M., Banday, A. J., Hivon, E., and Wandelt, B. D. (2002). HEALPix — a Framework for High Resolution, Fast Analysis on the Sphere. In Bohlender, D. A., Durand, D., and Handley, T. H., editors, *Astronomical Data Analysis Software and Systems XI*, volume 281 of *Astronomical Society of the Pacific Conference Series*, page 107.

Griest, K. (1993). The Search for the Dark Matter: WIMPs and MACHOs. In Akerlof, C. W. and Srednicki, M. A., editors, *Texas/PASCOS '92: Relativistic Astrophysics and Particle Cosmology*, volume 688 of *Annals of the New York Academy of Sciences*, page 390.

Gupta, N., Srianand, R., Baan, W., Baker, A., Beswick, R., Bhatnagar, S., Bhattacharya, D., Bosma, A., Carilli, C., Cluver, M., Combes, F., Cress, C., Dutta, R., Fynbo, J., Heald, G., Hilton, M., Hussain, T., Jarvis, M., Jozsa, G., Kamphuis, P., Kembhavi, A., Kerp, J., Klöckner, H.-R., Krogager, J., Kulkarni, V., Ledoux, C., Mahabal, A., Mauch, T., Moodley, K., Momjian, E., Morganti, R., Noterdaeme, P., Oosterloo, T., Petitjean, P., Schröder, A., Serra, P., Sievers, J., Spekkens, K., Väisänen, P., van der Hulst, T., Vivek, M., Wang, J., Wong, O. I., and Zungu, A. R. (2017). The MeerKAT Absorption Line Survey (MALS). *ArXiv e-prints*.

Hamilton, A. J. S. (1998). Linear Redshift Distortions: a Review. In Hamilton, D., editor, *The Evolving Universe*, volume 231 of *Astrophysics and Space Science Library*, page 185.

Haslam, C. G. T., Salter, C. J., Stoffel, H., and Wilson, W. E. (1982). A 408 MHz all-sky continuum survey. II - The atlas of contour maps. A&AS, 47:1.

Heywood, I., Armstrong, R. P., Booth, R., Bunker, A. J., Deane, R. P., Jarvis, M. J., Jonas, J. L., Jones, M. E., Kloeckner, H., Kneib, J., Knudsen, K. K., Levrier, F., Obreschkow, D., Rigopoulou, D., Rawlings, S., Smirnov, O. M., Taylor, A. C., Verma, A., Dunlop, J., Santos, M. G., Stanway, E. R., and Willott, C. (2011). MESMER: MeerKAT Search for Molecules in the Epoch of Reionization. *ArXiv e-prints*.

Hotelling, H. (1933). Analysis of a Complex of Statistical Variables into Principal Components. *Journal of Educational Psychology*, **24**(6 & 7), 417–441 & 498–520.

Hu, W. (2003). CMB temperature and polarization anisotropy fundamentals.

Hubble, E. (1929). A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15:168–173.

Hyvärinen, A. (1999). Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. *IEEE Trans. on Neural Network.*, 10:626–634.

Hyvärinen, A. and Oja, E. (2000). Independent Component Analysis: Algorithms and Applications. *Neural Netw*, 13:411–430.

Impellizzeri, C. M. V., McKean, J. P., Castangia, P., Roy, A. L., Henkel, C., Brunthaler, A., and Wucknitz, O. (2008). A gravitationally lensed water maser in the early Universe. Nature, 456:927–929.

Jarvis, M. J., Taylor, A. R., Agudo, I., Allison, J. R., Deane, R. P., Frank, B., Gupta, N., Heywood, I., Maddox, N., McAlpine, K., Santos, M. G., Scaife, A. M. M., Vaccari, M., Zwart, J. T. L., Adams, E., Bacon, D. J., Baker, A. J., Bassett, B. A., Best, P. N., Beswick, R., Blyth, S., Brown, M. L., Bruggen, M., Cluver, M., Colafranceso, S., Cotter, G., Cress, C., Dave, R., Ferrari, C., Hardcastle, M. J., Hale, C., Harrison, I., Hatfield, P. W., Klockner, H.-R., Kolwa, S., Malefahlo, E., Marubini, T., Mauch, T., Moodley, K., Morganti, R., Norris, R., Peters, J. A., Prandoni, I., Prescott, M., Oliver, S., Oozeer, N., Rottgering, H. J. A., Seymour, N., Simpson, C., Smirnov, O., Smith, D. J. B., Spekkens, K., Stil, J., Tasse, C., van der Heyden, K., Whittam, I. H., and WIlliams, W. L. (2017). The MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE) Survey. *ArXiv e-prints*.

Jelić, V., Zaroubi, S., Labropoulos, P., Bernardi, G., de Bruyn, A. G., and Koopmans, L. V. E. (2010). Realistic simulations of the Galactic polarized foreground: consequences for 21-cm reionization detection experiments. MNRAS, 409:1647–1659.

