UNIVERSITY OF THE WESTERN CAPE

MASTERS' THESIS

Detecting Galaxies in the Epoch of Reionization

Author:
Siyanda MATIKA

Supervisor:
Prof. Mario SANTOS
Prof. Roy Marteens
Dr. J. Zwart

UNIVERSITY *of the*
WESTERN CAPE

A thesis submitted in fulfillment of the requirements
for the degree of MSc Astrophysics
in the

Astronomy group
Department of Physics and Astronomy

August 17, 2017

Declaration of Authorship

I, Siyanda MATIKA, declare that this thesis titled, "Detecting Galaxies in the Epoch of Reionization" 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: 04/04/17

"Statisticians, like artists, have the bad habit of falling in love with their models."

George Box

UNIVERSITY OF THE WESTERN CAPE

Abstract

FACULTY NATURAL SCIENCE
Department of Physics and Astronomy

MSc Astrophysics

Detecting Galaxies in the Epoch of Reionization

by Siyanda MATIKA

The aim of this project was to develop a technique that allows us to measure the redshift of radio galaxies in the Epoch of Reionization. We achieved this by constructing a parameterization of the 21-cm absorption feature which is expected to be seen in the spectra of high redshift radio galaxies. A Bayesian framework for parameter estimation and model selection was applied on the candidate models found. The method's proficiency is demonstrated with two data sets: SIMFAST21-generated & model-generated data. The SIMFAST21-generated data was constructed using a simulation of the Epoch of Reionization. The model-generated data was synthesized from our best analytical model for the 21-cm absorption feature. A spectrum template from radio galaxies at low redshifts was then combined with the data sets. The resulting spectra are fitted using a model that accounts for 21-cm absorption feature. The parameter fitting was done using the expected SKA1-LOW noise. A strong detection can be achieved for galaxies with redshifts $z_g > 12$ and luminosities $L > 10^{26}$[W/Hz]. Objects at lower z, although with brighter flux, are harder to detect due to the weak absorption feature (e.g. less intervening HI). A detection proxy grid space was assembled to quantify model evidence found when fitting data with the best model and null hypothesis (no absorption). To interpret the detection proxy, the Jeffreys scale was used. The grid for both data sets was overlaid with the SKADS simulated sources. In a 25 deg^2 sky area, a total of 62206 "SKADS" radio galaxies above redshift 7 are expected to be detected in continuum by SKA-LOW. From these, about 1% should be confirmed to be in the epoch of reionization by SKA-LOW (with the same survey) through the detection of the 21 cm absorption feature using the proposed technique.

Acknowledgments

First and foremost I would like to thank my supervisors Mario Santos, Roy Maartens, Jon Zwart for their support and guidance over the past two years, for giving me the opportunity to work on a interesting project. I'm especially grateful to JZ who has proofread my work several times and added a great depth and breadth to my research project. Thanks to Marta, Ze, Song and Prina for coding Café sessions. I extend special thanks to my fellow students Sibonelo, Clemence, Eliab, Siyambonga, Sheean. I thank the National astrophysics and space science program (NASSP) and SKA Africa for providing financial assistance during my studies. I would like to thank all those in the Astrophysics Group and at my institution, University of the Western Cape. I am grateful to the many people who have given me feedback and advice on my research either directly or indirectly in conversations. I would like to acknowledge the use of the Centre for High Performance Computing (CHPC) facility, I thank Sean February for his care and efficiency in maintaining the facility and assisting me with many enquiries. I am very grateful to CENTRA/IST for the computer resources. The authors thankfully acknowledge the computer resources, technical expertise and assistance provided by CENTRA/IST. Computations were performed at the cluster "Baltasar-Sete-Sóis" and supported by the H2020 ERC Consolidator Grant "Matter and strong field gravity: New frontiers in Einstein's theory" grant agreement no. MaGRaTh-646597.

Contents

Declaration of Authorship iii

Abstract vii

Acknowledgments ix

1 Introduction 1
- 1.1 The Universe in a nutshell . . . 2
- 1.2 Essential Cosmology . . . 4
- 1.3 Epoch of Reionization . . . 7
  - 1.3.1 High redshift Radio galaxies . . . 8
- 1.4 21 cm signal . . . 12
  - 1.4.1 Brightness Temperature . . . 14
  - 1.4.2 Gunn Peterson trough . . . 17
  - 1.4.3 CMB experiments . . . 19
- 1.5 21 cm Forest . . . 21
- 1.6 Experiments . . . 24
- 1.7 Outline . . . 28

2 Model Fitting 29
- 2.1 Bayesian Inference . . . 30
- 2.2 Markov chain Monte Carlo methods . . . 32
- 2.3 Nested Sampling . . . 33

3 Simulation Software 35
- 3.1 SIMFAST21 . . . 36
- 3.2 SKADS . . . 39
- 3.3 21cmNEST . . . 40

4 21 cm Absorption 42
- 4.1 Galaxy Flux . . . 44
- 4.2 Absorption Models . . . 46

5 Fitting the 21 cm Absorption Feature **52**
5.1 Systematic tests in Modelling 52
5.1.1 Flat Noise Profile 53
5.1.2 SKA1-LOW Noise Profile 75
5.2 Results 84
5.2.1 Posterior Distributions 85
5.2.2 Reconstructions and redshift histograms 89
5.2.3 Detection space 99

6 Conclusions and future work **108**
6.1 Future work 110

Bibliography **111**

Chapter 1

Introduction

The first luminous objects formed roughly half a billion years after the Big Bang. Observing this epoch would give insight as to how the first galaxies formed after the Big Bang. Probing the early Universe would answer questions about the driving forces that had a significant impact on the rapid evolution of the star formation rate density which peaks at a redshift of 2.5 (Burgarella et al., 2013). Recent studies have attributed this to a decline in the production of stars in the Universe as a whole (Reddy and Steidel, 2009). Studies have uncovered that the star formation rate has been continuously declining over the last 11 billion years, being 30 times lower today than at its likely peak (Madau and Dickinson, 2014).

Studying this period would undoubtedly help us in better understanding as to how the Universe has evolved over time. This would clear up many mysteries in one of the Universe's vital stages in it's evolution which is known as the Epoch of Reionization. The period where neutral hydrogen (HI) in the Universe began to ionize. The photons responsible for ionizing the Universe are believed to have come mostly from star formation (Springel, Frenk, and White, 2006; Bromm et al., 2009).

Reionization is the last phase change: the Universe moves from a neutral state that contains clouds of hydrogen to an ionized state with protons and electrons. Understanding the mechanisms at play at this stage is key to tracing out how the Universe evolved. As a consequence, the most distant galaxies at high redshift appear to an observer as faint objects, even if their luminosity is large. Only a very small fraction of galaxies can be observed in a deep image of the sky. Due to sensitivity limitations in current telescopes our knowledge about the evolution of the Universe, decreases as we go to higher redshifts. A wide variety of instruments are currently under construction, these include the European Extremely Large Telescope (EELT), James Webb Space Telescope (JWST) and Square Kilometer Array (SKA), a radio telescope which aims to detect the redshifted 21 centimeter (cm) radiation emitted by neutral hydrogen in the early Universe.

Currently, there are many simulations of the EoR, which are a mixture of numerics and analytics, to get the benefits of both. Numerical simulations are needed to keep track of many different physical processes, which often have feedback processes as well as non-linear behaviour. Numerical simulations are however computationally expensive, especially with the large dynamic range needed for reionization. The simulation should go from roughly 100 Mpc to kpc scales to model both the large scale structure as well as galaxies. This large dynamic range is computationally very expensive. This is where it is useful to add analytic and semi-analytic methods to a simulation. They contain physical and astronomical arguments as well as assumptions to make it possible to relate different quantities. This assignment of luminosities is usually done using observations not simulations

1.1 The Universe in a nutshell

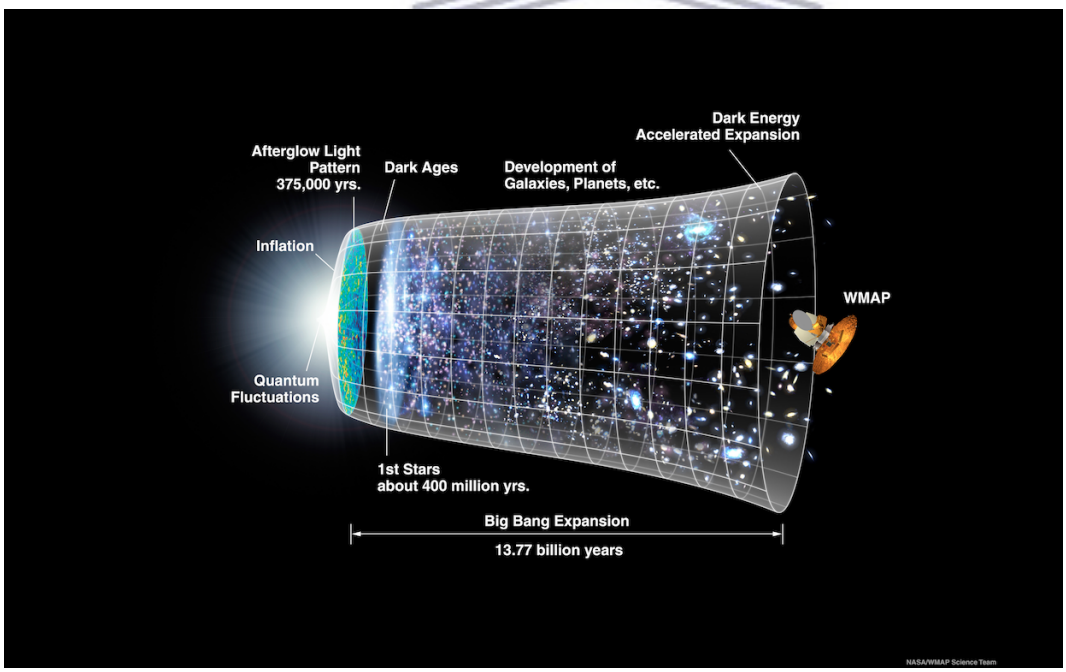


Fig. 1.1: Schematic showing the evolution of the Universe. After the radiation era, the Universe passed through the matter-radiation equality and entered the matter dominated era. Matter and radiation are diluted as the Universe expanded. The density remains constant to a good approximation. It is only in the present era that dark energy has begun to dominate the Universe. Image credit: NASA / WMAP Science Team

It is assumed that the Universe underwent rapid expansion in all directions at the beginning, in an era known in a era known as Inflation (Linde, 1982). This rapid expansion expansion froze quantum fluctuations in the vacuum which formed primordial

density fluctuations which are understood today to be the seeds of large scale structure (Gawiser and Silk, 1998). After that the Universe enters the standard phase called the Hot Big Bang, composed of an extremely hot and dense plasma state [1].

This plasma was dominated by interactions between baryonic matter and photons. Photons were scattering off electrons in a process known as Thomson scattering, hence they were unable to travel a significant distance. During this era the Universe was opaque to electromagnetic radiation, precluding any possibility for direct imaging of its evolution. This stage was dominated by two competing forces: the force of gravity, with matter trying to form clumps, and the radiation pressure from photons trying to escape. These competing opposing forces, induced oscillations which arose in the plasma. The oscillations moved at the speed of sound giving a rise to a preferred scale defined as the sound horizon (r_s); which is the distance a sound wave could have travelled since the very early Universe (Eisenstein, Zehavi, and Hogg, 2004; Jiang et al., 2006).

Over time the Universe had expanded enough to allow electrons and protons to combine ('recombination'). Photons were no longer scattering off electrons and the plasma waves came to a halt. Photons propagated freely for the first time in what we call the last scattering surface (Hu, 2003), after more expansion of the Universe this later become the Cosmic Microwave Background (CMB) (Penzias and Wilson, 1965). As the baryons wave stalled, a distinctive radius of the spherical shell formed (Eisenstein, 2005). An imprint on the distribution of matter as a density excess was left over. Hence today Baryonic acoustic oscillation are understood to be the remnants of a primordial photon-baryon plasma interaction that left a distinctive imprint in the distribution of matter in the Universe today (Eisenstein, Zehavi, and Hogg, 2004).

The imprint this process left in the Universe can be used to constrain the geometry and dynamics of the Universe. Half a million years after the Big Bang, the constituents of hydrogen recombined and the Universe became transparent (Komatsu et al., 2009). Currently, the furthest we can electromagnetically observe is up until a snapshot of the Universe at recombination from the CMB, with detailed images of its evolution starting from an age of a billion years until the present time (Larson et al., 2010)[2].

[1] Kolb and Turner, 1990; Weinberg, 2008 are some notable sources that explain in depth the BIg Bang theory.

[2] We probe the evolution of the Universe using detailed images of galaxies.

1.2 Essential Cosmology

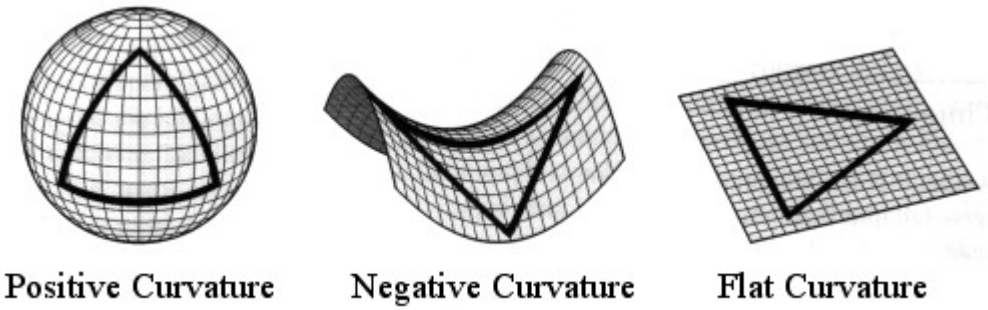


Fig. 1.2: The possible curvatures of space time in a Robertson Walker Universe.

We now describe the properties of adiabatic cold dark matter (CDM) model. The metric for space which is considered spatially isotropic and homogenous is the Friedmann Lemaître Robertson-Walker metric:

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

where (R, θ, ϕ) are spherical comoving coordinates, and $a(t)$, the scale factor which measures expansion of the Universe over time. k determines the geometry of the metric or simply the curvature of space time, (+: closed Universe, 0: flat Universe, $-$: open Universe). Observers at rest remain at some fixed (R, θ, ϕ), but their separation increases in time in proportion to $a(t)$. If an observer sees a nearby source emitting light at distance D, it moves away with the Hubble velocity $H(t)D$, where the Hubble constant at time t is defined as $H(t) \equiv da(t)/dt$. The light emitted by a source at time t is observed today with redshift $z \equiv 1/a(t) - 1 \equiv \frac{\lambda_0 - \lambda}{\lambda}$, where we set a(t= today)=1. λ_0 is the wavelength of the photon on arrival at the observer today, λ is the emission wavelength from the source.

From Einstein's field equations, we are able to derive the Friedman equation for an expanding Universe, which is defined as:

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2}. \tag{1.2}$$

This equation describes how the scale factor evolves, given its contents. For a given density, ρ, we can define the density parameter $\Omega = \rho/\rho_C$, where ρ_C is the critical density

defined by $\rho_C \equiv \frac{3H_0^2}{8\pi G}$. The Friedman equation can then be expressed as:

$$H^2(z) = H_0^2\Big[\Omega_m(1+z)^3 + \Omega_r(1+z)^4 + \Omega_k(1+z)^2 + \Omega_\Lambda\Big]; \tag{1.3}$$

where $H_0 \equiv H(z=0)$, is the Hubble expansion rate today. We also define another quantity known as the Hubble distance today, $D_H \equiv \frac{c}{H_0} \sim 2997.9\ h^{-1}Mpc$.

The contribution of spatial curvature to the expansion rate is $\Omega_k = 1 - \sum_i \Omega_i$ and Ω_Λ is the cosmological constant. The matter density is $\Omega_m = \Omega_c + \Omega_b$, where Ω_c represents cold dark matter and Ω_b represents baryonic matter .

The comoving distance towards an object at redshift z, is:

$$\chi(z) = D_H \int_0^z \frac{dz'}{H(z')}. \tag{1.4}$$

We also need two other distances frequently used in cosmology, angular-diameter distance and the luminosity distance. The angular diameter distance D_A is the ratio of an object's physical transverse size to its angular size (in radians). The luminosity distance D_L is defined by the relationship between an objects bolometric flux S to its intrinsic luminosity L:

$$\begin{aligned} D_A(z) &\equiv \frac{1}{1+z} S_k(\chi), \\ D_L(z) &\equiv (1+z) S_k(\chi), \end{aligned} \qquad S_k \equiv \begin{cases} \frac{D_H}{\sqrt{\Omega_k}} \sinh(\sqrt{\Omega_k}\frac{\chi}{D_H}), & \text{for } \Omega_k > 0, \\ \chi, & \text{for } \Omega_k = 0, \\ \frac{D_H}{\sqrt{\Omega_k}} \sin(\sqrt{\Omega_k}\frac{\chi}{D_H}), & \text{for } \Omega_k < 0. \end{cases}$$

According to the latest measurements $\Omega_k \approx 0$ (Planck Collaboration et al., 2015). The above equations can then be simplified to:

$$D_A(z) = \frac{1}{1+z}\chi = \frac{1}{1+z}\int_0^z \frac{dz'}{H(z')}, \tag{1.5}$$

$$D_L(z) = (1+z)\chi = (1+z)\int_0^z \frac{dz'}{H(z')}. \tag{1.6}$$

Using the previous expressions for luminosity and angular distance this leads to the relation:

$$D_L = D_A(1+z)^2 \tag{1.7}$$

Only if the Universe is not expanding, are these equal.
Let's assume that a radiant flux S is observed for a source of light with known bolometric luminosity L. [3].

Lets consider a galaxy whose rest-frame spectral energy given is by $L(\nu)$, which is defined so that $L(\nu)\mathrm{d}\nu$ is the energy emitted by the frequency range $\nu \rightarrow \nu+\mathrm{d}\nu$ per unit time and the observed flux is denoted by $S(\nu_0)\ \mathrm{d}\nu_0$. We account for the k-corrections by considering a scenario that shows two (the observer's frame is denoted by a subscript o and rest-frame is denoted by o's absence) frames can be related as $\nu_0 S_0 = \frac{\nu L_\nu}{4\pi D_L^2}$. This expression is re-arranged using the relation between the emitted and observed frequency $\nu = (1+z)\nu_0$. Combining this parameter eventually leads us to:

$$S(\nu_0) = \frac{L([1+z]\nu_0)}{4\pi\chi^2(z)[1+z]}. \tag{1.8}$$

$L([1+z]\nu_0)$ is measured in units of [W/Hz]. This equation is key for probing the expansion history of the Universe. For a class of objects of fixed intrinsic luminosity, distributed over various redshifts, the distance can be predicted for a given cosmology and compared with observed measurements to identify which cosmological model gives the best description of the observed Universe. Our main interest here was to derive the spectral energy distribution (SED) equation used to describe radio galaxies.

Our main cosmological interest in the work presented in this thesis is in verifying whether is possible to use existing low redshift templates along with a expected reionization feature to get redshift estimates for galaxies located within the Epoch of reionization.

[3] For a thorough derivation of these crucial cosmological parameters see *Cosmological Physics* by John A. Peacock

1.3 Epoch of Reionization

The Epoch of Reionization (EoR) is a period where neutral hydrogen (HI) in the Universe was ionized. The process is thought to have occurred about a billion years after the Big Bang. Current theory suggests that the first galaxies began heating up and ionizing nearby regions, forming disconnected ionized patches. As more stars formed, more X-ray and UV radiation was readily available to heat and ionize regions further away and the ionized patches began to combine with the neutral regions disappearing. This era is filled with complex cosmological and astrophysical processes such as the formation of large scale structures and the birth and evolution of stars. The epoch of reionization is a vital stage in the Universe's development.

Reionization is thought to take place between redshifts of $6 < z < 10$. This range is within reach of high-redshift experiments that are currently being built. Reionization may have begun at least as early as z$\sim$11 Planck Collaboration et al. (2015), halfway around z$\sim$9, and finished no later than $z \sim 6$ (Fan et al., 2005; Ouchi et al., 2010; McGreer, Mesinger, and Fan, 2011).

Details of how reionization occurred are still poorly understood. Even the duration of reionization is uncertain. The how and when the Universe was reionized, are major questions in modern cosmology that remain unanswered. Due to limitations in the sensitivity of our current telescopes in doing galaxy surveys, our knowledge declines as we approach higher redshifts. Arguably the most promising probe of this epoch is the signal uniquely found in hydrogen due to its hyperfine state, which emits radiation at a frequency of 1420 MHz (this state is better known with its corresponding wavelength which is 21 cm).

The currently accepted theory suggests that star formation in galaxies was the cause for reionization. Both stars and black holes contribute photons responsible for ionizing the intergalactic medium (IGM), but the early Universe is dominated by small galaxies which, in the local Universe, have disproportionately small central black holes. Observations of quasars at z$>$ 6 are quite rare which would support the theory suggesting that most of the photons ionizing IGM came from galaxies (Zaroubi, 2012).

Recent observations in optical and near infra-red have detected galaxies at redshifts 8 and 12 (Robertson et al., 2013a). According to Barkana and Loeb (2001), these observations infer that a large number of galaxies had already developed and star formation had started some time before EoR. However, these detected galaxies could not single

handily caused reionization. The current horizon of observations at z∼7, the observed population of galaxies fails by only a few factors to reionize the IGM. It is natural to attribute the missing ionizing emissivity to fainter galaxies, just below the current detection threshold. In support of such an extrapolation, there is a firm upper limit on the contribution from faint (individually undetectable) quasars to reionize the Universe by z∼ 6–7 (Labbe et al., 2010).

Observational surveys of galaxies in the distant Universe, therefore, provide us with our best guide to mapping how the volume density of ionizing radiation likely varies with both redshift and cosmic time. The sensitive observations from the Hubble Space Telescope and wide field observations are key to obtaining these improved constraints on the volume densities. Quantifying the volume density of ultra-faint galaxies is particularly important for determining the impact of galaxies in reionizing the Universe.

Currently modelling this period seems to be the only tool we able to use in testing existing theories on reionization. The use of numerical simulations have become imperative in probing more effectively the cosmos for low redshifts. At high redshifts, numerical simulations are used to better understand the underlying physics in the early stages of the Universe. They not only provide us with accurate and precise measurements of how the Universe has evolved, they are equally crucial to data analysis and the verification of current theoretical models. These simulations are vital in the planning of observations and can be used to constrain the reionisation processes by comparing to data.

1.3.1 High redshift Radio galaxies

The spectra of radio galaxies and quasars are characterised by the emission of synchrotron radiation from relativistic electrons having a power law distribution of electron energies with a magnetic field strength $B \sim 10^{-5}$ gauss. The free-free emission emerges directly from H II regions containing the ionizing stars.

Thermal re-radiating of starlight by dust overwhelms these components above $\nu \sim$ 200 GHz, defining a practical upper bound to the frequencies of "radio" observations. The radio emission is divided into two categories: the extended structure, which is transparent, and the compact structure, where the density of relativistic electrons is so great that the source becomes opaque to its own radiation. Most compact sources are identified with quasi-stellar objects (QSOs) or with active galactic nuclei (AGN). The extended sources are typically associated with galaxies, but many are quasars with no

visible optical extent. Most extended sources, particularly quasars, when examined with sufficient sensitivity and resolution, are found to contain a compact central radio component. Because the compact sources are affected by self-absorption, their spectra are flat. They are therefore most easily detected by radio surveys made at short wavelengths, whereas the steep-spectrum extended sources with their transparent spectra are characteristic of long-wavelength surveys.

Radio galaxies at the highest redshifts, particularly in the Epoch of Reionization (EoR), have the potential to be important probes of cosmology. Detecting radio galaxies is an important tool to study the formation and evolution of massive galaxies and large-scale structure in the Universe. The problem that has hindered the discovery of more radio sources at the highest redshifts is not the paucity of deep radio data, but the difficulty of obtaining targeted follow-up spectroscopy of the many candidate sources within a given radio survey. This is due to the fact that powerful radio sources at $z >$ 5 make up $\ll 1\%$ of the total radio source population (Wilman et al., 2008). Studying these radio galaxies will be vital in our understanding and may hold some clues on how the process of reionization unfolded.

A considerable number of surveys have been conducted to find very high-redshift radio galaxies over the last two decades. Searching for these high redshift radio galaxies (HzRGs), has led to the discovery of more than hundred radio galaxies at redshifts z > 2 (see summaries Rawlings, 2000; Miley and De Breuck, 2008). However by z > 3, their numbers become increasingly sparse, and using flux limited radio surveys such as the 3CR (S178 > 10 Jy; Laing, Riley, and Longair, 1983), or the MRC strip (S408 > 0.95 Jy; McCarthy, Patrick J., 1996) and more recent surveys (e.g. De Breuck et al., 2000; De Breuck et al., 2002; McLure et al., 2004; Cruz et al., 2006; Cruz et al., 2007; Brookes et al., 2008) may suggest that there are few bright radio sources around reionization era.

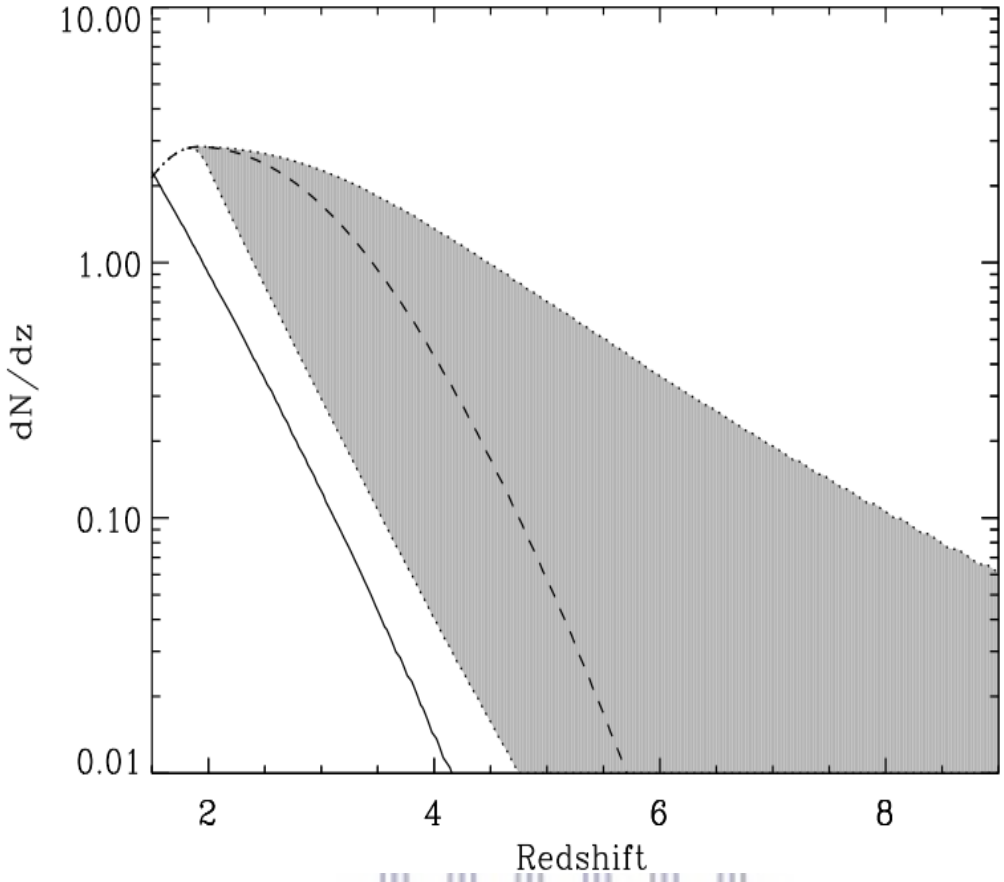


Fig. 1.3: The expected number of radio sources above 10 mJy as a function of redshift per square degree. The dark solid line is the expected number of low luminosity radio sources which would have low-luminosity emission lines and be very difficult to obtain redshifts for . Grey banded area shows the degree of uncertainty in the high-redshift evolution of the FRII sources. Image source (Jarvis and Rawlings, 2000)

It has been discovered that there is a statistical decrease of the angular size of radio sources with redshift (Miley, 1968). It has been noted in the literature that radio sources with very steep spectral indices at low frequencies < 1 GHz tend to be associated with galaxies at high redshift (Blumenthal, G.; Miley, 1979). This empirical correlation between radio spectral steepness and redshift has proved to be an efficient method for finding distant radio galaxies (See Fig.(3.2) for illustration). An accepted explanation for the correlation between z and α is that the correlation is the result of a concave radio spectrum coupled with a radio K-correction. A key mechanism for making the radio spectra concave are synchrotron and inverse Compton losses. Synchrotron and inverse Compton scattering are analogous processes, in inverse Compton scattering the relativistic electrons interact with photons and Thomson scatter them to high energies (Klamer et al., 2006).

At one point the most distant radio galaxy was 6C0140+326 at z = 4.41. But this was later overtaken again by another ultra-steep spectrum search which found a radio galaxy at $z = 5.19$ (Breugel et al., 1999). Attempts to detect higher redshifts radio galaxies have been futile. This may be due to methods such as steep-spectral index and angular size filtering not being 100 per cent efficient and a large number of radio sources at z > 5 may also be filtered out (Cruz et al., 2007). This is alarming since the space density of the z > 5 radio source population is extremely low to start with.

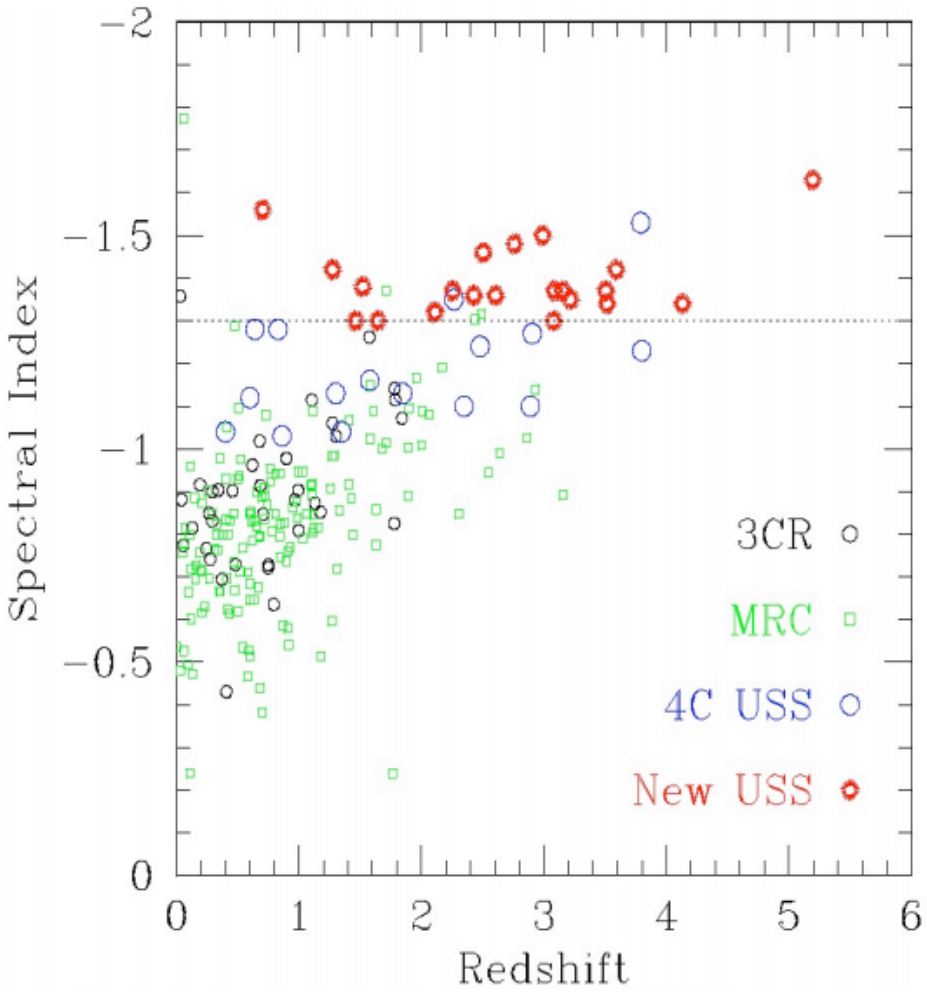


Fig. 1.4: Radio spectral index versus redshift, showing that more distant sources have steeper spectra. From (De Breuck et al., 2000).

1.4 21 cm signal

According to Quantum theory, elementary particles have a property called spin, which is allowed to be integer or half-integer. The ground state of neutral hydrogen is divided into two close states, commonly referred as the hyperfine states. The upper energy state (triplet) corresponds to the case where electron and proton are parallel, and the lower state (singlet), where electron and proton are anti-parallel. The transition from triplet to singlet is accompanied by the emission of radiation corresponding 21 cm. Conversely, the transition from singlet to the triplet corresponds to the absorption of the photon with a wavelength of 21 cm.

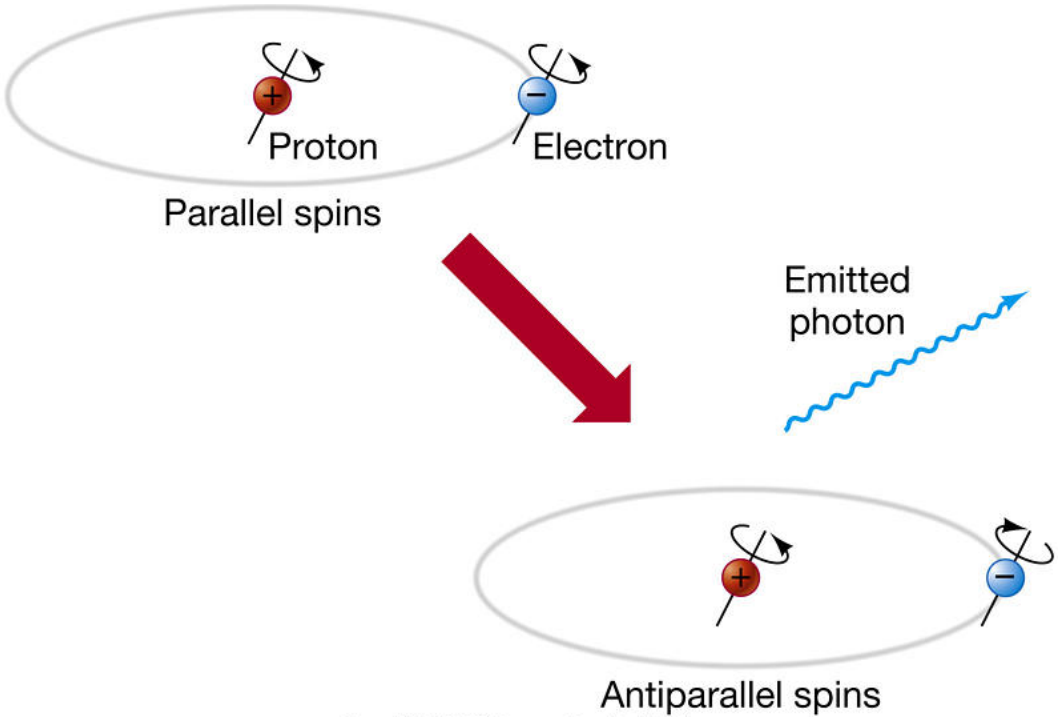


Fig. 1.5: An artist's impression of ground state of neutral hydrogen which are split into two states and makes it detectable to radio telescopes.

The 21 cm transition was theoretically predicted by Hendrick van de Hulst in 1944, and observation of the 21 cm (or 1420MHz) line of hydrogen signalled the beginning of spectral-line radio astronomy. First observed Holland in 1945 Van de Hulst, 1945 followed by observers Harold Ewen and Edward M. Purcell at Harvard and Australia (Ewen and Purcell, 1951). At a frequency of 1420 MHz, the incoming radiation from hydrogen is able to pass through dust clouds and thus provide a complete map of the structure of our galaxy as opposed to stars.

The relative occupancy of the spin levels is usually described in the terms of the hydrogen spin temperature T_S, defined by:

$$n_1/n_0 = 3\exp(-T_*/T_S). \tag{1.9}$$

Where n_0 and n_1 refer respectively to the singlet and triplet hyperfine levels in the atomic ground state (n=1) and $T_* = 0.068K$ is defined by $k_B T_* = E_{21}$, where the energy of the 21 cm is E_{21}=5.9$\times 10^{-6}$eV, corresponding to a photon frequency of 1420 MHz.

At early enough redshifts all the hydrogen in the Universe was basically neutral. This offers the prospect of detecting the 21 cm emission or absorption relative to some background source. Measuring of the redshifted HI 21 cm signal from IGM is a powerful probe and is likely to provide us with a large amount of information about the epoch of reionization (Liu et al., 2012; Mellema et al., 2012) . Using the 21 cm signal we can measure the transitions and from that we can infer the amount of neutral hydrogen in the Universe. The 21 cm transition is the most direct probe of the distribution and evolution of neutral hydrogen in the Universe (Ciardi et al., 2012).

Below we show the radiative transfer equation for a source emitting radiation that passes through clouds. We make use of radiative transfer for a light source with intensity I_ν, while the scattering is assumed to be negligible;

$$\frac{dI_\nu}{ds} = -\alpha_\nu I_\nu + j_\nu, \tag{1.10}$$

where the source term j_ν is called the emissivity and α_ν is the coefficients for absorption and emission along the path (i.e the extinction coefficient)[4]. s is the proper distance along the path.

The derivation is simplified by working in the Rayleigh-Jeans limit. This is appropriate since the Rayleigh-Jeans law applies at low frequencies (in the radio region it almost always applies). This allows us to relate the intensity I_ν to a brightness temperature T, given by $I_\nu = 2k_B T\nu^2/c^2$. We introduce the optical depth which is defined as $\tau = \int ds \propto \alpha_\nu(s)$. Using optical depth and brightness temperature can be expressed using the radiative transfer equation. For a light source along a line of sight. The radiative transfer equation for a source emitting radiation passing through clouds can be

[4]A more in depth derivation of eqn.(1.10) can be found in (Rybicki and Lightman, 2008).

written in terms of the optical depth for absorption and is defined as:

$$T_b(\nu) = T_{\rm S}(1 - e^{-\tau}) + T_{\rm CMB}(\nu)e^{-\tau}. \tag{1.11}$$

Where T_b is the observed brightness temperature. $T_{\rm S}$ is the spin temperature, which has already been defined[5]. $T_{\rm CMB}$ is the CMB temperature. The factors $e^{-\tau}$ and (1-$e^{-\tau}$) represent the emission and absorption probability respectively. τ represents the optical depth along the line of sight. The full expression for the optical depth for diffused IGM is given by (Furlanetto, Oh, and Briggs, 2006):

$$\tau_{\nu_{21\,\rm cm}}(z) = \frac{3hc^3 A_{21\,\rm cm}}{32\pi k_B \nu_{21\,\rm cm}^2} \frac{x_{HI} n_H}{T_{\rm S}(1+z)(dv_{||}/dr_{||})}, \tag{1.12}$$

h is Planck's constant, c the speed of light, k_B the Boltzmann constant, x_{HI} is the neutral hydrogen fraction . $n_{\rm H}$ is the number density of HI gas. $A_{21\,\rm cm}$ is the spontaneous decay rate of the 21 cm transition. $\nu_{21\,\rm cm}$ is the 21 cm rest frequency, $dv_{||}/dr_{||}$ is the gradient of the proper velocity along the line of sight which includes both the Hubble expansion and the peculiar velocity (see Sobolev, 1960 for details).

1.4.1 Brightness Temperature

We can observe the 21 cm signal against the Cosmic Microwave Background,

$$\delta T_b(\nu_{21\,\rm cm}) = \frac{T_{\rm S} - T_{\rm CMB(z)}}{1+z}(1 - e^{-\tau_{\nu_{21\,\rm cm}}}). \tag{1.13}$$

The parameters $T_{\rm S}$, $\tau_{\nu_{21\,\rm cm}}$, have already been defined, while $T_{\rm CMB(z)}$ is the CMB temperature as a function of redshift. We use equation (1.13) for the 21 cm signal but there are important assumptions that need to be made. We assume a optically-thin 21 cm line (Mellema et al. 2012). In cases where $\tau_{21\text{-}\rm cm} \ll 1$, eqn.(1.13) becomes (Field, 1959);

$$\delta T_b = \frac{3hA_{21\,\rm cm}c^3}{32\pi k_B \nu_{21\,\rm cm}^2} \frac{n_{\rm HI}}{(1+z)H(z)} \left(1 - \frac{T_{\rm CMB(z)}}{T_{\rm S}}\right) \left(1 + \frac{1}{H(z)} dv_{||}/dr_{||}\right)^{-1}. \tag{1.14}$$

[5]We have implicitly assumed that T_S is constant here.

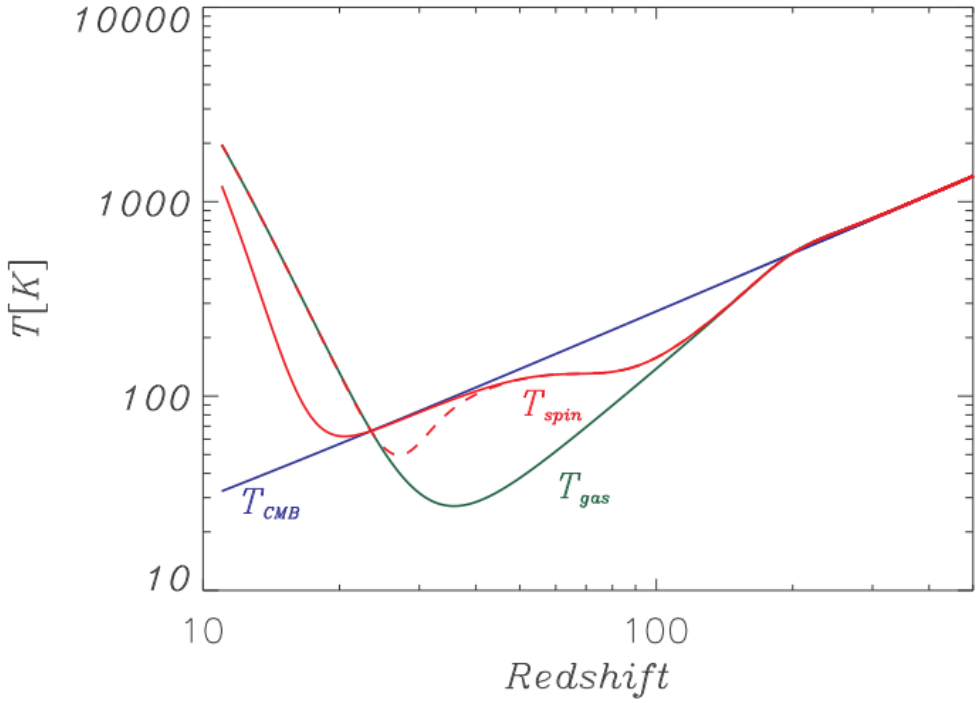


Fig. 1.6: Global temperature evolution as a function redshift from numerical simulation: gas temperature (green line), spin temperature (red line and dashed line) and CMB temperature (blue line). Figure from (Zaroubi, 2012).

At high redshifts (above 150) there is a coupling between the gas kinetic temperature (T_K) and the microwave background due to electrons left over from recombination (Fialkov and Loeb, 2016). After that, the gas decouples from the CMB and starts cooling adiabatically, so that its temperature drops below the CMB one. At the same time, collisional coupling kept the hydrogen gas spin temperature in equilibrium with its gas temperature (z $\sim$ 80) . At this stage, hydrogen can be observed through absorption against the microwave background since the hydrogen clouds are cooler. Collisional coupling becomes ineffective below a redshift $\sim$ 40, except in high density regions such as collapsed minihalos . Fig.(1.6), shows a numerical simulation exploring how the evolution of gas temperature, spin temperature and CMB temperature at different characteristic densities as the Universe cools due to expansion. he spin temperature can be defined in terms of the cosmic microwave background temperature (T_{CMB}) and the gas kinetic temperature (T_K). This coupling arises due to the Wouthuysen-Field mechanism, the coupling of the spin temperature of neutral hydrogen to Lyman-α radiation (Wouthuysen, 1952; Field, 1959). At low redshifts (z$\backsimeq$ 10) collisional coupling is less important but the radiative part plays an important role at high redshifts (up to

z ∼26). Note that,

$$T_s{}^{-1} = \frac{T_{CMB}^{-1} + x_\alpha T_\alpha^{-1} + x_c T_K^{-1}}{1 + x_\alpha + x_c}. \tag{1.15}$$

x_c, x_α are coupling coefficients for collisions and scattering of Lyman alpha photons (see Furlanetto, Oh, and Briggs, 2006 for more details on x_c, x_α). The collisional coupling coefficient depends on the number densities of the intervening (hydrogen atoms and electrons), as well as in their kinetic energy which is set by the CMB and the gas temperatures. Fig.(1.7) shows the brightness temperature evolution, the colour code indicates the strength of the 21 cm brightness as it transitions from absorption. The 21 cm signal can be observed from emission depending whether the spin temperature is higher or less than the background temperature. For the 21 cm contribution to be observable, the spin temperature T_S must of course differ from T_{CMB}. The HI would be detected in absorption or in emission depending on whether T_S is lower or higher than T_{CMB}. Around z∼ 15 local heating from X-Ray sources and quasars rises spin temperatures above the CMB and the brightness temperature is observed in emission.