Jolliffe, I. T. (2002). *Principal Component Analysis.* Second ed. Springer Series in Statistics., New York: Springer-Verlag New York.

Kaiser, N. (1987). Clustering in real space and in redshift space. MNRAS, 227:1–21.

Keihänen, E., Kurki-Suonio, H., and Poutanen, T. (2005). MADAM- a map-making method for CMB experiments. MNRAS, 360:390–400.

Kitaura, F.-S., Jasche, J., and Metcalf, R. B. (2010). Recovering the non-linear density field from the galaxy distribution with a Poisson-lognormal filter. MNRAS, 403:589–604.

Lahav, O. and Liddle, A. R. (2014). The Cosmological Parameters 2014. *ArXiv e-prints*.

Lewis, A., Challinor, A., and Lasenby, A. (2000). Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models. ApJ, 538:473–476.

Loeb, A. and Wyithe, J. S. B. (2008). Possibility of Precise Measurement of the Cosmological Power Spectrum with a Dedicated Survey of 21cm Emission after Reionization. *Physical Review Letters*, 100(16):161301.

Maartens, R., Abdalla, F. B., Jarvis, M., Santos, M. G., and SKA Cosmology SWG, f. t. (2015). Cosmology with the SKA – overview. *ArXiv e-prints*.

Masui, K. W., Schmidt, F., Pen, U.-L., and McDonald, P. (2010). Projected constraints on modified gravity cosmologies from 21 cm intensity mapping. Phys. Rev. D, 81(6):062001.

Masui, K. W., Switzer, E. R., Banavar, N., Bandura, K., Blake, C., Calin, L.-M., Chang, T.-C., Chen, X., Li, Y.-C., Liao, Y.-W., Natarajan, A., Pen, U.-L., Peterson, J. B., Shaw, J. R., and Voytek, T. C. (2013). Measurement of 21 cm Brightness Fluctuations at z ˜ 0.8 in Cross-correlation. ApJ, 763:L20.

Mauch, T., Murphy, T., Buttery, H. J., Curran, J., Hunstead, R. W., Piestrzynski, B., Robertson, J. G., and Sadler, E. M. (2003). SUMSS: a wide-field radio imaging survey of the southern sky - II. The source catalogue. MNRAS, 342:1117–1130.

Miville-Deschênes, M.-A., Lagache, G., Boulanger, F., and Puget, J.-L. (2007). Statistical properties of dust far-infrared emission. A&A, 469:595–605.

National Aeronautics and Space Administration (2016). Active Galaxies. https://imagine.gsfc.nasa.gov/science/objects/active_galaxies1.html. Accessed 2017-06-27.

NRAO (2008). Some aerial views of the vla. http://www.aoc.nrao.edu/epo/puente/views/vlaviews.index.html. Accessed: 2017-10-01.

Nunhokee, C. D., Bernardi, G., Kohn, S. A., Aguirre, J. E., Thyagarajan, N., Dillon, J. S., Foster, G., Grobler, T. L., Martinot, J. Z. E., and Parsons, A. R. (2017). Constraining Polarized Foregrounds for EOR Experiments II: Polarization Leakage Simulations in the Avoidance Scheme. *ArXiv e-prints*.

Olivari, L. C., Remazeilles, M., and Dickinson, C. (2016). Extracting H I cosmological signal with generalized needlet internal linear combination. MNRAS, 456:2749–2765.

Pacholczyk, A. G. (1970). *Radio astrophysics. Nonthermal processes in galactic and extragalactic sources*. Freeman.

Padovani, P., Alexander, D. M., Assef, R. J., De Marco, B., Giommi, P., Hickox, R. C., Richards, G. T., Smolčić, V., Hatziminaoglou, E., Mainieri, V., and Salvato, M. (2017). Active galactic nuclei: what's in a name? A&A Rev., 25:2.

Patanchon, G., Ade, P. A. R., Bock, J. J., Chapin, E. L., Devlin, M. J., Dicker, S., Griffin, M., Gundersen, J. O., Halpern, M., Hargrave, P. C., Hughes, D. H., Klein, J., Marsden, G., Martin, P. G., Mauskopf, P., Netterfield, C. B., Olmi, L., Pascale,

E., Rex, M., Scott, D., Semisch, C., Truch, M. D. P., Tucker, C., Tucker, G. S., Viero, M. P., and Wiebe, D. V. (2008). SANEPIC: A Mapmaking Method for Time Stream Data from Large Arrays. ApJ, 681:708–725.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine, Series*, **2**(11), 559-572.

Penzias, A. A. and Wilson, R. W. (1965). A Measurement of Excess Antenna Temperature at 4080 Mc/s. ApJ, 142:419–421.