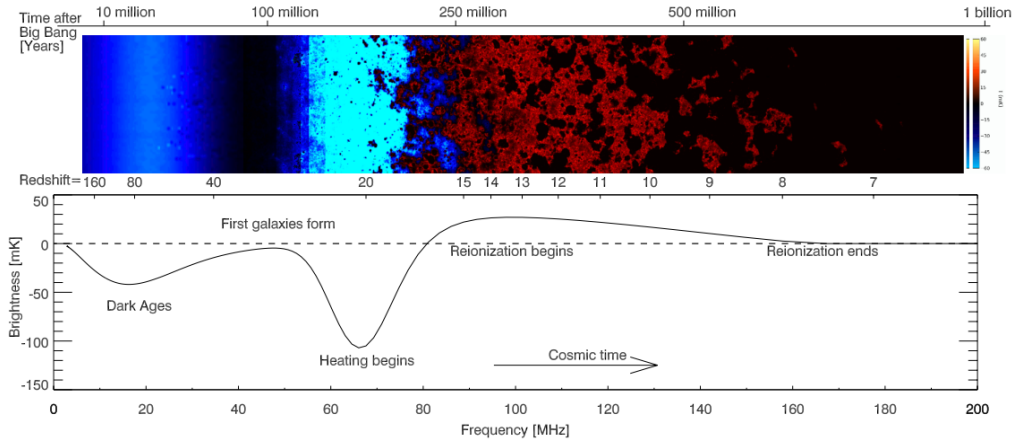


Fig. 1.7: Brightness temperature, which measures the average properties of the HI as a function of redshift. Upper panel shows a spatial view of the fluctuations and lower panel shows different epochs. Figure from (Pritchard and Loeb, 2012) .

1.4.2 Gunn Peterson trough

The Lyman-α line was discovered by Theodore Lyman in 1905 at Harvard University (Paxson and Thayer, 1920). The IGM can be studied by analysis of Lyman alpha forest; an arrangement of absorption lines of hydrogen through its excitation states, that appear in the spectra of galaxies and quasars alike. This feature was first hypothesised by Gunn and Peterson in 1965 and is now known as the Gunn-Peterson (GP) trough (Gunn and Peterson, 1965). The energy of a quasar will be absorbed if its energy is less than that of ultraviolet in the electromagnetic spectrum.

Even as the Universe started to ionize at redshifts z > 6, the scattering cross-section of photons with energies ear or above the Lyman-α line was still very high (White et al., 2003). A fraction of neutral hydrogen as small as (10^{-3}) will make the GP optical depth in the IGM high enough to cause the suppression of observed emission and saturates the IGM absorption. In quasars (redshifts $z < 6.4$) the lack of complete absorption is viewed as evidence for reionization, which means the Universe had already ionized by then (Becker et al., 2001). The Gunn-Peterson trough is used as an indicator of reionization in high-redshift quasars. Fig(1.8) illustrates this feature in objects above redshift 6, while it's not visible in the quasar's below redshift 6.

Basically, as photons from from distant quasars travel through the IGM, they may pass through regions of neutral hydrogen which scatter photons that have redshifted into the frequency corresponding to the Lyman-α transition. As a result, there would be an absorption feature bluewards of the Lyman-α in the quasar spectra. The absorption feature becomes stronger as one observes quasars at higher and higher redshifts. This is because the IGM is dense at higher redshifts. Eventually, one would expect to observe a trough in the spectrum as the absorption is so strong. This was first observed at a redshift of 6.28 (Becker et al., 2001). Numerous detections of quasars have brought more data for analysis (Becker et al., 2001; Fan et al., 2003; Jiang et al., 2007).

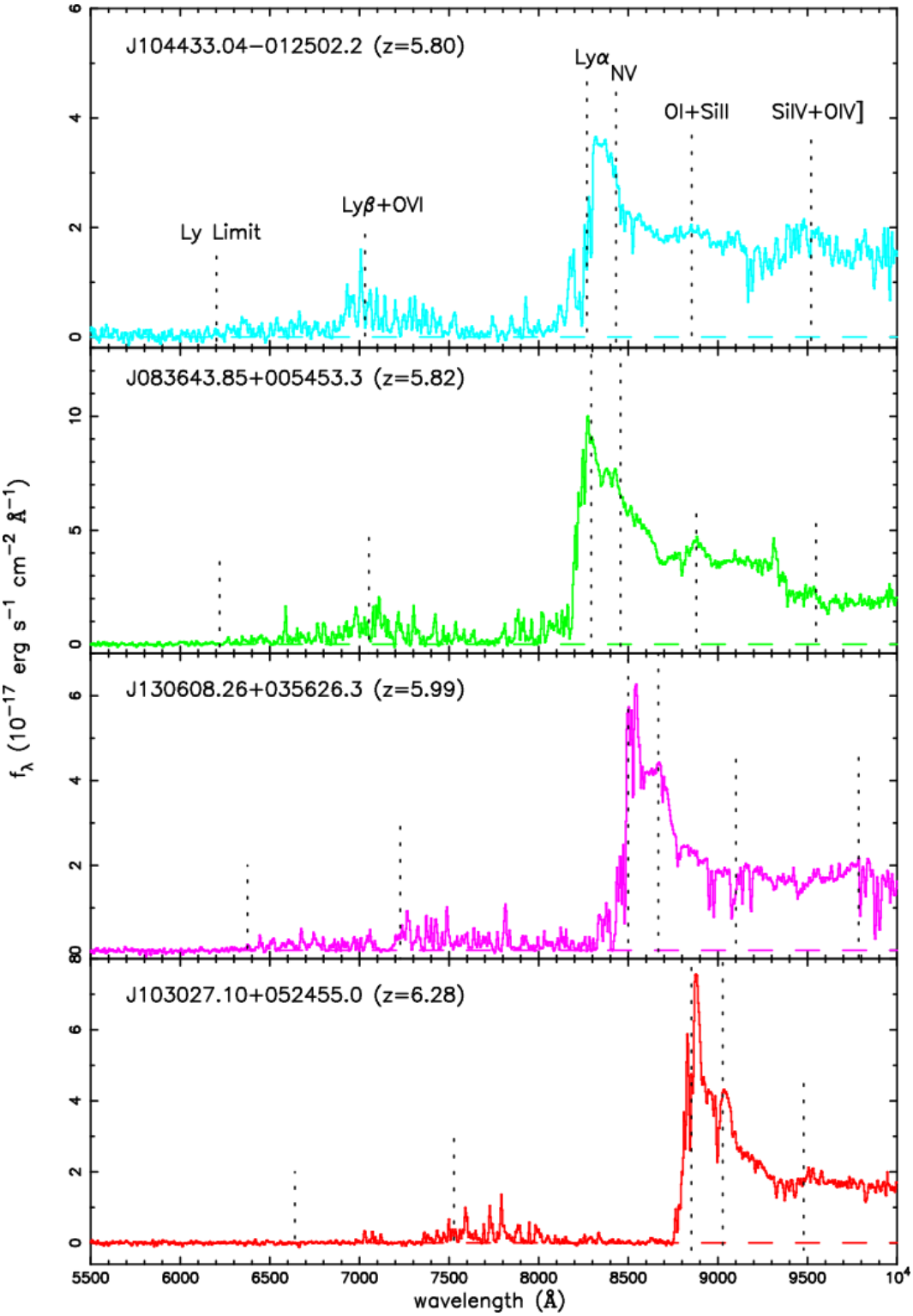


Fig. 1.8: Optical spectra for four high-redshift quasars observed with Keck/ESI Telescope. The expected prominent emission lines, as well as the Lyman limit, are indicated by the dashed lines. The GP trough is only observed in the spectra of quasars at z >6. This result provides one of the most important pieces of evidence of cosmic reionization.(Becker et al., 2001).

1.4.3 CMB experiments

The most compelling observational evidence to support a Universe that reionized come from measurements of the optical depth of CMB photons. The vast majority of the Universe's ordinary matter today is in the form of free electrons and protons. A key difference between the CMB and other proposed probes of reionization such as 21 cm surveys is that the CMB experiments are only a 2-dimensional measurement with observables integrated along the line of sight (Pritchard, Loeb, and Wyithe, 2010). This would rule out techniques like redshift tomography. Despite their integral nature, CMB measurements have yielded significant constraints on cosmic reionization to date (Komatsu et al., 2009).

The detection of the reionization bump in large-scale CMB polarisation by the WMAP satellite. The free electrons scatter the light coming from the CMB and produce a polarised signal (the polarization is generated when free electrons scatter a radiation field with quadruple anisotropy).

Recent experiments have exploited this property and observations suggest that the Universe was reionized early, in the first several hundred million years. CMB experiments provide information on the epoch of reionization through the measurement of the total Thomson optical depth. The scattering of CMB photons with free electrons is measured by means of the optical depth:

$$\tau = \int \sigma_T n_e dt = \int_0^{z_{reion}} \frac{1}{H(z)(1+z)} c\sigma_T n_e(z) d\mathbf{z} \tag{1.16}$$

where c is the speed of light, σ_T, the Thomson cross-section and n_e is the electrons' mean density. z_{reion} is the redshift of reionization. The main stumbling block when it comes to CMB experiments is Galactic foregrounds which pose a significant challenge in measuring polarization on these 10s of degree angular scales, and multiple frequencies are essential to disentangling the signals. Using measurements of the l <20 TE and EE power spectra from the Planck satellite [6], find $\tau = 0.066 \pm 0.012$.

However, CMB photons scattering in a reionized medium greatly enhance the polarization signal making it very likely that such a signal will be easily detected by CMB satellite experiments (WMAP, Planck) satellite. The results suggest that reionization

[6] where T and E stand for temperature anisotropy and polarized E-mode (zero curl) anisotropy.

began as early as $z \sim 10$ (Hinshaw, G., Larson, D., 2013). Current estimates from Planck collaboration put $z_{\rm reion} = 8.8^{+1.2}_{-1.1}$ (Planck Collaboration et al., 2015; Collaboration, 2016).

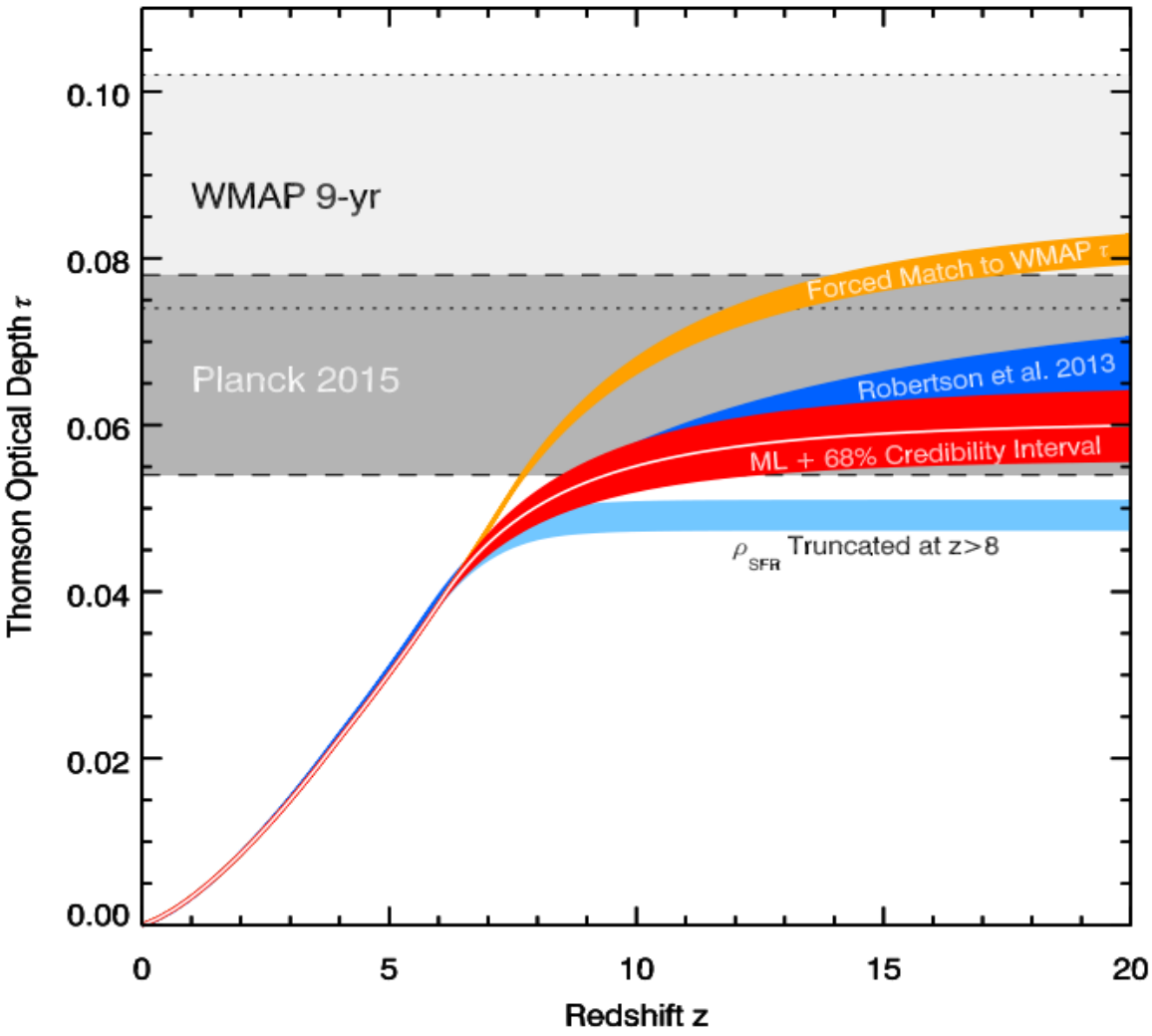


Fig. 1.9: Thomson optical depth measurements from WMAP & Planck data (grey region), model comparisons are shown ; (red region) model computed using the SFR histories from (Robertson et al., 2015), (blue region; (Robertson et al., 2013b)) is based on the WMAP results , (orange region) a model forced to reproduce the 9-year WMAP τ constraints, and (light blue region) a model truncated at z > 8 from (Robertson et al., 2013b).

1.5 21 cm Forest

In the spectra of high-redshift radio sources, we expect an absorption feature produced by the intervening neutral hydrogen along the line of sight. These absorption features are called the 21 cm forest. The emission from these sources is expected to be resonantly absorbed by the neutral IGM and form a system of absorption features just like the Lyman-α forest in the spectra of distant quasars.

The 21 cm forest is quite analogous to the case of the optical Lyman-α forest, which results in a suppression of the source flux due to diffuse neutral hydrogen along the line of sight. There are a few differences between these two phenomena; such as the several orders of magnitude in the corresponding Einstein coefficients. The regimes in which the 21 cm forest and the Lyman-α forest are relevant, are significantly different (Ciardi et al., 2012). Figure(1.10) shows a schematic of how light from a quasar in the early Universe passes through clouds of hydrogen gas in the IGM. The schematic also shows that when the light is observed it would be redshifted to account for the expansion of the Universe.

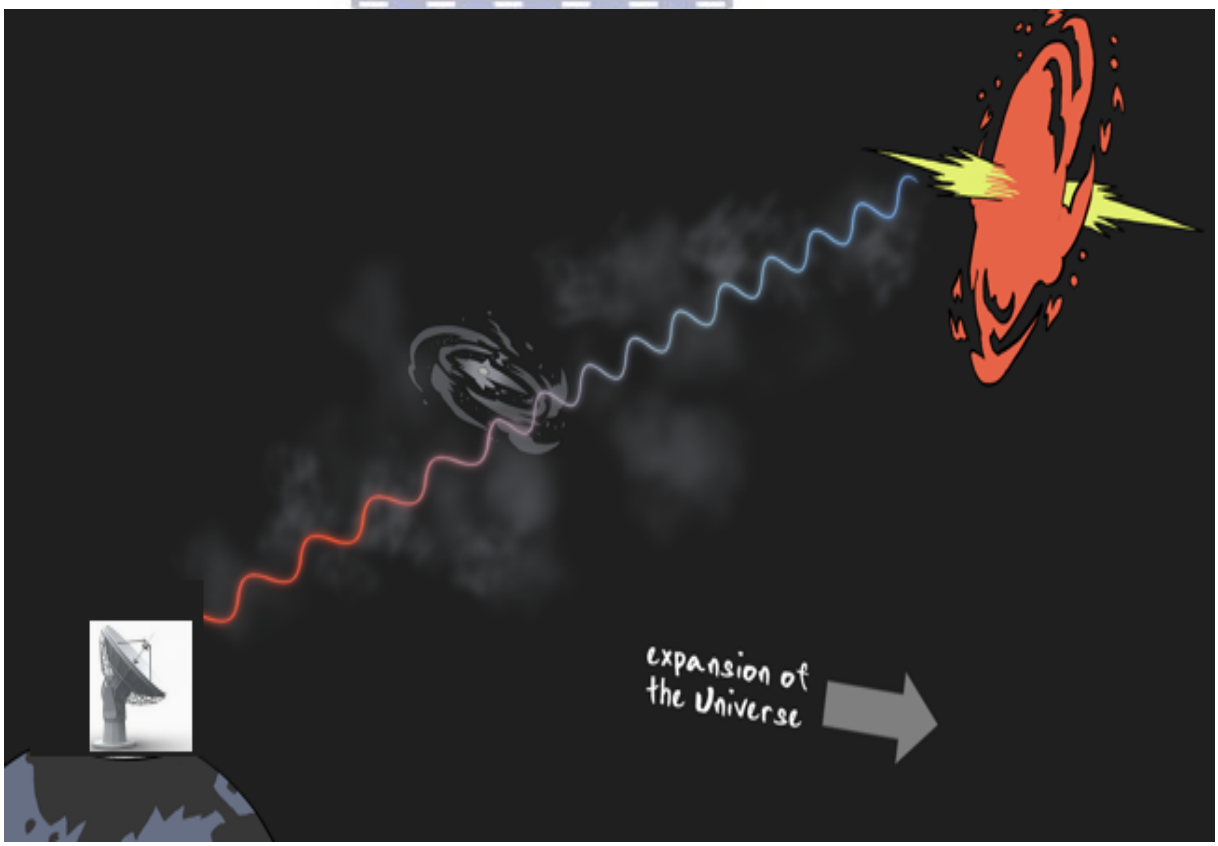


Fig. 1.10: An artist's impression showing light from a high redshift source passing through gas clouds before being observed by a ground-based telescope.

The main contrast with the Lyman-α forest is that the 21 cm forest occurs at significantly higher redshifts (z $>$ 6) and contains a great deal of information about the IGM even when the It is largely neutral. At high redshifts, neutral hydrogen in the IGM is more abundant hence the number of lines in the Lyman-α forest multiplies, above z$\geq$ 6. This causes a suppression of the forest, this phenomenon is known as the Gunn-Peterson through (see Fig.(1.8)).

This idea was first initially investigated by Carilli, Gnedin, and Owen 2002 and subsequently by other authors (Furlanetto and Loeb 2002; Furlanetto, Oh, and Briggs 2006, Xu et al. 2009, Mack and Wyithe 2012a, Ciardi et al. 2012). Detecting the 21 cm absorption requires a source population at z>6. The 21 cm forest is more advantageous in comparison to its counterparts because it is likely to probe very small structures due to the frequency resolution of the instrument. The 21 cm forest could be used as a diagnostic probe of the IGM and could be used to study the evolutionary history of the IGM heating and shed light on the nature of ionizing sources and the abundance of these high-redshift emitting sources.

The 21 cm forest is appealing since it bypasses the main limitations expected for 21 cm tomographic measurements counterparts (Xu et al., 2009; Mack and Wyithe, 2012b), which tend to suffer from a range of difficulties including foreground and ionospheric contamination (Ciardi et al., 2015).

Fig.(1.11) shows a synthetic spectrum constructed using a simulation by (Carilli, Gnedin, and Owen, 2002). The spectrum has the 21 cm forest expected to appear in the spectrum of a galaxy at z=10.0. The 21 cm absorption feature on the spectrum is signalled by a significant dip at a frequency of 129 MHz which corresponds to redshift $z \approx 10$. The synthetic spectrum is based on the powerful well studied radio galaxy Cygnus-A.

One worrying issue with the 21 cm forest is that the estimates of the abundance of radio-loud sources at z>6 are uncertain; this could be a problem if the sample size of sources detected is not at least in the hundreds. Fewer than fifty high-redshift radio sources have been detected (Mortlock et al., 2011), (Pentericci et al., 2014) and (Momjian et al., 2014) but the existence of higher redshift quasars is uncertain.

If a sample of the hundreds of these systems could be detected, then that could provide valuable information about the reionization process and the IGM's physical properties during reionization. Currently, we know of a few bright high redshift sources observed

(J163912.11+405236.5 with a radio luminosity of $\log_{10} L_{1400\mathrm{MHz}}/[\mathrm{W/Hz\ sr}]) \sim$26.5 at 1.4 GHz discovered by Jarvis et al., 2009), but with the imminent availability of highly sensitive radio telescopes like the Square Kilometre Array, the prospects for detecting such sources are promising.

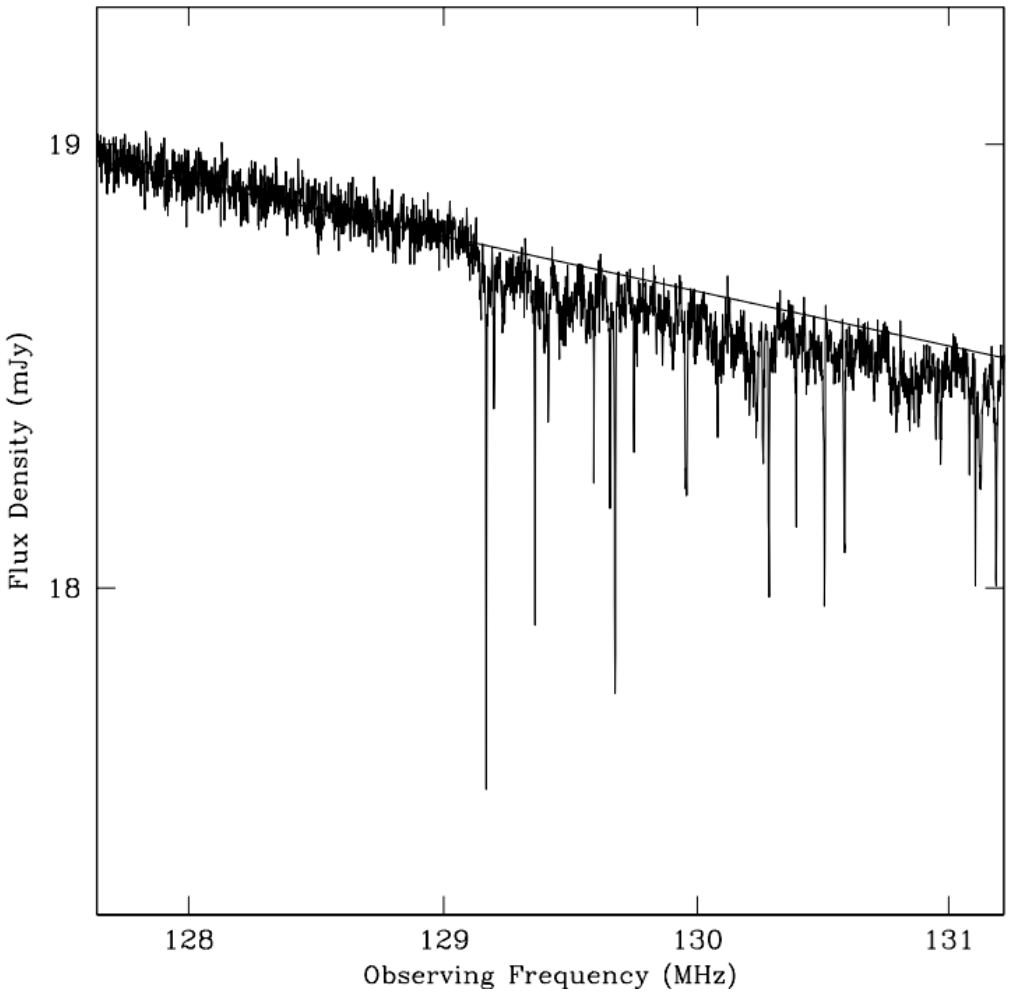


Fig. 1.11: Simulated spectrum of a source with $S_{120\,\mathrm{MHz}}$ = 20 mJy at z = 10 using a spectral model based on Cygnus-A and assuming HI 21 cm absorption by the IGM, the straight line is represents the case where there is no reionization hence no 21 cm absorption feature . Figure from (Carilli, Gnedin, and Owen, 2002).

1.6 Experiments

Reionization together with the Dark Ages after recombination, are a few of the most speculative epochs in the evolution of the Universe. This will, however, change because of a wave of upcoming observations should produce a rapid advance in our understanding.

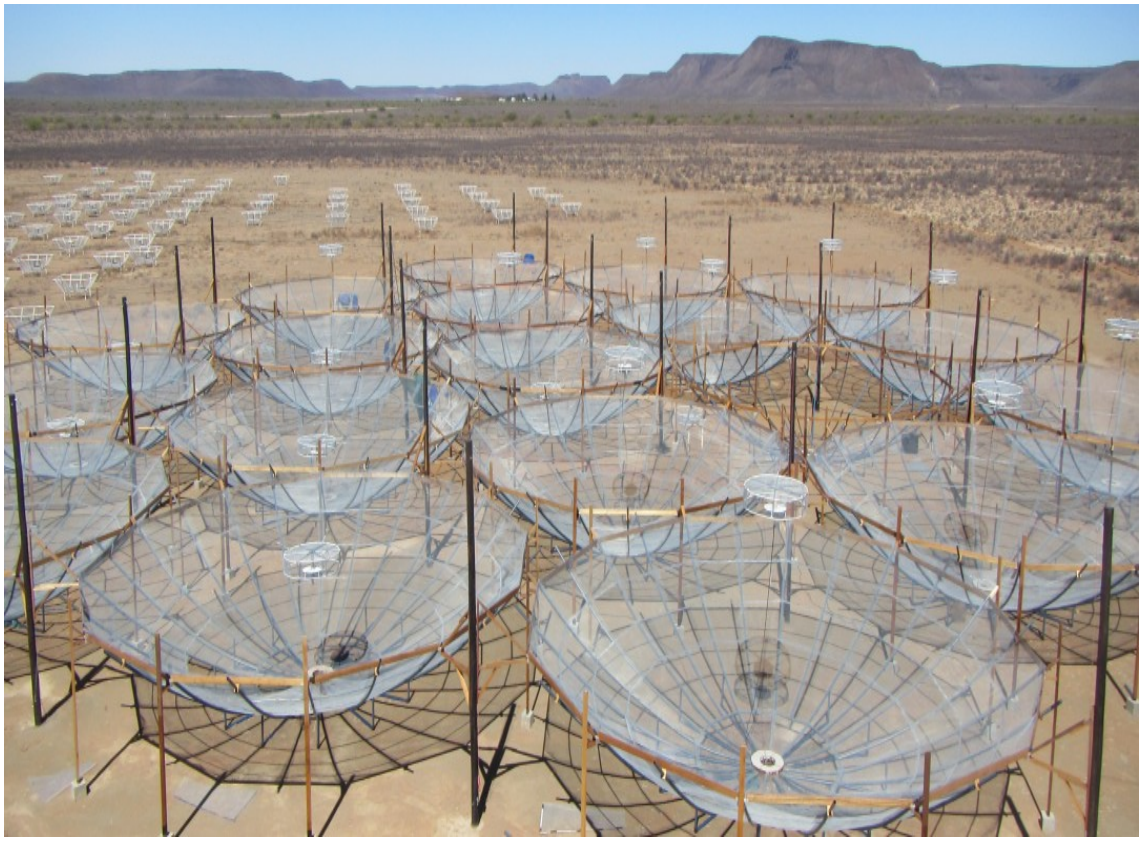


Fig. 1.12: HERA's first stage of development (HERA-19) has 19 dishes 14-m, zenith-pointing currently deployed in South Africa. From : www.reionization.org.

HERA [7]: The Hydrogen Epoch of Reionization Array (HERA) is a telescope which will measure redshifted 21 cm line of neutral hydrogen in order to better characterise the Epoch of Reionization (EOR). The telescope is currently under construction in South Africa, HERA has been granted status as a SKA Precursor. The fully functional HERA will be a 350-element interferometer with 14-m parabolic dishes observing a frequency range from 50 to 250 MHz. The first 19-element array has already been built as of September 2016 and it us expected that a total of 37 elements, will be completed by the end 2017. Estimates suggest this will provide about five times the sensitivity of PAPER-128 (DeBoer et al., 2016). A fully built HERA will potentially produce the first

[7] http://reionization.org/

images of the EOR (Beardsley et al., 2015).

PAPER [8]: The Precision Array for Probing the Epoch of Reionization (PAPER) is a low-frequency radio interferometer designed to measure the power spectrum of high redshifted 21 cm emission to improve our understanding of cosmic reionization. PAPER operates in the 100–200 MHz band, which is roughly between redshifts 6-12. PAPER consists of two different arrays locations: a 32 antenna at the National Radio Astronomy Observatory (NRAO) site in Socorro, New Mexico USA, which is mainly used for engineering investigations and field testing, and a 64 element telescope in the Square Kilometre Array South Africa (SKA-SA) site in the Karoo. The South African based PAPER is used for science observations. The PAPER team were able to set the tightest limits (2 σ upper limits on the 21 cm power spectrum (Parsons et al., 2014)) to date on the 21 cm EoR power spectrum (Parsons et al., 2010; Parsons et al., 2014).

LOFAR [9]: The LOw-Frequency ARray (LOFAR) is a radio telescope designed and constructed by ASTRON to probe the low-frequency regime of 10-240 MHz for a broad range of astrophysical studies (Falcke et al., 2006). LOFAR consists of stations distributed throughout the Netherlands and Europe. In the Netherlands, a total of 40 LOFAR stations has been deployed with additional eight international stations currently built throughout Europe. This experiment has a good chance of making the first positive detection of the 21 cm line from the EoR (Haarlem et al., 2013).

SKA1-LOW [10]: The most anticipated telescope in modern radio Astronomy is the Square Kilometre Array (SKA). The SKA will have two phases. Phase 1 (SKA1) is due to begin construction in 2018 and be completed by 2023, while the construction of phase 2 is between 2023 and 2030. Phase 1 has two components: component one is the mid-frequency array of about 200 dishes antennas in South Africa, when incorporated with MeerKAT (SKA1-MID) there will be roughly 260 dishes.

The second component of phase 1, is the SKA1-LOW, a low-frequency array in Australia made up of 130, 000 dipole antennas. The science goal of particular interest is, to provide observational data from the so-called Dark Ages (300,000 years after the Big Bang when the ionized plasma became neutral and the Universe cools), this period marks a time where first light from a billion years when young galaxies are seen to form for the first time. To see back to first light requires a telescope 100 times more

[8] http://eor.berkeley.edu/
[9] http://www.lofar.org/
[10] www.skatelescope.org

powerful than the biggest radio telescopes currently in the world.

MWA [11]: The Murchison Widefield Array (MWA) is one of three Square Kilometre Array Precursor telescopes along with MeerKAT(Booth et al., 2009), and the Australian SKA Pathfinder (ASKAP; (Johnston, Feain, and Gupta, 2009)). The MWA is located at the Murchison Radio-astronomy Observatory in Western Australia. MWA has four key science goals, searching for redshifted 21 cm emission from the epoch of reionization; Galactic and extragalactic surveys; time-domain astrophysics; and solar, heliospheric, and ionospheric science and space weather (Bowman et al., 2013). The MWA operates at frequency range 80-300 MHz and consists of 128 aperture arrays distributed over a 3-km diameter area (Tingay et al., 2013). The MWA plans to produce deeper limits with on 21 cm power spectrum (Dillon, Liu, and Tegmark, 2013).

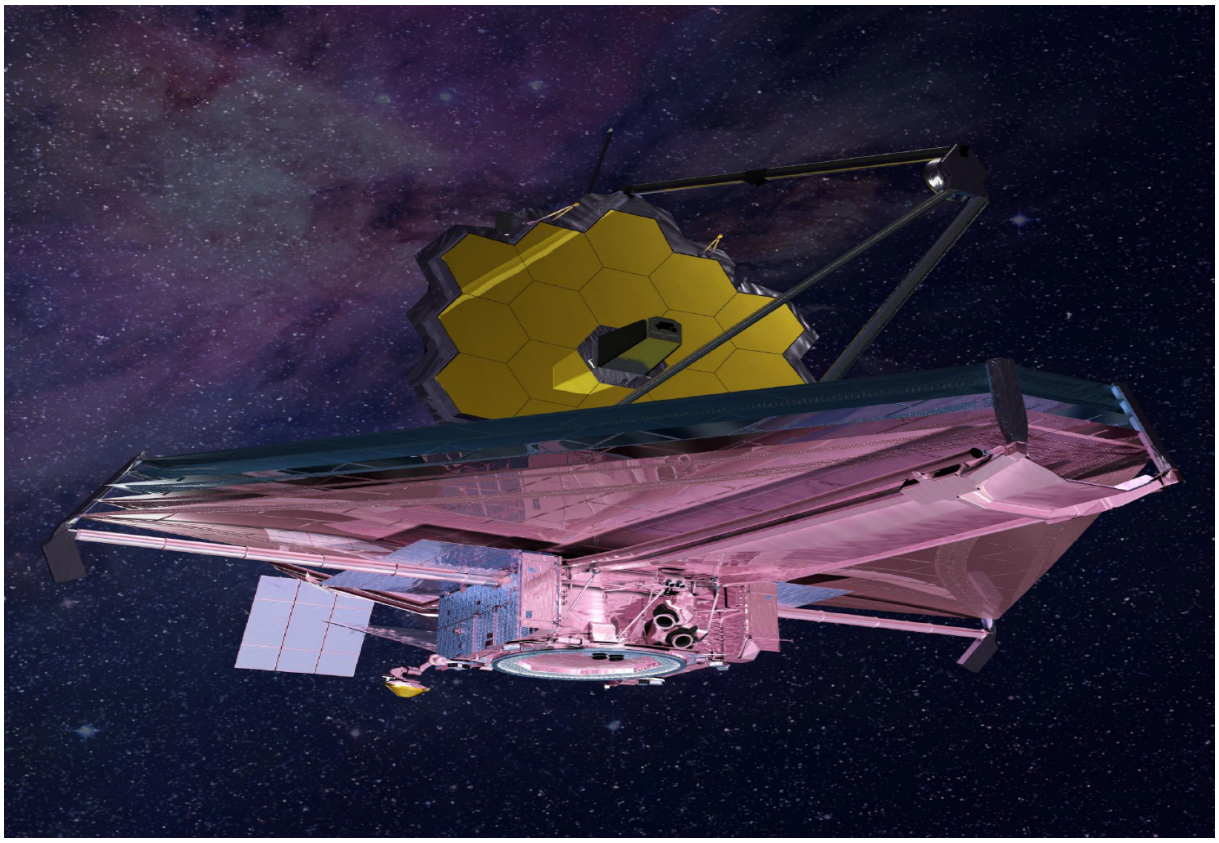


Fig. 1.13: An artist's impression of the James Webb Space Telescope, the successor to the Hubble Space Telescope, image from: (www.nasa.gov).

Near-infrared surveys: In recent times near infrared observation have had a crucial role in trying to find the sources that reionized the Universe. Many of these observations carried out using the Hubble Space Telescope (HST), which will soon be succeeded by the James Webb Space Telescope (JWST). This will be beneficial in studying

[11] www.mwatelescope.org

the Gunn-Peterson trough (see Sec. (1.4.2)) in quasars seen at z $\geq$6.

The JWST includes a primary mirror 6.5 metres in diameter, and offers instrument sensitivity (forecasts place the signal to noise ratio of 10 after a 10,000 second observation) across the infrared wavelength range of 0.6-28μ m (Glasse et al., 2015), which will allow detection of the first galaxies, Fig.(1.13) shows a schematic of how the JWST will look. The size of the Sun shield (the large flat screen in the image) is 22 meters $\times$10 meters.

The requirement for a significant contribution of ionizing photons from z > 10 galaxies remains an important uncertainty which is perceived as a major goal for the James Webb Space Telescope.

An added advantage of using an infrared telescope, only at infrared wavelengths that we can see the first stars and galaxies forming after the Big Bang. Infrared wavelengths are used at this epoch, because the light shortward of Lyα is absorbed by the GP trough/Lyα forest. You have to use near- infrared at z>6 as this is where the Lyα is redshifted to.

Some of the goals for JWST are to study galaxy, star and planet formation in the Universe. Probing very first stars and galaxies that formed in the early Universe, JWST will look deep into space to look back in time. An emphasis will be made to study the earliest stars and galaxies formation of the Universe, JWST will be beneficial since it will be observing in infrared light band.

1.7 Outline

The aim of this thesis was to develop a novel technique that allows us to the measure the redshift of radio galaxies in the early Universe. We achieved this by constructing a parameterization for the absorption feature which is expected to be seen in the spectra of high-z radio galaxies. In the past, this has been unattainable, because of the limitations in the sensitivities of radio telescopes.

Throughout the work unless stated otherwise, I employed the ΛCDM Cosmology, assuming the Planck parameters (Collaboration, 2016) i.e. { Ω_m: 0.308 , Ω_Λ: 0.6911, H_0: 67.7 km/s/Mpc, n_s: 0.96, σ_8: 0.815 }. Chapter 2, explores the statistical methods (Bayesian) utilized. This background chapter is intended to give an exhaustive description of the statistics is used to develop a robust technique parameter fit spectra of galaxies in EOR. Chapter 3 presents and gives description in some detail of the SIMFAST21, a numerical implementation of exercusion-set principle. I give a brief description of the code developed to parameter fit the spectra of galaxies in the EOR. In Chapter 4. I applied the statistics formalism discussed in Chapter 5 to the analytical models while parameter estimating the 21 cm absorption feature in the spectra of galaxies in reionization. In Chapter 6 summarises the main conclusions of the work presented in this thesis and gives an overview of the future directions of this work.

Chapter 2

Model Fitting

In order to probe the epoch of reionization using the 21 cm absorption feature, we need to model the 21 cm feature precisely and choose among potential viable models for the fitting. We model the 21 cm absorption feature using using Bayesian framework for model selection and parameter estimation. There are 3 contender models being considered for the feature and all 3 are polynomials models.

In this kind of analysis the "best" model is one that achieves the best compromise between quality of fit and simplicity. Indeed, the Bayesian model comparison is a very powerful tool to reward the models that fit well the data exhibiting predictability, while models with large number of parameters, not required by the data, tend to be less preferred since some of the parameter may be free parameters (Gordon and Trotta, 2007). Model selection is vital because it is used to decide whether a introducing an additional parameter is required to better estimate. The classical approach accepts additional parameters on the basis that they improve χ^2, this has been shown to be a pitfall (Trotta, 2008). A key method for exploring the posterior is to use a Markov Chain Monte Carlo (MCMC) method, for example the Metropolis-Hastings algorithm(Metropolis et al., 1953; Hastings, 1970), which is the MCMC algorithm other method used in this thesis to explore parameter space, which crucially also computes the Bayesian evidence, is the Nested Sampling technique.

Let us consider a model, $\mathcal{M}_i$ and where the model's parameters are represented by a vector $\overrightarrow{\theta}$. We find the probability of models using the parameters vector $\overrightarrow{\theta}$ with the given data $\mathcal{D}$ for Model $\mathcal{M}_i$. Combining all the aforementioned we find an expression for the posterior using Bayes' Theorem [1].

$$p(\overrightarrow{\theta}|\mathcal{D}, \mathcal{M}_i) = p(\mathcal{D}|\overrightarrow{\theta}, \mathcal{M}_i)\frac{\pi(\overrightarrow{\theta}|\mathcal{M}_i)}{p(\mathcal{D}|\mathcal{M}_i)} \tag{2.1}$$

[1] A noteworthy source that explores the brief history & mathematical background of Bayesian Statistics is (Loredo, 1992)

The term $p(\mathcal{D}|\overrightarrow{\theta}, \mathcal{M}_i)$ is the likelihood $\mathcal{L}(\overrightarrow{\theta})$. The likelihood describes the probability of the data given the parameters. Since the value of the likelihood can become very small, it is often computed as the log-likelihood. $\pi(\overrightarrow{\theta}|\mathcal{M}_i)$ is the prior distribution function and allows us to quantify the belief about the parameters $\overrightarrow{\theta}$ without referencing the data.

There are various ways to set up the prior conditions. In cases where the parameters are physical, which means we are able to measure them directly using telescope, this limits the range for prior range. The prior grants the experimenter opportunity to incorporate useful information into the computation about the nature of the model parameters, which becomes effective when marginalizing over single parameters with multiple techniques. $p(\mathcal{D}|\mathcal{M}_i)$ is the model evidence and serves as a normalization constant.

2.1 Bayesian Inference

The term $p(\mathcal{D}|\mathcal{M}_i)$ in eqn.(2.1) is referred to as Bayesian evidence. In the case of parameter estimation we tend to neglect $p(\mathcal{D}|\mathcal{M}_i)$, since it has no influence constraining the model parameters $\overrightarrow{\theta}$. However, this is not case in model selection, where $p(\mathcal{D}|\mathcal{M}_i)$ takes center stage, because of the obvious model dependence on priors, which is a very appealing feature for model selection. We simply define it as factor required to normalize the posterior:

$$Z = \int \mathcal{L}(\overrightarrow{\theta})\pi(\overrightarrow{\theta}|\mathcal{M}_i)d\overrightarrow{\theta} \tag{2.2}$$

As we intuitively expect, the evidence is larger for a model if it has many parameters and smaller evidence for a model with few parameters. Hence Bayesian evidence invokes Occam's razor: a simple model with few parameters will have a larger evidence than a complicated model with many parameters, unless the latter model is much more suitable at predicting the behaviour of data points.

If we consider two competing models, $\mathcal{M}_i$ and $\mathcal{M}_j$, where the models share some common parameters. We find the probability of models using the parameters vector $\overrightarrow{\theta}$ with the given data $\mathcal{D}$ for Models $\mathcal{M}_i$ and $\mathcal{M}_j$. Combining all the aforementioned we find an expression for the evidence using Bayes' Theorem:

$$\frac{p(\mathcal{D}|\mathcal{M}_i)}{p(\mathcal{D}|\mathcal{M}_j)} = \frac{Z_i}{Z_j}\frac{p(\mathcal{M}_i)}{p(\mathcal{M}_j)} \tag{2.3}$$

Where $\frac{p(\mathcal{M}_i)}{p(\mathcal{M}_j)}$ is the prior probability ratio for models $\mathcal{M}_i$ and $\mathcal{M}_j$. We again use the logarithmic scale to present the ratio of the posterior model probabilities to avoid cases where the number is too small and is rounded off to zero.

$$\Delta \ln Z \equiv \ln \left[\frac{Z_i}{Z_j} \frac{p(\mathcal{M}_i)}{p(\mathcal{M}_j)} \right] \tag{2.4}$$

Model evidence are not meaningful by themselves and require further interpretation, it is important to find a useful scale to judge differences in evidence. $\Delta \ln Z$, is called the Bayes factor, which was first introduced by Jeffreys, 1961 it effectively compares between two competing models ($\mathcal{M}_i$, $\mathcal{M}_j$), using the model's Bayesian evidence to workout $\Delta \ln Z_{ij}$= $\ln Z_i$-$\ln Z_j$, this determines which model better estimates the data. The interpretation scale (Jeffreys' scale) are in Table (2.1).

Essentially the Jeffreys' scale is an empirically calibrated scale, with thresholds. A neat analogy of thinking about the the Jeffreys' scale is in terms of betting odds at a game; one understands the odds of 150:1 mean a there's slim a chance of winning hence betting with a large sum of money on the outcome would be considered unwise. Notice from Table (2.1) that the quantities are expressed in the logarithmic scale of the Bayes factor. This informs us that evidence changes are minuscule when they occur, significant changes in the evidence strength scale require about an order of magnitude which means something drastic would need to happen if we are considering the same models (e.g changes in noise scale used when comparing two competing models).

$\Delta \ln \mathbf{Z}$	**ratio**	**strength of evidence**
< 1.0	$\lesssim$ 3:1	inconclusive
1.0	$\sim$ 3:1	weak evidence
2.5	$\sim$ 12:1	Moderate evidence
5.0	$\sim$ 150:1	Strong evidence

Table 2.1: Jeffreys' scaling to evaluate the strength of evidence when comparing two models. The scale is empirical . Column one shows the logarithm of the base factor, column two is a representation of column one as a ratio. The last column shows the interpretation of having evidence at a particular region (Trotta, 2008).

2.2 Markov chain Monte Carlo methods

The name "Monte Carlo" was initially associated with gambling because of the most famous casino at Monte Carlo, but it soon became a synonymous term for describing random processes. The first major Markov chain Monte Carlo (MCMC) algorithm was devised by Metropolis et al., 1953, the so called Random Walk Metropolis algorithm, which creates a Markov Chain: a series of random samples whose values are determined only by the values of previous points in the series. Hastings, 1970 found a more elegant generalization of the sampling method. The next value only depends on the current value. θ is thought of as a vector in multidimensional space.

Algorithms for generating Markov chains are numerous and greatly vary in complexity and applicability. Many of the most important ideas were generated in physics,especially in the context of statistical mechanics, thermodynamics. Below I show the steps needed to construct a simple MCMC using Metropolis-Hastings algorithm; the most popular MCMC algorithm:

1. Start from some initial parameter value θ^c.
2. Evaluate the unnormalized posterior $p(\theta^c)$.
3. Propose a new parameter value θ^*. This new parameter is drawn from a jump distribution center on the current parameter value (the most common jump distribution is $\mathcal{N}(0, \sigma^2)$).
4. Evaluate the new unnormalized posterior $p(\theta^*)$
5. Decide whether or not to accept the new value. Accept new value with probability R, if $r < R$, where r is the acceptance rate.
6. Repeat steps 3-5

2.3 Nested Sampling

MCMC methods are far more useful in exploring uni-modal parameter spaces and parameter inference. When it comes to computing Bayesian model evidence MCMC are inefficient. Generally numerical calculations of Bayesian model evidence in the cosmology context involve multi-dimensional integrals that are very difficult to compute. Coincidentally, several codes have been developed to combat this. These codes (PolyChord (Handley, Hobson, and Lasenby, 2015), Nested sampling and affine-invariant (Allison and Dunkley, 2013) and Diffusive nested sampling (Brewer, Pártay, and Csányi, 2009)) are based on a nested sampling algorithm by Skilling (Skilling, 2004).