Perley, R. A., Dreher, J. W., and Cowan, J. J. (1984). The jet and filaments in Cygnus A. ApJ, 285:L35–L38.

Perlmutter, S., Turner, M. S., and White, M. (1999). Constraining Dark Energy with Type Ia Supernovae and Large-Scale Structure. *Physical Review Letters*, 83:670–673.

Peterson, J. B., Aleksan, R., Ansari, R., Bandura, K., Bond, D., Bunton, J., Carlson, K., Chang, T.-C., DeJongh, F., Dobbs, M., Dodelson, S., Darhmaoui, H., Gnedin, N., Halpern, M., Hogan, C., Le Goff, J.-M., Liu, T. T., Legrouri, A., Loeb, A., Loudiyi, K., Magneville, C., Marriner, J., McGinnis, D. P., McWilliams, B., Moniez, M., Palanque-Delabruille, N., Pasquinelli, R. J., Pen, U.-L., Rich, J., Scarpine, V., Seo, H.-J., Sigurdson, K., Seljak, U., Stebbins, A., Steffen, J. H., Stoughton, C., Timbie, P. T., Vallinotto, A., and Teche, C. (2009). 21-cm Intensity Mapping. In *astro2010: The Astronomy and Astrophysics Decadal Survey*, volume 2010 of *ArXiv Astrophysics e-prints*.

Petroff, E. (2017). Fast radio bursts: recent discoveries and future prospects. *ArXiv e-prints*.

Planck Collaboration, Ade, P. A. R., Aghanim, N., Alves, M. I. R., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., Aumont, J., Aussel, H., and et al. (2014a). Planck 2013 results. I. Overview of products and scientific results. A&A, 571:A1.

Planck Collaboration, Ade, P. A. R., Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., Aumont, J., Baccigalupi, C., Banday, A. J., and et al. (2014b). Planck 2013 results. XVI. Cosmological parameters. A&A, 571:A16.

Planck Collaboration, Ade, P. A. R., Aghanim, N., Arnaud, M., Ashdown, M., Aumont, J., Baccigalupi, C., Banday, A. J., Barreiro, R. B., Bartlett, J. G., and et al. (2015). Planck 2015 results. XIII. Cosmological parameters. *ArXiv e-prints*.

Porcas, R. W., Booth, R. S., Browne, I. W. A., Walsh, D., and Wilkinson, P. N. (1981). VLBI structures of the images of the double QSO 0957+561. Nature, 289:758–762.

Reich, P. and Reich, W. (1988). A map of spectral indices of the Galactic radio continuum emission between 408 MHz and 1420 MHz for the entire northern sky. A&AS, 74:7–23.

Reich, W. (1990). The large-scale magnetic field structure near the Galactic centre. In Beck, R., Wielebinski, R., and Kronberg, P. P., editors, *Galactic and Intergalactic Magnetic Fields*, volume 140 of *IAU Symposium*, pages 369–372.

Riess, A. G., Filippenko, A. V., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P. M., Gilliland, R. L., Hogan, C. J., Jha, S., Kirshner, R. P., Leibundgut, B., Phillips,

M. M., Reiss, D., Schmidt, B. P., Schommer, R. A., Smith, R. C., Spyromilio, J., Stubbs, C., Suntzeff, N. B., and Tonry, J. (1998). Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. AJ, 116:1009–1038.

Rybicki, G. B. and Lightman, A. P. (1979). *Radiative processes in astrophysics*. Wiley-Interscience.

Sancisi, R., Fraternali, F., Oosterloo, T., and van der Hulst, T. (2008). Cold gas accretion in galaxies. A&A Rev., 15:189–223.

Santos, M., Alonso, D., Bull, P., Silva, M. B., and Yahya, S. (2015a). HI galaxy simulations for the SKA: number counts and bias. *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, page 21.

Santos, M., Bull, P., Alonso, D., Camera, S., Ferreira, P., Bernardi, G., Maartens, R., Viel, M., Villaescusa-Navarro, F., Abdalla, F. B., Jarvis, M., Metcalf, R. B., Pourtsidou, A., and Wolz, L. (2015b). Cosmology from a SKA HI intensity mapping survey. *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, page 19.

Santos, M. G., Bull, P., Alonso, D., Camera, S., Ferreira, P. G., Bernardi, G., Maartens, R., Viel, M., Villaescusa-Navarro, F., Abdalla, F. B., Jarvis, M., Metcalf, R. B., Pourtsidou, A., and Wolz, L. (2015c). Cosmology with a SKA archivePrefix = "arXiv", HI intensity mapping survey. *ArXiv e-prints*.