Nested sampling is Monte Carlo method for evaluating Bayesian evidence but also has a byproduct which is the posterior inference (Skilling, 2004). Nested sampling is lot more powerful and efficient than the conventional MCMC algorithms. It utilize relations between likelihood and prior volume, which moves from a multi-dimensional integral into one-dimensionl integral. Prior volume (several authors define it as 'the cumulative prior mass' Skilling (2004), Feroz and Hobson (2008), and Feroz et al. (2013)) is defined as:

$$X(\lambda) = \int_{L(\Theta)>\lambda} \pi(\Theta) d^D\Theta \tag{2.5}$$

where D is the dimensionality of the parameter space. In model selection the evidence takes the central role and is simply the factor required to normalize the posterior over $\pi(\Theta)$. $X(\lambda)$ is defined in such a way that it decreases from 1 to 0 as λ increases. This enables integration of the evidence from a multi-dimensional space to a one-dimensional space. The Nested Sampling method draws a number of random points with respect to the prior $\pi(\Theta)$. A critical assumption made is that the integral extends over regions of parameter space within the iso-likelihood contour $L(\Theta) = \lambda$. Also we assume that $L(X)$ is the inverse of eqn.(2.5).

In short, nested sampling begins with a sampling of the entire prior volume. Samples are then drawn from successively more likely regions of parameter space until the posterior bulk, which is over-sampled with respect to the full prior, is located and explored. The samples are weighted appropriately and the posterior and evidence may be estimated.

$$Z = \int_0^1 L(X) dX \tag{2.6}$$

A neat ploy is used to calculate $L(X)$ without computing the likelihood $\&$ posterior over a given grid in parameter space. Nested sampling draws random points with respect to a given prior. $L(X)$ values are rejected based on the condition $L(X)_{\text{new}} < L(X)_{\text{rejected}}$, this insures a new set of points in the range are covered. This is gradually carried out until the maximum likelihood converges. Thus this is used to evaluate the likelihood at $L_s = L(X_s)$, where the sequence X_s is a sequence of decreasing values

$$0 < X_m < ... < X_2 < X_1 < X_0 = 1$$

In turn this is used to numerically approximate the evidence, with weighted sums:

$$Z = \sum_{i=1} L_i w_i \tag{2.7}$$

The algorithm based on the method discussed above is :

1. Set up a sample of N points in parameter space $\{\theta_1, ...\theta_N\}$ taken from the prior $\pi(\theta)$ (the prior generally used is a simply uniform distribution over the prior range) and set initial values $Z = 0$, $X_0 = 1$
2. Loop for i=1,2,...,m
   (a). find a point θ_L with the lowest likelihood from current $\{\theta_1, ...\theta_N\}$ and set L_i=$L(\theta_L)$
   (b). Set X_i= $\exp(-i/N)$
   (c). Set w_i=X_{i-1} -X_i
   (d). We then update Z using Z_i=Z_{i-1}+w_i L_i
   (e). One point θ_k is sampled from $\pi(\theta)$ with the condition $L(\theta) > L_i$ and update θ_L with θ_k to continue to next N-point in the sequence.
3. For the last step: Z is updated with addition of X_m $[L(\theta_1)$+...+$L(\theta_N)]/N$

Chapter 3

Simulation Software

In this section, we present the outputs of a consistent hydrogen reionization simulation with our novel implementation to simulate the epoch of reionization.

Simulations are vital as we attempt to probe the high redshift Universe that has complex astrophysical and cosmological properties. Numerous theoretical studies have focused on modelling the EoR process and its 21 cm signal. The main requirement is that simulations should be large enough so comparison with available data can be made (Ocvirk et al., 2016). A key limiting factor is the need for high dynamic range, e.g. it still remains a computational challenge to simulate large scales while maintaining the necessary resolution.

Simulations give us accurate and precise measurements of how the universe has evolved, they are equally crucial to data analysis and the verification of current theoretical models. They enable us to probe the underlying processes occurring in galaxies and vast information about the star formation and interstellar medium (Gluscevic and Barkana, 2010).

Simulations are also vital in the planning of observations. We can use them to estimate for parameters such as; sensitivities and predict the parameters of interest in Cosmology and Astrophysics (Davé, Finlator, and Oppenheimer, 2006).

The route with numerical simulations seems to be the better option in terms of modelling the epoch of reionization, but there are several challenges that need to be overcome. Due to a large amount of volumes observed in real observations, the task of constructing such simulations is considered not only strenuous but also computational expensive (Santos et al., 2008). The second challenge is the simulation of the complex, sometimes non-linear, astrophysical phenomena that drive reionisation.

There are two categories of simulations available: numerical Simulations and semi-numerical simulations. Semi-numerical simulations are very fast and versatile which makes them more useful than full numerical simulations to test different models even though they are not as accurate (Ciardi et al., 2000; Alvarez et al., 2009; So et al., 2014; Iliev et al., 2015). The effectiveness of semi-numerical simulation is mostly exploited in large scale simulations like a simulation of the formation galaxies. The disadvantage of semi-numerical simulations is that they can only provide approximate calculations.

Current semi-numerical simulations include (Mesinger and Furlanetto, 2007; Zahn et al., 2007; Santos et al., 2009; Battaglia et al., 2012; Zhou et al., 2013; Majumdar et al., 2014) both of attempt to model the early stages of the universe. The simulation of interest in our case is SIMFAST21, which models the epoch of reionization using approximations from smaller numerical simulations.

3.1 SIMFAST21

SIMFAST21 is a semi-numerical code which is basically a 'hybrid' (Santos et al., 2009); a combination of numerical simulations and analytical techniques so it can minimize computing time while maintaining a high level of efficiency in results. The code is designed to predict and simulate 3-dimensional intensity maps of the 21 cm line during the Epoch of Reionization. The code can easily be modified since it is divided into modules which can be run separately. Below we briefly describe SIMFAST21 by (Santos et al., 2009).

SIMFAST21 has a large number of input parameters used, they are usually taken from full numerical simulations or current observations. SIMFAST21 can be a tool used for planning observations. The output files are represented as 3-dimensional boxes with a length L and a total number of cells N^3 which are set initially by the user.

The prescription to generate the ionization field is based on the algorithm first described in detail by (Mesinger and Furlanetto, 2007). The set of instructions followed by the code are as follow:

1. A linear density field is generated at redshift $z = 0$. Density fluctuations for each of the cells in the 3-dimensional box are created using a random distribution. This is the initial state, these fluctuation are expected to gravitational clump and form collapsed structures.

2. The linear density fields which originate from density fluctuations at $z = 0$ are evolved linearly to other z. These linear density fields are then used to obtain dark matter halos. The halos are found using criteria based on the excursion set formalism (Sheth and Tormen, 1999); the code uses a sphere with varying radii and locates halos by determining whether there is enough mass within the sphere centred at a given pixel (for there to be a dark-matter halo). This condition finds halos down to a radius defined by the resolution of the box. Second condition check; in each cell for the probability of containing one halo or more with a given radius, for a decreasing radius. The number of halos within a given radius follows a random Poisson distribution.

3. Velocity field, created from the density field, accounts for peculiar velocities in the 21 cm signal. An extra module is used to calculate first-order nonlinear corrections to the density field and dark matter-halos, from the Zel'dovich approximation (Zel'Dovich, 1970; Yoshisato et al., 2005).

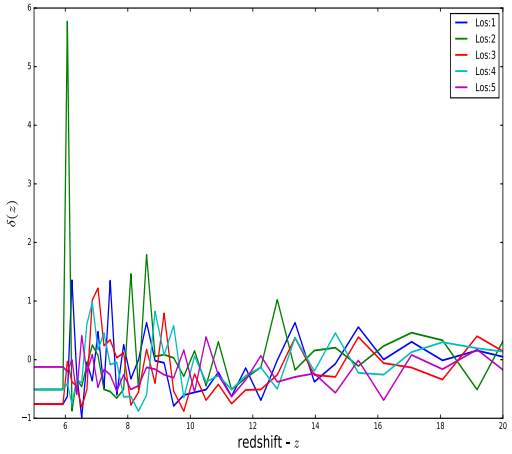
4. Ionized regions are identified using a criteria called the excursion-set algorithm, based on the assumption of a constant efficiency parameter of ionising photon per unit halo mass.

5. The ionization field is produced from the density field and halo catalogue. The simulation accounts for physically ionizing power in each halo and adds recombination to each halo using values gathered from recently detailed simulations or recent observations. The simulation assumes that gas ionizes at a temperature of 10^4 K. An extra module estimates the optical depth to reionization from the ionization field.

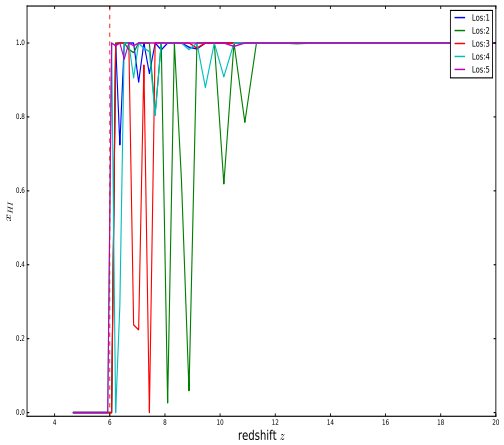
6. Extra modules approximate the coupling of the 21 cm line to the gas temperature due to radiative and collisional emission. These account for gas heating and also calculate corrections to the 21 cm line fluctuation using the already simulated dark matter density field, dark matter halos, star formation rate density and ionized box.

The Fig.(3.1) shows some of the simulation outputs used throughout the work, in creating the 21 cm absorption feature. Fig.(3.1) shows 5 different lines of sight from SIMFAST21 simulation output runs. The reason for taking these 5 lines of sight, is to the compare how large are variations from one another, and looking at Fig.(3.1) we note that the variations are substantial but the general trend in all 5 lines of sight agrees. One way to think of the cause of the variations, consider that the light from EOR passes through different objects at different position hence one should expect differences in the spatial attributes from simulation. The Cosmology parameters used in

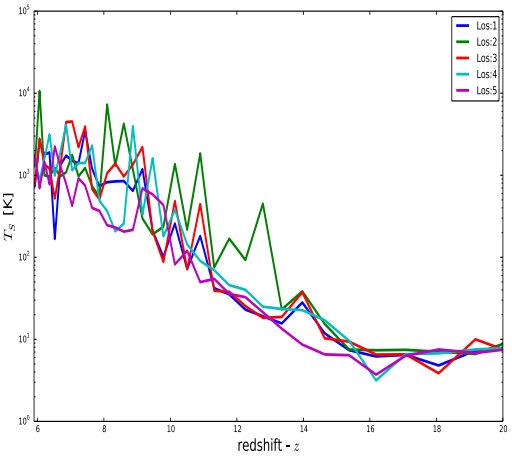
SIMFAST21 are { Ω_m: 0.308 , Ω_Λ: 0.6911, H_0: 67.7 km/s/Mpc, n_s :0.96, σ_8: 0.815 } while Simulation attributes are L(length of simulation boxes)=100 Mpc/h, N^3=300^3.



(a)



(b)



(c)

Fig. 3.1: A SIMFAST21 simulation (Santos et al., 2009) outputs as function of redshift in the range $5 \leq z \leq 20$: (a) overdensity δ, (b) neutral hydrogen fraction x_{HI} and (c) spin temperature (T_S). The code was developed by (Santos et al., 2009). I ran the simulation using a box size of 100Mpc/h and assuming the cosmolgy from the Planck parameters (Collaboration, 2016) . To recover the above data multiple slices through each box were used. We took 5 different lines of sight (los).

3.2 SKADS

The SKA Design Study Semi-empirical Extragalactic Simulation (SKADS) was developed by Wilman et al., 2008 as a semi-empirical simulation for modelling the cosmological distribution of radio galaxies. It combines together our understanding of the evolution of the radio source population and extrapolates it out to very high redshift ($z = 20$) and to 10-nJy flux-density limits in preparation for the continuum surveys possible with the Square Kilometre Array (SKA) and its Pathfinder facilities. I only give a brief overview of this simulation here and the reader is referred to Wilman et al., 2008 for full details. A total of 107128 sources were found above $z > 6$ and $S_{151\,\mathrm{MHz}} >$ 0.45 mJy when querying SKADS.

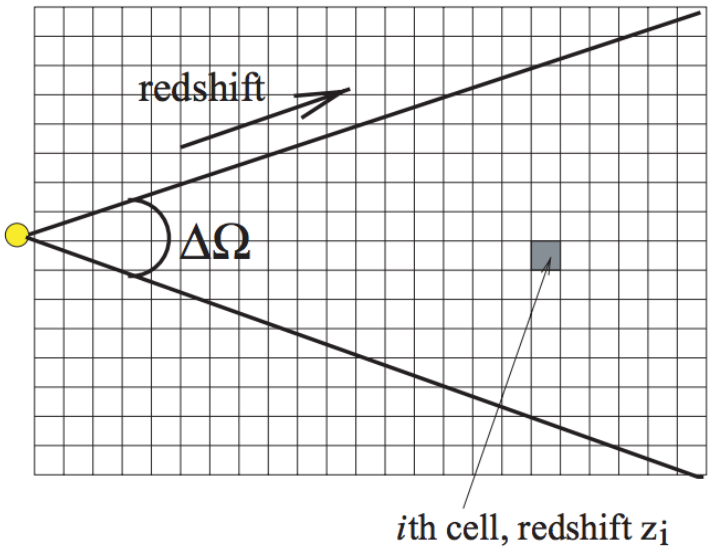


Fig. 3.2: 2D illustration of the geometry of the simulation. An imaginary observer is situated at the centre of one face of the grid and looks out into the simulation volume over a solid angle $\Delta\Omega$. The simulation consists in looping over all cells contained within $\Delta\Omega$. Image source (Wilman et al., 2008).

The simulation randomly draws its sources from the observed radio continuum luminosity function of the extragalactic populations, covering a sky area 20 $\times$20 deg^2 at 151, 610, 1400, 4860 MHz. The simulation starts at $z = 0$, using dark matter density field $\delta\rho/\rho$ defined on a cuboid grid of 5Mpc h^{-1} comoving cells which make 550 $\times$ 550 $\times$ 1550 array. Radio sources are then distributed on this underlying matter distribution according to a prescription of how the various sub-populations of radio source (FRIIs, FRIs,radio-quiet quasars, star-forming galaxies and star-burst galaxies) relate to their typical dark-halo mass. I will use this simulation to compare my results on the space

density of high-redshift radio galaxies in of high-redshift radio galaxies in this thesis to the best available models of the expected evolution of high redshift radio sources.

3.3 21cmNEST

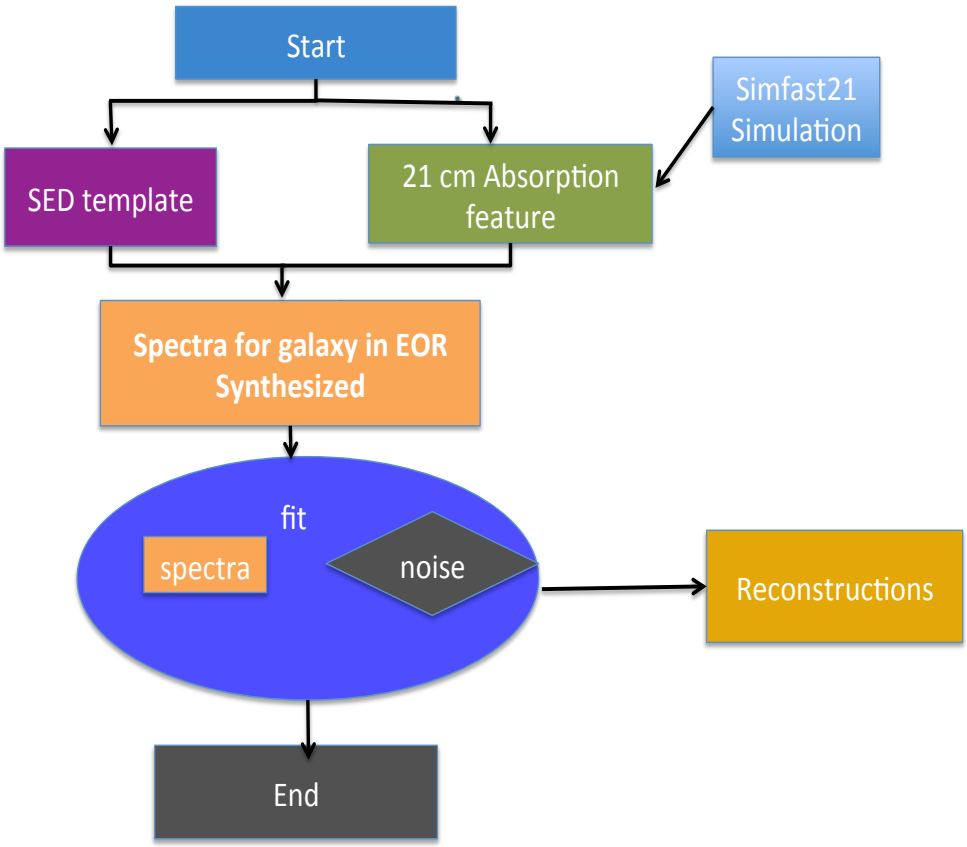


Fig. 3.3: The 21cmNEST algorithm. initialized inputs are given at "start" where the user gives redshift, luminosity and spectral index which is converted to spectra. The MultiNest algorithm is only implemented later on when it comes to the fitting stage; where recovery initial input parameters is done, by computing the likelihood using the best model and using the generated spectra as input data.

21cmNEST is the code I have developed to parameter estimate the crucial information such as redshift and luminosity from the spectra of radio galaxies form the Epoch of Reionization. 21cmNEST is designed to help solve statistical problems using a Bayesian inference. It is implemented in both FORTRAN and Python which allows for

a flexible definition of mathematical models while keeping the reliability and speed requirements of the numerical operation. The algorithm is parallelized using OpenMPI.

The code uses information provided by the user to synthesise the expected spectra for radio galaxies with 21 cm absorption feature. This involves accessing the evolved spatial attributes like ionization fraction, spin temperature and others from SIMFAST21 to construct the 21 cm absorption feature. The code assumes a power law for the radio luminosity and take a value at 1.4GHz taken from low z galaxies since we do not have a model for radio emission at high z. Note however that different luminosity amplitudes will be considered later. SIMFAST21 is very efficient in noise scales comparable to the expected SKA1-LOW telescope. The frequency range explored for this work is between 50 MHz-250 MHz. The template is projected to high redshift, to give the expected flux values for high-redshift galaxies. The 21 cm absorption feature is combined with the high-z projected SED to give the expected spectra. Fig.3.3 shows the steps undertaken in developing the code. The schematic is shown in Fig.3.3 is a single script that executes all the "steps" seen in Fig.3.3.

For the parameter estimation, the code attempts to recover the input information initially provided to construct spectra by the user. The parameter estimation is achieved through a Bayesian Inference. For the Bayesian inference, the code has incorporated Multinest (Feroz, Hobson, and Bridges, 2009; Feroz et al., 2013) to parameter estimate. 21cmNEST is relatively fast, on average a run takes roughly 5 minutes to complete. SIMFAST21 code will be available on Github as soon I publish the results. Eventually 21cmNEST will be performed on actual observations. However, for this proof of concept, we first create mock observations of the spectra of galaxies in the EOR and generate sensitivities for SKA1-Low. In this thesis I have ran 21cmNEST using a high performance computing facility. Only a single node with 24 cores while the RAM used was 2 GB. The code takes roughly 4 minutes to compute a single run.

Chapter 4

21 cm Absorption

Consider the incident radiation emitted by a radio-luminous source at some arbitrary redshift z_g. We observe the spectral intensity I_ν at a frequency ν; the signal can be measured in the radio as an effective temperature of a blackbody, defined by Rayleigh Jean's limit of the Planck formula, $I_\nu \equiv 2Tk_B\lambda^2$, where λ the wavelength of the radiation. The existence of neutral hydrogen prior to reionization offers the possibility of probing reionization and dark ages because HI atoms in the (IGM) absorb incoming radiation from the first luminous sources. Hence we would have a better understanding of the distribution of matter and IGM temperature.

At precisely the frequency ν_{21} =1420 MHz (which corresponds to λ =21cm), the incident radiation is absorbed by any neutral HI atoms. There are two other contributing terms, one from the 21 cm emission and the other from the CMB so that we end up with:

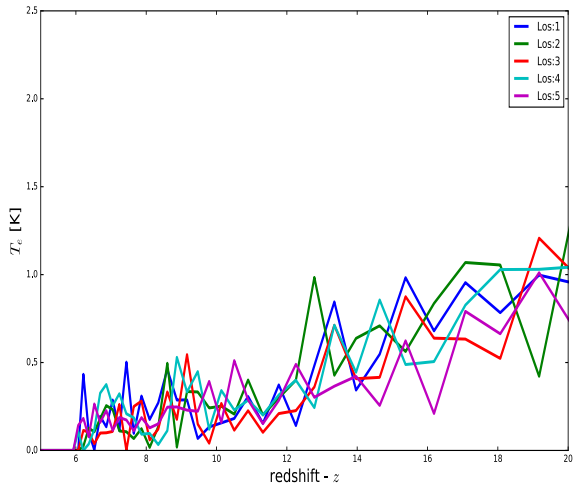
$$T_f = T_i\left[1 - \frac{T_e(z)}{T_{\rm S}(z)}\right] + T_{CMB}\left[1 - \frac{T_e(z)}{T_{\rm S}(z)}\right] + T_e(z). \tag{4.1}$$

what I call T_e (see eqn.(4.3)) is the emission temperature, while T_{CMB} is the CMB temperature and $T_{\rm S}$, the spin (excitation) temperature related to the relative population of the hyperfine levels. Note that all these quantities in eqn.(4.1) depend on the line of sight. Due to a negligible contribution from both 21 cm emission and from the CMB in eqn.(4.1), we can calculate to a good approximation the amount of absorption expected:

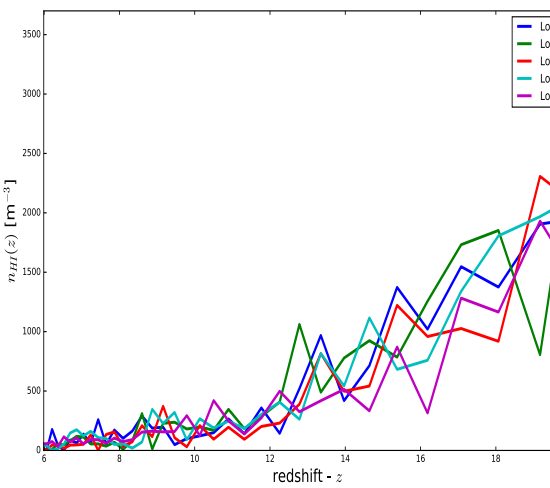
$$T_f = T_i\left[1 - \frac{T_{\rm e}(z)}{T_S(z)}\right], \tag{4.2}$$

where T_i is the initial temperature associated with the incident radiation emitted by the radio source and T_f is the observed temperature. For our case, we assume the spin temperature is coupled to the gas temperature, which should be an accurate approximation for z<25. For $T_{\rm e}$, we have:

$$T_{\rm e} \approx 1.82 \times 10^{-28}\frac{n_{HI}(z)\cdot c}{H(z)}\,{\rm k}\,, \tag{4.3}$$



(a)



(b)

Fig. 4.1: Simulation outputs in the redshift range $6 \leq z \leq 20$. The panels show the emission temperature and the number density for Hydrogen atoms derived. We took 5 different lines of sight (los). The size of the simulation box is 100Mpc/h. (a) T_e emission temperature from equation(4.3) (b) n_{HI} number density for Hydrogen atoms from equation (4.4). Cosmological parameters used correspond to those in (Collaboration, 2016).

where c is the speed of light, $n_{HI}(z)$ is the neutral hydrogen number density;

$$n_{HI}(z) \approx 90\Omega_b h^2 x_{HI}(z)[1+z]^3[1+\delta(z)]\ \mathrm{m}^{-3}. \tag{4.4}$$

Ω_b, is the baryonic matter density and $h \equiv \frac{H_0}{100}$. x_{HI} is the neutral fraction, $\delta(z)$ is the HI overdensity and $H(z)$ is the Hubble rate. We note that z is the redshift along the line-of-sight, so that, for a given observation frequency ν_0, $z \equiv \nu_{21\,\mathrm{cm}}/\nu_0 - 1$.

4.1 Galaxy Flux

For a given galaxy with intensity I_0 observed with a radio telescope, the observed flux S_0 is defined as:

$$S_0 = \int W(\Omega - \Omega_0) I_0(\Omega) d\Omega. \tag{4.5}$$

W is some dimensionless window function describing the telescope point spread function (PSF)and Ω is the solid angle on the sky. Assuming $T_{\rm e}$, $T_{\rm CMB}$, $T_{\rm S}$ are the same across the PSF:

$$S(\nu_0) = S_i(\nu_0)\left[1 - \frac{T_e(z)}{T_S(z)}\right] + \Delta\Omega T_{CMB}\left[1 - \frac{T_e(z)}{T_S(z)}\right] + \Delta\Omega T_e(z). \tag{4.6}$$

Where S_i, is the flux received from the source if there was no absorption, $\Delta\Omega$ is the full width at half maximum (FWHM) PSF of the telescope. Since $\Delta\Omega T_{\rm CMB}$ and $\Delta\Omega T_{\rm e}$ are small, eqn.(4.6) can be approximated by:

$$S_{\rm EOR}(\nu_0) = S_i(\nu_0)\left[1 - \frac{T_e(z)}{T_S(z)}\right] \tag{4.7}$$

The eqn.(4.7) gives us the underlying physics behind the flux expected to be observed from galaxies in the EOR: the ratio of $T_e(z), T_S(z)$ informs as to how much of the incident light from the galaxy is absorbed by the HI clouds during reionization. The factor $1 - \frac{T_e(z)}{T_S(z)}$ is crucial because, intuitively we know that is T_e =0 in eqn.(4.7) before the incident light interacts with the neutral hydrogen. This means eqn.(4.7) reduces to $S_0(\nu_0) = S_i(\nu_0)$ without the presence of neutral hydrogen[1]. A radio galaxy at redshift z_g can be assumed to have a luminosity that follows a power law in frequency to a good approximation, e.g.[2]

$$L(\nu) \approx L_0\left(\frac{1.4\,\text{GHz}}{\nu}\right)^{\alpha} [\text{W/Hz}]. \tag{4.8}$$

Typically the spectral index is $\alpha \simeq 0.75$. This relates to the observed flux density through:

$$S(\nu_0) = \frac{L(\nu)}{4\pi(1+z_g)r^2}, \tag{4.9}$$

where r is the comoving distance to z_g and $\nu = \nu_0/(1+z_g)$. In order to give an idea of the values associated with this observation, let's assume a galaxy at $z_g = 8$ with

[1] Eqn.(4.7) is true only if we are assuming the background source brightness temperature is much higher than the absorbing galaxies.

[2] It is currently impossible to distinguish these high-z radio bright quasars from low-z radio galaxies. Upcoming large area surveys with NIR photometry, such as EUCLID or WFIRST could enable the identification of these high-z quasars providing targets for the SKA (Willott et al., 2010).

$L_0 = 10^{26}$[W/Hz];

$$S(\nu_0) \simeq 0.2\left(\frac{1.4\,\text{GHz}}{\nu_0}\right)^{\alpha}[\text{mJy}]. \tag{4.10}$$

This corresponds to about 1 mJy at the frequency 150-MHz (this frequency roughly corresponds to z=8). Now consider the radiation emitted by a radio galaxy at high-z at some redshift z_g. Then the observed flux is:

$$S(\nu_0) = 0.214\left\{\frac{L_0 \cdot 46.8}{10^{26}[\text{W/Hz}]\,(1+z_g)^{1+\alpha}}\right\}\left(\frac{9142.7\,\text{Mpc}}{r(z_g)}\right)^2\left(\frac{1.4\text{GHz}}{\nu_0}\right)^{\alpha}\ [\text{mJy}] \tag{4.11}$$

Where ν_0 is the frequency at which we are observing the flux S, α is the spectral index for a galaxy and L_0 [W/Hz] is the galaxy's luminosity.

If this radio galaxy is found within the epoch of reionization one should account for the expected 21 cm absorption in its observed spectral energy distribution. The way to account for the 21 cm absorption feature in the spectra of galaxies in reionization would be treating the incident signal from galaxy before being absorbed by the HI, then a separate component that would account for when the incident signal is absorbed. This would ideally be given by the product of a piecewise function and template for galaxies in EOR,

$$S_{EOR}(\nu_0) = S(\nu_0)A(z)\ [\text{mJy}], \quad A(z) \equiv \begin{cases}\left[1 - \dfrac{T_e(z)}{T_S(z)}\right] & \text{if } z_g > z, \quad (4.12a)\\ 1 & \text{if } z_g < z, \quad (4.12b)\end{cases}$$

where$A(z)$ is the 21 cm absorption feature expected to be seen in the spectra of radio galaxies during reionization.[3]. It is key to note that $T_S > T_e$. Our aim is to detect if the galaxy has some a portion of $A(z)$ in its spectra. In order to achieve that, we will try to find a simple model for $A(z)$ that fits the data (which in our case is taken from simulations). We attempt to model the behaviour of $A(z)$, since we will not be able to observe it alone, but will rather observe it as a product as in seen eqn.(4.12). If $z > z_g$, then there is no 21 cm absorption. After reionization $T_e = 0$, hence there is no absorption feature again. Fig.(4.2) shows the effect of the 21 cm absorption feature for a galaxy placed in different redshifts but with the same luminosity. Note, the plots in Fig.(4.2) are in agreement with intuition; the flux for lower redshift galaxies (of same luminosity) should be higher than it's corresponding counterpart located at a higher redshift. Another key point: the 21 cm absorption feature on high redshift galaxies should be more apparent than for a lower redshift galaxy because at high (z >12) redshifts there

[3]Note that $A(z)$ is a dimensionless factor.

is more neutral hydrogen and conversely in the lower redshifts (z <11) there is little neutral hydrogen left hence the 21 cm absorption would not be as significant in the spectra of low redshift galaxies and T_S is also higher.

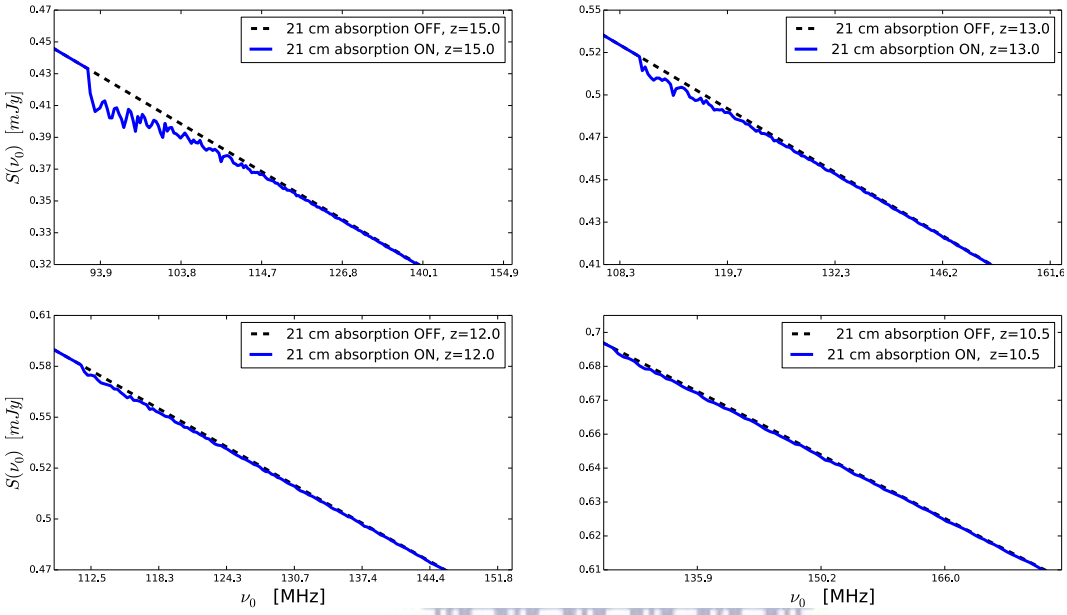


Fig. 4.2: An illustration of the impact the 21 cm absorption feature has on the spectrum of a galaxy with a luminosity, $L_0 = 10^{26}$[W/Hz]. The galaxy is placed on several different redshifts, z_g =15, 13, 12, 10.5. The black dotted lines show galaxy's spectrum with no 21 cm absorption feature (eqn.(4.11)), the blue line shows the galaxy's spectrum when the 21 cm absorption feature is accounted for (eqn.(4.12)).

4.2 Absorption Models

In the previous section, we have constructed the expected spectra (see Fig.(4.2)) for galaxies in EOR. The task that remains is finding candidate models for the expected 21 cm absorption feature, $A(z)$, in the spectrum of galaxies during reionization (see Sec.4.1). We find that fitting with polynomials is the most effective way to model $A(z)$. We could have chosen more exotic models for the feature, the only problem is that these exotic models may require additional parameters. When choosing the models we attempt to keep polynomial order as low as possible. Arbitrary fitting of higher order polynomials can be ineffective in a regression analysis.

$$\ln(1 - A_n) = a_0 + a_1 \ln(1+z) + a_2 \ln(1+z)^2 + ... + a_n \ln(1+z)^n = \sum_{i=0}^{n} a_i \ln(1+z)^i, \tag{4.13}$$

The polynomial general expression is expressed in a logarithmic scale because the trend is not as obvious in a linear scales. The first four polynomials suffice; to approximate the behaviour seen in galaxies during and after EoR[4]. These models can be expressed as:

$$A_0(z) = 1, \forall z \tag{4.14}$$

$$A_1(z; z_g, a_0, a_1) = \begin{cases} 1 - \left[e^{a_0}(z+1)^{a_1}\right], & \text{if } z_g > z, \\ 1, & \text{else} \end{cases} \tag{4.15}$$

$$A_2(z; z_g, a_0, a_1, a_2) = \begin{cases} 1 - \left[e^{a_0}(z+1)^{a_1} e^{a_2[\ln(1+z)]^2}\right], & \text{if } z_g > z, \\ 1, & \text{else} \end{cases} \tag{4.16}$$

$$A_3(z; z_g, a_0, a_1, a_2, a_3) = \begin{cases} 1 - \left[e^{a_0}(z+1)^{a_1} e^{a_3[\ln(1+z)]^3 + a_2[\ln(1+z)]^2}\right], & \text{if } z_g > z, \\ 1, & \text{else} \end{cases} \tag{4.17}$$

a_0, a_1, a_2 ... a_n are dimensionless coefficients. These three models $\{A_1, A_2, A_3\}$ describe the shape of data being fitted (i.e 21 cm absorption feature constructed from SIMFAST21 spatial data). In furthering the investigation steps are ran by rigorously finding the best model using Bayesian inference. As explained in Sec.2; Bayesian inference determines whether to pursue a complicated model with many parameters or a simple one with fewer parameters; essentially invoking Occam's razor. The parameter estimation is performed at different noise scales, this serves as a an efficiency test of SIMFAST21.

Fig.(4.3) shows the resulting reconstructions found when fitting the 21 cm absorption feature, $A(z)$, using the three polynomials. It is found that models A_1, A_2 and A_3 (i.e 1st-, 2nd- and 3rd- order polynomials) reasonably fit the distribution of data points of $A(z)$ (as seen in Fig.(4.3). Based only on Fig.(4.3) we cannot adequately judge which of these models is superior. The models in Fig.(4.3) are very accurate, there is

[4]The three polynomials are accompanied by the zeroth order polynomial to account for the case where the galaxy is found after the reionization era

a strong correlation between the predictions from model and the actual data. Fig.(4.3) shows corresponding models with a 68 percent confidence interval. Note that the labels on Fig.(4.3) are $\ln(1+z)$ on x-axis and y-axis $\ln[1-A(z)]$, respectively. The data in Fig.(4.3) only accounts for a single line of sight from SIMFAST21, a different line of sight would be similar the one seen in Fig.(4.3) the only differences would be in the structure since the density distribution is not the same throughout simulated space.

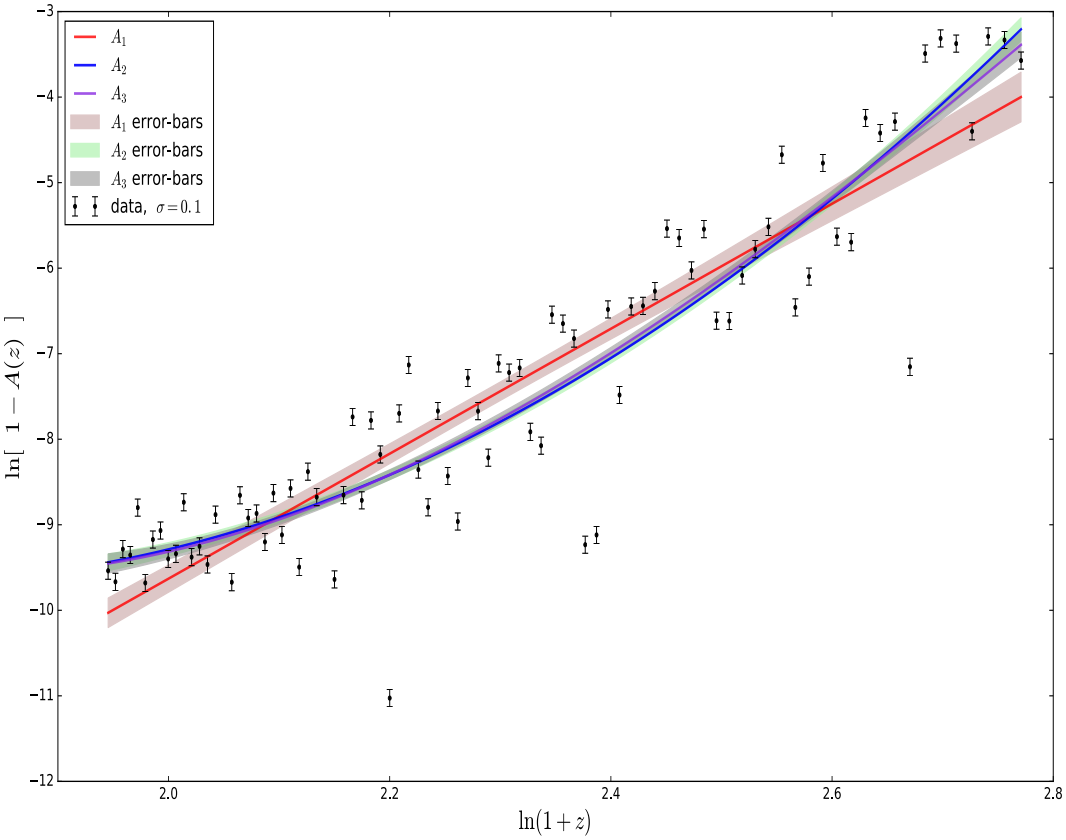


Fig. 4.3: SIMFAST21 constructed 21 cm absorption feature ,$A(z)$, shown as a function of redshift (black points). The error-bars shown correspond to a simplistic noise case, σ =0.1 mJy. The lines (purple, blue, red) are the resulting from fitting models A_3, A_2, A_1 to the 21 cm absorption feature $A(z)$. The shaded regions for each of the lines represent 68 percent confidence levels. The data used is for a single line-of-sight from SIMFAST21.

The coefficients found when fitting the these models are: $A_3 : \{a_3 = -2.58, a_2 = 23.577, a_1 = -60.37, a_0 = 37.758\}$, $A_2 : \{a_2 = 6.187, a_1 = -21.63, a_0 = 9.224\}$, $A_1 : \{a_1 = 7.3048, a_0 = -24.24\}$. It is important to not that we should not put an emphasis on the parameters found for the fits (Showed in Fig.(4.18)), because that alone is not enough to inform us on which of these models best fit the given data. In Fig.(4.18) we only use a single line of sight, because the general trend of the 21 cm absorption feature is the same for every line of sight, there are differences in the feature which

are due to randomness of structure appearing throughout the SIMFAST21 simulation. Since the modelling is no way able model the behaviour of the structure, or predict where it will appear, we thus ignore the structure and assume that it is the same for all lines of sight, since its effect negligible to an extent. For these reasons mentioned above we proceed with the modelling using only a single line of sight from SIMFAST21.

As mentioned above, the modelling approach pursued assumes we can decompose the 21 cm absorption feature from the luminous object's spectrum,

$$S_{\text{model}}(\nu_0) = S(\nu_0)A_{\text{model}}(z) \ \ [\text{mJy}], \tag{4.18}$$

$A_{\text{model}}(z)$, denotes a function that models the 21 cm absorption feature (where $\{A_1, A_2, A_3\}$ are the candidates models for $A_{\text{model}}(z)$), in Fig.(4.18) we have shown how these models approximate 21 cm absorption feature. The spatial attributes are extracted from SIMFAST21 used construct $A(z)$, is the assumed true data. Since no observation of the 21 cm absorption feature have been made. In the modelling, a difficulty arises when parameter-estimating the 21 cm absorption using these contender models, since $A(z)$ and the models have a sharp edge (similar to Heaviside functions). A solution to this problem is found by integrating over bins in z and similarly for the models, only after that we begin to parameter estimate. If we use models and parameter-fit directly we encounter numerical issues with the sharp edge of the models; since they are discontinuous, which makes parameter estimation a strenuous task. The integration is used because the experiment does not have infinite frequency resolution. The equation should be of the form: $\widetilde{A} \equiv \int A(\nu)\, \mathrm{d}\nu$. Another requirement is that $\widetilde{A}$ be dimensionless just like $A(z)$. We consider the following equation as the basis for the parameter estimation:

$$\widetilde{A}(z) = \frac{1}{\Delta\nu_0} \int_{\nu_i}^{\nu_{i+1}} A(z)\, \mathrm{d}\nu_0, \tag{4.19}$$

ν_i, ν_{i+1} is the frequency bin interval in the line -of-sight of the incident light ray, and $\Delta\nu_0 \equiv \nu_{i+1} - \nu_i$ [5]. All the equations defined using the form A_i can now be described by eqn.(4.19). The 21 cm absorption feature $A(z)$ is transformed to $\widetilde{A}(z)$, while A_0, A_1, A_2 and A_3 become $\widetilde{A}_0, \widetilde{A}_1, \widetilde{A}_2$ and $\widetilde{A}_3$ respectively.

Now the flux eqn.(4.18), it can now be re-defined to include $\widetilde{A}$. A re-defined eqn.(4.18) compares well with eqn.(4.19), since they are related through $z = \frac{\nu_{21\,\text{cm}}}{\nu_0} - 1$, this means the flux density of a galaxy located in reionization can be modelled using eqn.(4.21)

[5] The term $\frac{1}{\Delta\nu_0}$ ensures that the equation is dimensionless.

which incorporates integration over frequency bin interval which is based on the bandwidth $\Delta\nu$ used for a specific telescope used, because of the finite resolution which is dependent on telescope. The equations for the expected spectra S_{EOR} and modeling are defined as:

$$S_{EOR}(\nu_{i0}) = \frac{1}{\Delta\nu_0}\int_{\nu_i}^{\nu_{i+1}} S(\nu_0)A(z)\,\mathrm{d}\nu_0\ [\mathrm{mJy}], \tag{4.20}$$

$$S_{\mathrm{model}}(\nu_{i0}) = \frac{1}{\Delta\nu_0}\int_{\nu_i}^{\nu_{i+1}} S(\nu_0)A_{\mathrm{model}}(z)\,\mathrm{d}\nu_0\ [\mathrm{mJy}], \tag{4.21}$$

where A_{model} is any of these models A_0 A_1,A_2,A_3. Note that eqn.(4.14) and $S_{\mathrm{model}=A_0}(\nu_{i0})$ represent scenario where galaxy is found after EoR or the effect absorptin is minimal to it's spectrum (i.e the 21 cm absorption feature is not visible, this serve as the null hypothesis, in the sense that they are the worst case scenario we expect in the spectra of objects in EOR). Using eqn.(4.14) and $S_{\mathrm{model}=A_0}(\nu_0)$ the 21 cm absorption feature seen on a particular object can qualitatively measured to determine the level confidence about the detection. Tables(4.1,4.2) provide the priors used for model $\widetilde{A}_1$ and $S_{A_1}(\nu_0)$. Note eqn.(4.21) will only be used in the results (section 5.2), where the best model would have already been found.

I have specified the priors for each of these models used. This is because the prior quantifies existing knowledge and may be based on past theoretical predictions and for the parameter inference step. The prior for a parameter might be from previous observations or estimates from simulation. Sometimes the choice of priors on the parameters is taken to be uninformative. There are two types of priors I have used in this thesis, uniform and log-uniform prior.

Parameter	Prior
a_0	uniform $\in$ [0, 100]
a_1	uniform $\in$ [-100, 0]

Table 4.1: Priors used in model $\widetilde{A}_1$.

The priors showed in Table(4.1 & 4.2) were chosen because some are non-physical (a_1 & a_2)[6], and are others are physical but they all agree with our intuitive understanding of problem, these proposed priors agree with the beliefs for all the given prior distribution and I have checked if each prior really fits the purpose and that it is not sensitive to arbitrary choices (i.e robustness or sensibility analysis have been check for).

[6]The parameters a_1 & a_2 are non-physical because they have not yet been observational measured yet, hence we treat them as non-physical for the time being. However the rest of the parameters used are physical, their priors are based on SKADs and other available studies.

Parameter	Prior
z_g	uniform $\in$ [5.5, 22]
$\frac{L_0}{10^{26}][\mathrm{W/Hz}]}$	log- uniform $\in$ [10^{-6}, 10^{6}]
α	uniform $\in$ [0, 2]

Table 4.2: Priors used in model $S_{A_1}(\nu_0)$. Note that the priors for $\widetilde{A}_1$ are included when parameter estimating $S_{A_1}(\nu_0)$ since $S_{A_1}(\nu_0)$ incorparates $\widetilde{A}_1$.

In summary, $\widetilde{A}_i$ models, are the best route to model the 21 cm absorption feature. Since the goal was to find ways to model for the 21 cm absorption feature, leading us to be able to create a technique to estimate crucial parameters that include redshift and luminosity for galaxies found in the EOR.

I will often refer to the models $\{\widetilde{A}_0, \widetilde{A}_1, \widetilde{A}_2, \widetilde{A}_3\}$ as $\{0, 1, 2, 3\}$ to make the notation easier. The parameters a_0, a_1, a_2 and a_3 in the models are treated as dimensionless and fiducial because no observations have been done yet. In the near future when observations of the EOR are available, these parameters could be used to better understand the IGM environment.

Chapter 5

Fitting the 21 cm Absorption Feature

5.1 Systematic tests in Modelling

In this section, I run a series of tests checking for the robustness of the modelling algorithm. This is done using the 21 cm absorption feature and the synthesised data from the models $\{0, 1, 2, 3\}$. From the systematics, I aim to find the best model for the 21 cm absorption feature. This best model will be crucial in investigating further as to whether some systematics are due to nature of data from the 21 cm absorption feature or from the estimation itself. The main priority parameter in the fits is the redshift, because if can constrain the redshift then we can have limits to reionization. We assumed a frequency resolution of $\Delta\nu$= 4MHz for this simple noise case. Note that in order to determine the best model I use the null hypothesis as the point of reference (i.e the worst possible fitting since it does not account for the absorption feature).

The likelihood function $\mathcal{L}(\overrightarrow{\theta})$ also needs to be specified, it describes the probability of obtaining the data for given parameters and it depends on the experimental technique used to gather the data. In this case, the data were gathered subject to a small amount of Gaussian noise, so a Gaussian likelihood is appropriate:

$$\mathcal{L}(\overrightarrow{\theta}) = p(\mathcal{D}|\overrightarrow{\theta}) = \prod_{i=1}^{n} \sqrt{\frac{1}{2\pi\sigma_i}} \exp\left(-\frac{1}{2}\left(\frac{y_{\mathcal{D}} - y_{\overrightarrow{\theta}}}{\sigma_i^2}\right)^2 \right), \tag{5.1}$$

where $y_{\mathcal{D}}$, $y_{\overrightarrow{\theta}}$ correspond to the observed data and the model of choice. $y_{\mathcal{D}}$ is subject to a small amount of Gaussian noise of variance σ_i^2 such that: $y_{\mathcal{D}} \sim \mathcal{N}(y_{\overrightarrow{\theta}}, \sigma_i)$.

In these systematics tests, we also parameter fit the added noise. Assuming that the foreground cleaning/RFI (radio-frequency interference) complexities are solved we could be able to predict the expected observational noise, to an extent that noise models are developed. There are obvious advantages to parameter fitting the noise (by fitting the expected noise we have more prior information), it can be used as an indicator of how strong a certain model at estimating the given data. A successful fit should

have a recovered noise similar to the added noise. Reasons for this not being the same could be due too poor choice of models or issues in the structure of the input data. A crucial detail keep in mind is that in Subs.(5.1.1 and 5.1.2) I only fit the parameters: σ: noise, `Amplitude`: confirmation constant [1], z_g: the approximate position of source emitting light, a_0, a_1, a_2, a_3: fitting constants from the various models.

5.1.1 Flat Noise Profile

I first consider the simplest noise case, by using the constant (flat, i.e. one constant in freq) noise, which is generated from a Gaussian distribution with $\mathcal{N}(0, \sigma)$. We start out the parameter estimation with the simplistic case, as a check if whether a technique works properly in recovering added noise. Note that I used noise levels σ =0.1 mJy, 0.01 mJy and 0.001 mJy. The prior chosen for all three noise levels is σ [mJy]: uniform $\in$ [0.0,1.0].

The tables in subsection 5.1.1 and 5.1.2 are set up as follows; the first column represents the log-evidence "ln Z " found when fitting with some model "$\widetilde{A}_i$". While the first row represents the various data-sets available for models to parameter fit. In each column, the cell with the highest log-evidence value is used as a basis; i.e each column $\widetilde{A}_i$ represents data set used and each row represents the log evidence found when fitting with some model @$\widetilde{A}_i$. This means the cell with highest log-evidence in a column is set to zero, while the other cells in the column are subtracted with the highest evidence in a column, hence the rest of cells in the column are negative. This subtraction per column is done using quadrature error propagation.

Tables (5.1), (5.2) and (5.3) show the findings to model selection at different noise scales. While column four represents the 21 cm absorption feature data generated from SIMFAST21, row three to six represents the model used to find the evidence. In the three tables columns two, three, four and five represent simulated data created using each model. These serve as a sanity check for the parameter estimation method.

Model $\widetilde{A}_0$ corresponds to the case where there is no 21 cm absorption feature (a null hypothesis), hence the null model evidence is used as a relative evidence scale by finding its ratio with the 21 cm absorption feature contender models. Based on this null relative Evidence, we are able to find the best model.

[1] The Amplitude, It was basically used to check if the modelling is working properly. The value recovered for the amplitude should always be 1. The constant is create in way that it is dimensionless

For the parameter-fitting, we explore 2-dimensional triangle plots, which show the posterior distribution of the parameters estimated by the various models. The red lines on the triangle plot indicate the input parameters (i.e true parameters used to synthesize the input data). The statistical uncertainty in the parameters should be increasing as the noise is increased in the parameter estimation. Also note the numerical sampling of the posterior is carried out via 21cmNEST, a code I have developed, which incorporates MULTINEST. The stopping criterion is described in the MULTINEST (see Feroz, Hobson, and Bridges, 2009). It is set in terms of a target quantity in the evidence value 'evtol', which is the tolerance on the evidence value. Once this reaches the target, the stopping criterion is triggered, this is very different from a conventional MCMC code.

σ=0.001 mJy, $\Delta\nu$=4 MHz					
lnZ \ data	$\widetilde{A}_0$	$\widetilde{A}_1$	$\widetilde{A}_2$	$\widetilde{A}_3$	$\widetilde{A}$
$\ln Z@\widetilde{A}_0$	**0**	-88.43 $\pm$ 0.014	-106.3 $\pm$ 0.22	-2.66 $\pm$ 0.18	-57.44 $\pm$ 0.12
$\ln Z@\widetilde{A}_1$	-7.42 $\pm$ 0.14	**0**	-3.35 $\pm$ 0.28	-3.32 $\pm$ 0.21	**0**
$\ln Z@\widetilde{A}_2$	-19.2 $\pm$ 0.15	-17.84 $\pm$ 0.24	**0**	-1.27 $\pm$ 0.21	-7.44 $\pm$ 0.18
$\ln Z@\widetilde{A}_3$	-30.34 $\pm$ 0.15	-87.67 $\pm$ 0.19	-63.3 $\pm$ 0.23	**0**	-59.29 $\pm$ 0.15

Table 5.1: The log evidence found when fitting models $\widetilde{A}_0$,$\widetilde{A}_1$, $\widetilde{A}_2$ and $\widetilde{A}_3$ to the synthesized data generated from $\widetilde{A}_0$, $\widetilde{A}_1$, $\widetilde{A}_2$, $\widetilde{A}_3$ as mock data. In the last column all the models are fitted to $\widetilde{A}$. Note all of the parameter estimation is done at noise(σ)=0.001 mJy.

Table(5.1) shows each model's evidence found from the various sets of data. This table shows model evidence found for the smallest noise case. The model evidence highlighted by bold cells in the table is very large compared to the evidence from other cells in the first columns $\{\widetilde{A}_0,\widetilde{A}_1\}$. This suggests that bold cells are parameter estimated with the best model for the distribution of the respective data. In column $\{\widetilde{A}_2,\widetilde{A}_3\}$ the evidence in bold cells are not significantly larger than neighbouring evidence from other cells in the same column, this could be an indicator that these two models $\widetilde{A}_2$,$\widetilde{A}_3$ have too many parameters.

σ=0.01 mJy, $\Delta\nu$=4 MHz					
lnZ \ data	$\widetilde{A}_0$	$\widetilde{A}_1$	$\widetilde{A}_2$	$\widetilde{A}_3$	$\widetilde{A}$
$\ln Z@\widetilde{A}_0$	**0**	-6.08 $\pm$ 0.14	-16.37 $\pm$ 0.18	-0.71 $\pm$ 0.16	-35.74 $\pm$ 0.12
$\ln Z@\widetilde{A}_1$	-3.12 $\pm$ 0.13	**0**	-9.0 $\pm$ 0.23	-0.21 $\pm$ 0.18	**0**
$\ln Z@\widetilde{A}_2$	-1.0 $\pm$ 0.13	-2.29 $\pm$ 0.19	**0**	-0.45 $\pm$ 0.18	-5.55 $\pm$ 0.18
$\ln Z@\widetilde{A}_3$	-1.25 $\pm$ 0.13	-5.36 $\pm$ 0.16	-15.62 $\pm$ 0.2	**0**	- 36.66 $\pm$ 0.16

Table 5.2: Natural log evidence found when fitting models $\widetilde{A}_0$,$\widetilde{A}_1$, $\widetilde{A}_2$ and $\widetilde{A}_3$ to the synthesized data generated from $\widetilde{A}_0$, $\widetilde{A}_1$, $\widetilde{A}_2$, $\widetilde{A}_3$ mock data and $\widetilde{A}$ at noise(σ)=0.01 mJy.

In Table (5.1) where parameter estimation is done at noise level 0.001 mJy, the evidence is highest diagonally (pattern is highlighted with a bold type face). For the first four columns which represent $\{\widetilde{A}_0, \widetilde{A}_1, \widetilde{A}_2, \widetilde{A}_3\}$ this diagonal pattern is observed. Only in the last column, which is for 21 cm absorption data, we do see a break in pattern, Model $\widetilde{A}_1$ has the highest evidence, hence it parameter estimates better than models $\widetilde{A}_2$ and $\widetilde{A}_3$. In Table (5.1) & (5.2), we also observe the same diagonal pattern for the first four columns and in the last column for Table (5.2) & (5.3) model $\widetilde{A}_1$ continues to out-perform model $\widetilde{A}_2$ and $\widetilde{A}_3$. Note another pattern visible on the three tables, the evidence decreases with an increase in the noise level.

σ=0.1 mJy, $\Delta\nu$=4 MHz					
lnZ \ data	$\widetilde{A}_0$	$\widetilde{A}_1$	$\widetilde{A}_2$	$\widetilde{A}_3$	$\widetilde{A}$
$\ln Z@\widetilde{A}_0$	**0**	-0.91 $\pm$ 0.11	-0.76 $\pm$ 0.14	-0.68 $\pm$ 0.3	-0.52 $\pm$ 0.11
$\ln Z@\widetilde{A}_1$	-1.22 $\pm$ 0.11	**0**	- 0.15 $\pm$ 0.16	-0.28 $\pm$ 0.11	**0**
$\ln Z@\widetilde{A}_2$	-1.94 $\pm$ 0.11	-0.49 $\pm$ 0.13	**0**	-0.48 $\pm$ 0.16	-1.22 $\pm$ 0.11
$\ln Z@\widetilde{A}_3$	-1.93 $\pm$ 0.11	- 0.93 $\pm$ 0.12	-0.27 $\pm$ 0.16	**0**	-0.64 $\pm$ 0.28

Table 5.3: The log evidence found when fitting models $\widetilde{A}_0$,$\widetilde{A}_1$, $\widetilde{A}_2$ and $\widetilde{A}_3$ to mock data generated from $\widetilde{A}_0$, $\widetilde{A}_1$, $\widetilde{A}_2$, $\widetilde{A}_3$ and whilst also fitting $\widetilde{A}$ at noise(σ)=0.1 mJy.

We define the convergence of a chain by looking at the posterior distributions recovered for all the parameters in a fit. If the posterior distribution for all the parameters each peaks at single point in the given range then this clearly indicates that the chain has converged. This convergence is only true if the posterior distribution is within the given prior range.

In Fig.(5.1) we show the model selection, based on parameter estimation using models

$\widetilde{A}_1$, $\widetilde{A}_2$, $\widetilde{A}_3$ on the 21 cm absorption feature data, $\widetilde{A}$. The evidence is represented as a ratio of model $\widetilde{A}_0$. Finding relative evidence of the models as ratio to $\widetilde{A}_0$, allows us to have a relative scale, since model $\widetilde{A}_0$ has no absorption (null hypothesis). The amplitude coefficient (seen in the 2-dimension posteriors) is dimensionless, it only serves to check if the model is biased in any way, it has no other role.

The null relative evidence is defined as the difference in Bayesian model evidence:

$$\Delta \ln Z \equiv \ln Z_i - \ln Z_0, \tag{5.2}$$

where $\ln Z_i$, is the Bayesian model evidence any of the 21 cm absorption feature models and $\ln Z_0$ represents the null hypothesis Bayesian model evidence. Eqn.(5.2) enables one to have a more quantitative scale since model A_0 is the model with no absorption (i.e the null hypothesis). This actually allows a particular object's 21 cm absorption feature to be quantified in strength relative to the null hypothesis.

Note the uncertainties found for $\Delta \ln Z$ appear to be very small, this undoubtedly makes it easier to choose the best model to use in the parameter fitting of the 21 cm absorption feature. $\Delta \ln Z$ is a fully Bayesian proxy for the signal to noise ratio (SNR) because we can apply Jeffrey's scale (2.1). Essentially, relative null Bayesian Evidence will inform us whether an object with some $\Delta \ln Z$ is observed to have the 21 cm absorption but also interpreted on the confidence of detection using Jeffrey's scale.

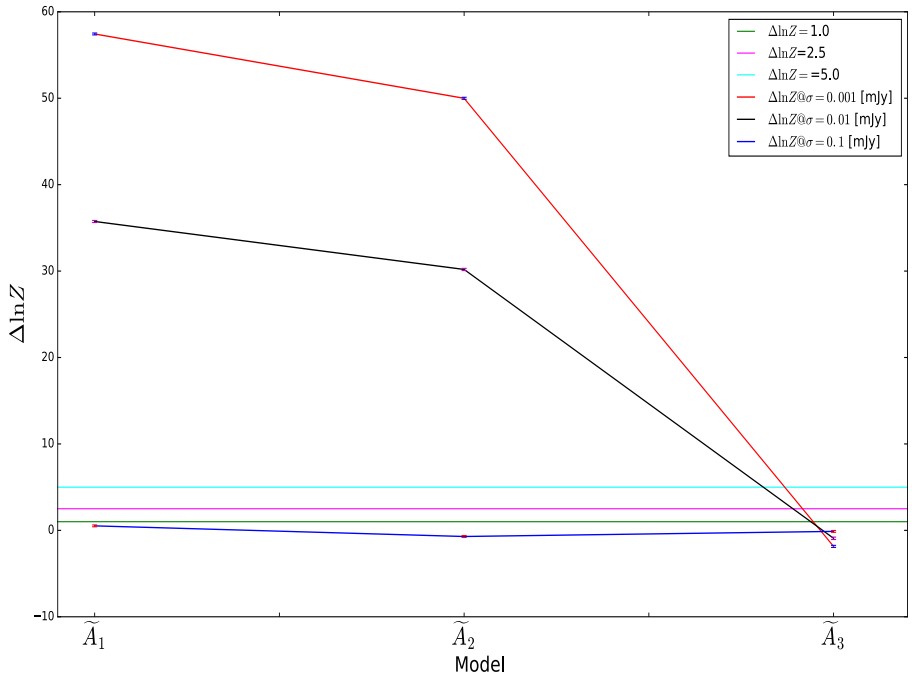


Fig. 5.1: Evidence, relative to the null hypothesis , $\Delta \ln Z$, found for each model ($\widetilde{A}_1$, $\widetilde{A}_2$, $\widetilde{A}_3$) fitting $\widetilde{A}$, each line represents different noise level, $\sigma = 0.001$ mJy (red line), $\sigma = 0.01$ mJy (black line), $\sigma = 0.1$ mJy (blue line). The three horizontal lines labelled 1, 2.5 and 5 correspond to the intervals in the Jeffreys scale.

The lines in Fig.(5.1) corresponding to a noise of 0.001 mJy (red line) and 0.01 mJy (black line) seem to have similar behaviour, both initially have high null relative evidence ratio values around model 1, then the null relative evidence ratio begins to decline around the models 2 and 3. These two lines correspond to $\sigma = 0.001$ mJy (red line) and $\sigma = 0.01$ mJy (black line). The line corresponding to 0.1 mJy noise (blue line) seems to be flat throughout, this most likely tells that our models are unable to fit at this noise level. Fig.(5.1) sheds light on the problem of finding the winning model out of $\{1, 2, 3\}$. Model 1 is the best model of out of the three based on the two simplistic noise levels $\{\sigma = 0.001, 0.01\}$ mJy. In the case of $\sigma = 0.1$ mJy, note that the parameter fitting fails to distinguish the better model of out of the three, the models are on an equal footing when noise used to is (σ =0.1 mJy). This instance shows we can use null relative evidence to infer whether an object with 21 cm absorption can be detected at a particular noise level, this means $\Delta Z \approx SNR$.

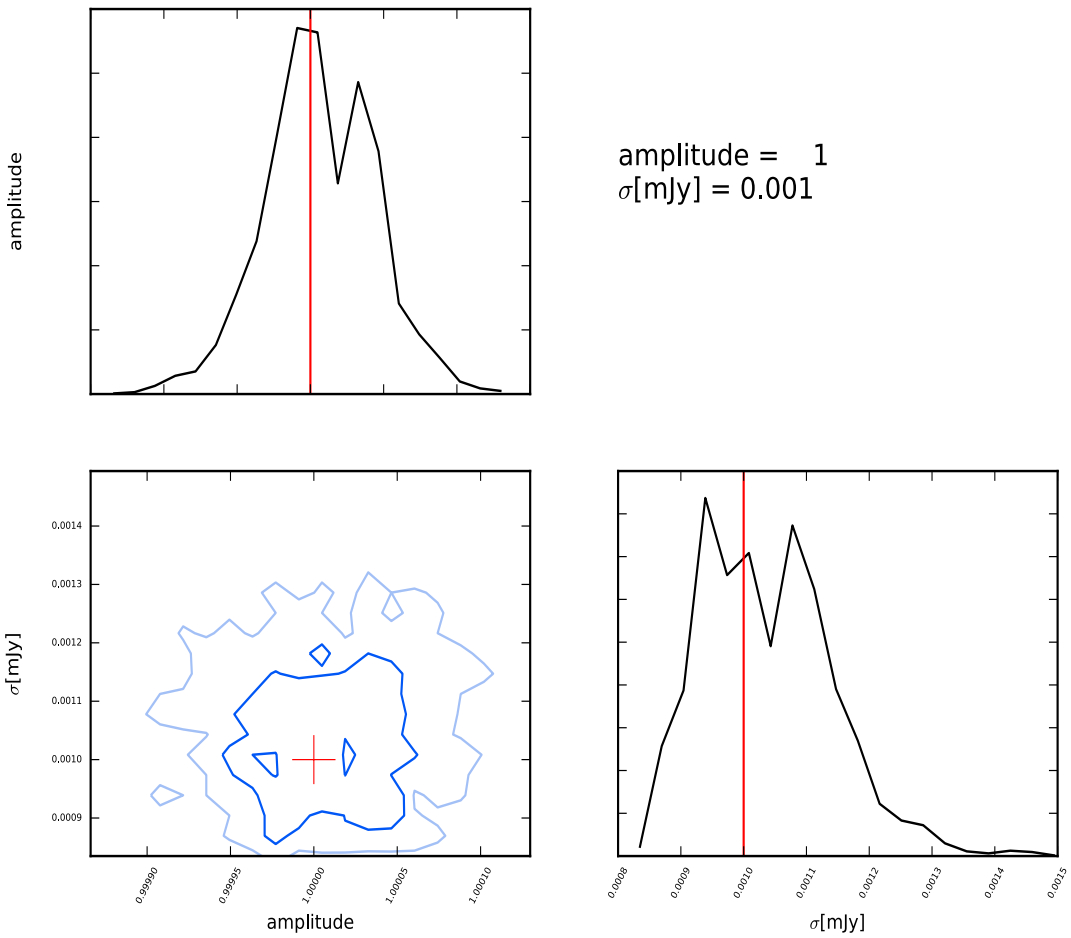
This behavior represented by the blue line Fig.(5.1) suggests the parameter estimation the data are too noise to constrain the model. In all the noise levels model 3 is around the same region.

Note that in the case of σ=0.1 mJy, the models have evidence values that are in the region as "inconclusive" on Jeffery's scale, even though the evidence in some cases slightly favours model with 21 cm absorption feature over the null hypothesis, but the confidence in the result is low, this may as well favour the null hypothesis (negative relative evidence).

Now we present the systematics in the form of triangle plots for each of the models. These are shown in the form of the posterior distribution function of the parameters in of each model.

The triangle plot has contours composed of combinations each pair of parameters. Ideally, the contour should be symmetric and circular in shape, which indicates consistent parameter estimation with data. In the contours, there are two, sometimes distinct contour lines. The inner one indicates 68 percent confidence interval while the outer line indicates the 98 percent confidence interval.

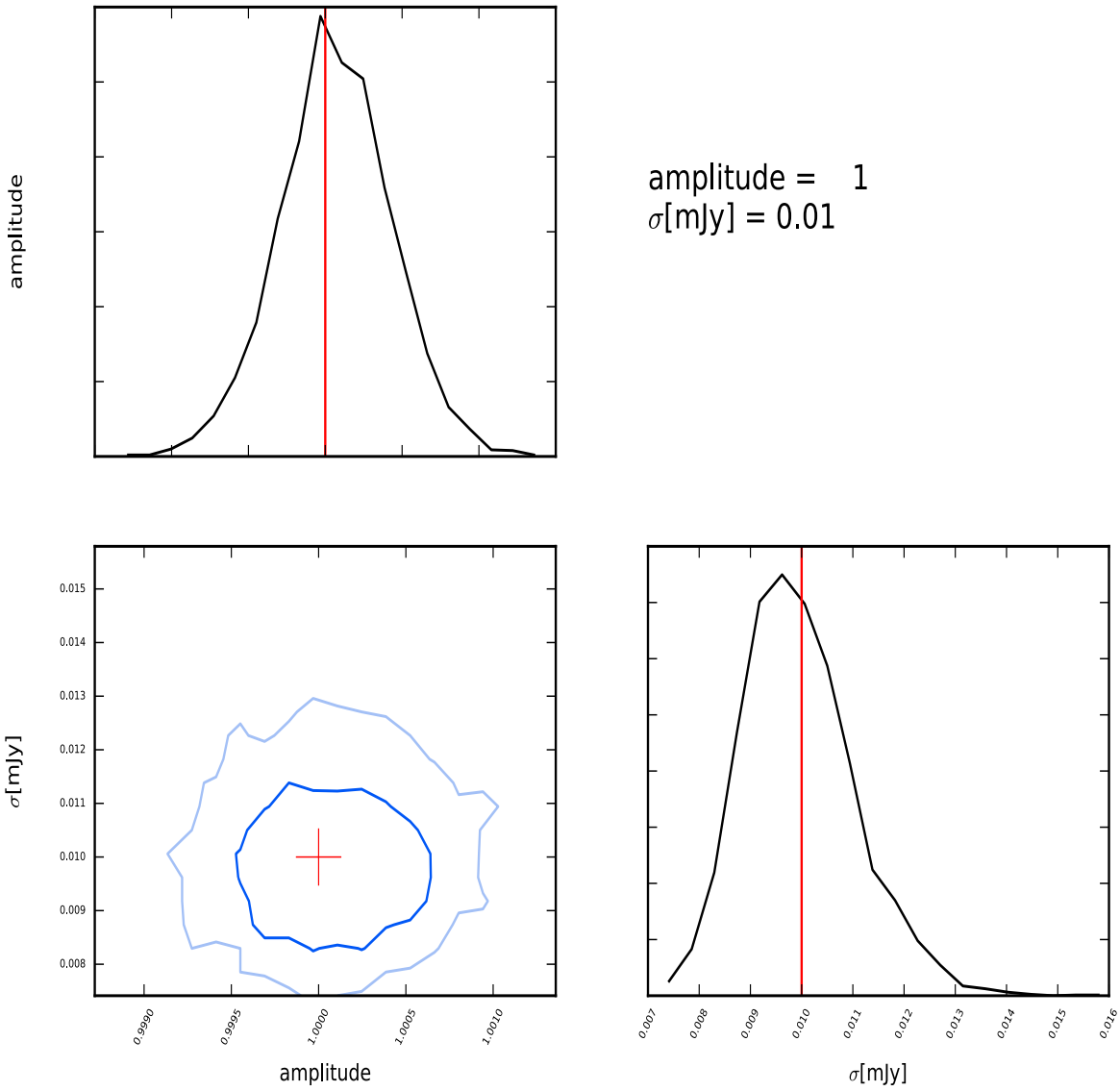
In Fig.(5.2) three triangle plots are shown for model $\widetilde{A}_0$, fitted to simulated data created also using model $\widetilde{A}_0$. For all three triangle plot each corresponding to noise levels of σ =0.001, 0.01, and 0.1 mJy, the posterior distributions are consistent compared to red lines which are the truth values, this pattern is observed in all three triangle plots. In Fig.(5.2a) the signal and noise are consistent since the fit is able to recover the input noise. The truth values as well are in agreement with the parameter fitting. Fig.(5.2b) shows the null-generated data fitted with null model. The posterior distributions for each of the parameters are narrow and centered on the true value, which shows the fitting is able to separately recover the signal the input noise. The probability distribution at σ =0.01 mJy is slightly distorted. In Fig.(5.2a) both σ and `amplitude` are well constrained by the posterior.



(a) σ=0.001 mJy

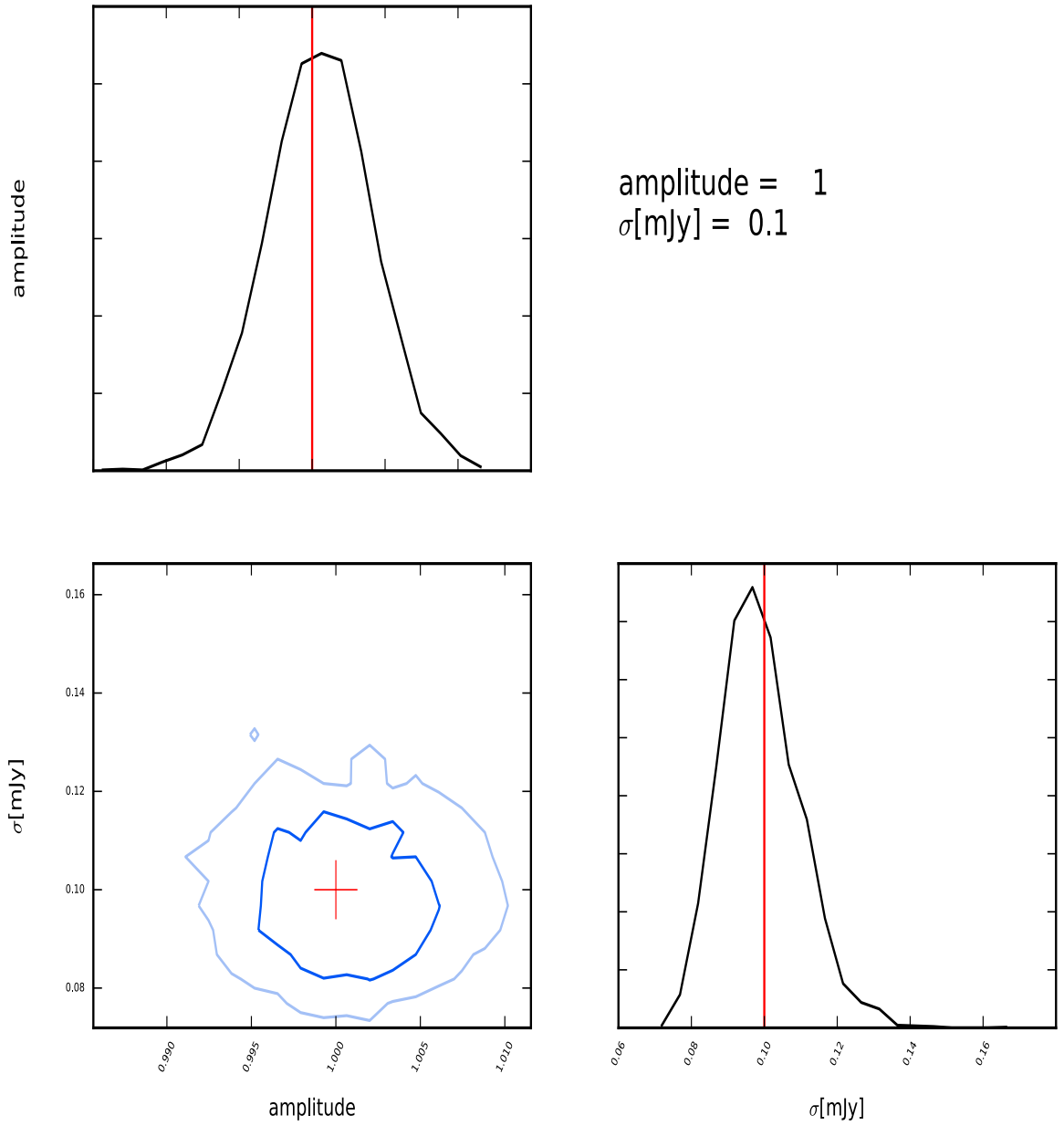
Fig. 5.2: The marginalized posterior distributions for the case where $\widetilde{A}_0$ is used to synthesize mock data and fitted with $\widetilde{A}_0$. The inner dark blue line indicate the 68 percent and the light region the 95 percent confidence levels. On **a**:$\widetilde{A}_0$ mock data is fitted with $\widetilde{A}_0$ at noise level (σ=0.001 mJy), **c**:$\widetilde{A}_0$ mock data is fitted with $\widetilde{A}_0$ at (σ=0.01 mJy), **c**: $\widetilde{A}_0$ mock data is fitted with $\widetilde{A}_0$ at (σ=0.1mJy). Red lines are the true values.

The recovered posterior lies within 1σ of the true values in Fig.(5.2c). Generally, an exact match between the posterior with the true values is not expected but the null is the trivial fitting. There are obvious differences in the triangle plots even though in all three triangle plots are consistently recovered. In Fig.(5.2a) the contours are slightly distorted even though parameters have been recovered. In Fig.(5.2b) and Fig.(5.2c) the contours are more distinctly defined and have symmetry.



(b) σ=0.01 mJy

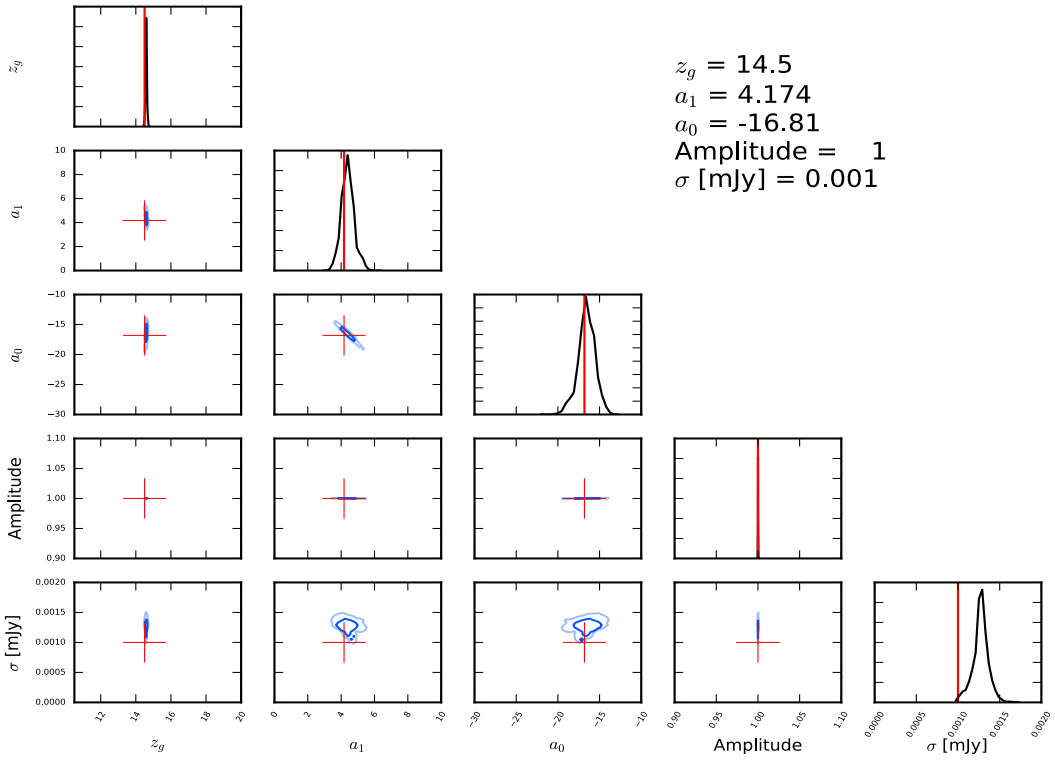
Fig. 5.2: continuation of Fig.(5.2)



(c) σ=0.1 mJy

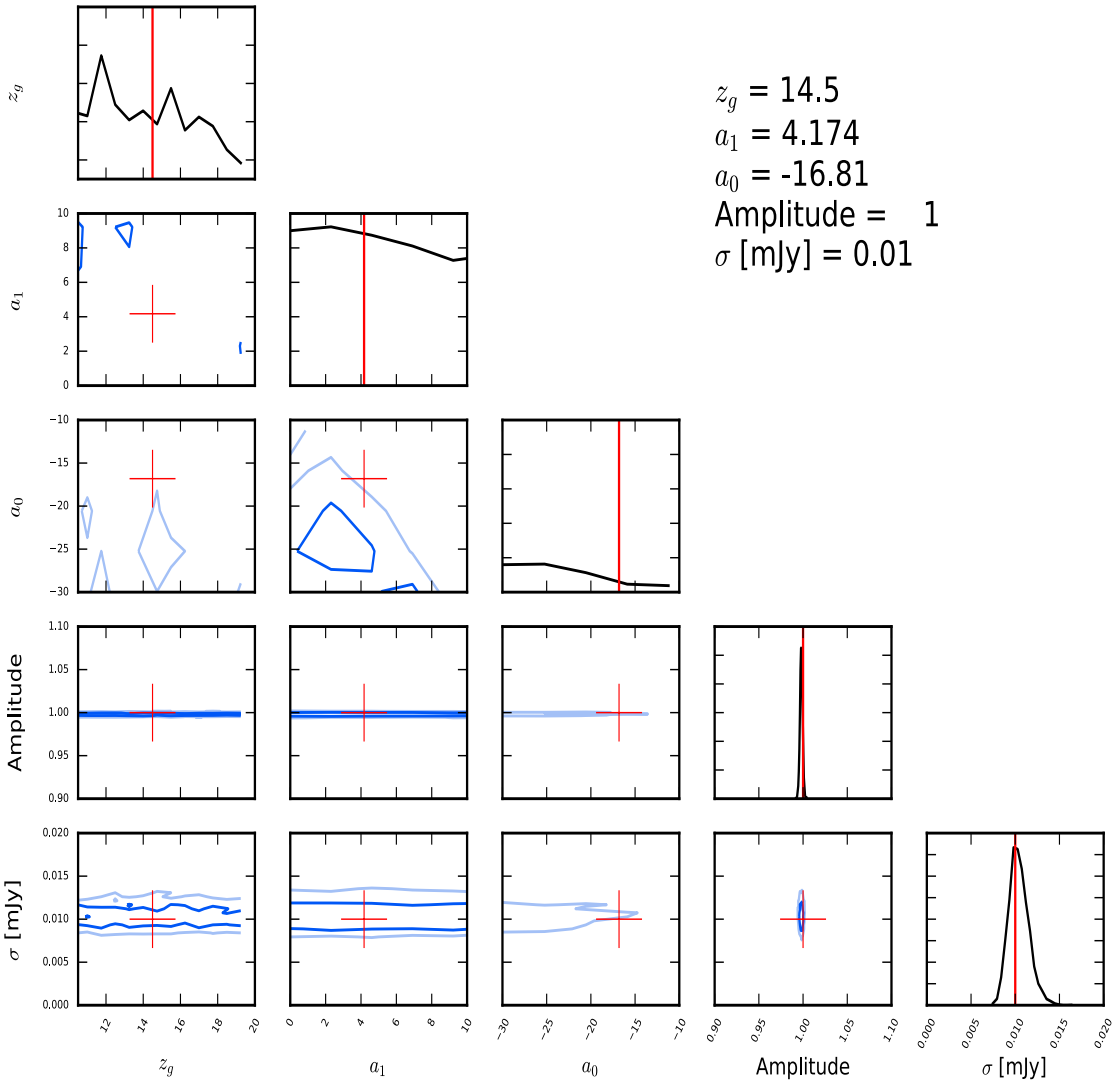
Fig. 5.2: continuation of Fig.(5.2)

In Fig.(5.3) we show parameter marginalization for posterior distribution for parameters from model $\widetilde{A}_1$. It is clear from Fig.(5.3a) (σ =0.001 mJy) parameter estimation for all the model parameters have been estimated to reasonable accuracy. The parameter fitting is in agreement with the true values in Fig.(5.3a) while the recovered noise is slightly biased. In Fig.(5.3) shows marginalization using model $\widetilde{A}_1$'s simulated data. The posterior for the redshift in Fig.(5.3a) seems to have accurately recovered the redshift. The fittings are done at noise levels σ =0.001, 0.01 and 0.1 mJy. The probability distribution for the parameters σ =0.001 mJy and 0.01 mJy is consistent with the true values and posteriors increase with increasing noise. For the case σ =0.1 mJy where some of the probability functions never converge. The most probable reason is the fact that the signal to noise ratio(SNR) is very low.



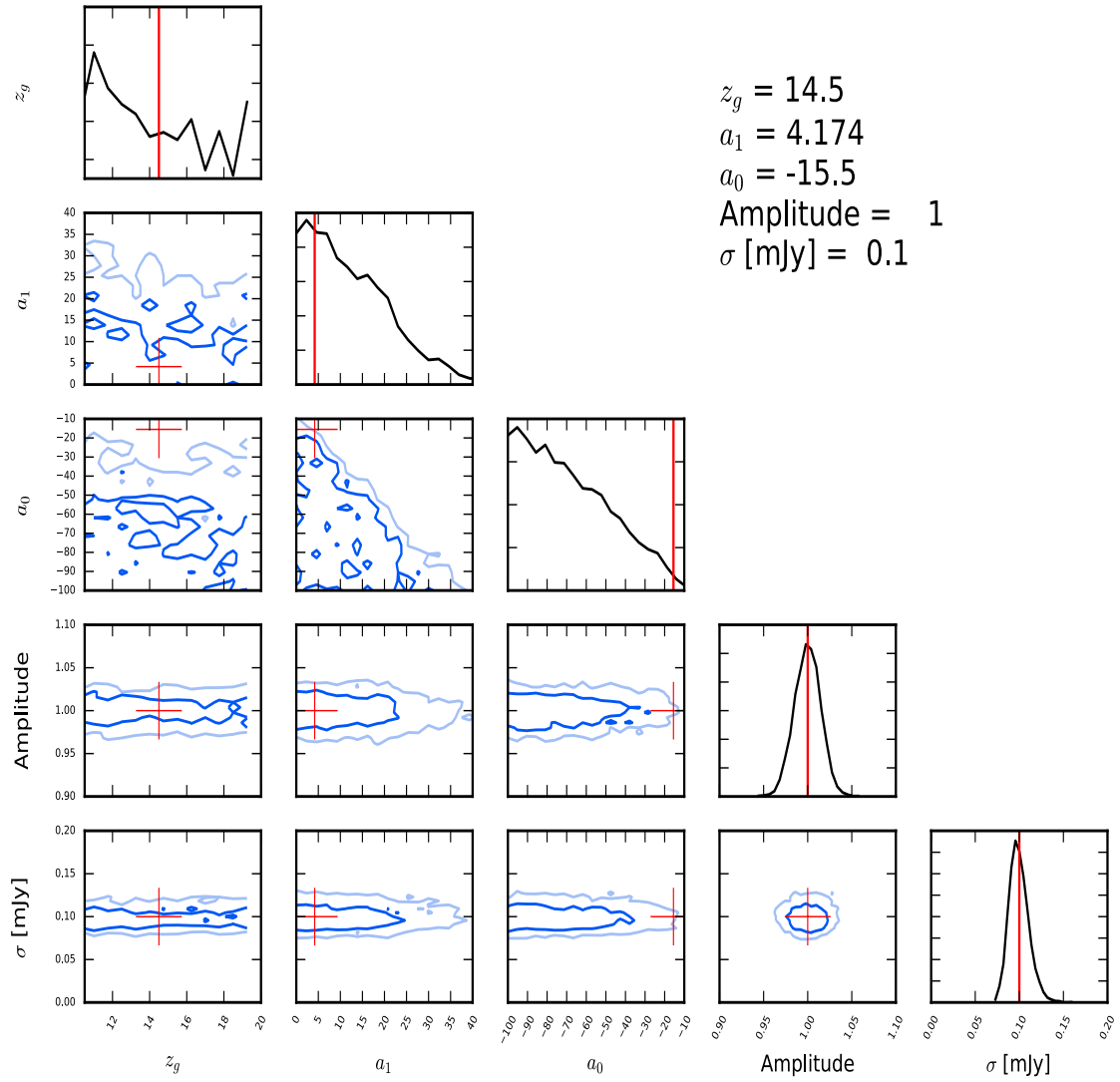
(a) σ=0.001 mJy

Fig. 5.3: The posterior distributions for the case where $\widetilde{A}_1$ is used to synthesize mock data and fitted with $\widetilde{A}_1$. The inner dark blue regions indicate the 68 percent and the light region the 95 percent confidence levels. On **(a)**:$\widetilde{A}_1$ mock data is fittted with $\widetilde{A}_1$ at noise level (σ=0.001 mJy), **(b)**:$\widetilde{A}_1$ mock data is fitted with $\widetilde{A}_1$ at (σ=0.01 mJy), **(c)**: $\widetilde{A}_1$ mock data is fittted with $\widetilde{A}_1$ at (σ=0.1 mJy). Red lines are the fiducial values.



(b) σ=0.01 mJy

Fig. 5.3: Contuination of Fig.(5.3)



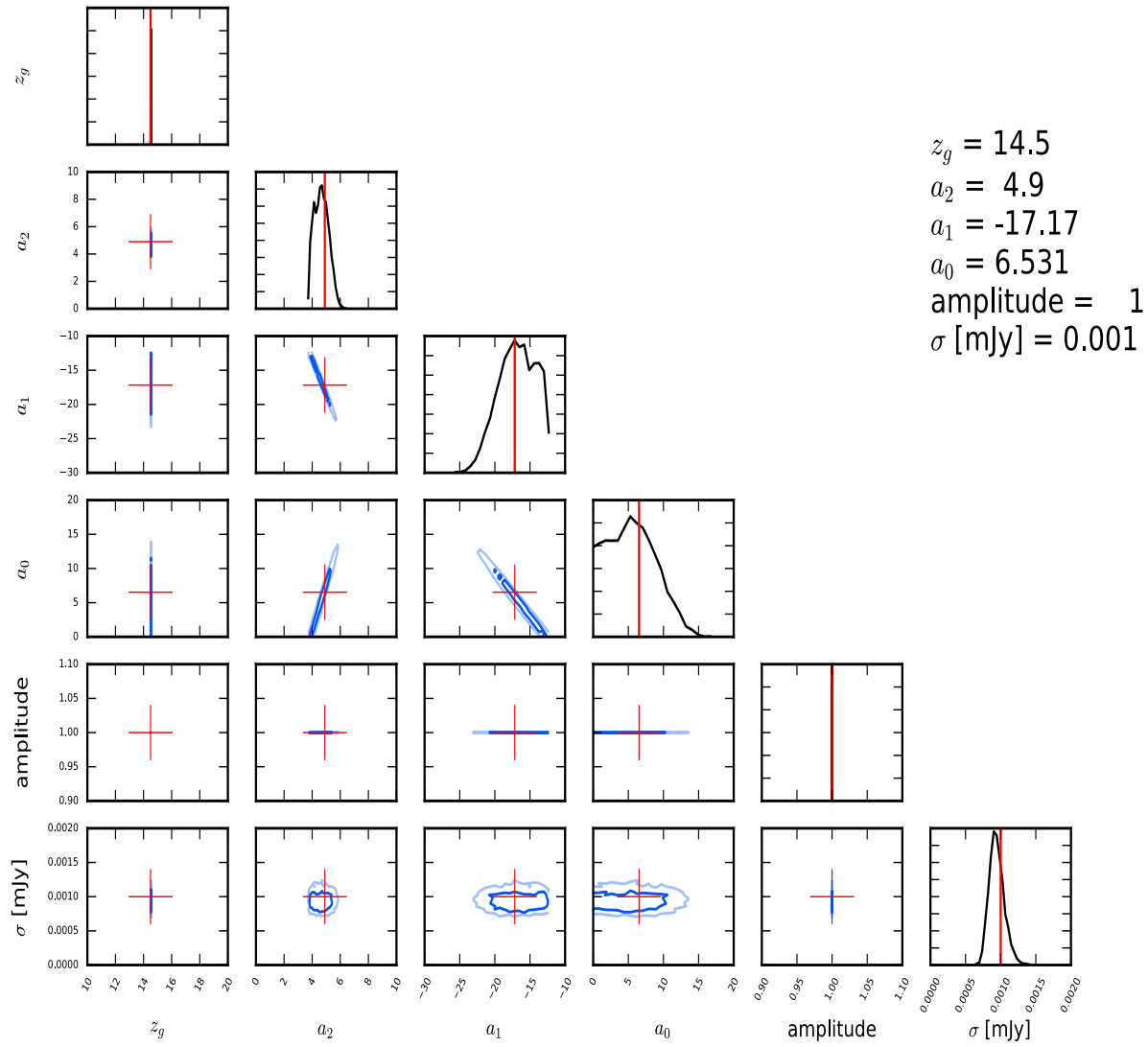
(c) σ=0.1 mJy

Fig. 5.3: Contuination of Fig.(5.3)

Fig.(5.4) shows a 2-D marginalization where the assumed data is simulated using model $\widetilde{A}_2$, this data is then re-fitted using $\widetilde{A}_2$ on different noise levels. Fig. (5.4a) shows posteriors found when using a simplistic noise level (σ =0.001 mJy), the posteriors found for this case are narrow and are aligned with the true values provided.