Santos, M. G., Cluver, M., Hilton, M., Jarvis, M., Jozsa, G. I. G., Leeuw, L., Smirnov, O., Taylor, R., Abdalla, F., Afonso, J., Alonso, D., Bacon, D., Bassett, B. A., Bernardi, G., Bull, P., Camera, S., Chiang, H. C., Colafrancesco, S., Ferreira, P. G., Fonseca, J., van der Heyden, K., Heywood, I., Knowles, K., Lochner, M., Ma, Y.-Z., Maartens, R., Makhathini, S., Moodley, K., Pourtsidou, A., Prescott, M., Sievers, J., Spekkens, K., Vaccari, M., Weltman, A., Whittam, I., Witzemann, A., Wolz, L., and Zwart, J. T. L. (2017). MeerKLASS: MeerKAT Large Area Synoptic Survey. *ArXiv e-prints*.

Santos, M. G., Cooray, A., and Knox, L. (2005). Multifrequency Analysis of 21 Centimeter Fluctuations from the Era of Reionization. ApJ, 625:575–587.

Seo, H.-J. and Eisenstein, D. J. (2007). Improved Forecasts for the Baryon Acoustic Oscillations and Cosmological Distance Scale. ApJ, 665:14–24.

SKA Organization (2017a). The basic physics of how radio emission works. https://skatelescope.org/radio-astronomy/. Accessed: 2017-10-01.

SKA Organization (2017b). The square kilometre array. http:www.skatelescope.org. Accessed: 2017-10-19.

Smidt, J., Whalen, D. J., Johnson, J. L., and Li, H. (2017). The Formation of the First Quasars in the Universe. *ArXiv e-prints*.

Smoot, G. (1994). Background Radiation: Space Missions, Cosmic Microwave Background. *Evolution of the Universe and its Observational Quest*, pages 115–136.

Smoot, G. F. (1991). COBE - New sky maps of the early universe. In Sato, K. and Audouze, J., editors, *Primordial Nucleosynthesis and Evolution of Early Universe*, volume 169 of *Astrophysics and Space Science Library*, pages 281–294.

Smoot, G. F. (1992). COBE DMR results and implications. In Signore, M. and Dupraz, C., editors, *NATO Advanced Science Institutes (ASI) Series C*, volume 359 of *NATO Advanced Science Institutes (ASI) Series C*, pages 331–344.

Smoot, G. F. and Debono, I. (2017). 21 cm intensity mapping with the Five hundred metre Aperture Spherical Telescope. A&A, 597:A136.

Spitler, L. G., Cordes, J. M., Hessels, J. W. T., Lorimer, D. R., McLaughlin, M. A., Chatterjee, S., Crawford, F., Deneva, J. S., Kaspi, V. M., Wharton, R. S., Allen, B., Bogdanov, S., Brazier, A., Camilo, F., Freire, P. C. C., Jenet, F. A., Karako-Argaman, C., Knispel, B., Lazarus, P., Lee, K. J., van Leeuwen, J., Lynch, R., Ransom, S. M., Scholz, P., Siemens, X., Stairs, I. H., Stovall, K., Swiggum, J. K., Venkataraman, A., Zhu, W. W., Aulbert, C., and Fehrmann, H. (2014). Fast radio burst discovered in the arecibo pulsar alfa survey. *The Astrophysical Journal*, 790(2):101.

Strom, R. G. (1973). Faraday Depolarization of Radio Galaxies and Quasars with Simple Spectra. A&A, 25:303.

Taylor, A. R., Stil, J. M., and Sunstrum, C. (2009). A Rotation Measure Image of the Sky. ApJ, 702:1230–1236.

Tegmark, M. (1997). How to Make Maps from Cosmic Microwave Background Data without Losing Information. ApJ, 480:L87–L90.

Tucci, M., Martínez-González, E., Toffolatti, L., González-Nuevo, J., and De Zotti, G. (2004). Predictions on the high-frequency polarization properties of extragalactic radio sources and implications for polarization measurements of the cosmic microwave background. MNRAS, 349:1267–1277.

Waelkens, A., Jaffe, T., Reinecke, M., Kitaura, F. S., and Enßlin, T. A. (2009). Simulating polarized Galactic synchrotron emission at all frequencies. The Hammurabi code. A&A, 495:697–706.

Wilman, R. J., Miller, L., Jarvis, M. J., Mauch, T., Levrier, F., Abdalla, F. B., Rawlings, S., Klöckner, H.-R., Obreschkow, D., Olteanu, D., and Young, S. (2008). A semi-empirical simulation of the extragalactic radio continuum sky for next generation radio telescopes. MNRAS, 388:1335–1348.

Zaroubi, S. (2013). The Epoch of Reionization. In Wiklind, T., Mobasher, B., and Bromm, V., editors, *The First Galaxies*, volume 396 of *Astrophysics and Space Science Library*, page 45.

Zwicky, F. (1933). Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6:110–127.