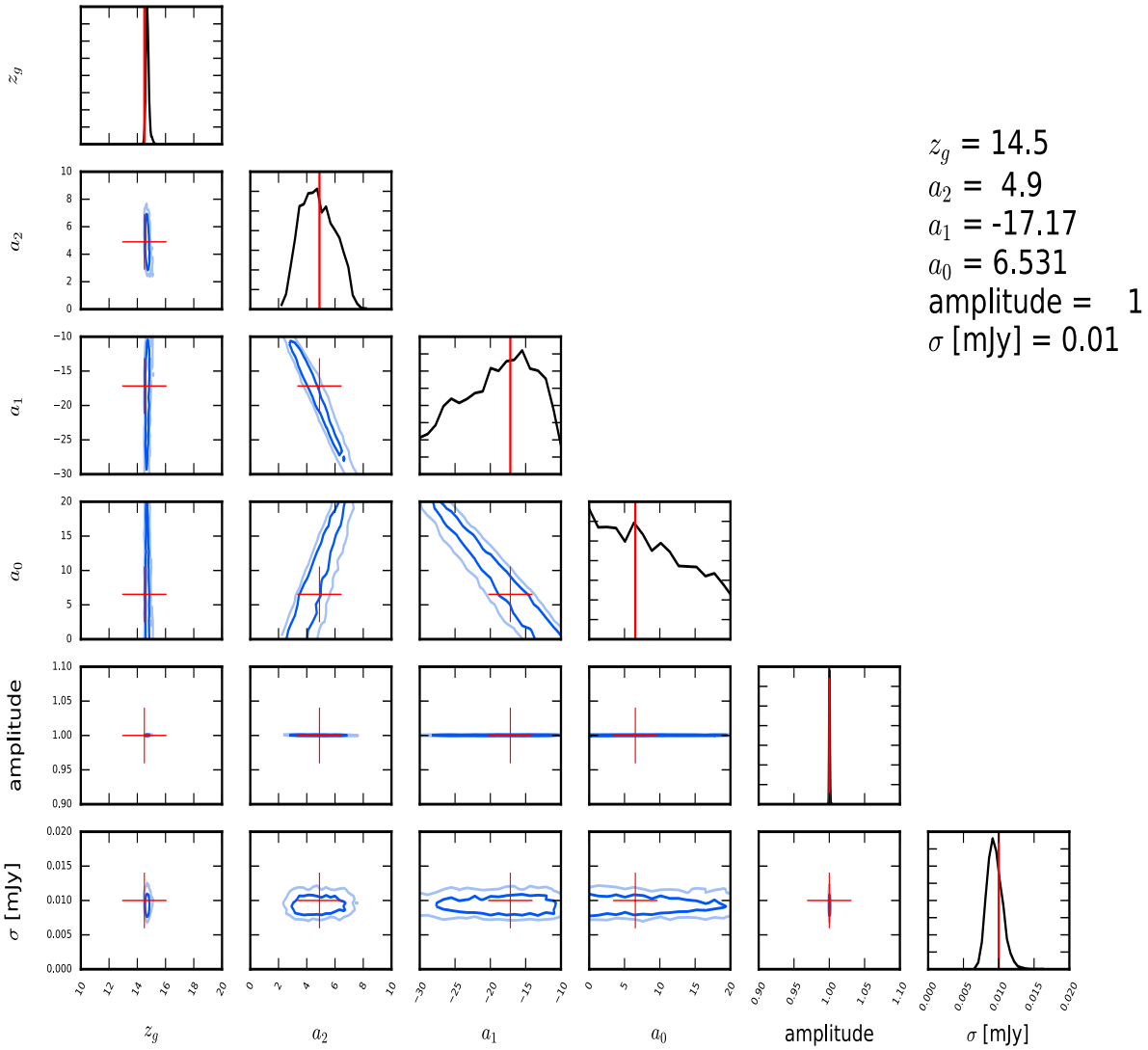
Fig.(5.4b) shows posterior distribution for parameters found when using the simplistic noise case σ =0.01 mJy. The probability distributions are slightly broader, many

of parameters found are in agreement with the true values. The effect of added noise is clearly evident, as contours are slightly broader than those seen in Fig.(5.4c). Note that there is no apparent bias observed which agrees with intuition since it is a re-fitting with the same model used to synthesise the data.



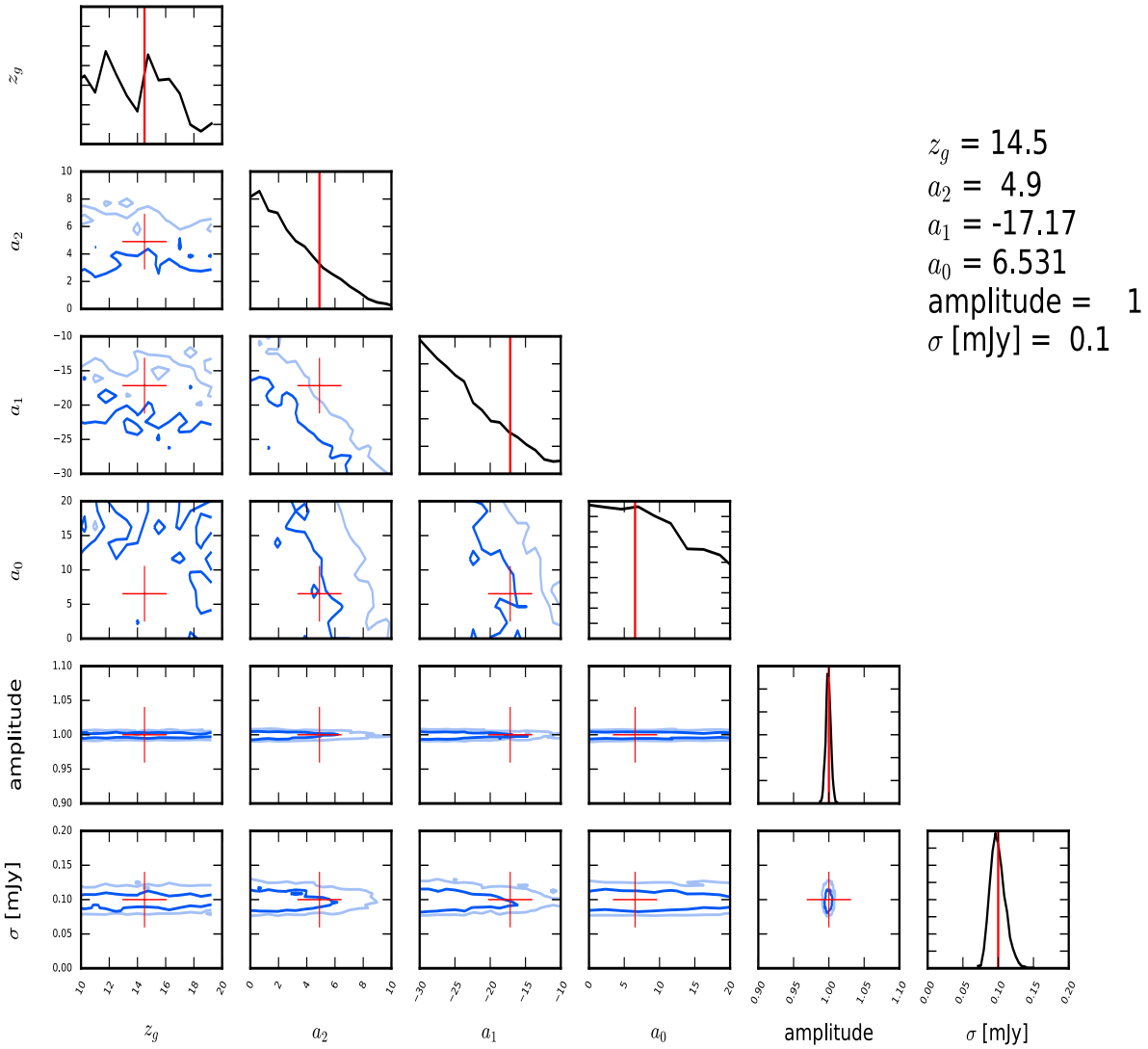
(a) σ=0.001 mJy

Fig. 5.4: The posterior distributions for the case where $\widetilde{A}_2$ is used to synthesize mock data and fitted with $\widetilde{A}_2$. The inner dark blue regions indicate the 68 percent and the light region the 95 percent confidence levels. On **(a)**:$\widetilde{A}_2$ mock data is fittted with $\widetilde{A}_2$ at noise level (σ=0.001 mJy), **(b)**:$\widetilde{A}_2$ mock data is fitted with $\widetilde{A}_2$ at (σ=0.01 mJy), **(c)**: $\widetilde{A}_2$ mock data is fittted with $\widetilde{A}_2$ at (σ=0.1 mJy). Red lines are the true values.



(b) σ=0.01 mJy

Fig. 5.4: Continuation of Fig.(5.4)



(c) σ=0.1 mJy

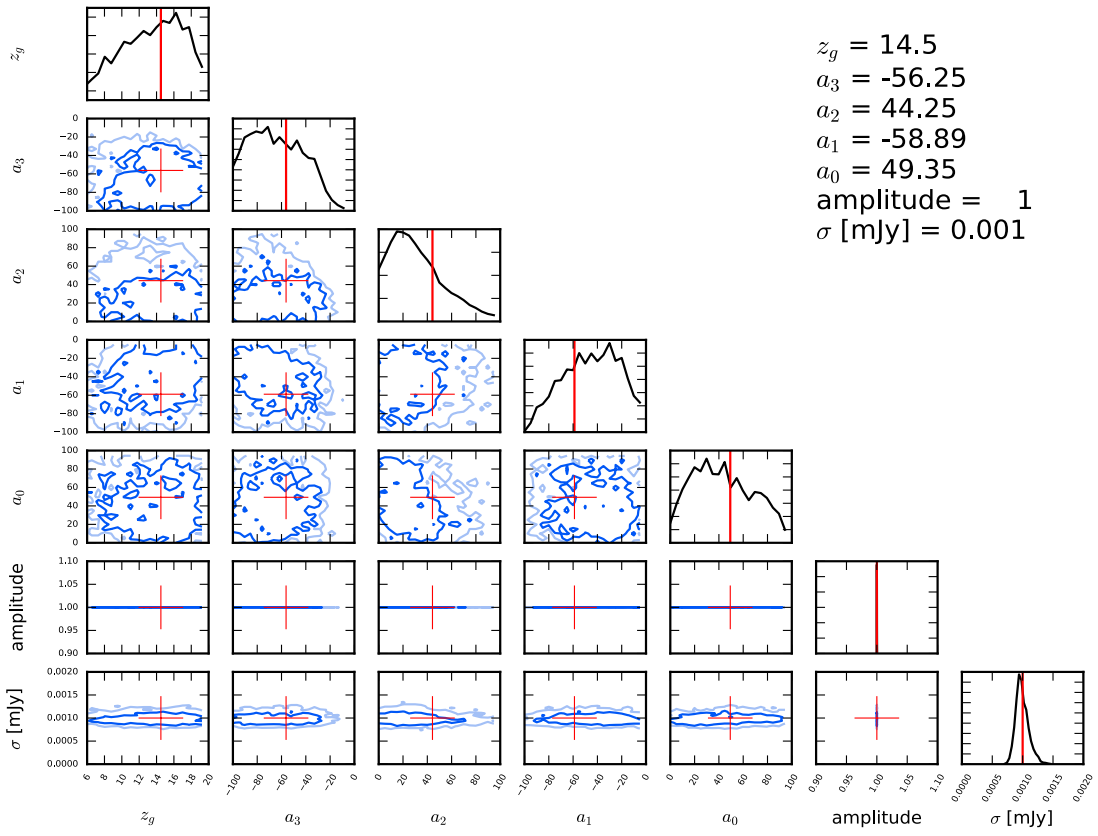
Fig. 5.4: Continuation of Fig.(5.4)

Fig.(5.4c) shows parameter estimation at noise level (σ =0.1 mJy), the model, in this case, the data is too noise to constrain the model at the level of noise added, the model fails to determine signal and noise, this is seen in the flat posteriors in most of the parameters. The quality of the fittings decreases with an increase in the noise the three triangle plots.

Fig.(5.5) shows posterior distributions for the case where synthetic data from model

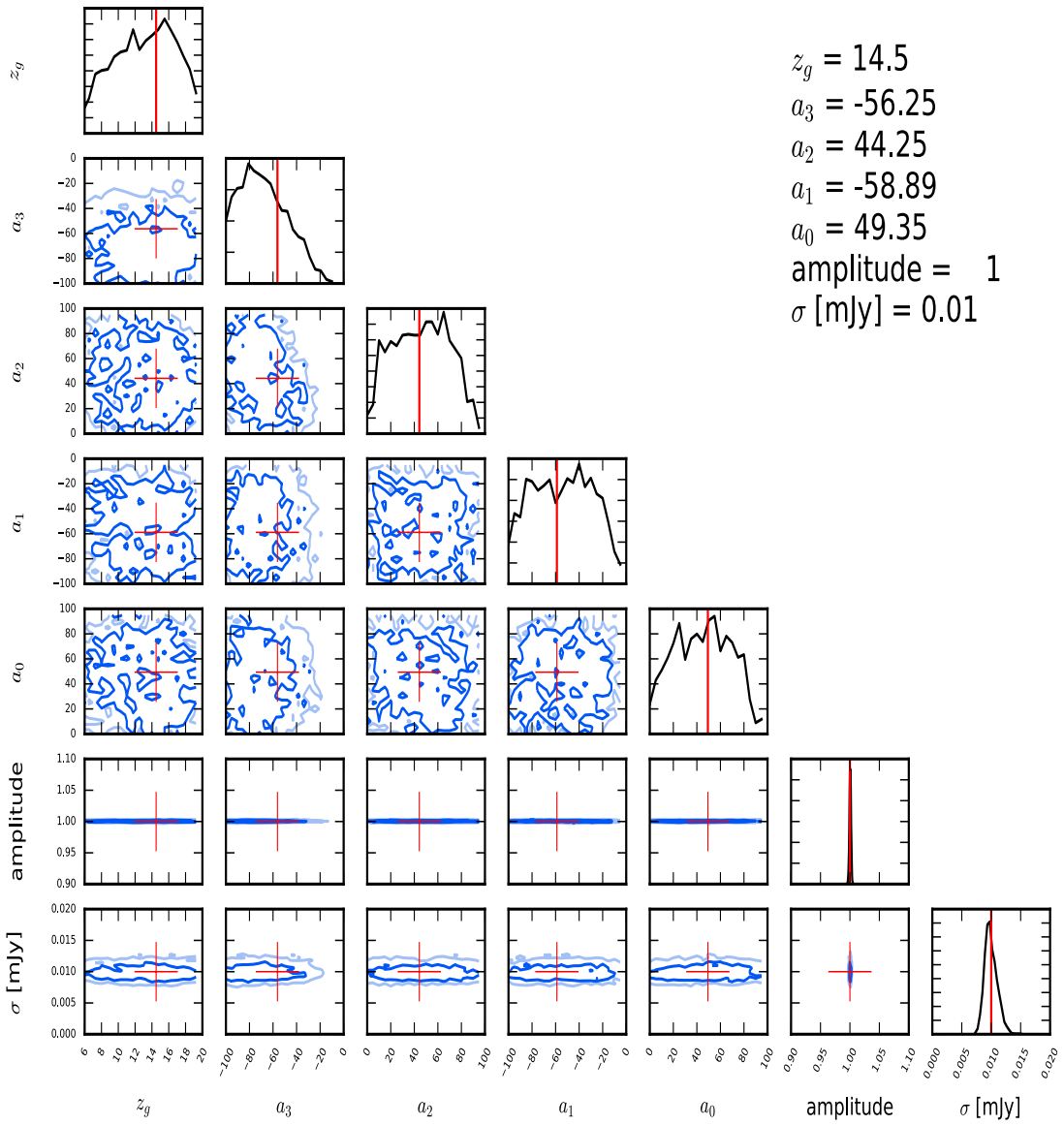
$\widetilde{A}_3$ is used and then re-fitted using $\widetilde{A}_3$. This type of parameter estimation is done for the three noise cases: σ =0.001 mJy, σ =0.01 mJy, σ =0.1 mJy. Fig.(5.5a) shows the posterior distribution found from the parameter estimation at noise level σ =0.01 mJy. Even at the noise level, many of the estimated parameters appear to be flat, and the true values are not agreement with the posterior distribution peaks.

Fig.(5.5b) shows the parameter estimation done at noise level σ=0.01 mJy. The behaviour of the posteriors found is similar to that of Fig.(5.5a) where the posteriors are slightly chaotic, the posteriors appear to be flat and are not peaking at true values. Fig.(5.5c) shows parameter estimation at σ =0.1 mJy, the posteriors are not peaking at true values.



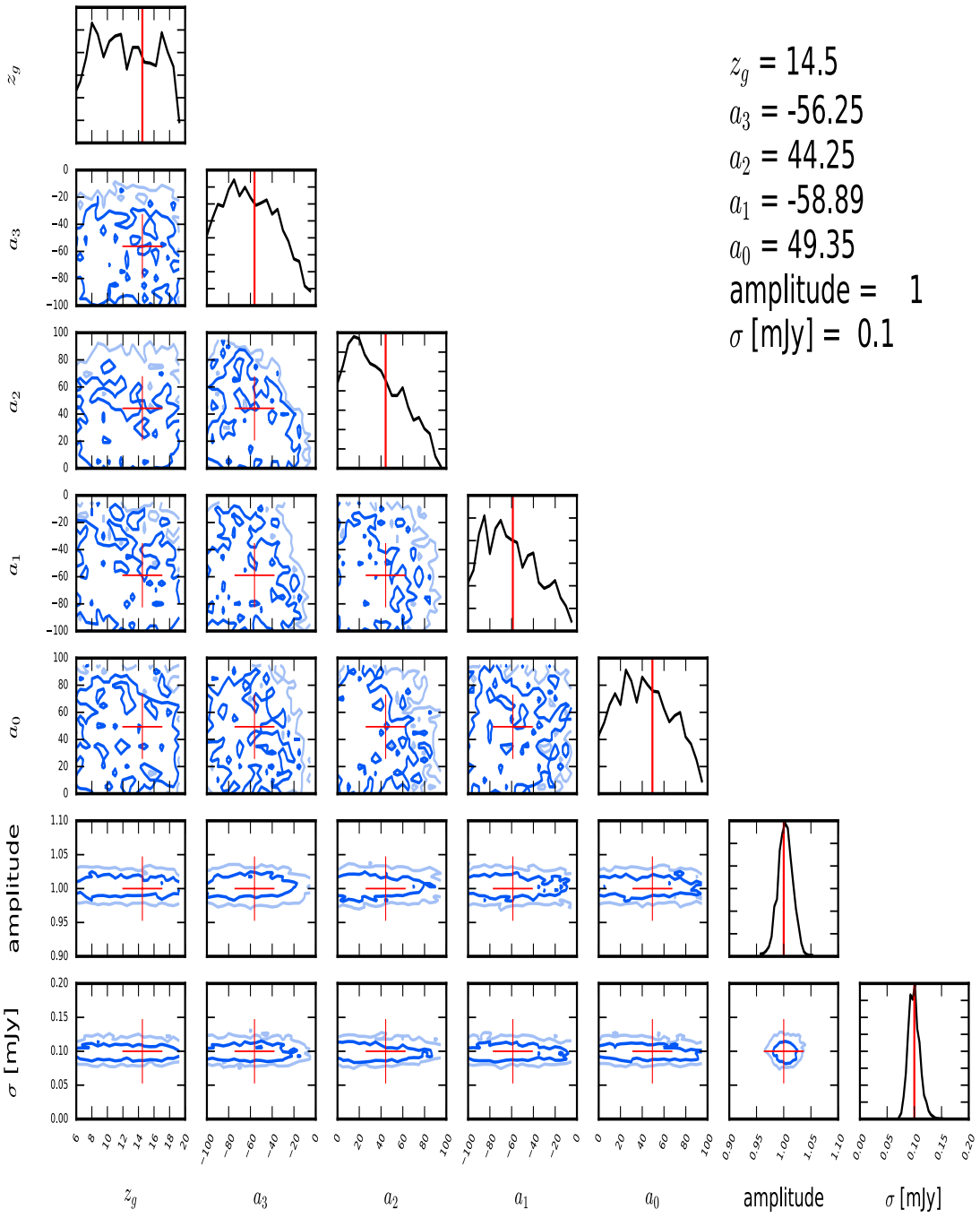
(a) σ=0.001 mJy

Fig. 5.5: The posterior distributions for the case where $\widetilde{A}_3$ is used to synthesize mock data and fitted with $\widetilde{A}_3$. The inner dark blue regions indicate the 68 percent and the light region the 95 percent confidence levels. On **(a)**:$\widetilde{A}_3$ mock data is fittted with $\widetilde{A}_3$ at noise level (σ=0.001 mJy), **(b)**:$\widetilde{A}_3$ mock data is fitted with $\widetilde{A}_3$ at (σ=0.01 mJy), **(c)**: $\widetilde{A}_3$ mock data is fittted with $\widetilde{A}_3$ at (σ=0.1 mJy). Red lines are the fiducial values.



(b) σ=0.01 mJy

Fig. 5.5: Continuation of Fig.(5.5)



(c) σ=0.1 mJy

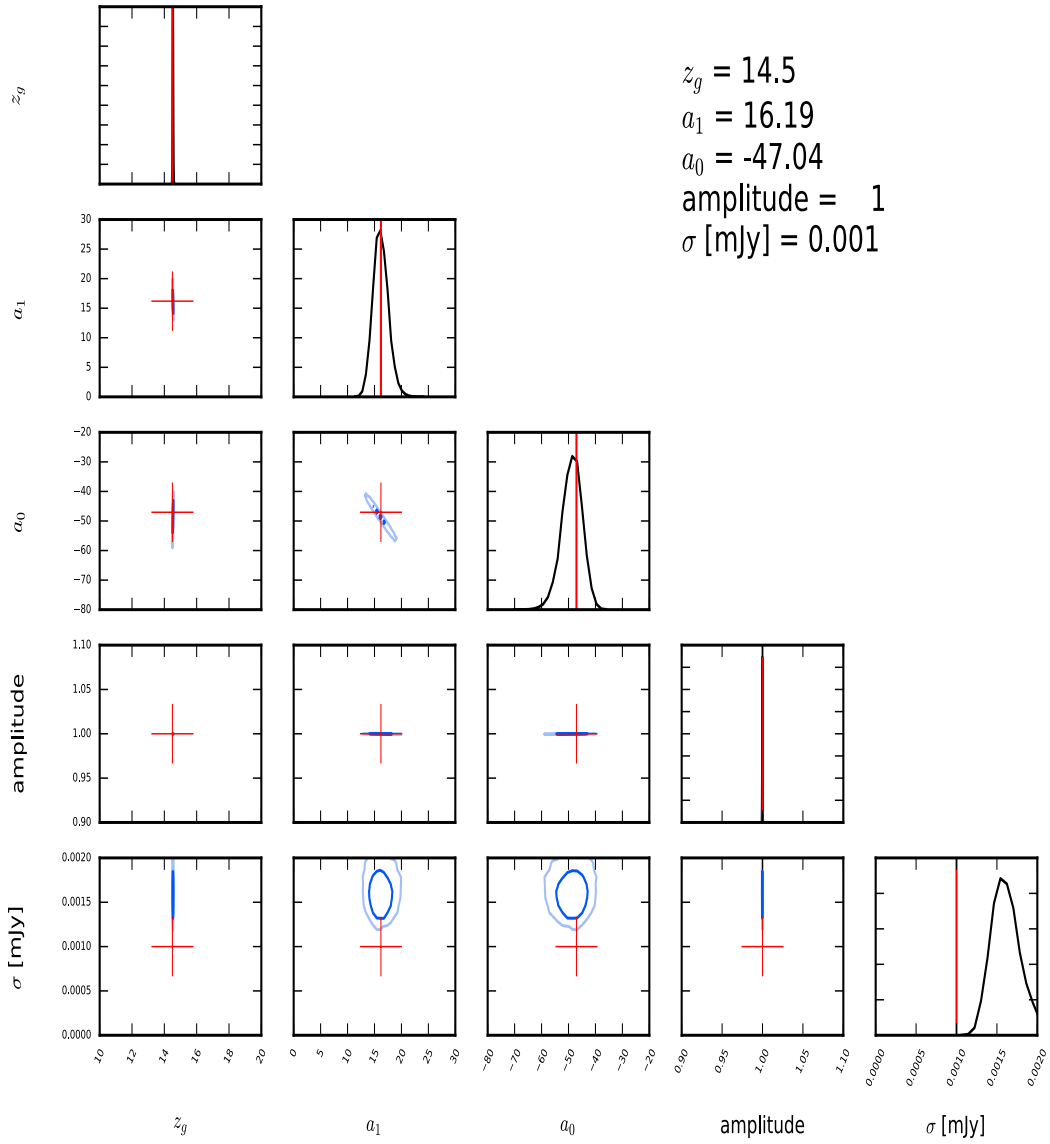
Fig. 5.5: Continuation of Fig.(5.5)

Fig.(5.6) shows the posterior distribution found when fitting 21 cm absorption feature with $\widetilde{A}_1$, this is shown in the Tables(5.1,5.2,5.3). It seems to best fit the data better than its counterpart models (that is $\widetilde{A}_2$, $\widetilde{A}_3$). The panel (5.6a) corresponds to σ =0.001 mJy. The model accurately fits at this noise level hence the true values (indicated with red lines) and posteriors peaking in the same region.

Note also the probability distribution functions for the parameters are very narrow in shape. When looking at the parameters in Fig.(5.6a) it's apparent that the data has a bias recovery in the parameter σ, which corresponds to noise. By bias, I mean that the recovered value for this parameter is shifted by several standard deviations away from the true value. This effect for this parameters is large and equivalent to several standard deviations. The bias is attributed to the structure of the 21 cm absorption data from SIMFAST21 since it is not seen when fitting the synthesised data using the corresponding model.

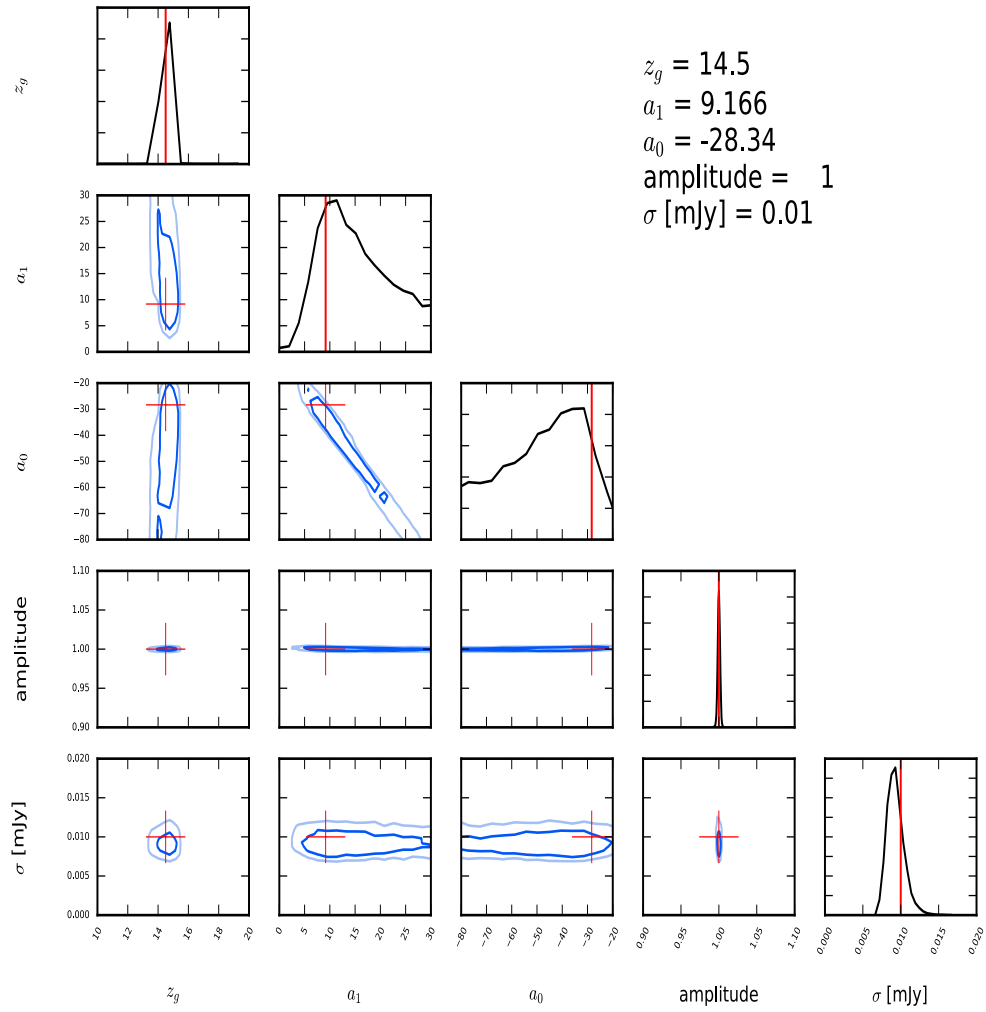
In Fig.(5.6b) the posterior distribution is broader than those of Fig.(5.6b), the bias in these parameters is still apparent but less heightened than the case of σ =0.001 mJy. The parameter estimation for σ =0.01 mJy is still accurate with posteriors peaking on regions near the true values. Fig.(5.6c) shows parameter estimation for the 21 cm absorption feature data fitted with $\widetilde{A}_1$. From Fig.(5.6c) the parameter estimation seems to fail, due to level of noise added to the input data, this case very different from the two previous cases (5.6a,5.6b), because the posteriors for some of the parameters seem to become flat and the posteriors peak for some of the parameters but do not match the true values. In terms of parameter estimating the redshift, which is the main priority parameter mode, the posteriors for the redshift from $\widetilde{A}_1$ model seem to best because they are appear a lot more narrow and centred at the true value redshift than those of $\widetilde{A}_2$ or $\widetilde{A}_3$.

the galaxy are very good, and if this is the main goal of the study then this looks good. This should be emp



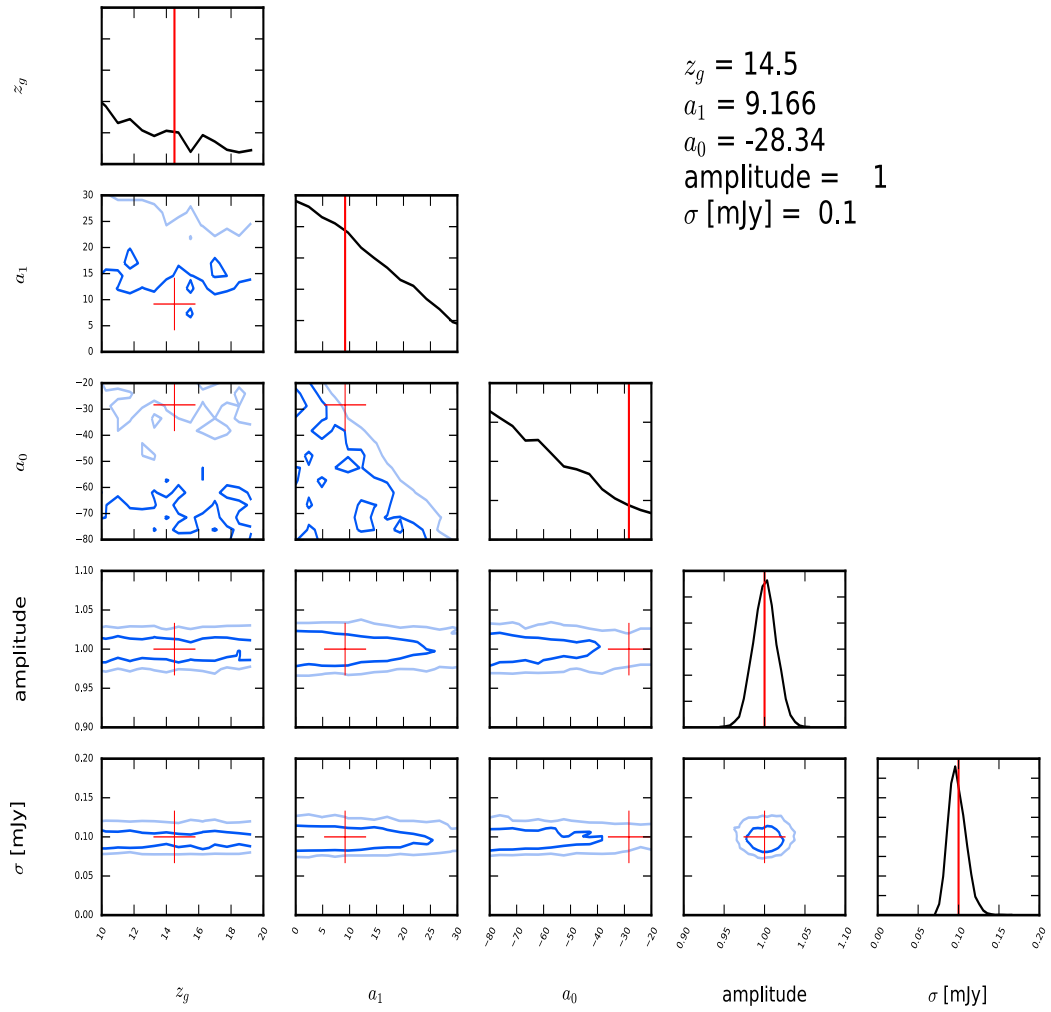
(a) σ=0.001 mJy

Fig. 5.6: The posterior distributions for the case where $\widetilde{A}$ is fitted with $\widetilde{A}_1$. The inner dark blue regions indicate the 68 percent and the light region the 95 percent confidence levels. On **a**:$\widetilde{A}$ is fitted with $\widetilde{A}_1$ at noise level (σ=0.001 mJy), **b**:$\widetilde{A}_{\rm data}$ is fitted with $\widetilde{A}_1$ at (σ=0.01 mJy), **c**: $\widetilde{A}$ is fitted with $\widetilde{A}_1$ at (σ=0.1 mJy). Red lines are the true values.



(b) σ=0.01 mJy

Fig. 5.6: continuation of Fig.(5.6)



(c) σ=0.1 mJy

Fig. 5.6: continuation of Fig.(5.6)

5.1.2 SKA1-LOW Noise Profile

Detection of individual luminous sources galaxies, during the epoch of reionization and cosmic dawn through their signature on the HI 21 cm signal is one of the direct approaches to probe the epoch. Upcoming low-frequency radio telescopes such as the HERA, MWA, LOFAR aim to detect the HI 21 cm signal. However, it has been proposed that these experiments may be able to detect these individual reionization sources. I use the currently expected design specifications for the SKA1-LOW instrument outlined in the SKA System Baseline Design (Mesinger et al., 2014; Greig, Mesinger, and Koopmans, 2015) [2]. SKA1-LOW is expected to include a total of 911 distributed stations in the order below:

- 433 stations in a 1.2 km diameter core.
- 650 stations in a 2 Km diameter core. stations in a 3 Km diameter core.
- 866 stations in a 4 Km diameter core.

The noise we expect from observations of S_{EOR}, using the Square Kilometer Array (SKA1-LOW) telescope is given by:

$$\sigma(\nu) = \frac{2k_B T_{sys}}{A_t \sqrt{\Delta\nu\, t_{int}}} \quad \text{[mJy]} \tag{5.3}$$

$\Delta\nu$ is the bandwidth, t_{int} is the integration time and k_B has already been defined. A_t is the area of the telescope given by:

$$A_t = \begin{cases} 400525.0\left(\frac{100\ \text{MHz}}{\nu}\right)^2\ \text{m}^2, & \text{if } \nu > 110\ \text{MHz} \\ 400525.0\ \text{m}^2, & \text{else} \end{cases}$$

T_{sys} is the system temperature given by $T_{sys} = T_{rcvr} + T_{sky}$, and T_{rcvr} is the receiver temperature: $T_{rcvr} = 0.1 \cdot T_{sky} + 40$ [K], where T_{sky} is given by:

$$T_{sky} = 60\left(\frac{300\ \text{MHz}}{\nu}\right)^{2.55} \quad \text{[K]}$$

In Fig.(5.7) the noise profiles used throughout the parameter estimation are shown, this includes both the simplistic noise and Gaussian random noise. For this analysis, we assumed a frequency resolution of $\Delta_n u$= 4MHz (although the resolution can be smaller, we need to keep the number of channels small in order to make the analysis

[2] https://www.astron.nl/~broekema/papers/Infieri-15/ska-sdp.pdf

fast enough) and while the integration time t_{int}=1000 hrs.

In order to fit for the noise in the realistic case. We define a dimensionless parameter r, that acts as a scaling factor throughout the curve for SKA1-LOW noise.

$$\sigma_i^{'}(\nu) = \sigma_i(\nu) \cdot r \quad \text{[mJy]}, \; r = 1, \tag{5.4}$$

where $\sigma_i(\nu)$ is SKA1-LOW noise. We use $\sigma_i^{'}(\nu)$, in the SKA1-LOW noise and parameter fit for r. Since $\sigma_i^{'}$ is dependent on frequency per bin, r is essentially a measure of the shape of $\sigma_i^{'}$ the frequency range. The prior chosen for the SKA1-LOW noise is r: uniform $\in$ [0.0,5.0].

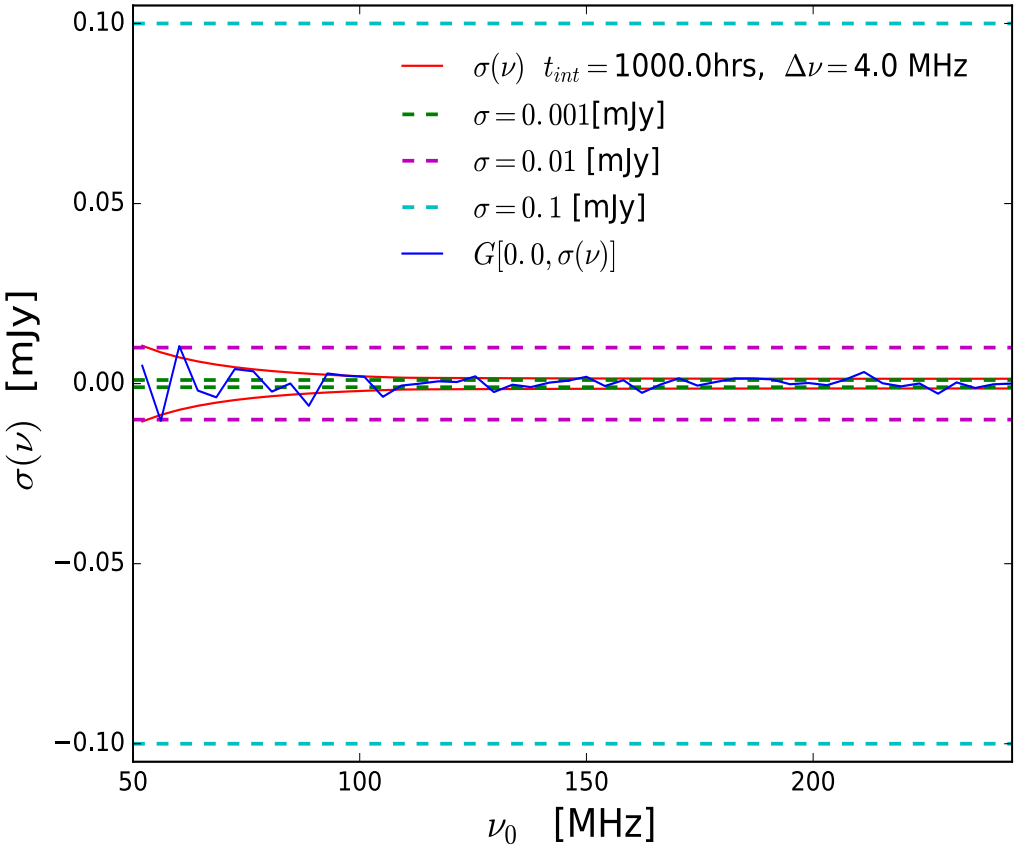


Fig. 5.7: Noise expected when observing galaxies in the Epoch of Reionization with the SKA1-LOW. The red line is eqn. (5.3) and the blue line is a random Gaussian of the blue line with a mean of zero and with eqn.(5.3) being the variance. The green line corresponds to $\sigma = 0.001$ mJy, magenta line to $\sigma = 0.01$mJy and cyan to $\sigma = 0.1$mJy.

I now use the expected noise for the SKA1-LOW telescope, giving a more realistic setup of the noise. The noise testing is very similar to the one for the simplistic noise case.

$\sigma(\nu)$ at t_{int} = 1000 hrs, $\Delta\nu$ = 4MHz					
lnZ \ data	$\widetilde{A}_0$	$\widetilde{A}_1$	$\widetilde{A}_2$	$\widetilde{A}_3$	$\widetilde{A}$
$\ln Z@A_0^*$	**0**	-59.27 $\pm$ 0.14	-74.26 $\pm$ 0.18	-2.10 $\pm$ 0.14	-58.81 $\pm$ 0.19
$\ln Z@A_1^*$	-1.15 $\pm$ 0.14	**0**	-2.51 $\pm$ 0.23	-1.73 $\pm$ 0.16	**0**
$\ln Z@A_2^*$	-2.26 $\pm$ 0.15	-5.28 $\pm$ 0.21	**0**	-2.11 $\pm$ 0.16	-16.04 $\pm$ 0.20
$\ln Z@A_3^*$	-1.39 $\pm$ 0.14	-42.40 $\pm$ 0.22	-56.50 $\pm$ 0.26	**0**	-62.68 $\pm$ 0.17

Table 5.4: The log evidence found when fitting models $\widetilde{A}_0$,$\widetilde{A}_1$, $\widetilde{A}_2$ and $\widetilde{A}_3$ to the $\widetilde{A}_0$, $\widetilde{A}_1$, $\widetilde{A}_2$, $\widetilde{A}_3$ mock data and $\widetilde{A}$ for the realistic noise $\sigma(\nu)$.

Table(5.4) shows the model evidence recovered from the various models used to parameter estimate synthesised data using models themselves and data from the expected 21 cm absorption feature from SIMFAST21. Note the diagonal zeros pattern from column two to column five. The diagonals are all found to have the highest evidence values in their respective columns.

The model fitting is shown in Table(5.4) also serves as a test, because the diagonals correspond to the synthetic data estimated with the very same model used to synthesise the same data being fitted. Intuitively one should expect the corresponding model to have the highest evidence where it is fitting its own synthetic data.

The one other notable similarity Table(5.4) has with Tables(5.1,5.2,5.3) (i.e simplistic noise case), model 1 is found to be the best when it comes to parameter estimating the 21 cm absorption feature. Finding seen Table(5.4) and Tables(5.1,5.2,5.3) seem to give an indication which of the models is best suited to parameter estimate 21 cm feature, this based on many analysis test checked. The best model will also be used as the ideal 21 cm absorption feature and run tests and compare it with the actual 21 cm absorption from SIMFAST21.

In Fig.(5.8) the disparities between the models are large. Based on Fig.(5.7) model 1 is the best model when considering the null relative evidence. The results seen when using the SKA1-LOW noise profile are in agreement with those seen in when using the simple noise case. The are similarities between the relative null Bayesian model from the simple noise and SKA1-LOW noise profile. For instance the shape in Fig(5.8) agrees with shape seen in σ=0.01 mJy and σ= 0.001 mJy. There is one slight contrast, the null relative Bayesian evidence for the SKA1-LOW noise profile, the connecting lines for relative evidence seems to more steeper than those in σ= 0.001 mJy and σ= 0.01 mJy.

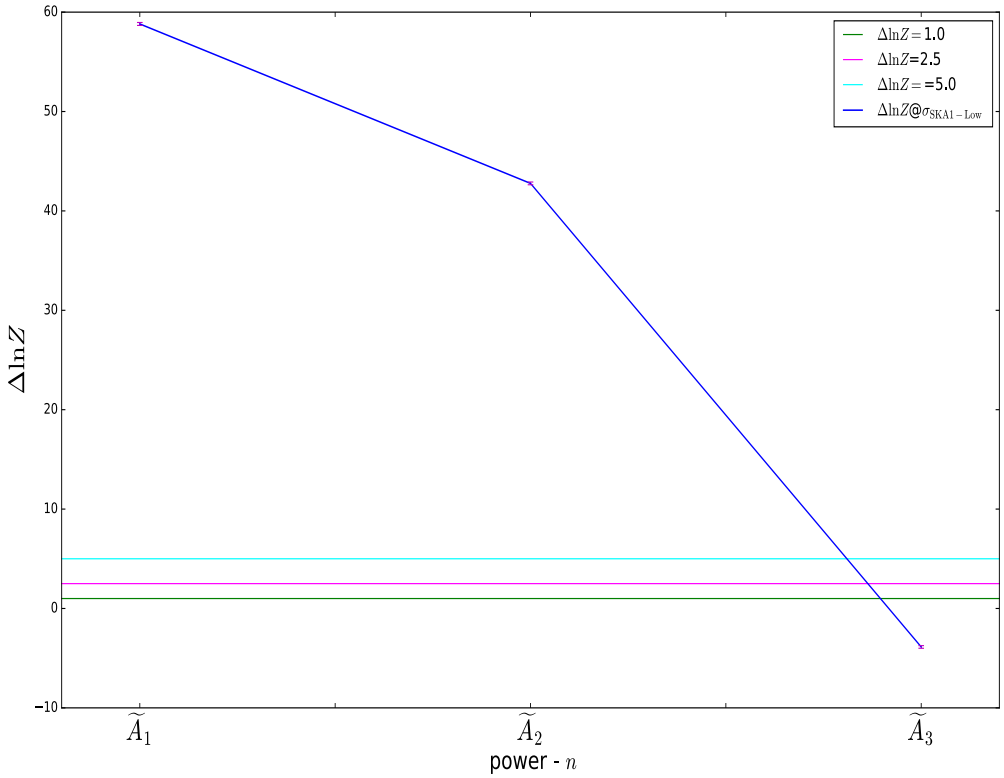


Fig. 5.8: $\Delta \ln Z_i$ found for each model ($\widetilde{A}_1$, $\widetilde{A}_2$, $\widetilde{A}_3$) fitting A_{data}, when using the realistic noise case where $\sigma(\nu, \Delta\nu = 4\text{MHz}, t_{int} = 1000\text{ hrs})$. Note in $\Delta \ln Z_i = \ln Z_i - \ln Z_0$, $\ln Z_i$ corresponds to 21 cm absorption models, while $\ln Z_0$ corresponds to the null hypothesis. The three horizontal lines labelled 1, 2.5 and 5 correspond to the intervals in the Jeffrry scale.

Fig.(5.9) has 2-D marginalization posteriors that correspond to that of winning evidence values for fitting 21 cm absorption data $\widetilde{A}$ in Table(5.4). This analysis is done using the expected SKA1-LOW noise. Fig.(5.10a) shows the simulated data generated from $\widetilde{A}_0$ and then re-fitted with $\widetilde{A}_0$. The posteriors in the triangle plot are in agreement with the provided true values. The contours are very precise and clearly defined, the model accurately parameter estimates. Fig.(5.10b) 2D marginalization posterior probability distributions for the parameters when fitting synthesized data using model $\widetilde{A}_1$ and refitted with $\widetilde{A}_1$ with the SKA1-LOW noise. Parameters found are aligned with the true values, the parameter estimation accurately recovers parameters. Fig.(5.10c) shows posteriors found when fitting synthesized data from $\widetilde{A}_2$ then parameter estimated using $\widetilde{A}_2$. Fig.(5.10d) shows posteriors found when simulated data from model $\widetilde{A}_3$ is fitted using model $\widetilde{A}_3$, the behaviour seen just like in the simplistic noise is still chaotic for the posteriors.

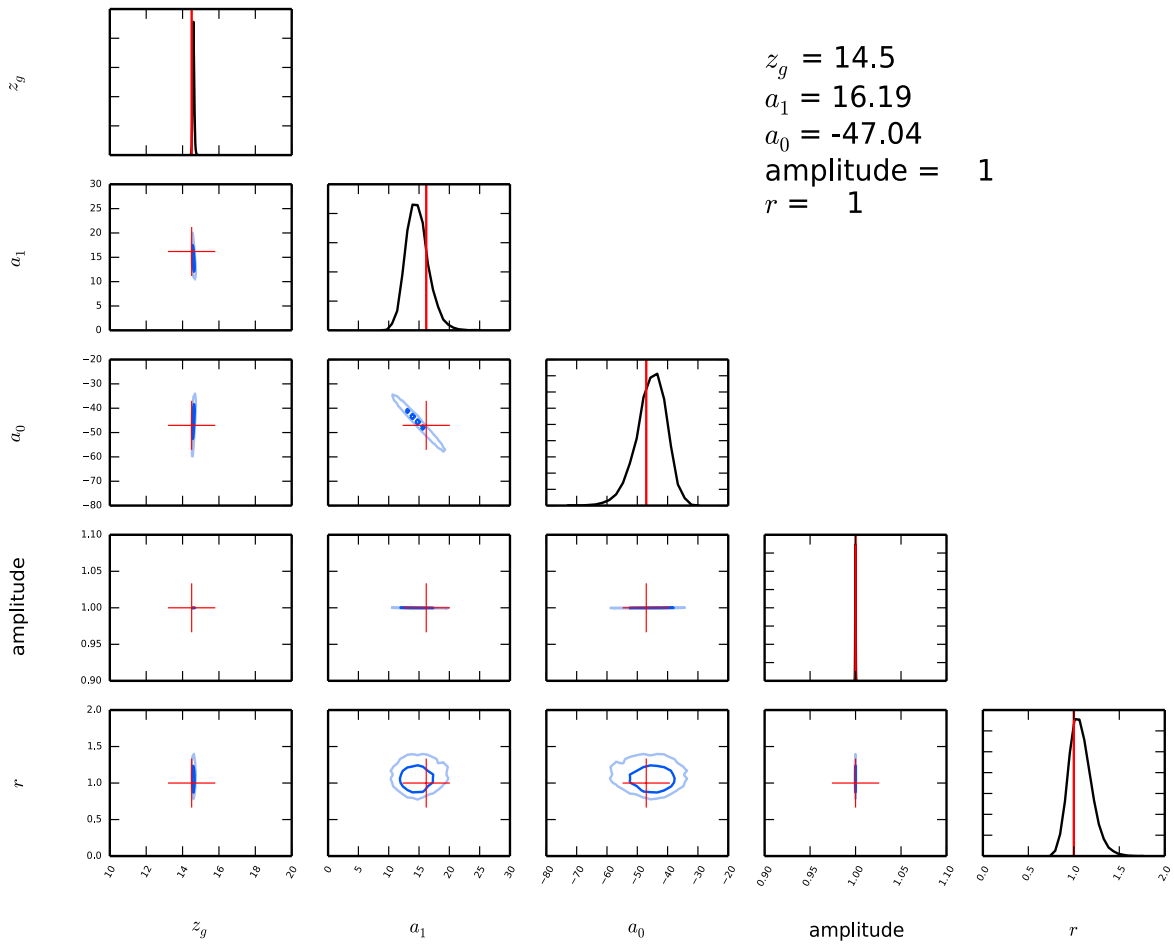
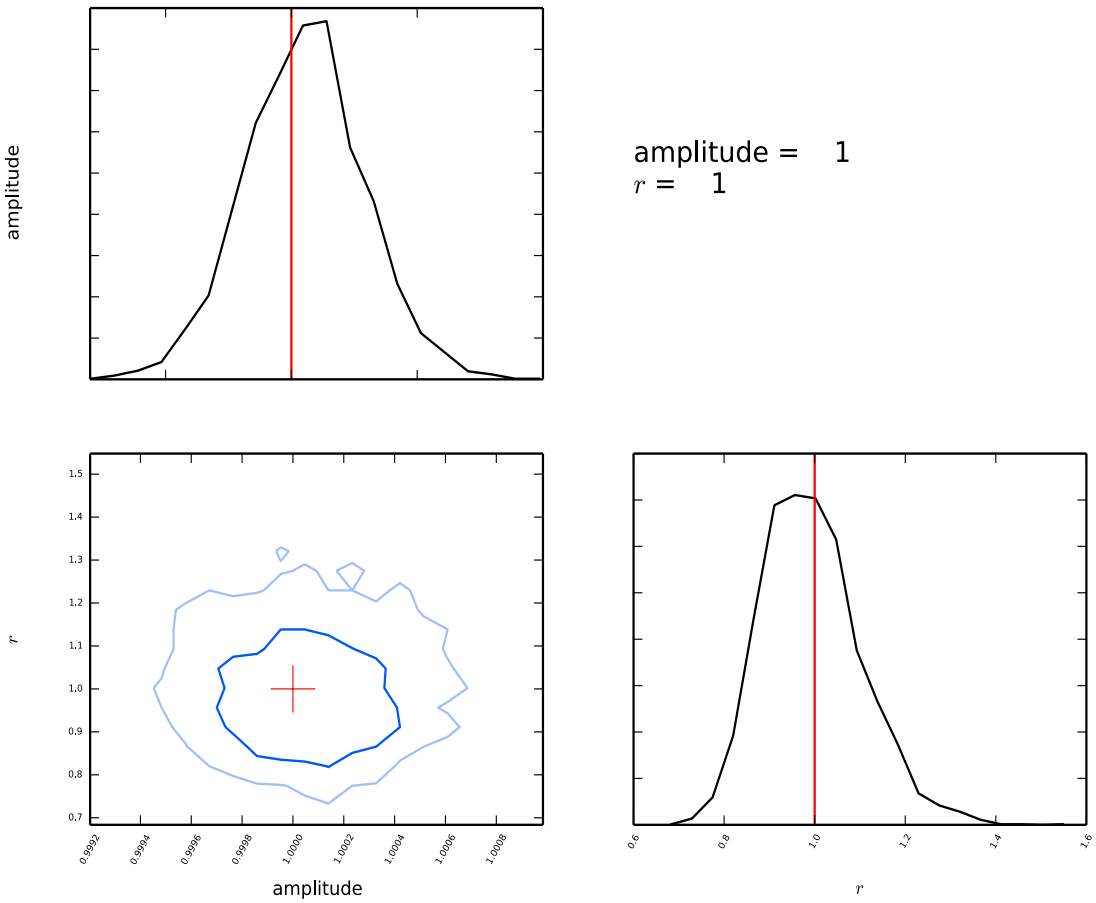


Fig. 5.9: Posterior distributions found for the case where $\widetilde{A}$ is fitted with $\widetilde{A}_1$ using the SKA1-LOW noise profile. The inner dark blue regions indicate the 68 percent and the light region the 98 percent confidence levels.

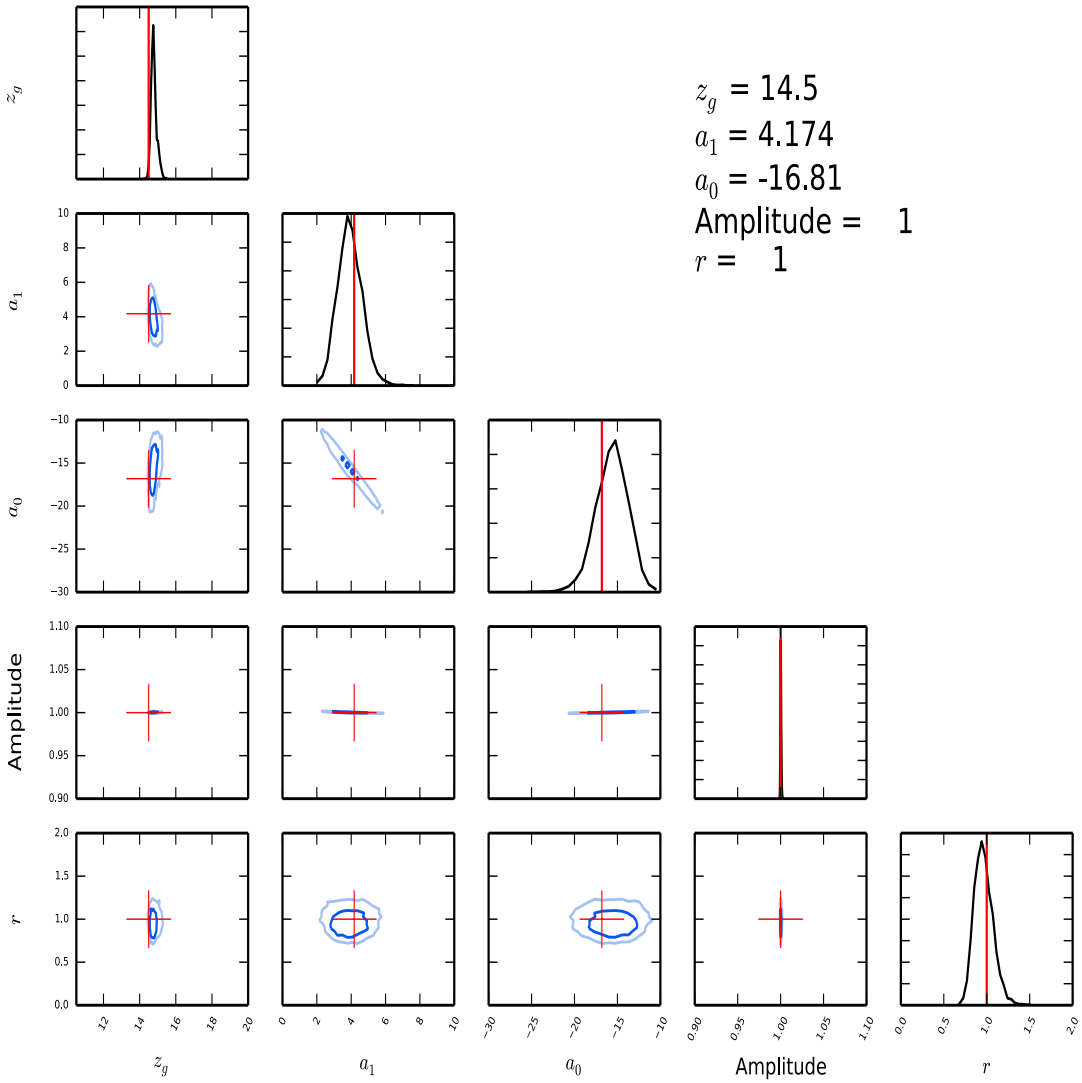
Fig.(5.9) shows posteriors found when fitting the 21 cm absorption data from SIMFAST21 using model $\widetilde{A}_1$. The posteriors appear smooth for this case, the posteriors are narrow and are in agreement with the true values. The posterior distribution for z_g is so narrow that it appears as a line. In some of the parameters, there is a deviation between posteriors and true values, but these deviations are within 68 percent confidence interval. These deviations are not as apparent as the one seen in subsection 5.1.1. The bias observed in subsection 5.1.1 seems to be within the SKA1-LOW noise regime hence it is contribution is obvious in Fig.(5.9). The effect of structure within the data is manifested as the bias in some of the parameters. The bias that is be attributed to the structure from the 21 cm absorption is due to the structure within line-of-sight in SIMFAST21. Using

these two different noise types (in subsection(5.1.1&5.1.2)) to estimate redshift position using the 21 cm absorption feature is a clear success. What has most notable; even worst possible cases where the rest of parameters are inconclusively recovered redshift posterior seems to be better parameter estimated. The next task is test parameter estimation power of 21cmNEST using a galaxy's complete spectrum, where parameters like spectral index and luminosity are included. This task should be more tedious because of the coupling between an object's luminosity to its redshift position.

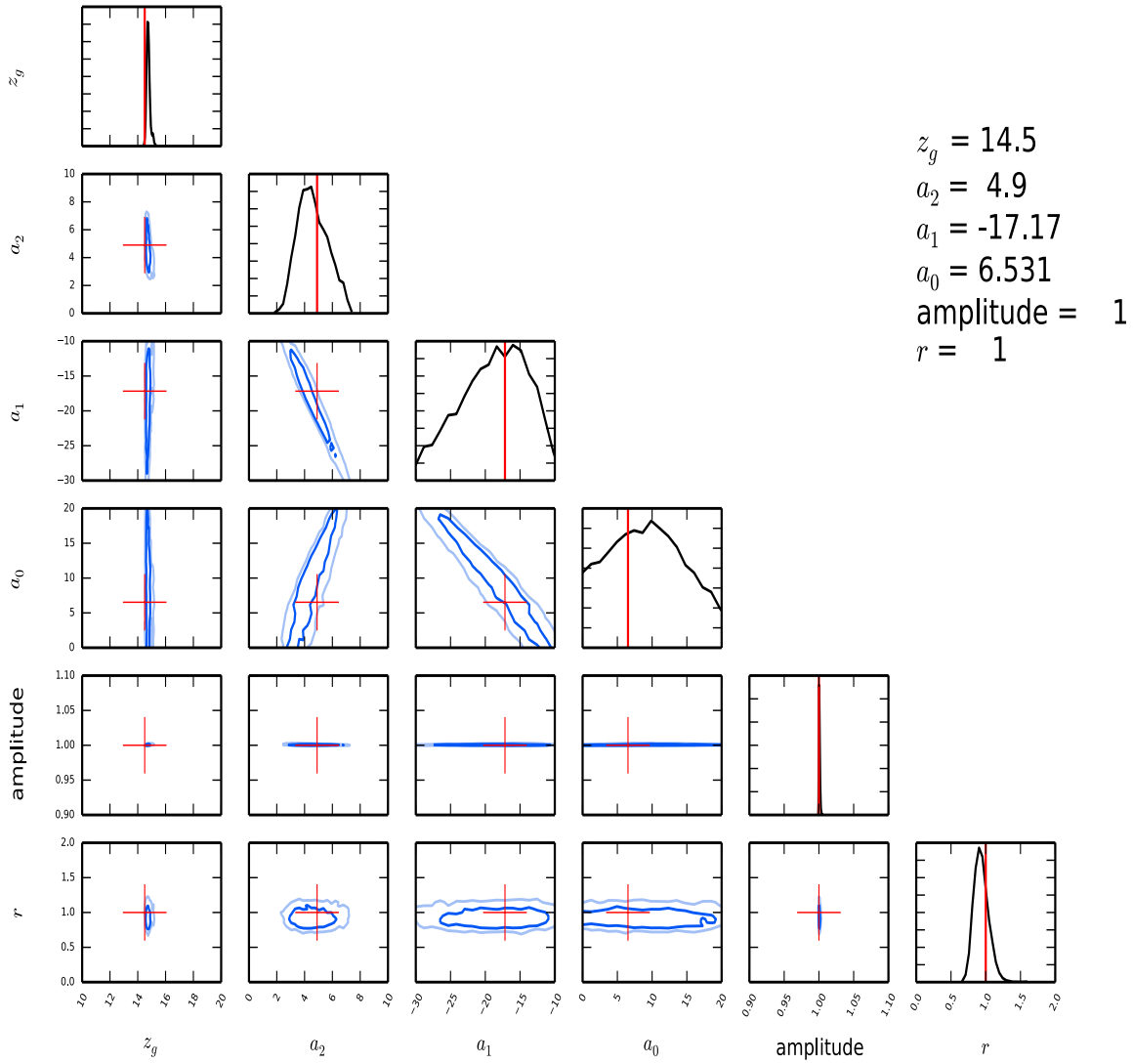


(a)

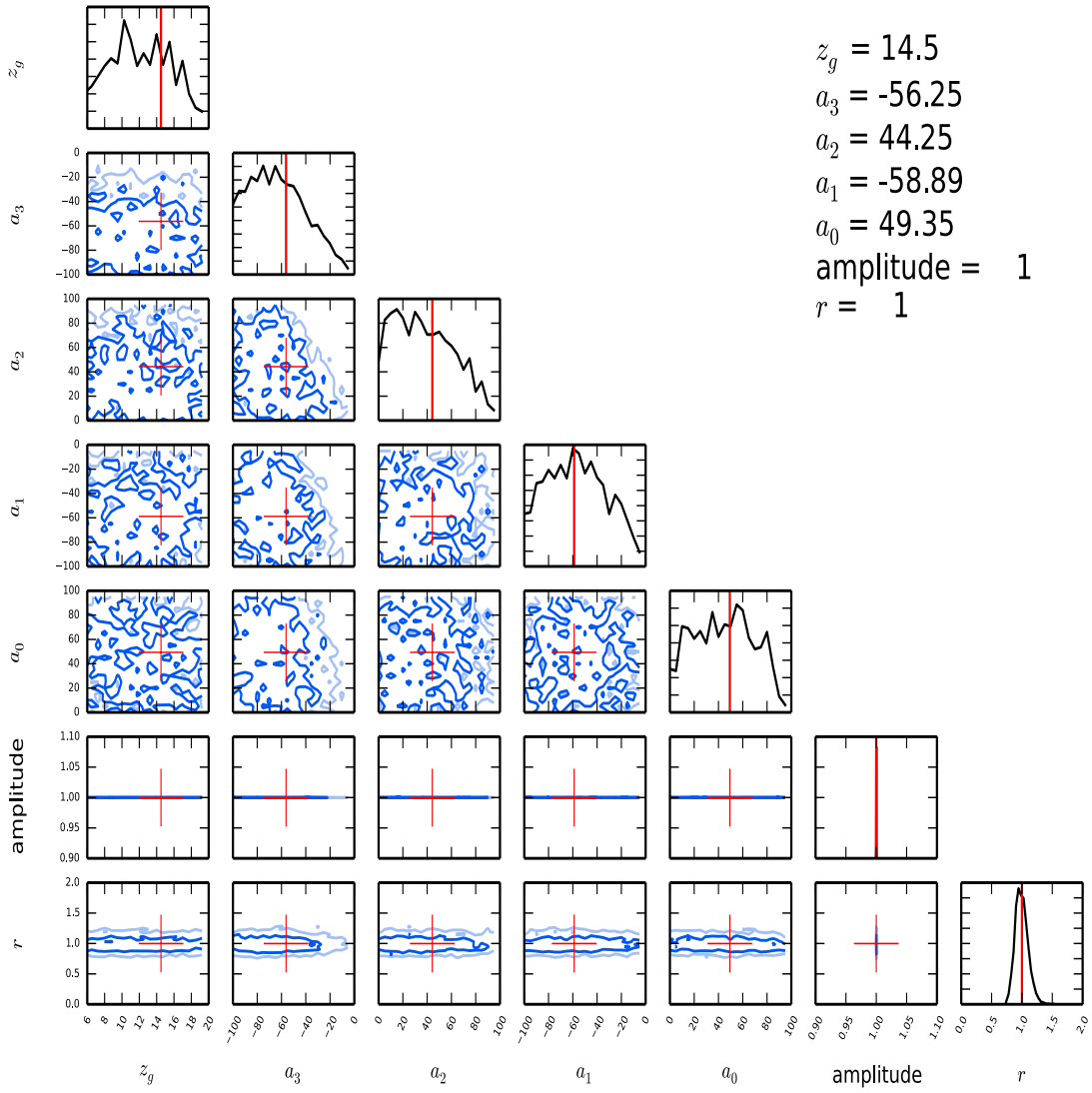
Fig. 5.10: Posterior distributions found when using the SKA1-LOW noise profile. The inner dark blue regions indicate the 68 percent and the light region the 98 percent confidence levels. **(a)**: Simulated data using $\widetilde{A}_0$ is fitted with $\widetilde{A}_0$ at noise level of SKA1-LOW, **(b)**: simulated data from $\widetilde{A}_1$ is fitted with $\widetilde{A}_1$, **(c)**: $\widetilde{A}_2$ is fitted with $\widetilde{A}_2$. **(d)**: Simulated data using $\widetilde{A}_3$ is fitted with $\widetilde{A}_3$. Red lines are the true values.



(b)



(c)



(d)

5.2 Results

The aim is to design a fitting technique that allows us to parameter estimate crucial information from the spectra of radio galaxies observed during the Epoch of Reionization. I present the results from the parameter estimation using the tools I have extensively discussed in the previous sections. I defined two types of data where the parameter estimation was applied to; there is the 21 cm absorption feature data extracted from SIMFAST21 (I will refer to this as the SIMFAST21-generated data i.e polynomial model generated data). The second data set is the simulated data from the best model (i.e found from this analysis) at parameter estimating the 21 cm absorption feature (I will refer to this data as model-generated data).

I refer to 21 cm absorption feature data as structured because the data is not smooth as its seen in Fig.(4.3), this structure in the data is due to the SIMFAST21 modelling environment in EOR which is not expected to be smooth when it is eventually observed.

Note also in Fig.(4.3) model 1 is smooth and well behaved, hence I decided to define it as the model-generated data. This means there are differences in the two data sets, this is advantageous because we can use the data set we have an analytical formula to (model-generated data) and compare the findings with those of SIMFAST21-generated data. This comparison is crucial in determining whether some of the effects we observe in findings are due to the structure from the simulation which mimics the Universe's environment during the Epoch of Reionization or the observed effects could be a consequence of limitations in the parameter estimation. The parameter fitting for the two data sets uses the following equation,

$$S_{A_1}(\nu_i; z_g, L_0, a_0, a_1) = \frac{1}{\Delta\nu_0} \int_{\nu_i}^{\nu_{i+1}} S(\nu_0; z_g, L_0) \cdot A_1(z; z_g, a_0, a_1)\, \mathrm{d}\nu \text{ [mJy]}. \tag{5.5}$$

$S(\nu_0)$ and A_1 are equations (4.15,4.11) and have been substantially discussed. z_g, L_0, a_0, a_1, r are the parameters estimated.

Another key component in the parameter estimation is that I have fitted both data sets (model-generated and SIMFAST21-generated data) with another flux model that associates the null hypothesis case where there is no 21 cm absorption observed in galaxy spectra during EOR. I have used this null hypothesis flux (power law) to infer whether a radio galaxy in EOR will be detected to have the 21 cm absorption feature in its spectra or the feature will be drowned out the by noise which ultimately means the galaxy spectra will be observed as a power law with no 21 cm absorption feature.

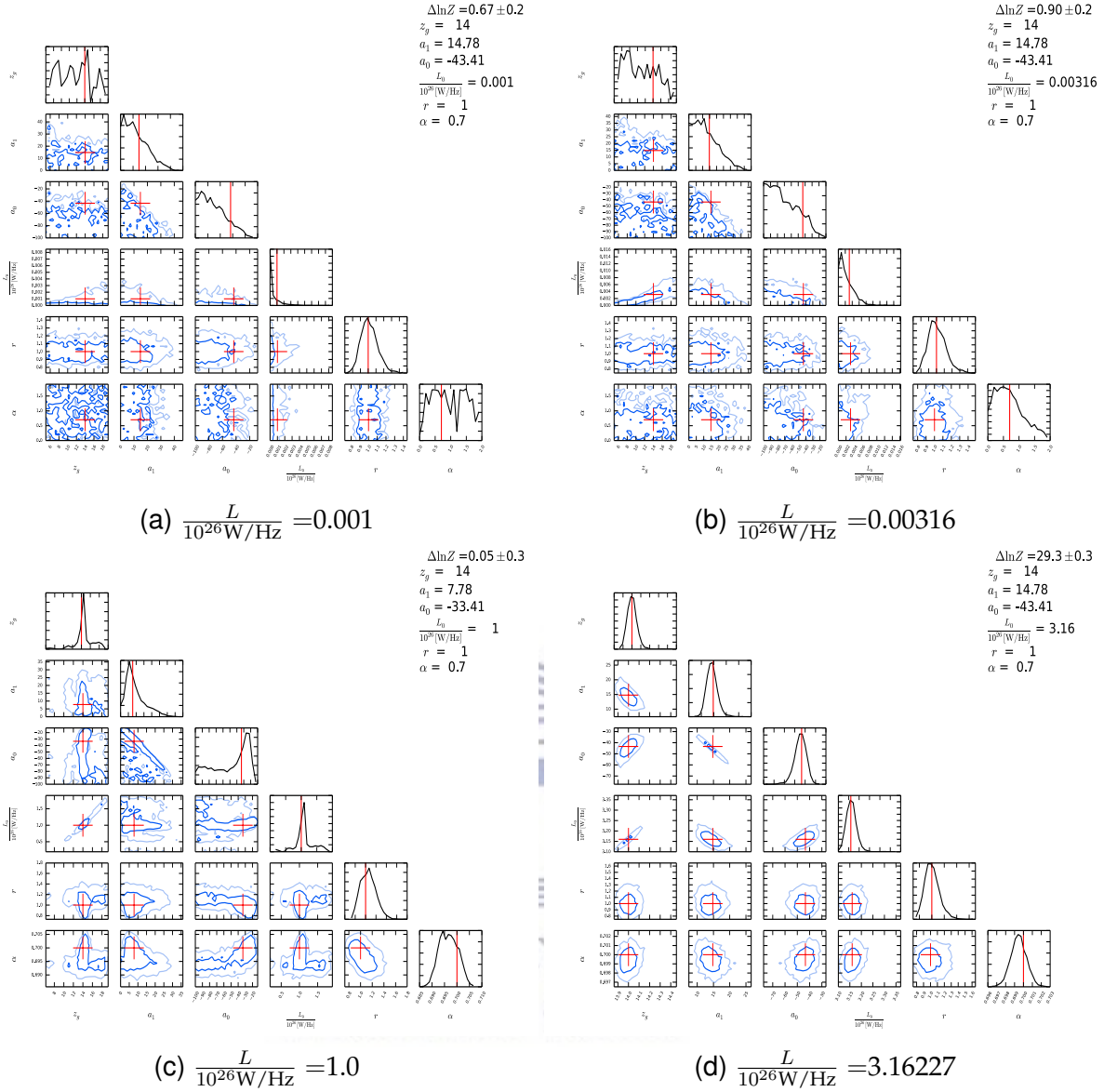
We parameter fit the null hypothesis flux using eqn.(4.15) but instead of unnecessarily punishing the model with the addition of other parameters z_g and L_0; which will not be required since the point is to find the simplest model to fit the data which serves as the null hypothesis flux, so it is rather crucial to use a power law:

$$S(\nu_0; S_0, \alpha) = S_0 \left(\frac{1.4\text{GHz}}{\nu_0} \right)^{\alpha} \text{ [mJy]} \tag{5.6}$$

5.2.1 Posterior Distributions

The 21cmNEST algorithm produces various outputs, these include posterior distribution, Bayesian model evidence and statistical estimators such as the mean for given parameter in the model. The scheme used to test 21cmNEST with the SIMFAST21-generated and model-generated data is as follows: four spectral energy distributions are generated at the same redshift z=14.0, spectral index, α =0.7, but they have different luminosities, L_0. All these objects are assumed to be observed with a telescope that has SKA1-LOW specifications. Fig.(5.11) shows posteriors when fitting spectra generated from the model-generated data, while the SIMFAST21-generated data parameter-estimations are shown in Fig.(5.12). Both Fig.(5.12) and Fig.(5.11) are ordered by increasing luminosity. The priors used for the parameter estimation can be found in Table(4.2). In all the triangle plots true values for the parameter estimations are provided in the top left corner. Note all the triangle plots have parameter $\Delta \ln Z$, this parameter which acts as a detection proxy since $\Delta \ln Z \approx SNR$. $\Delta \ln Z$ is found by taking the difference in Bayesian model evidence from eqn.(5.5) and eqn(5.6) when these models parameter fit either SIMFAST21-generated or model-generated data. To interpret how the proxy detection confidence scale works I use the Jeffreys scale for model selection. In cases where $0 < \Delta \ln Z <$1.0 these are regarded as non-detections because the margin is too small to be considered a detection. In objects where $\Delta \ln Z$ is found to be $<$0, this indicates the 21 cm absorption feature in the spectra of the object is too small/obscured by noise hence the null hypothesis is favoured. Cases where $\Delta \ln Z$ $<$0 are regarded as non-detections. See Table(2.1) for an in depth interpretation of the scaling applied. Fig. (5.11) Fig.(5.12) show the posterior distributions found when parameter fitting both model-generated and SIMFAST21-generated data. In Fig.(5.11a) and Fig.(5.12a) the posterior distributions for both SIMFAST21-generated and model-generated data at low luminosities $\frac{L}{10^{26}\text{W/Hz}}$ =0.001 and 0.003162, appear chaotic for many parameters and it is only the noise r and spectral index α that are being recovered at this level, this may very well be an indication that the signal to noise ratio is very low. Even when looking at these objects' corresponding proxy detection confidence scaling, we note they are non-detections. This is probably due to the amount of noise for these faint

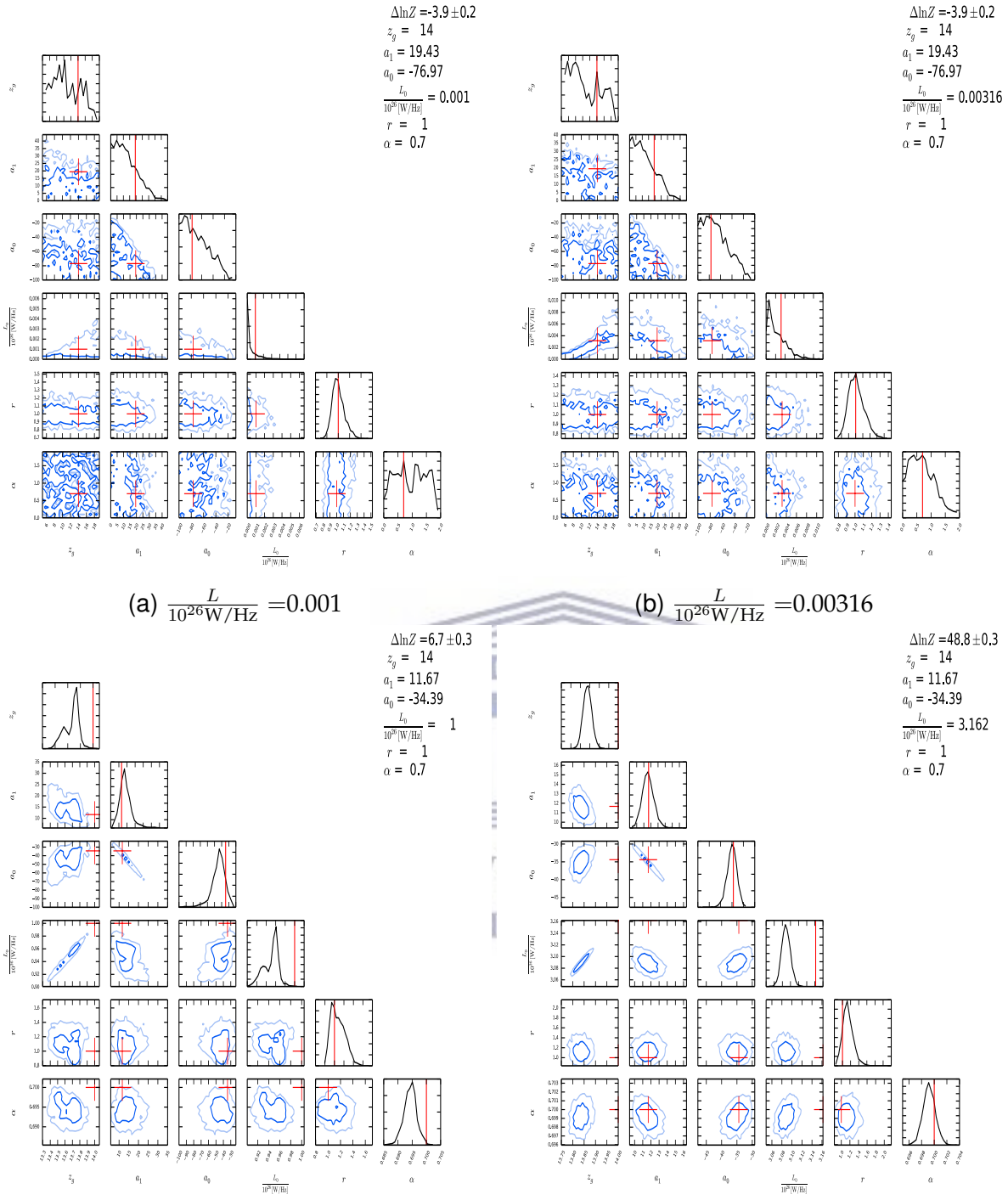
sources that makes it difficult for our models to parameter fit the 21 cm absorption feature. When looking at objects with $\frac{L}{10^{26}\mathrm{W/Hz}}$ =1.0 in Fig.(5.11c) and Fig.(5.12c), the parameter fitting has slightly improved as all parameters in both SIMFAST21-generated and model-generated are peaked around the provided true values. One notable difference is the bias observed in the triangle plots for the SIMFAST21-generated data, for a few of the parameters the bias appears in the same direction. In the last set of triangle plots at luminosity $\frac{L}{10^{26}\mathrm{W/Hz}}$ =3.1622 in Fig.(5.11d) and Fig.(5.12d), the posteriors are well defined and smooth. The posteriors for the model-generated case are in agreement with the provided true values, this is not the case in the SIMFAST21-generated data parameter estimation. The bias seems more apparent now at this luminosity for SIMFAST21-generated data. The bias is most obvious in parameters z_g and $\frac{L}{10^{26}\mathrm{W/Hz}}$, this shift is off by a few standard deviations from the peak of the posterior distributions. The redshift posteriors for the bright galaxies are more clear, this demonstrates that bright objects are much easier to parameter redshift position. This, however, is not case when it comes to faint objects, because of the signal to noise which drowns the signal has the model is not be able to recover much information about the source of interest. A key feature to note in the the triangle plots of objects detected, the redshift parameter posteriors are well defined and look very good when comparing their positions to the corresponding true values.



(a) $\frac{L}{10^{26}\mathrm{W/Hz}}$ =0.001 (b) $\frac{L}{10^{26}\mathrm{W/Hz}}$ =0.00316

(c) $\frac{L}{10^{26}\mathrm{W/Hz}}$ =1.0 (d) $\frac{L}{10^{26}\mathrm{W/Hz}}$ =3.16227

Fig. 5.11: 2D marginalized posterior probability distributions for the parameters for the model-generated data case. The red lines correspond to the true values. The objects have the same redshift, z_g, spectral index, α, but have different luminosties, the luminosities increase with descending order, $\frac{L}{10^{26}\mathrm{W/Hz}}$: (a) 0.001, (b) 0.00316, (c) 1.0 and (d) 3.16227.



(a) $\frac{L}{10^{26}\mathrm{W/Hz}}$ =0.001

(b) $\frac{L}{10^{26}\mathrm{W/Hz}}$ =0.00316

(c) $\frac{L}{10^{26}\mathrm{W/Hz}}$ =1.0

(d) $\frac{L}{10^{26}\mathrm{W/Hz}}$ =3.16227

Fig. 5.12: 2D marginalized posterior probability distributions for the parameters for the SIMFAST21-generated data case. The red lines correspond to the true values. The objects have the same redshift, z_g , spectral index, α, but have different luminosties, the luminosities increase with descending order, $\frac{L}{10^{26}\mathrm{W/Hz}}$: (a) 0.001, (b) 0.00316, (c) 1.0 and (d) 3.16227.

5.2.2 Reconstructions and redshift histograms

The Figs. (5.13,5.14) show the reconstructions of the flux for the parameter estimation discussed in 5.2.1. The reconstruction procedure uses one of the statistical estimators found for parameters fitted. The MAP (maximum a posteriori probability) is the chosen estimator for the reconstructions since it was found to be more accurate than other statistical estimators which include Maximum Likelihood (ML) estimation and Mean estimation. An advantage of using the MAP estimation procedure incorporates our prior beliefs about parameter values into the new estimate. The MAP estimation is the value of the parameter that maximizes the entire posterior distribution (which is calculated using the likelihood). A MAP estimate is the mode of the posterior distribution. Note that the reconstructions only show the residuals, which are calculated by finding the difference between constructed spectra using the MAP parameters and simple power law spectra constructed using the input parameters, therefore:

$$\text{Residual} = \text{SED(MAP parameters)-Power law(Input parameters)} \tag{5.7}$$

For clarity, the Input parameters used to correspond to those used to construct input data, for both model-generated and SIMFAST21-generated data. The reasons for finding the residual and instead of using the recovered spectra become obvious when one considers, very little to nothing will be visible in the region of interest i.e 21 cm absorption feature, because the feature is relatively small when compared to the rest of spectra. Another reason is that with a residual plot it is easier to access the accuracy of the parameter estimation compared to the data. These are key things in terms of determining how successful the parameter estimation is.

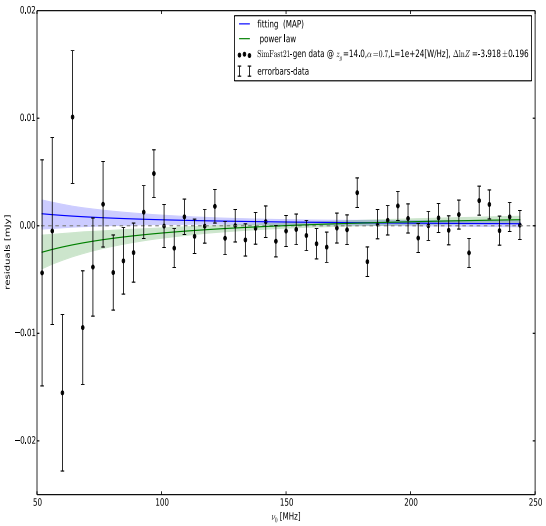
In the attempt to better understand how the parameter estimation behaves, both SIMFAST21-generated data and model-generated data are fitted with a simple power law; which naively assumes spectra for objects in EOR have no 21 cm absorption feature. This is used to test if whether an objects spectra in EOR is better parameter estimated with a SED that accounts for 21 cm absorption feature or if a simple power law would be better suited. This technique is used to infer whether the objects being parameter estimated would be observed as plain power law or as spectra that have 21 cm absorption feature. Using the model evidence recovered from the power law and model that accounts for 21 cm absorption, to devise a detection proxy for each object parameter estimated.

In the lowest luminosity case in Fig.(5.13, 5.14) $\frac{L}{10^{26}\mathrm{W/Hz}}$=0.01, the MAP fitting picks up the absorption in the model-generated case which seems remarkable considering the

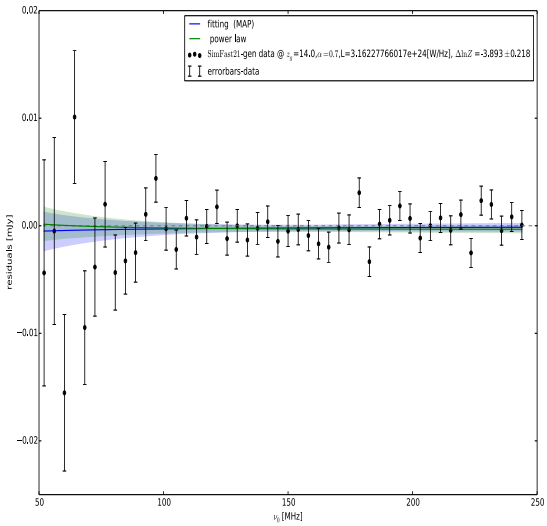
signal to noise scale. Looking at the corresponding scale for the SIMFAST21-generated data, the MAP does not seem to detect an absorption feature. This disparity is likely due to the nature of the data set (SIMFAST21-generated data & model-generated data) used. When looking at detection proxy ($\Delta \ln Z$) for this model-generated data object at $\frac{L}{10^{26}\mathrm{W/Hz}}$=0.01 in Fig.(5.14a), it reads as $\Delta \ln Z$ =-0.66±0.179, which means it is regarded as a non-detection (since it favours the simple power law fitting) probably because the feature is not strong enough at such signal to noise scales.

In the corresponding object in Fig.(5.13a), from the SIMFAST21-generated data the detection proxy is -3.918±0.196, it is also a non-detection. Moving on to objects at $\frac{L}{10^{26}\mathrm{W/Hz}}$=0.0316 in Fig.(5.13a), on the SIMFAST21-generated data case, the MAP fitting still does not pick up the absorption feature and $\Delta \ln Z$<0, it is therefore a non-detection, the model-generated case pick up the 21 cm absorption feature but $\Delta \ln Z$<0, it falls in the category of non-detection.

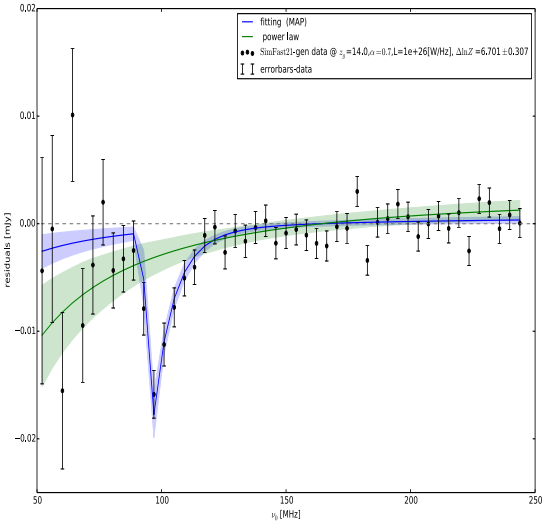
On $\frac{L}{10^{26}\mathrm{W/Hz}}$=1.0 level both SIMFAST21-generated (see Fig.(5.13c)) and model-generated (see Fig.(5.14c)) data the fittings pick up the 21 cm absorption feature. This suggests the signal to noise is reasonable and the objects are both detected with $\Delta \ln Z$>0. The objects at $\frac{L}{10^{26}\mathrm{W/Hz}}$=3.16 for both SIMFAST21-generated data in Fig.(5.13d) and model-generated data in Fig.(5.14d) appear to be accurately modelled, since MAP fitting matches the data in many occasions on both sets of data. Note for both SIMFAST21-generated data and model-generated data $\Delta \ln Z$ >>0, hence these objects are easily categorized as detections. Note that even in the cases where there is a bias in the redshift, the probability of detecting of these galaxies is at high, these objects could be the first to be confirmed detections with upcoming radio telescope.



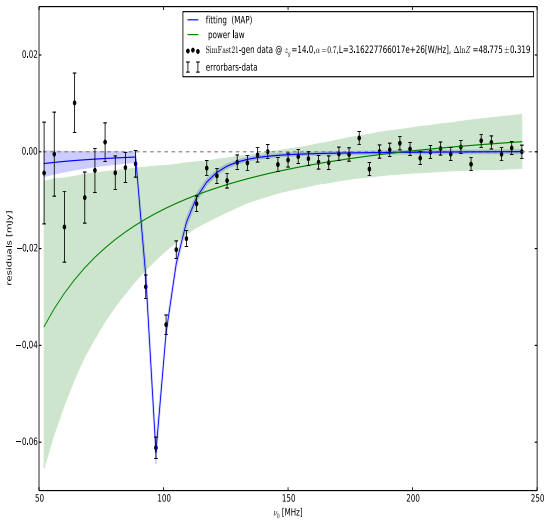
(a) $\frac{L}{10^{26}\mathrm{W/Hz}}$=0.01



(b) $\frac{L}{10^{26}\mathrm{W/Hz}}$=0.0316

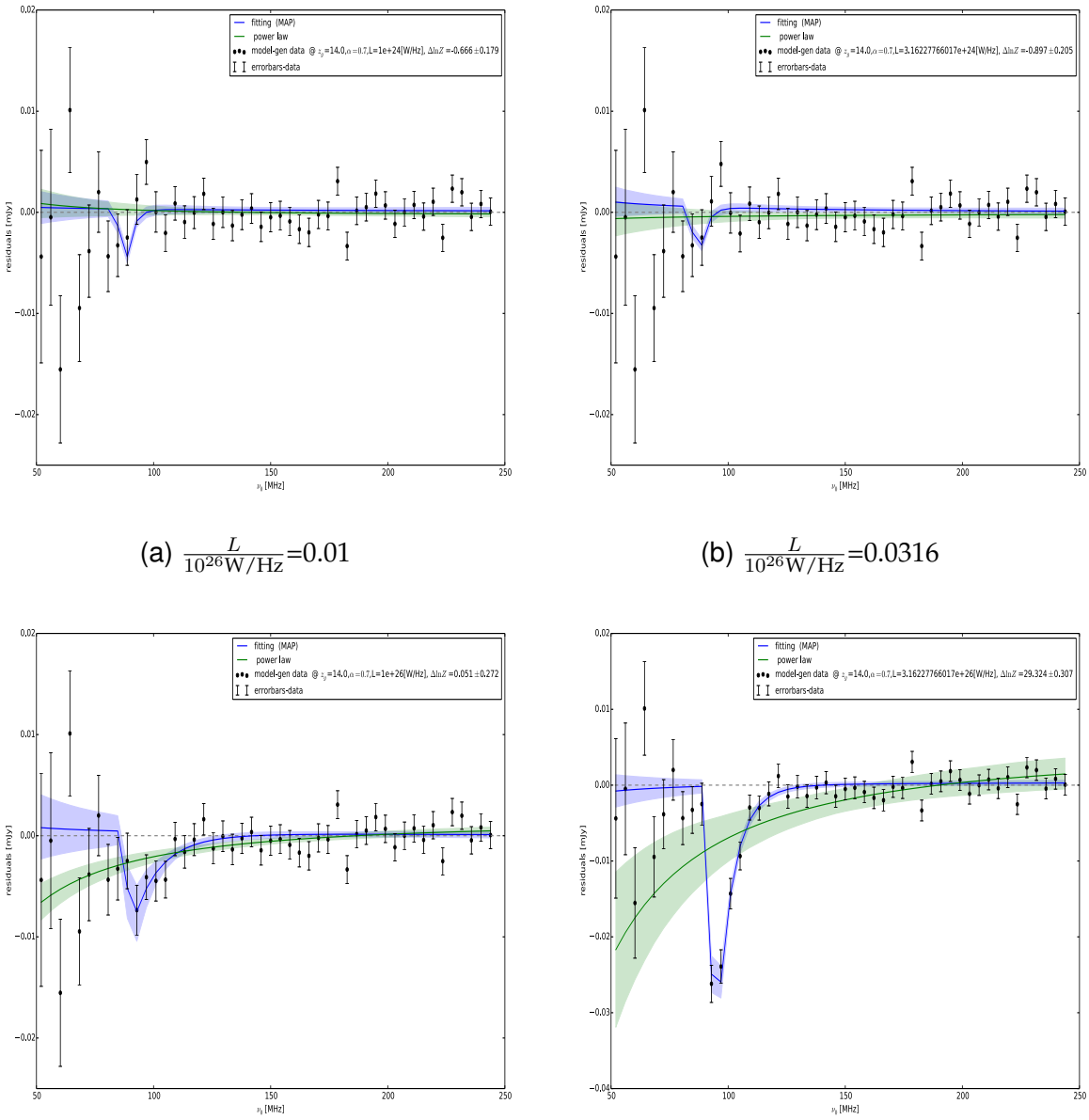


(c) $\frac{L}{10^{26}\mathrm{W/Hz}}$=1.0



(d) $\frac{L}{10^{26}\mathrm{W/Hz}}$=3.16

Fig. 5.13: Residual reconstructions of 4 objects from SIMFAST21-generated data. The black points represent the SIMFAST21-generated data. The blue line is the reconstruction from the parameter estimation using eqn.(5.5). The green line is the null power law fitting using eqn.(5.6) .



(a) $\frac{L}{10^{26}\mathrm{W/Hz}}$=0.01

(b) $\frac{L}{10^{26}\mathrm{W/Hz}}$=0.0316

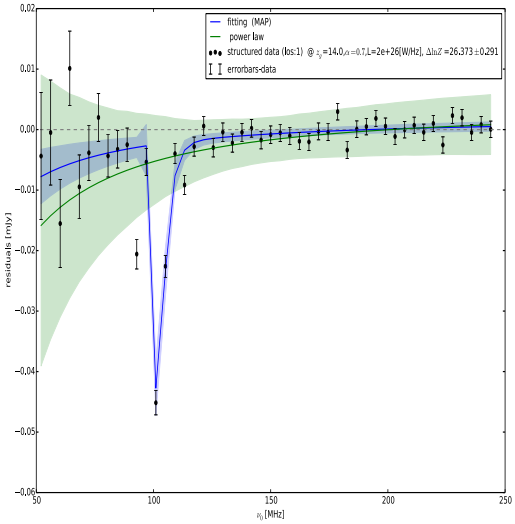
(c) $\frac{L}{10^{26}\mathrm{W/Hz}}$=1.0

(d) $\frac{L}{10^{26}\mathrm{W/Hz}}$=3.16

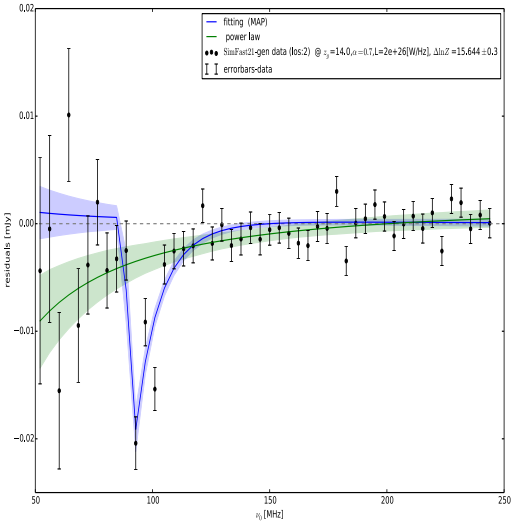
Fig. 5.14: Residual reconstructions of 4 objects from the model-generated data. The black points represent the model-generated data. The blue line is the reconstruction from the parameter estimation using eqn.(5.5). The green line is the null power law fitting using eqn.(5.6).

Fig.(5.15) shows reconstructions from three different lines of sight from the SIMFAST21-generated data. These reconstructions are from the same object (i.e same z_g,L and α), but observed at different lines of sight. The trend of the data from all three lines of sight is the same, Fig.(5.15a) and Fig.(5.15b) seem almost identical but Fig.(**??**) appears to be slightly different from these other two. These reconstructions serves as evidence that even randomness of the structure the overall behaviour is the same in all three. This

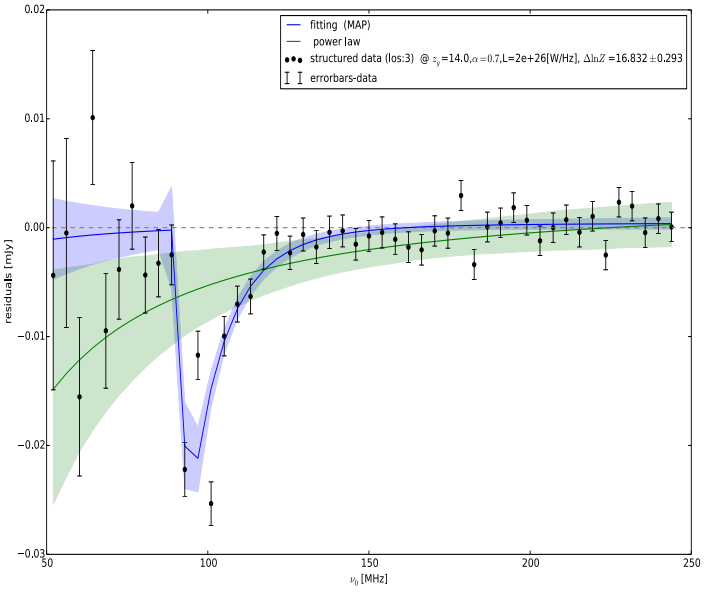
means we can use a single LOS to instead of many since the trend is identical. The log evidence found though does show more variation than the reconstructions themselves.



(a) LOS 1



(b) LOS 2



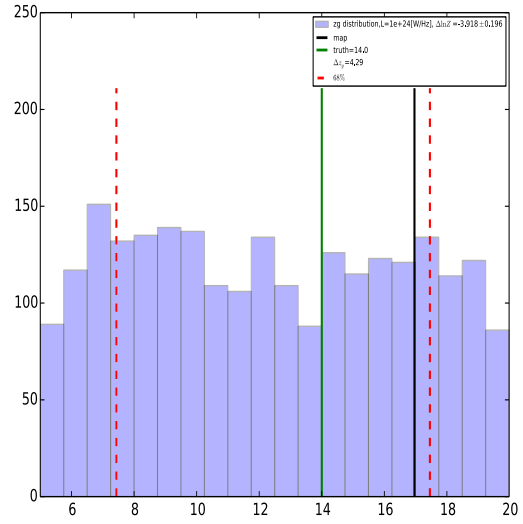
(c) LOS 3

Fig. 5.15: Reconstructions of same the object (z_g=14.0, L=1 $\times$ 10^{26}W/Hz and α =0.7) using the SIMFAST21-generated data. The black points represent the model-generated data. The blue line is the reconstruction from the parameter estimation using eqn.(5.5). The green line is the null power hypothesis using eqn.(5.6).

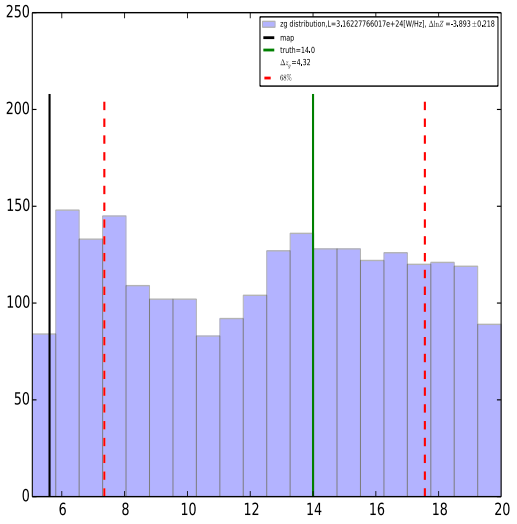
Now the analysis focuses on a single parameter for the four objects that been already been analysed using posterior distributions and reconstructions. Probably the most significant and interesting parameter to look at using a histogram is z_g the object's estimated redshift. Looking at the histogram for the estimated redshift for both SIMFAST21-generated and model-generated data will give insight into the apparent bias discussed, which is visible in the few SIMFAST21-generated data in a few of the estimated parameters. These histograms give a critical overview as to whether the bias seen in the SIMFAST21-generated data increases with an increase in luminosity.

A key feature of the analysis is whether the histogram distributions differ much for the two data sets. The histograms in Fig.(5.16) and Fig.(5.17) certainly give a different perspective and a rather more in-depth understanding of the conditions at play with parameters being estimated. It will give insight into the differences expected in z_g from faint sources and bright sources. The object at in Fig.(5.16a), with $\frac{L}{10^{26}\mathrm{W/Hz}}$=0.01 histogram distribution from the SIMFAST21-generated case are very chaotic, there is no obvious peak and distribution does not converge which could be due to (one of) two reasons; poor model or the signal to noise ratio, but since the models have been checked and tested, the signal to noise is the only plausible cause.

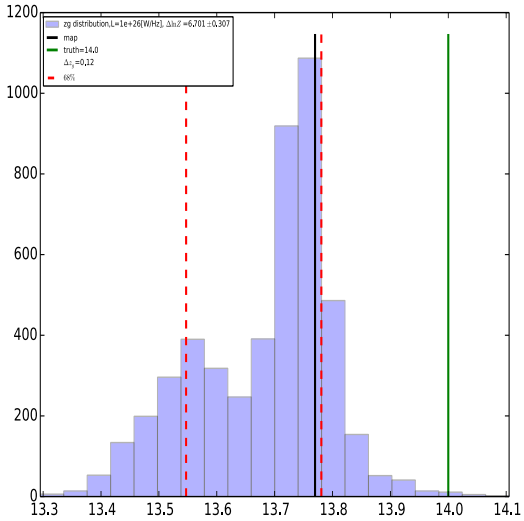
When looking at the corresponding object for the model-generated data case in Fig.(5.17a), the histogram distribution is skewed toward redshift 6. The histogram at $\frac{L}{10^{26}\mathrm{W/Hz}}$=0.0316 for the model-generated data in Fg.(5.17b), appears flat with no definitive peak appearing in the range. The SIMFAST21-generated histogram at $\frac{L}{10^{26}\mathrm{W/Hz}}$=0.0316 with $\Delta \ln Z = -0.897 \pm 0.205$ seems to suggest that z_g>14, but the histogram is still chaotic, hence the redshift estimate is not definitive based on the distributions. This is due to noise drowning the signal, hence it has impact on the parameter fitting. For the SIMFAST21-generated data object at $\frac{L}{10^{26}\mathrm{W/Hz}}$=1.0, the histogram is narrow and has a clear visible peak around z_g =13.8±0.12 , this value does not match the true value used. but its very close to the true value. This instance exposes the bias in the SIMFAST21-generated data. For the corresponding object in the model-generated data case, the distribution is narrow and converges around z_g =14.1±2.07, which is right on the true value. We note both these objects from SIMFAST21-generated & model-generated data are still being detected to have the 21 cm absorption feature. In summary, even though high redshift sources may be slightly biased, they are detected with ease to have the 21 cm absorption in their spectra.



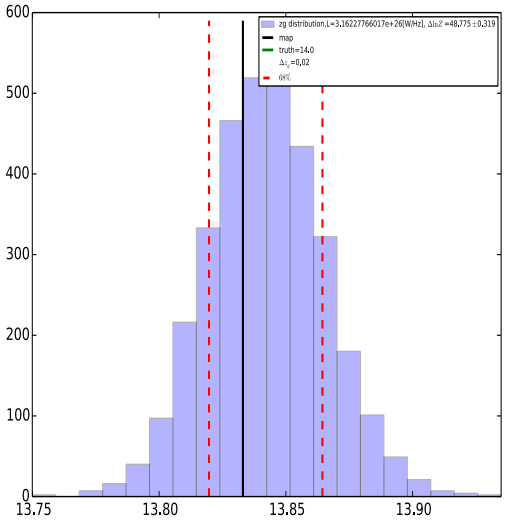
(a) $\frac{L}{10^{26}\mathrm{W/Hz}}$=0.01



(b) $\frac{L}{10^{26}\mathrm{W/Hz}}$=0.0316

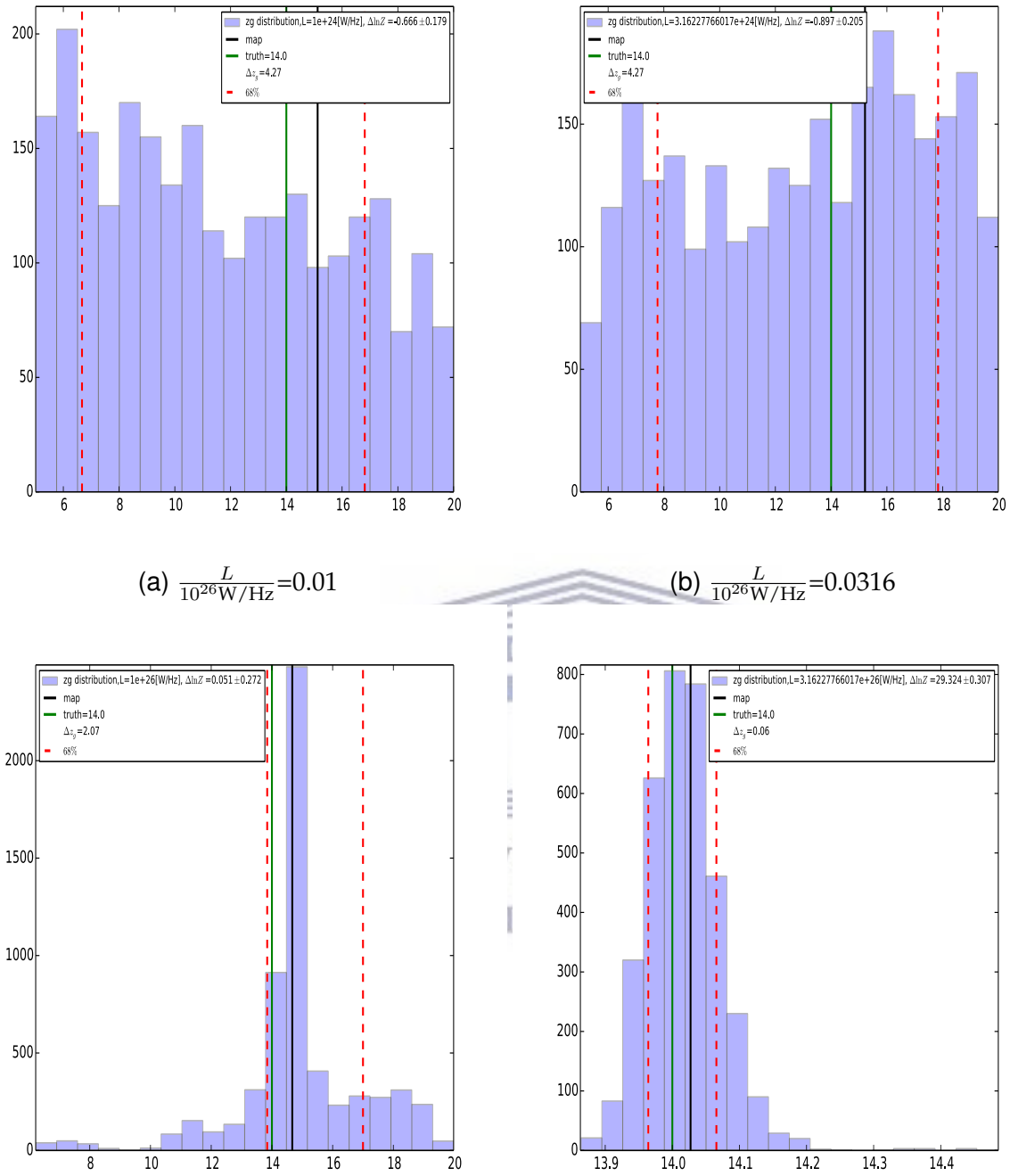


(c) $\frac{L}{10^{26}\mathrm{W/Hz}}$=1.0



(d) $\frac{L}{10^{26}\mathrm{W/Hz}}$=3.16

Fig. 5.16: Histogram distribution for the estimated redshift (z_g) of 4 objects from the SIMFAST21-generated data. The green line represents the true value, the red dotted lines indicate the 68 percent confidence interval for the data distribution, black line the MAP.

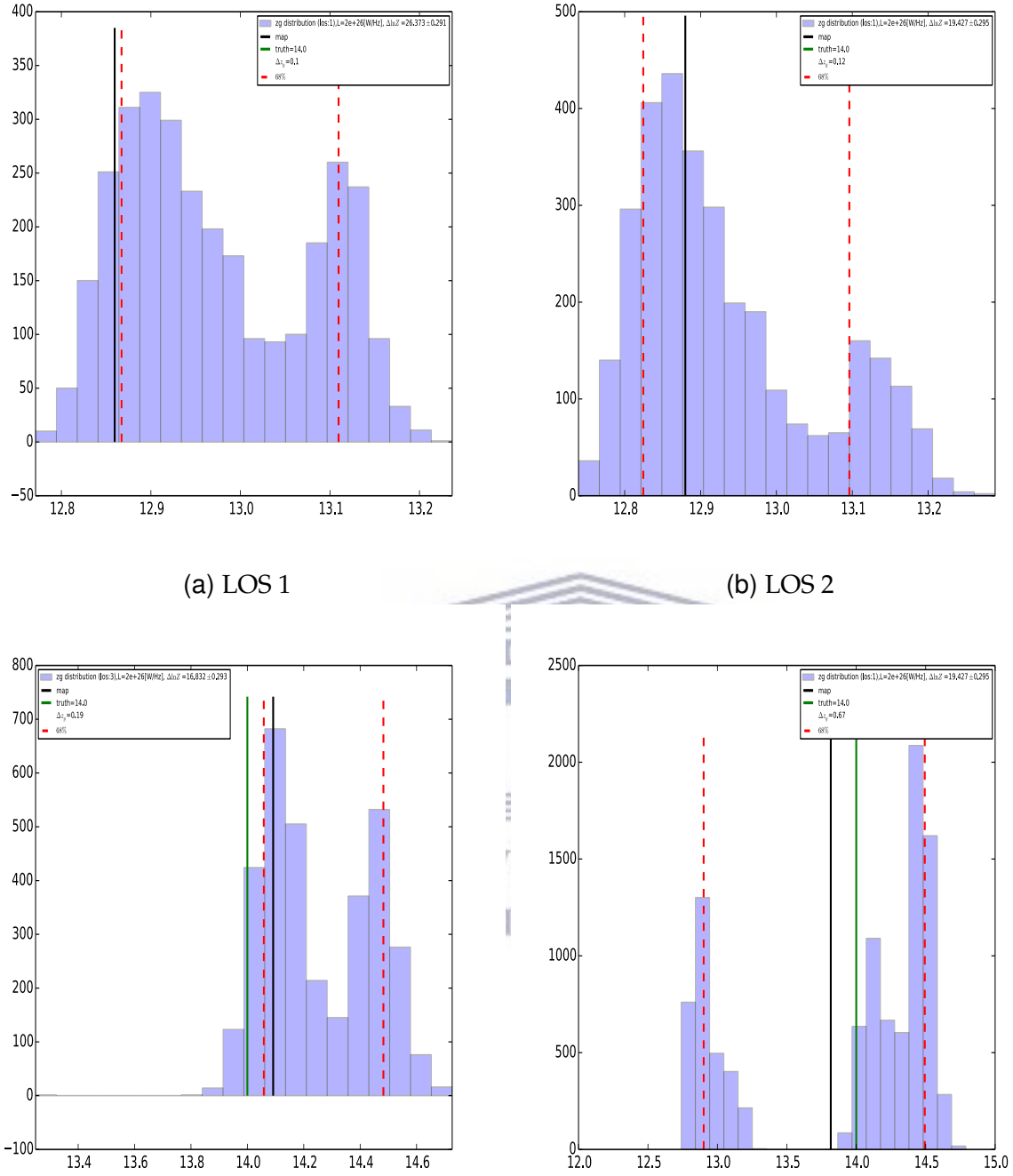


(a) $\frac{L}{10^{26}\mathrm{W/Hz}}$=0.01 (b) $\frac{L}{10^{26}\mathrm{W/Hz}}$=0.0316

(c) $\frac{L}{10^{26}\mathrm{W/Hz}}$=1.0 (d) $\frac{L}{10^{26}\mathrm{W/Hz}}$=3.16

Fig. 5.17: Histogram distribution for the estimated redshift (z_g) of 4 objects from the model-generated data. The green line represents the true value, the red dotted lines indicate the 68 percent confidence interval for the data distribution, black line the MAP.

Fig.(5.18) shows the redshift distribution (corresponding to the reconstructions showed in Fig.(5.15)) from the same source viewed from three different lines of sight in SIMFAST21. We see that indeed for each line of sight, the z fit will be slightly different, although always showing a bias due to the structure that the model does not capture. Ideally we would want to run fits for many lines of sight and combine the posteriors in order to take into account the possible variations between lines of sight. In any case, even with real data, such variation would have to be taken into account with simulations when constructing such "error bars". However, fitting the redshift is not our final objective. Our aim is to determine if there is some absorption, e.g. if the galaxies "lives" in the Epoch of Reionization. Although the evidence for that does change between lines of sight, they all show a detection of the absorption feature (as compared to the null hypotheses). Running this for many lines of sight would improve the statistics for the "false-positives" or "false-negatives" but should not impact on most strong detections. This is an issue that we refer to follow up work.



(a) LOS 1

(b) LOS 2

(c) LOS 3

(d) Combined LOS1,LOS2 and LOS3

Fig. 5.18: Histogram distribution for the estimated redshift (z_g) of the same objecct in 3 lines of sight from the SIMFAST21-generated data and Combined histogram from these 3 lines of sight. The green line represents the true value, the red dotted lines indicate the 68 percent confidence interval for the data distribution, black line the MAP.

5.2.3 Detection space

We introduce a space where the confidence of the parameter estimation is shown in the form of a grid space. The grid is a tool to explore the functional space for the selection function by constructing a regular grid in luminosity space and redshift space. This technique gives a general idea to where parameter fit is most and least efficient. Note the grid in most stances appears yellow and red, this indicates that we are likely to detect an object with a 21 cm absorption feature with high confidence.

Ideally, the space used find how confident are the detections, should be in all the 3 physical parameters: redshift, z_g, luminosity, L_0, and the spectral index, α. Using all 3 of these parameters in constructing a gridded detection space would undoubtedly be expensive computationally. Due to these aforementioned reasons, the grid for the detection space is restricted to only 2-dimensions. The parameter dimensions used are luminosity and redshift, the spectral index is kept constant at 0.7 throughout the grid [3]. In the gridding, the objects are placed in a range of luminosity and redshift along with a constant spectral index of 0.7. The grid is constructed by using the same line of sight and ran to different redshifts. The grid is populated with fits found using 21cmNEST.

All the objects placed in the grid are parameter estimated using both the best model found (i.e. eqn.(5.5)) and the power law (i.e. eqn.(5.6)). Using the two models' Bayesian evidence, a detection proxy confidence is found. This proxy determines the strength of a model that predicts a 21 cm absorption feature favoured over a power law that does not account for 21 cm absorption in galaxies observed during Reionization. The detection proxy is interpreted using the Jeffreys' scaling.

In the constructed grid for the SIMFAST21-generated and model-generated data, the detection scale is colour coded with a probabilistic detection scale. The probabilistic detection scale is the ratio of a galaxy with a 21 cm absorption feature and a galaxy with no 21 cm absorption feature being detected at some redshift and luminosity. This is used to better understand the limitations within the best model used in the parameter estimation of galaxies that have the 21 cm absorption feature. The detection space grid is crucial in finding the signal to noise ratio because the 21 cm absorption model and power law Bayesian mode difference also acts as a proxy for the signal to noise ratio (SNR).

The ranges used in grid are; luminosity: $[10^{23}, 10^{28}]$ W/Hz and redshift: [5.5,20.5].

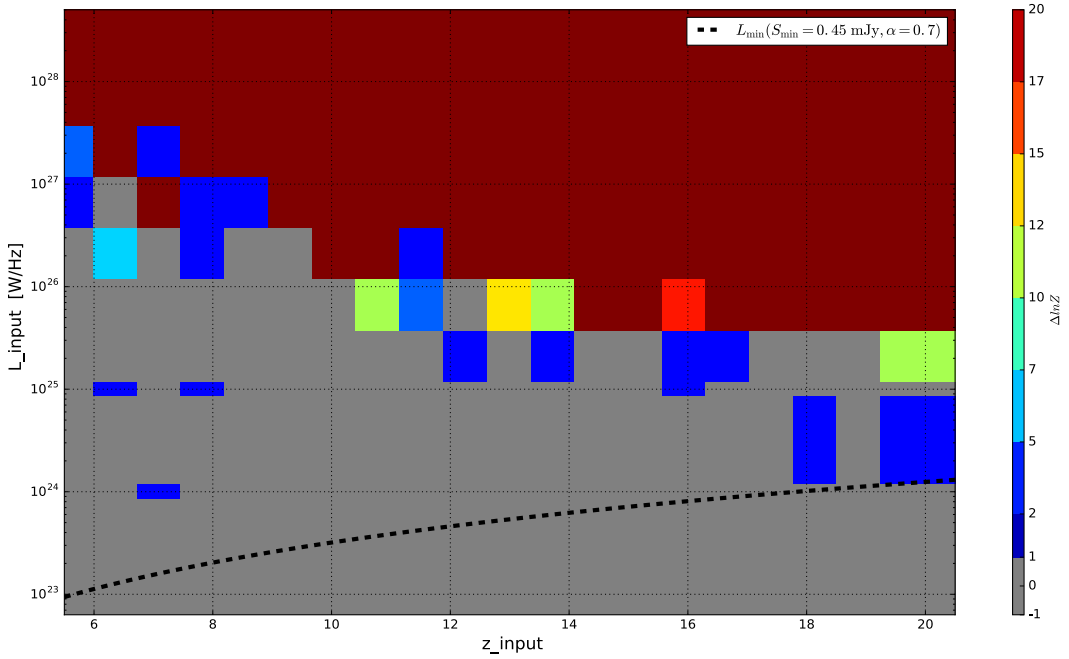
[3] Note In the fits we include the spectral index as part of our analysis.

Note that the range of redshifts is spanned beyond the observationally estimated end of reionization redshift ($z_{\text{reion}} = 6$ Collaboration (2016)). The reason for extending beyond the current z_{reion} in the grid, is to establish if there are any differences in the non-detected of objects in regions within EOR and to those found post EOR, where there is nothing expected to be observed because Hydrogen in the IGM has fully ionized.

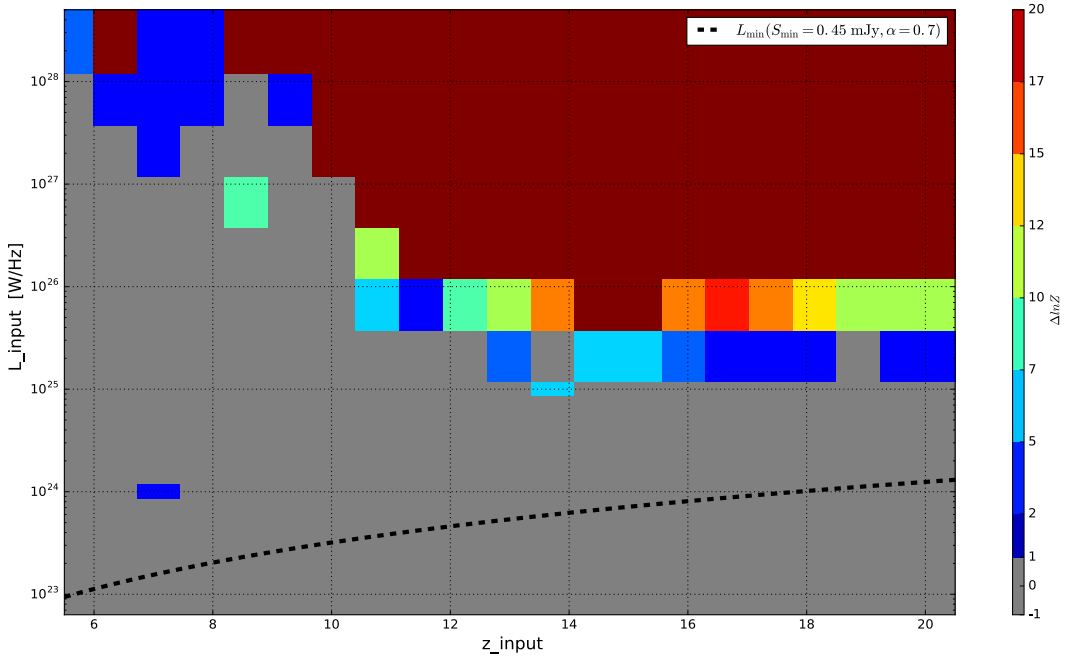
Fig.(5.19) shows grid plots for detection space constructed using SIMFAST21-generated and model-generated data shown side by side. The are a few differences between the two detection space plots. In the model-generated data grid at redshifts lower than 10.5, there are a lot more objects detected, nearly as the same number of objects detected at high redshifts.

The pattern observed in the model-generated data is almost uniform, the colour code correlates with a very high detection confidence indicated by the colour red which is on the colour bar interpreted using Jeffreys' scale. The bluish region in the model-generated data grid appears only on the outskirts of the detected objects. These blue colour coded objects form a barrier separating detected objects and undetected objects. Based on the objects detected on the grid, the selection function for the model seems to detect objects at higher redshifts, but at lower redshifts, the number of objects declines slightly. Comparing with the SIMFAST21-generated data grid, there is an obvious behaviour seen, at low redshifts very few objects are detected at all.

In Fig.(5.19) the vast area covered by this detection grid is at high redshift and high luminosity. Below z=12, we note that detection area is significantly small.This is indicating that few low redshift objects will be detected. The dotted line in the grid represents the luminosity minimum, L_{min}, I have chosen L_{min} based on the sensitivity (S_{min}) of the SKA1-LOW noise. I have defined S_{cut} as ten times the averaged SKA1-LOW noise continuum, $\bar{\sigma}(\nu)$ SKA1-LOW noise. We chose 10 times sigma because below that scale, a noise fluctuation might be confused with a galaxy.



(a) model-generated data



(b) SIMFAST21-generated data

Fig. 5.19: Grid space constructed from luminosity and redshift. The colour coded regions indicate confidence scale for parameter fit. The dotted line in the grid represents the luminosity minimum which is based on the SKA1-LOW sensitivity.

Looking at Fig.(5.19), this implies that objects at low redshifts, need to bright in order to be detected, since there is very little HI absorption. Note the higher redshifts objects can have lower luminosity values and can be categorised as detections with 21 cm absorption feature. Based on this we can conclude that objects with 21 cm signal are easier to detect as increases redshift.

The most likely cause for the disparities in SIMFAST21-generated and model-generated data grids lies in the fact that, when parameter fitting SIMFAST21-generated data using the best model, the model does not account for the structure found in the signal when using SIMFAST21-generated data (data extracted from SIMFAST21). The model assumes that the neutral hydrogen is a smooth function of z which fails at when you have intervening bubbles.

There is a sudden rise in the number of detected objects at z>12.0, this relates to the fact that the absorption feature in the SIMFAST21-generated data approaches 1.0 as we decrease in redshift. Another real indicator of structure in the data is the bias seen in the triangle plots, this would be different if the modelling compensated for the structure from the respective line of sight. The two grid plots would be in agreement 100 percent if the structure was not an issue.

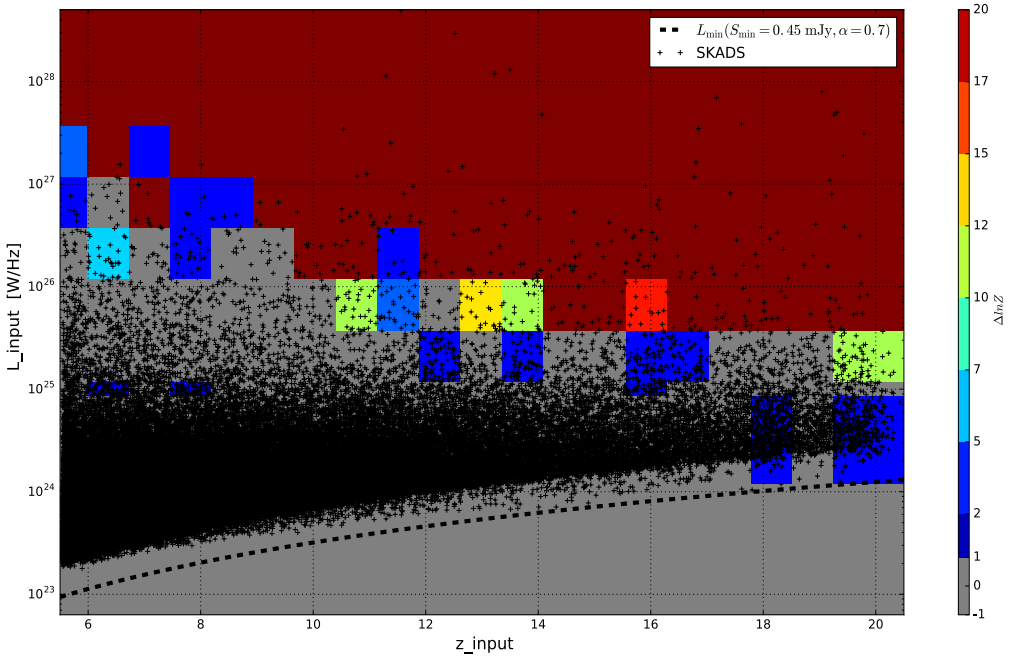
I have used a semi-empirical simulation of the extragalactic radio continuum sky. The simulation is a 20×20 deg ^{2} patch filled with extragalactic sources and is one of a suite of Square Kilometre Array Design Study (SKADS) simulations developed to serve as theoretical precursor experiments to the upcoming biggest radio telescope to ever be built, the SKA. The SKADS Simulated Skies Source Extractor or mostly known as S ^{3}-SEX (Wilman et al., 2008).

I have queried for the flux density information for objects above $z > 6$ and $S_{\text{cut}} > 0.45$ mJy at 151 MHz (objects below this flux are ignored since they are likely not bright enough) from S^3-SEX. Setting a redshift threshold of 6 ensures we avoid computational costs by parameter estimating objects which are located outside EoR. An interesting follow-up test could that we could do in the future; would be randomly parameter estimating objects without having prior knowledge about their respective locations. 107128 sources were recovered and have been overlaid on the grid in Fig.(5.20). This is done to get an approximate number of the objects that might be detected. Since S^3-SEX simulation is based on low redshift observations which are which are extrapolated to high-z since there are no observations available. By overlying S^3-SEX the simulated objects on top of these grids developed, I will be essentially be obtaining a

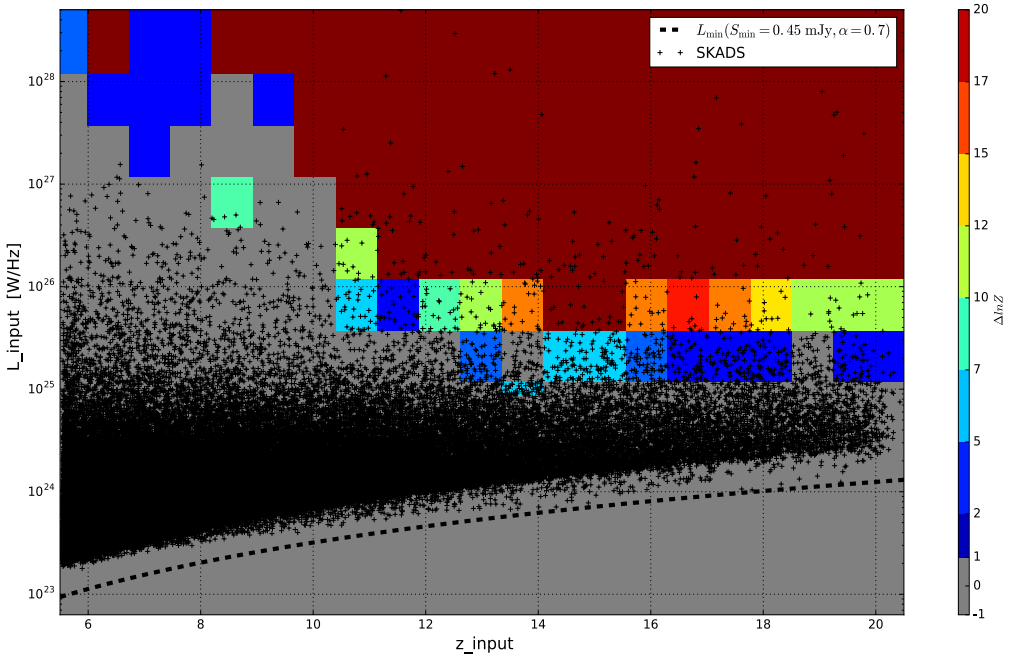
selection function. This selection function estimates the number of objects detected by 21cmNEST.

Fig.(5.20a) shows the S ^{3}-SEX overlaid in the simulation-generated data grid. There are very few objects detected below z=10, because of the sudden depletion of neutral Hydrogen in SIMFAST21. This pattern will obviously be carried over to model-generated data since the model is based on data from SIMFAST21. On higher redshifts, there are far more objects detected.

The S^3-SEX objects that fall into any of the colour coded region (excluding those in the grey colour coded region) are regarded as detection because 21cmNEST would be able to detect those objects to have the 21 cm absorption feature. In Fig.(5.20b) S ^{3}-SEX simulation objects are overlaid on top of the model-generated data. 1342 (roughly around 1 percent) were SKADS objects detected in Fig.(5.20b), while in Fig.(5.20a) 889 SKADS objects were detected which accounts for around 0.8 percent, the disparity in the number of objects mostly like due differences in the data, the data in Fig.(5.20a) is simulated-generated which contains structure from simulation while data in Fig.(5.20b) has no structured since it is constructed from the analytic model.



(a) model-generated data



(b) SIMFAST21-generated data

Fig. 5.20: Grid space constructed from luminosity and redshift. The colour coded regions indicate confidence scale for parameter fit. The dotted line in the grid represents the luminosity minimum which is based on the SKA1-LOW sensitivity. Sources from S^3-SEX are overlaid in the grid range.

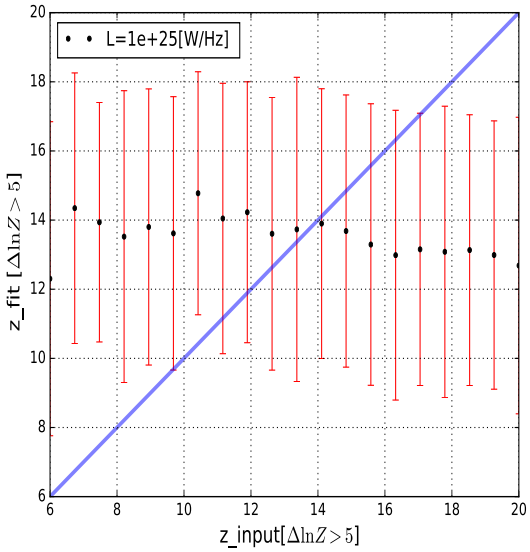
Taking into consideration the objects that were used to create the grid in the detection space, we compare their estimated redshift to the input redshift. These objects are separated according to luminosity and vary in redshift. This categorization is pursued to check if whether faint sources can be estimated accurately as the bright sources. This would mean, the redshift estimation is independent of the luminosity of source.

In all the objects that are placed at the luminosities: $L = 1 \times 10^{25}$ W/Hz and $L = 3.16 \times 10^{25}$ W/Hz, we find that none of the estimated redshifts are in agreement with the input redshift. Note that the errorbars for these estimated redshifts are excessively large. We can conclude that objects at $L = 1 \times 10^{25}$ W/Hz and $L = 3.16 \times 10^{25}$ W/Hz are too faint. Hence in this sources redshift can not be accurately estimated.

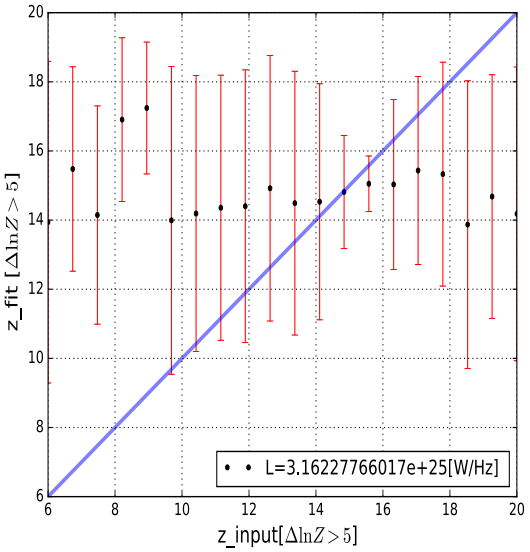
For the sources with the luminosities: $L = 1{\times}10^{26}$ W/Hz and $L = 3.16{\times}10^{26}$ W/Hz, we find that a lot of the estimated redshifts are in agreement with the input redshift. The error-bars are very small, which suggest `z-fit` for these objects is well constrained in bright sources.

I have imposed a condition on detected objects; I only considered objects that have $\Delta \ln Z$>5. This is because objects that have lower $\Delta \ln Z$ are considered to be weak to inconclusive detections when applying the Jeffreys scale.

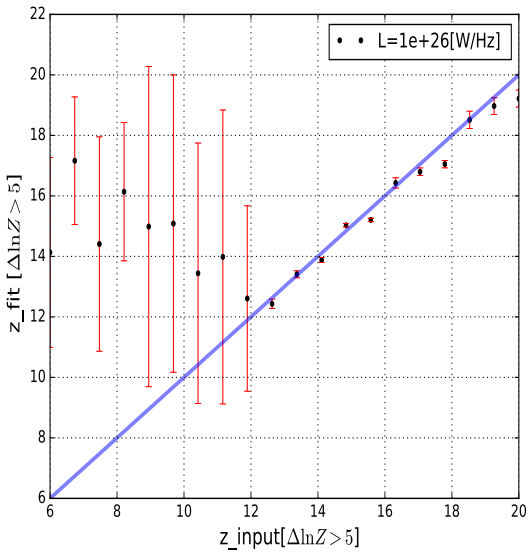
The Fig.(5.21) shows `z-input` plotted against the recovered redshift `z-fit` for objects used to construct the L-z grid. Fig.(5.21) is from the SIMFAST21-generated data, while Fig.(5.22) is from the model-generated data. In both Fig.(5.21) and Fig.(5.22) the errorbars get bigger as brightness decreases. As seen from Fig.(5.19) the possibility of detecting objects with 21 cm absorption at luminosities below 10^{26}W/Hz drastically decline as 21 cm feature is immersed in the noise thus the fitting technique detects these objects as power law. The opposite is seen as luminosity is increased, and redshift errorbars become very small.



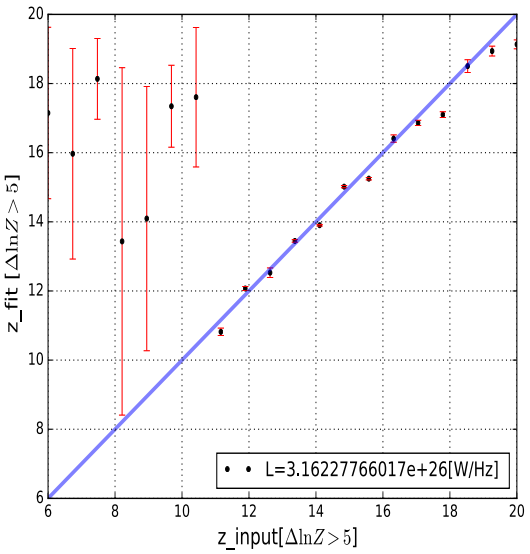
(a) $L = 1 \times 10^{25}$[W/Hz]



(b) $L = 3.16227 \times 10^{25}$[W/Hz]

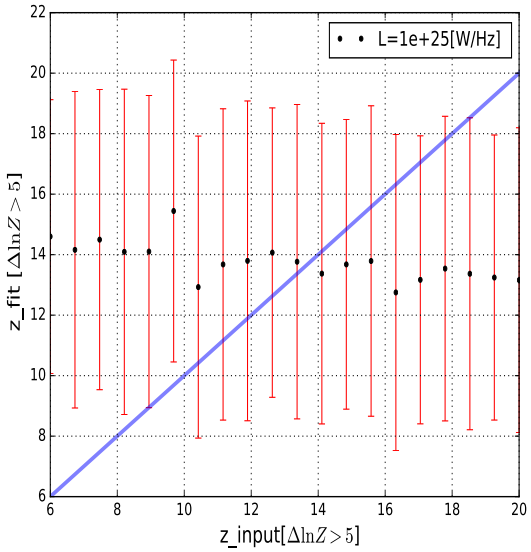


(c) $L = 1 \times 10^{26}$[W/Hz]

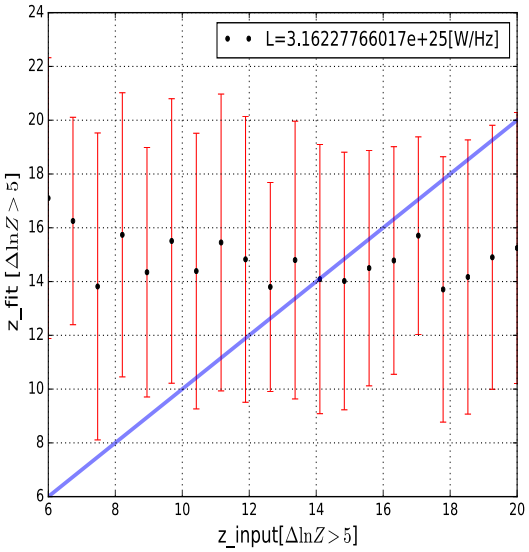


(d) $L = 3.16227 \times 10^{26}$[W/Hz]

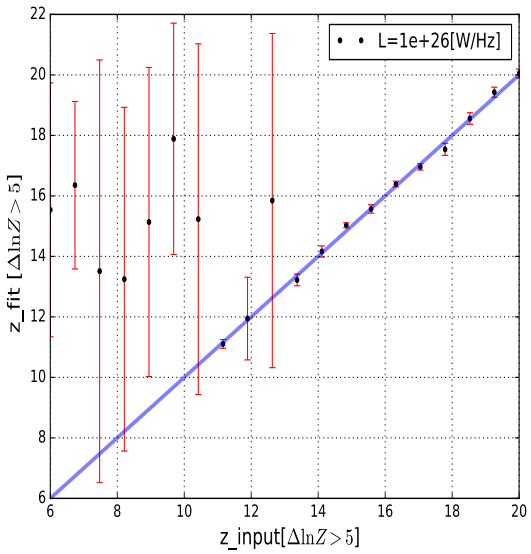
Fig. 5.21: The input redshifts (z-input) for galaxies placed on the same luminosity are estimated redshift (z-fit), these objects are constructed from the SIMFAST21-generated data. The estimated redshifts are accompanied by the corresponding errorbars.



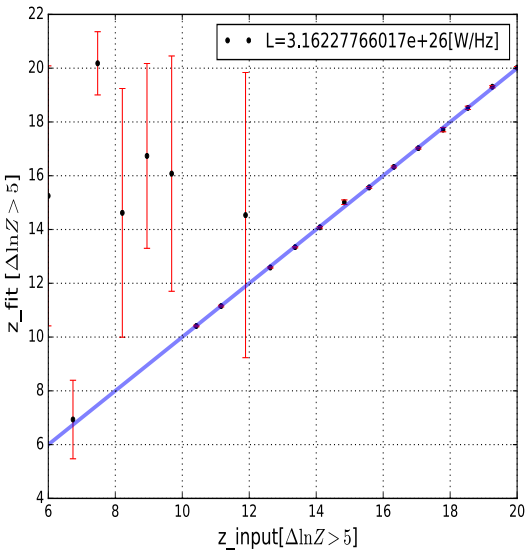
(a) $L = 1 \times 10^{25}$[W/Hz]



(b) $L = 3.16227 \times 10^{25}$[W/Hz]



(c) $L = 1 \times 10^{26}$[W/Hz]



(d) $L = 3.16227 \times 10^{26}$[W/Hz]

Fig. 5.22: The input redshifts (z-input) for galaxies placed on the same luminosity are estimated redshift (z-fit), these objects are constructed from the model-generated data. The estimated redshifts are accompanied by the corresponding error-bars.

Chapter 6

Conclusions and future work

The primary aim of this work was to design a technique to parameter-estimate spectral models for galaxies found in the epoch of reionization. My technique will be the first of its kind to detect galaxies in the epoch of reionisation for those of low signal to noise ratio. So far, we have not been able to probe the Universe above redshift 6 in the radio, but radio telescopes like SKA1-Low will address this. The method relies on the expected 21 cm absorption feature in the spectra of galaxies observed from reionization, is vital in estimating the redshift of sources in reionization.

The first task was to find analytical models that could mimic the expected 21 cm absorption feature using SIMFAST21. In addressing this, three models $\{1, 2, 3\}$ (i.e $\{\widetilde{A}_1, \widetilde{A}_2, \widetilde{A}_3\}$) were fitted. I applied a Bayesian framework for the model selection, in order to determine the most suitable model. Initially, the modelling did not work because I did not incorporate the experiment's resolution; this was solved by integrating beam over the frequency. Two different noise distributions (i.e. One constant in frequency and the other based on SKA1-LOW noise) were used for the parameter estimation. Model 1 was found to be the most suitable model in both cases. The method's proficiency is demonstrated with two data sets: SIMFAST21-generated $\&$ model-generated data. The SIMFAST21-generated data was constructed using a simulation of the Epoch of Reionization. Model-generated data are synthesised from the best analytical model 1 (i.e. $\widetilde{A}_1,$). These data sets were vital for comparing findings; to check whether the behaviour is random or appears on both data sets used. I found that SIMFAST21-generated data is structured and this structure is unique to each line -of- sight from SIMFAST21. The model-generated data is more smoother. In summary, SIMFAST21-generated and model-generated data each were combined with the high redshift galaxy template to form spectra of galaxies in the EOR. The resulting spectra from SIMFAST21-generated and model-generated data were fitted using a SED template that accounts for the 21 cm absorption feature in galaxies during EOR. I found that these templates were very effective when account for the or the 21 cm absorption feature in galaxies during EOR. The modelling and fitting technique were developed and combined into a code called

21cmNEST.

To better understand the modelling capabilities 21cmNEST, I ran a fit for objects at different luminosities but at the same redshift. Objects at luminosities below 10^{26} W/Hz had flat posterior distributions indicating poor quality of fits because of low signal-to-noise ratio. I used these recovered parameters to reconstruct the expected spectra. The residuals were a lot more conclusive, because objects with luminosities below 10^{26} W/Hz in favour a model which is a power law (i.e the null hypothesis- for a case where the galaxy is in an already-ionized Universe) instead of a model that accounts for the 21 cm absorption feature. In summary, all the faint objects found in the EOR will favour the null hypothesis (no 21 cm absorption feature) because the feature is hidden in the noise. The fitted redshift z_g distribution of faint sources is rather flat and uninformative: many have large uncertainties (roughly $\Delta z_g \sim 4.0$) in their redshift position are biased.

I found in the case of fitting the SIMFAST21-generated data that there is a bias in. But when fitting the toy model (model-generated) data there was no bias visible. The structure along the line-of-sight from the simulation leads to the bias observed. Note, however, that in the low signal-to-noise regime that we will generally encounter, such features will be washed out. In the very-high signal-to-noise regime, a biased detection will be made but that can always be corrected later by specifically observing at such sources with spectroscopic surveys. Objects at higher redshifts (z>12) and brighter (L> 10^{26} W/Hz) are easier to model. The opposite is seen in the case of low redshift and faint sources.

A method was formulated for a detection proxy in the form of a grid to quantify regions where the absorption feature can be detected. In L-z space, the detection proxy is quantified using the Bayesian evidence found when fitting data with the best model versus that fitted with a power law. The Jeffreys scale was used to interpret the detection proxy. The resulting grids look very similar for both SIMFAST21-generated and model-generated data. However, the structure in the SIMFAST21-generated data has fewer sources detectable at low redshifts ($z < 10.5$) .

The grids for both SIMFAST21-generated and toy data were overlaid with SKADS simulated sources found using the expected specifications for SKA1-LOW; objects below a flux $S_{\text{cut}} > 0.45$ mJy at 151 MHz were ignored. Another constraint condition on the SKADS sources was a redshift threshold of 6 , to ensures I avoid computational costs in modelling objects located outside the EoR. This gave a rough selection function. A total

of 889 (0.8 %) SKADS objects were detected $\ln \Delta Z > 1.0$ in the grid constructed from the 21cmNEST, while 1342 (1%) were detected when using the toy data. In the grids, there are regions above the minimum sensitivity of the SKA1-LOW noise continuum limit where no detections are made. Some objects may be observable with SKA1-LOW, but 21cmNEST may not detect them because their 21 cm absorption feature is obscured by the noise. I have described the concept used to construct 21cmNEST and applied it to model-generated and SIMFAST21-generated data for checking the detection of the 21 cm absorption feature.

6.1 Future work

I explore the possible future extension for this methodology.

(i) In this thesis I have used cosmological parameters from (Collaboration, 2016). We could compare different Cosmology models, and more importantly, look at other reionization models.

(ii) An exciting prospect would be comparing the results found when using other radio telescopes (HERA, MWA and LOFAR) that have the potential to detect these radio sources during the EOR, as discussed in Chapter 1.

(iii) I could try to model the 21 cm absorption feature with models other than power laws.

(iv) Finally, I plan on using this robust statistical method on the observational data from the LOFAR telescope in order to see if I can find radio galaxies in the data from the epoch of reionization.

Bibliography

Allison, R. and J. Dunkley (2013). "Comparison of sampling techniques for Bayesian parameter estimation". In: *Monthly Notices of the Royal Astronomical Society* 437.4, pp. 3918–3928. ISSN: 0035-8711. DOI: 10.1093/mnras/stt2190. arXiv: 1308.2675. URL: http://arxiv.org/abs/1308.2675.

Alvarez, Marcelo a. et al. (2009). "Connecting Reionization to the Local Universe". In: *The Astrophysical Journal* 703, pp. L167–L171. ISSN: 0004-637X. DOI: 10.1088/0004-637X/703/2/L167. arXiv: 0812.3405. URL: http://arxiv.org/abs/0812.3405.

Barkana, R and A Loeb (2001). "In the beginning: the first sources of light and the reionization of the universe". In: *Physics Reports* 349. URL: http://www.sciencedirect.com/science/article/pii/S0370157301000199.

Battaglia, N et al. (2012). "Reionization on Large Scales I: A Parametric Model Constructed from Radiation-Hydrodynamic Simulations". In: *eprint arXiv* 1211, p. 2821. ISSN: 0004-637X. DOI: 10.1088/0004-637X/776/2/81. arXiv: 1211.2821. URL: http://adsabs.harvard.edu/cgi-bin/nph-data{_}query?bibcode=2012arXiv1211.2821B{\&}link{_}type=ABSTRACT{\%}5Cnpapers://75fc0b17-327f-4862-966f-7602a0de4150/Paper/p3855.

Beardsley, A. P. et al. (2015). "ADDING CONTEXT TO JAMES WEBB SPACE TELESCOPE SURVEYS WITH CURRENT AND FUTURE 21 cm RADIO OBSERVATIONS". In: *The Astrophysical Journal* 800.2, p. 128. ISSN: 1538-4357. DOI: 10.1088/0004-637X/800/2/128. URL: http://adsabs.harvard.edu/abs/2015ApJ...800..128B.

Becker, Robert H. et al. (2001). "Evidence for Reionization at z ˜ 6: Detection of a Gunn-Peterson Trough in a z=6.28 Quasar". In: pp. 1–10. DOI: 10.1086/324231. arXiv: 0108097 [astro-ph]. URL: http://arxiv.org/abs/astro-ph/0108097.

Blumenthal, G.; Miley, G. (1979). "Spectral index dependent properties of steep spectrum radio sources". In: *Astronomy and Astrophysics* 80.1, pp. 13–21. URL: http://adsabs.harvard.edu/cgi-bin/bib{_}query?1979A{\&}A....80..13B.

Booth, R. S. et al. (2009). "MeerKAT Key Project Science, Specifications, and Proposals". In: arXiv: 0910.2935. URL: http://arxiv.org/abs/0910.2935.

Bowman, Judd D. et al. (2013). "Science with the Murchison Widefield Array". In: *Publications of the Astronomical Society of Australia* 30, e031. ISSN: 1323-3580. DOI: 10.1017/pas.2013.009. arXiv: 1212.5151. URL: http://www.journals.cambridge.org/abstract{_}S132335801300009X.

Breugel, W J M van et al. (1999). "A Radio Galaxy at z = 5.19". In: *The Astrophysical Journal* 518.2, pp. L61–L64. ISSN: 0004637X. DOI: 10.1086/312080. URL: http://adsabs.harvard.edu/cgi-bin/nph-data{_}query?bibcode=1999ApJ...518L..61V{\&}link{_}type=EJOURNAL{\%}5Cnpapers3://publication/doi/10.1086/312080.

Brewer, Brendon J., Livia B. Pártay, and Gábor Csányi (2009). "Diffusive Nested Sampling". In: arXiv: 0912.2380. URL: http://arxiv.org/abs/0912.2380.

Bromm, Volker et al. (2009). "The formation of the first stars and galaxies." In: *Nature* 459.7243, pp. 49–54. ISSN: 0028-0836. DOI: 10.1038/nature07990. arXiv: 0905.0929. URL: http://eds.a.ebscohost.com/eds/pdfviewer/pdfviewer?vid=10{\&}sid=5f2a4cf7-f58c-4cbd-9530-05c258cd73bc{\%}40sessionmgr4008{\&}hid=4203{\%}5Cnhttp://dx.doi.org/10.1038/nature07990.

Brookes, M. H. et al. (2008). "A Combined EIS-NVSS Survey of Radio Sources (CENSORS) - III. Spectroscopic observations". In: *Monthly Notices of the Royal Astronomical Society* 385.3, pp. 1297–1326. ISSN: 00358711. DOI: 10.1111/j.1365-2966.2008.12786.x. arXiv: 0802.1158 [astro-ph].

Burgarella, D. et al. (2013). "Herschel PEP/HerMES: the redshift evolution (0 <= z <= 4) of dust attenuation and of the total (UV+IR) star formation rate density". In: *Astronomy and Astrophysics* 554, A70. ISSN: 0004-6361. DOI: 10.1051/0004-6361/201321651. arXiv: 1304.7000. URL: http://ads.nao.ac.jp/abs/2013A{\&}A...554A..70B{\%}5Cnhttp://ads.nao.ac.jp/cgi-bin/nph-data{_}query?bibcode=2013A{\&}A...554A..70B{\&}link{_}type=ARTICLE.

Carilli, C. L., N. Y. Gnedin, and F. Owen (2002). "HI 21 Centimeter Absorption beyond the Epoch of Reionization". In: *The Astrophysical Journal* 577.1, pp. 22–30. ISSN: 0004-637X. DOI: 10.1086/342179. arXiv: 0205169 [astro-ph]. URL: http://arxiv.org/abs/astro-ph/0205169.

Ciardi, B et al. (2000). "Inhomogeneous reionization of the intergalactic medium regulated by radiative and stellar feedbacks". In: *Monthly Notices of the Royal Astronomical Society* 314, p. 611. ISSN: 1365-2966. DOI: 10.1046/j.1365-8711.2000.03365.x. arXiv: 9907189 [astro-ph]. URL: http://adsabs.harvard.edu/cgi-bin/nph-data{_}query?bibcode=2000MNRAS.314..611C{\&}link{_}type=ABSTRACT{\%}5Cnpapers://75fc0b17-327f-4862-966f-7602a0de4150/Paper/p3093.

Ciardi, B. et al. (2012). "Prospects for detecting the 21 cm forest from the diffuse intergalactic medium with LOFAR". In: *Monthly Notices of the Royal Astronomical Society* 428.2, pp. 1755–1765. ISSN: 0035-8711. DOI: 10.1093/mnras/sts156. arXiv: 1209.2615. URL: http://arxiv.org/abs/1209.2615.

Ciardi, Benedetta et al. (2015). "21cm Forest with the SKA". In: *arXiv preprint*, p. 10. arXiv: 1501.04425. URL: http://arxiv.org/abs/1501.04425.

Collaboration, Planck (2016). "Planck intermediate results. XLVII. Planck constraints on reionization history". In: *Astronomy & Astrophysics*, p. 19. ISSN: 0004-6361. DOI: 10.1051/0004-6361/201628897. arXiv: 1605.03507. URL: http://arxiv.org/abs/1605.03507.

Cruz, Maria J. et al. (2006). "The 6C** sample of steep-spectrum radio sources -I. Radio data, near-infrared imaging and optical spectroscopy". In: *Monthly Notices of the Royal Astronomical Society* 373.4, pp. 1531–1562. ISSN: 00358711. DOI: 10.1111/j.1365-2966.2006.11101.x.

Cruz, Maria J. et al. (2007). "The 6C** sample of steep-spectrum radio sources - II. Redshift distribution and the space density of high-redshift radio galaxies". In: *Monthly Notices of the Royal Astronomical Society* 375.4, pp. 1349–1363. ISSN: 00358711. DOI: 10.1111/j.1365-2966.2007.11390.x. arXiv: 0612268v2 [arXiv:astro-ph].

Davé, Romeel, Kristian Finlator, and Benjamin D. Oppenheimer (2006). "The physical properties and detectability of reionization-epoch galaxies". In: *Monthly Notices of the Royal Astronomical Society* 370.1, pp. 273–288. ISSN: 00358711. DOI: 10.1111/j.1365-2966.2006.10464.x. arXiv: 0511532 [astro-ph].

De Breuck, Carlos et al. (2000). "A Sample of 669 Ultra Steep Spectrum Radio Sources to Find High Redshift Radio Galaxies". In: *arXiv.org* astro-ph.2, pp. 303–333. ISSN: 0365-0138. DOI: 10.1051/aas:2000181. arXiv: 0002297 [astro-ph]. URL: http://adsabs.harvard.edu/cgi-bin/nph-data{_}query?bibcode=2000A{\%}26AS..143..303D{\&}link{_}type=EJOURNAL{\%}5Cnpapers3://publication/doi/10.1051/aas:2000181.

De Breuck, Carlos et al. (2002). "Optical and Near-Infrared Imaging of Ultra–Steep-Spectrum Radio Sources: The [ITAL]K[/ITAL]-[CLC][ITAL]z[/ITAL][/CLC] Diagram of Radio-selected and Optically Selected Galaxies". In: *The Astronomical Journal* 123.2, pp. 637–677. ISSN: 00046256. DOI: 10.1086/324632. arXiv: 0109540 [astro-ph]. URL: http://stacks.iop.org/1538-3881/123/i=2/a=637.

DeBoer, David R. et al. (2016). "Hydrogen Epoch of Reionization Array (HERA)". In: p. 26. arXiv: 1606.07473. URL: http://arxiv.org/abs/1606.07473.

Dillon, Joshua S., Adrian Liu, and Max Tegmark (2013). "A fast method for power spectrum and foreground analysis for 21 cm cosmology". In: *Physical Review D* 87.4,

p. 043005. ISSN: 1550-7998. DOI: 10.1103/PhysRevD.87.043005. arXiv: 1211.2232. URL: http://arxiv.org/abs/1211.2232.

Eisenstein, Zehavi, and Hogg (2004). "Baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies". In: *The Astrophysical Journal* 129.3, pp. 450–467. DOI: 10.1086/466512. arXiv: 0501171v1 [arXiv:astro-ph].

Eisenstein, D. J. (2005). "Dark energy and cosmic sound". In: *New Astronomy Reviews* 49.7-9, pp. 360–365. ISSN: 13876473. DOI: 10.1016/j.newar.2005.08.005.

Ewen, HI and EM Purcell (1951). "Observation of a Line in the Galactic Radio Spectrum: Radiation from Galactic Hydrogen at 1,420 Mc./sec." In: *Nature* 168.4270, pp. 356–356. ISSN: 0028-0836. DOI: 10.1038/168356a0. URL: http://scholar.google.com/scholar?hl=en{\&}btnG=Search{\&}q=intitle:No+Title{\#}0http://www.nature.com/doifinder/10.1038/168356a0http://www.nature.com/nature/journal/v168/n4270/abs/168356a0.html.

Falcke, H et al. (2006). "A very brief description of LOFAR - the Low Frequency Array". In: *Proceedings of the International Astronomical Union* 2.14, p. 2. ISSN: 1743-9213. DOI: 10.1017/S174392130701112X. arXiv: 0610652 [astro-ph]. URL: http://arxiv.org/abs/astro-ph/0610652.

Fan, Xiaohui et al. (2003). "A Survey of z>5.7 Quasars in the Sloan Digital Sky Survey II: Discovery of Three Additional Quasars at z>6". In: *arXiv.org* 125.April, pp. 1649–1659. DOI: 10.1086/368246. arXiv: 0301135 [astro-ph]. URL: http://arxiv.org/abs/astro-ph/0301135.

Fan, Xiaohui et al. (2005). "Constraining the Evolution of the Ionizing Background and the Epoch of Reionization with z ˜ 6 Quasars II: A Sample of 19 Quasars". In: DOI: 10.1086/504836. arXiv: 0512082 [astro-ph]. URL: http://arxiv.org/abs/astro-ph/0512082http://dx.doi.org/10.1086/504836.

Feroz, F. and M. P. Hobson (2008). "Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses". In: *Monthly Notices of the Royal Astronomical Society* 384.2, pp. 449–463. ISSN: 00358711. DOI: 10.1111/j.1365-2966.2007.12353.x. arXiv: 0704.3704. URL: http://arxiv.org/abs/0704.3704.

Feroz, F., M. P. Hobson, and M. Bridges (2009). "MultiNest: An efficient and robust Bayesian inference tool for cosmology and particle physics". In: *Monthly Notices of the Royal Astronomical Society* 398.4, pp. 1601–1614. ISSN: 00358711. DOI: 10.1111/j.1365-2966.2009.14548.x. arXiv: 0809.3437.

Feroz, F. et al. (2013). "Importance Nested Sampling and the MultiNest Algorithm". In: p. 28. arXiv: 1306.2144. URL: http://arxiv.org/abs/1306.2144.

Fialkov, Anastasia and Abraham Loeb (2016). "PRECISE MEASUREMENT OF THE REIONIZATION OPTICAL DEPTH FROM THE GLOBAL 21 cm SIGNAL ACCOUNTING FOR COSMIC HEATING". In: *The Astrophysical Journal* 821.1, p. 59. ISSN: 1538-4357. DOI: 10.3847/0004-637X/821/1/59. arXiv: 1601.03058. URL: http://arxiv.org/abs/1601.03058{\%}5Cnhttp://dx.doi.org/10.3847/0004-637X/821/1/59{\%}5Cnhttp://stacks.iop.org/0004-637X/821/i=1/a=59?key=crossref.c2c198698326a5a2dd4869a8e400d342.

Field, George B. (1959). "The Spin Temperature of Intergalactic Neutral Hydrogen". In: *The Astrophysical Journal* 129, p. 536.

Furlanetto, Steven, S. Peng Oh, and Frank Briggs (2006). "Cosmology at Low Frequencies: The 21 cm Transition and the High-Redshift Universe". In: February 2008, p. 207. DOI: 10.1016/j.physrep.2006.08.002. arXiv: 0608032 [astro-ph]. URL: http://arxiv.org/abs/astro-ph/0608032.

Furlanetto, Steven R. and Abraham Loeb (2002). "The 21 Centimeter Forest: Radio Absorption Spectra as Probes of Minihalos before Reionization". In: *The Astrophysical Journal* 579.1, pp. 1–9. ISSN: 0004-637X. DOI: 10.1086/342757. URL: http://stacks.iop.org/0004-637X/579/i=1/a=1.

Gawiser, Eric and Joseph Silk (1998). "Extracting Primordial Density Fluctuations". In: *Science* 280, p. 1405. ISSN: 00368075. DOI: 10.1126/science.280.5368.1405. arXiv: 9806197 [arXiv:astro-ph].

Glasse, Alistair et al. (2015). "The Mid-Infrared Instrument for the James Webb Space Telescope , IX: Predicted Sensitivity". In: *Publications of the Astronomical Society of the Pacific* 127.953, pp. 686–695. ISSN: 00046280. DOI: 10.1086/682259. arXiv: 1508.02427. URL: http://arxiv.org/abs/1508.02427.

Gluscevic, Vera and Rennan Barkana (2010). "Statistics of 21-cm fluctuations in cosmic reionization simulations: PDFs and difference PDFs". In: *Monthly Notices of the Royal Astronomical Society* 408.4, pp. 2373–2380. ISSN: 00358711. DOI: 10.1111/j.1365-2966.2010.17293.x. arXiv: 1005.3814. URL: http://arxiv.org/abs/1005.3814.

Gordon, Christopher and Roberto Trotta (2007). "Bayesian calibrated significance levels applied to the spectral tilt and hemispherical asymmetry". In: *Monthly Notices of the Royal Astronomical Society* 382.4, pp. 1859–1863. ISSN: 00358711. DOI: 10.1111/j.1365-2966.2007.12707.x. URL: http://mnras.oxfordjournals.org/cgi/content/long/382/4/1859.

Greig, B, A Mesinger, and L V E Koopmans (2015). "Optimal core baseline design and observing strategy for probing the astrophysics of reionization with the SKA". In: *eprint arXiv* 1509, p. 3312. arXiv: 1509.03312. URL: http://adsabs.harvard.edu/cgi-bin/nph-data{_}query?bibcode=2015arXiv150903312G{\&

}link{_}type=ABSTRACT{\%}5Cnpapers://ee00755c-a478-4d4e-a50b-ef01ee7b9957/Paper/p21199.

Gunn, James E. and Bruce A. Peterson (1965). "On the Density of Neutral Hydrogen in Intergalactic Space." In: *The Astrophysical Journal* 142, p. 1633. ISSN: 0004-637X. DOI: 10.1086/148444. URL: http://scholar.google.com/scholar?hl=en{\&}btnG=Search{\&}q=intitle:No+Title{\#}0http://adsabs.harvard.edu/doi/10.1086/148444http://articles.adsabs.harvard.edu/full/1965ApJ...142.1633G.

Haarlem, M. P. van et al. (2013). "LOFAR: The LOw-Frequency ARray". In: *Astronomy & Astrophysics* 556, A2. ISSN: 0004-6361. DOI: 10.1051/0004-6361/201220873. arXiv: 1305.3550. URL: http://arxiv.org/abs/1305.3550.

Handley, W. J., M. P. Hobson, and A. N. Lasenby (2015). "POLYCHORD: Nested sampling for cosmology". In: *Monthly Notices of the Royal Astronomical Society: Letters* 450.1, pp. L61–L65. ISSN: 17453933. DOI: 10.1093/mnrasl/slv047. arXiv: 1502.01856.

Hastings, W. K. (1970). "Monte carlo sampling methods using Markov chains and their applications". In: *Biometrika* 57.1, pp. 97–109.

Hinshaw, G., Larson, D., et. al (2013). "Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: cosmological parameter results". In: *The Astrophysical Journal Supplement Series*. URL: http://iopscience.iop.org/0067-0049/208/2/19.

Hu, Wayne (2003). *CMB temperature and polarization anisotropy fundamentals*. DOI: 10.1016/S0003-4916(02)00022-2. arXiv: 0210696 [astro-ph].

Iliev, Ilian T. et al. (2015). "Epoch of Reionization modelling and simulations for SKA". In: *arXiv preprint arXiv:1501.04213*. arXiv: 1501.04213. URL: http://arxiv.org/abs/1501.04213.

Jarvis, Matt J and Steve Rawlings (2000). "On the redshift cut-off for flat-spectrum radio sources". In: *Mon. Not. R. Astron. Soc* 319, pp. 121–136. ISSN: 00358711. DOI: 10.1046/j.1365-8711.2000.03801.x. arXiv: 0006081v1 [arXiv:astro-ph].

Jarvis, Matt J. et al. (2009). *The discovery of a typical radio galaxy at z = 4.88*. DOI: 10.1111/j.1745-3933.2009.00715.x. arXiv: 0907.1447.

Jeffreys, Harold (1961). *Theory of Probability*. 3rd. Vol. 2, p. 432. ISBN: 0-19-850368-7. URL: http://ocw.mit.edu/OcwWeb/Mathematics/18-175Spring-2007/LectureNotes/Index.htm.

Jiang, Linhua et al. (2006). "A Spectroscopic Survey of Faint Quasars in the SDSS Deep Stripe. I. Preliminary Results from the Co-added Catalog". In: *The Astronomical Journal* 131.6, p. 2788. URL: http://stacks.iop.org/1538-3881/131/i=6/a=2788.

Jiang, Linhua et al. (2007). "The Radio-Loud Fraction of Quasars is a Strong Function of Redshift and Optical Luminosity". In: *The Astrophysical Journal* 656.2, p. 680. URL: http://stacks.iop.org/0004-637X/656/i=2/a=680.

Johnston, S, I J Feain, and N Gupta (2009). "Science with the Australian Square Kilometre Array Pathfinder (ASKAP)". In: *The Low-Frequency Radio Universe ASP Conference Series* 407, p. 446. DOI: 10.1071/AS07033. arXiv: 0903.4011. URL: http://adsabs.harvard.edu/cgi-bin/nph-data{_}query?bibcode=2009ASPC..407..446J{\&}link{_}type=ABSTRACT.

Klamer, Ilana J. et al. (2006). "A search for distant radio galaxies from SUMSS and NVSS - III. Radio spectral energy distributions and the z-?? correlation". In: *Monthly Notices of the Royal Astronomical Society* 371.2, pp. 852–866. ISSN: 00358711. DOI: 10.1111/j.1365-2966.2006.10714.x. arXiv: 0606469 [astro-ph].

Kolb, E. W. and M. S. Turner (1990). "The early universe." In: *Front. Phys.*

Komatsu, Eiichiro et al. (2009). "Five-Year Wilkinson Microwave Anisotropy Probe Observations: Cosmological Interpretation". In: *The Astrophysical Journal Supplement Series* 180.2, pp. 330–376. ISSN: 0067-0049. DOI: 10.1088/0067-0049/180/2/330. arXiv: 0803.0547v2. URL: http://stacks.iop.org/0067-0049/180/i=2/a=330?key=crossref.355ec98555761e3f8a6a38d88962d16f.

Labbe, I et al. (2010). "Ultradeep Infrared Array Camera Observations of Sub-L* z ~ 7 and z ~ 8 Galaxies in the Hubble Ultra Deep Field: the Contribution of Low-Luminosity Galaxies to the Stellar Mass Density and Reionization". In: *The Astrophysical Journal Letters* 708.1, pp. L26–L31. ISSN: 2041-8205. DOI: 10.1088/2041-8205/708/1/L26. arXiv: 0910.0838. URL: http://stacks.iop.org/2041-8205/708/i=1/a=L26?key=crossref.61bcd581f2b3aa28d0f45e71dc2c77c6htt
//adsabs.harvard.edu/cgi-bin/nph-data{_}query?bibcode=2010ApJ...708L..26L{\&}link{_}type=ABSTRACT{\%}5Cnpapers2://publication/doi/10.1088/2041-8205/708/1/L26.

Laing, R. A., J. M. Riley, and M. S. Longair (1983). "Bright radio sources at 178 MHz: flux densities, optical identifications and the cosmological evolution of powerful radio galaxies". In: *Monthly Notices of the Royal Astronomical Society* 204.1, pp. 151–187. ISSN: 0035-8711. DOI: 10.1093/mnras/204.1.151. URL: http://articles.adsabs.harvard.edu/cgi-bin/nph-iarticle{_}query?1983MNRAS.204..151L{\&}data{_}type=PDF{_}HIGH{\&}whole{_}paper=YES{\&}type=PRINTER{\&}filetype=.pdf{\%}5Cnhttp://mnras.oxfordjournals.org/cgi/doi/10.1093/mnras/204.1.151.

Larson, D. et al. (2010). "Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Power Spectra and WMAP-Derived Parameters". In: p. 22. DOI: 10.

1088/0067-0049/192/2/16. arXiv: 1001.4635. URL: http://arxiv.org/abs/1001.4635.

Linde, Andrei D (1982). "A new inflationary universe scenario: a possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems". In: *Physics Letters B* 6, pp. 389–393.

Liu, Adrian et al. (2012). "Global 21cm signal experiments: A designer's guide". In: *Physical Review D*, p. 34. DOI: 10.1103/PhysRevD.87.043002. arXiv: 1211.3743. URL: http://arxiv.org/abs/1211.3743http://journals.aps.org/prd/abstract/10.1103/PhysRevD.87.043002.

Loredo, T J (1992). "Promise of Bayesian inference for astrophysics." In: *Conference on Statistical Challenges in Modern Astronomy*, p. 275. URL: http://adsabs.harvard.edu/cgi-bin/nph-data{_}query?bibcode=1992scma.conf..275L{\&}link{_}type=ABSTRACT{\%}5Cnpapers2://publication/uuid/8B41D9B4-CC77-4159-9664-B9A0E4820F18.

Mack, Katherine J. and J. Stuart B. Wyithe (2012a). "Detecting the redshifted 21 cm forest during reionization". In: *Monthly Notices of the Royal Astronomical Society* 425.4, pp. 2988–3001. ISSN: 00358711. DOI: 10.1111/j.1365-2966.2012.21561.x. URL: http://mnras.oxfordjournals.org/cgi/content/long/425/4/2988.

— (2012b). "Detecting the redshifted 21 cm forest during reionization". In: *Monthly Notices of the Royal Astronomical Society* 425.4, pp. 2988–3001. ISSN: 00358711. DOI: 10.1111/j.1365-2966.2012.21561.x. arXiv: 1101.5431.

Madau, Piero and Mark Dickinson (2014). "Cosmic Star Formation History". In: *Annual Review of Astronomy and Astrophysics* 52, pp. 415–486. ISSN: 0066-4146. DOI: 10.1146/annurev-astro-081811-125615. arXiv: 1403.0007. URL: http://arxiv.org/abs/1403.0007v3{\%}5Cnhttp://www.annualreviews.org/doi/abs/10.1146/annurev-astro-081811-125615.

Majumdar, Suman et al. (2014). "On the use of seminumerical simulations in predicting the 21-cm signal from the epoch of reionization". In: *Monthly Notices of the Royal Astronomical Society* 443.4, pp. 2843–2861. ISSN: 13652966. DOI: 10.1093/mnras/stu1342. arXiv: 1403.0941.

McCarthy, Patrick J., et al (1996). "The molonglo reference catalog/1 jansky radio source survey. i. radio galaxy identifications". In: *The Astrophysical Journal Supplement Series* 107, p. 19.

McGreer, Ian D., Andrei Mesinger, and Xiaohui Fan (2011). "The first (nearly) model-independent constraint on the neutral hydrogen fraction at". In: *Monthly Notices of the Royal Astronomical Society* 415.4, pp. 3237–3246. ISSN: 00358711. DOI: 10.1111/j.1365-2966.2011.18935.x. arXiv: 1101.3314.

McLure, Ross J. et al. (2004). "A sample of radio galaxies spanning three decades in radio luminosity -I. The host galaxy properties and black hole masses". In: *Monthly Notices of the Royal Astronomical Society* 351.1, pp. 347–361. ISSN: 00358711. DOI: 10.1111/j.1365-2966.2004.07793.x. arXiv: 0403106v1 [arXiv:astro-ph].

Mellema, Garrelt et al. (2012). "Reionization and the Cosmic Dawn with the Square Kilometre Array". In: arXiv: 1210.0197. URL: http://arxiv.org/abs/1210.0197.

Mesinger, Andrei and Steven Furlanetto (2007). "Efficient Simulations of Early Structure Formation and Reionization". In: *The Astrophysical Journal* 669.2, pp. 663–675. URL: http://stacks.iop.org/0004-637X/669/i=2/a=663{\%}5Cnpapers2://publication/doi/10.1086/521806.

Mesinger, Andrei et al. (2014). "Constraining the astrophysics of the cosmic dawn and the epoch of reionization with the SKA". In: *Proceedings of Science*. Vol. 9-13-June-. arXiv: 1501.04106.

Metropolis, Nicholas et al. (1953). "Equation of state calculations by fast computing machines". In: *Journal Chemical Physics* 21.6, pp. 1087–1092. URL: http://jcp.aip.org/resource/1/jcpsa6/v21/i6/p1087{_}s1?bypassSSO=1.

Miley, George and Carlos De Breuck (2008). "Distant radio galaxies and their environments". In: *Astronomy and Astrophysics Review*. ISSN: 09354956. DOI: 10.1007/s00159-007-0008-z. arXiv: 0802.2770.

Miley, George K. (1968). "Variation of the Angular Sizes of Quasars with Red-shift." In: *Nature* 5145.218, pp. 933–934.

Momjian, E. et al. (2014). "THE HIGHEST REDSHIFT QUASAR AT z = 7.085: A RADIO-QUIET SOURCE". In: *The Astronomical Journal* 147.1, p. 6. ISSN: 0004-6256. DOI: 10.1088/0004-6256/147/1/6. URL: http://stacks.iop.org/1538-3881/147/i=1/a=6?key=crossref.76b111e2246258a22cd1ea719fb18e6e.

Mortlock, Daniel J. et al. (2011). "A luminous quasar at a redshift of z = 7.085". In: *Nature* 474, pp. 616–619. URL: http://adsabs.harvard.edu/abs/2011Natur.474..616M.

Ocvirk, Pierre et al. (2016). "Cosmic Dawn (CoDa): the first radiation-hydrodynamics simulation of reionization and galaxy formation in the Local Universe". In: *Monthly Notices of the Royal Astronomical Society* 463.2, pp. 1462–1485. ISSN: 0035-8711. DOI: 10.1093/mnras/stw2036. arXiv: 1511.00011. URL: http://arxiv.org/abs/1511.00011.

Ouchi, Masami et al. (2010). "Statistics of 207 Lya Emitters at a Redshift Near 7: Constraints on Reionization and Galaxy Formation Models". In: DOI: 10.1088/0004-637X/723/1/869. arXiv: 1007.2961. URL: http://arxiv.org/abs/1007.2961http://dx.doi.org/10.1088/0004-637X/723/1/869.

Parsons, Aaron R. et al. (2010). "THE PRECISION ARRAY FOR PROBING THE EPOCH OF RE-IONIZATION: EIGHT STATION RESULTS". In: *The Astronomical Journal* 139.4, pp. 1468–1480. ISSN: 0004-6256. DOI: 10.1088/0004-6256/139/4/1468. arXiv: 0904.2334. URL: http://arxiv.org/abs/0904.2334.

Parsons, Aaron R. et al. (2014). "NEW LIMITS ON 21 cm EPOCH OF REIONIZATION FROM PAPER-32 CONSISTENT WITH AN X-RAY HEATED INTERGALACTIC MEDIUM AT z = 7.7". In: *The Astrophysical Journal* 788.2, p. 106. ISSN: 0004-637X. DOI: 10.1088/0004-637X/788/2/106. URL: http://stacks.iop.org/0004-637X/788/i=2/a=106?key=crossref.73a9f72b58541a467f748b648de854d4.

Paxson, Frederic L. and William Roscoe Thayer (1920). *Theodore Roosevelt, An Intimate Biography*. Vol. 25. 2. Grosset \& Dunlap, p. 306. DOI: 10.2307/1835384. URL: http://www.jstor.org/stable/10.2307/1835384?origin=crossref.

Pentericci, L. et al. (2014). "NEW OBSERVATIONS OF z = 7 GALAXIES: EVIDENCE FOR A PATCHY REIONIZATION". In: *The Astrophysical Journal* 793.2, p. 113. URL: http://adsabs.harvard.edu/abs/2014ApJ...793..113P.

Penzias, A. A. and R. W. Wilson (1965). "A Measurement of Excess Antenna Temperature at 4080 Mc/s." In: *The Astrophysical Journal* 142, p. 419. ISSN: 0004-637X. DOI: 10.1086/148307. URL: http://adsabs.harvard.edu/doi/10.1086/148307.

Planck Collaboration et al. (2015). "Planck 2015 results. XIII. Cosmological parameters". In: *arXiv*, p. 1502.01589. ISSN: 0717-6163. DOI: 10.1007/s13398-014-0173-7.2. arXiv: 1502.01589v2. URL: http://adsabs.harvard.edu/abs/2015arXiv150201589P.

Pritchard, Jonathan R and Abraham Loeb (2012). "21 cm cosmology in the 21st century." In: *Reports on progress in physics. Physical Society (Great Britain)* 75.8, p. 086901. ISSN: 1361-6633. DOI: 10.1088/0034-4885/75/8/086901. arXiv: 1109.6012. URL: http://arxiv.org/abs/1109.6012.

Pritchard, Jonathan R., Abraham Loeb, and J. Stuart B. Wyithe (2010). "Constraining reionization using 21-cm observations in combination with CMB and Lyα forest data". In: *Monthly Notices of the Royal Astronomical Society* 408.1, pp. 57–70. ISSN: 00358711. DOI: 10.1111/j.1365-2966.2010.17150.x. arXiv: 0908.3891. URL: http://arxiv.org/abs/0908.3891.

Rawlings, Steve (2000). "Cosmological Studies from Radio Source Samples". In: arXiv: 0008067 [astro-ph]. URL: http://arxiv.org/abs/astro-ph/0008067.

Reddy, Naveen A. and Charles C. Steidel (2009). "A STEEP FAINT-END SLOPE OF THE UV LUMINOSITY FUNCTION AT z ~ 2-3: IMPLICATIONS FOR THE GLOBAL STELLAR MASS DENSITY AND STAR FORMATION IN LOW-MASS HALOS". In: *The Astrophysical Journal* 692.1, pp. 778–803. ISSN: 0004-637X. DOI: 10.1088/0004-

637X/692/1/778. arXiv: 0810.2788. URL: http://stacks.iop.org/0004-637X/692/i=1/a=778?key=crossref.3bf516c819c70c4bc843ec21195c97b3.

Robertson, Brant E. et al. (2013a). "NEW CONSTRAINTS ON COSMIC REIONIZATION FROM THE 2012 HUBBLE ULTRA DEEP FIELD CAMPAIGN". In: *The Astrophysical Journal* 768.1, p. 71. ISSN: 0004-637X. DOI: 10.1088/0004-637X/768/1/71. URL: http://stacks.iop.org/0004-637X/768/i=1/a=71?key=crossref.f9e857aa25c7114e4e0d5fc6e241c7df.

— (2013b). "NEW CONSTRAINTS ON COSMIC REIONIZATION FROM THE 2012 HUBBLE ULTRA DEEP FIELD CAMPAIGN". In: *The Astrophysical Journal* 768.1, p. 71. ISSN: 0004-637X. DOI: 10.1088/0004-637X/768/1/71. URL: http://stacks.iop.org/0004-637X/768/i=1/a=71?key=crossref.f9e857aa25c7114e4e0d5fc6e241c7df.

Robertson, Brant E. et al. (2015). "Cosmic Reionization and Early Star-Forming Galaxies: a Joint Analysis of New Constraints From Planck and the Hubble Space Telescope". In: *Apj* 802.2, p. L19. ISSN: 2041-8213. DOI: 10.1088/2041-8205/802/2/L19. arXiv: 1502.02024. URL: http://arxiv.org/abs/1502.02024.

Rybicki, George B. and Alan P. Lightman (2008). *Radiative processes in astrophysics*. John Wiley & Sons, pp. 9–12. URL: http://www.bartol.udel.edu/{~}owocki/phys633/RadProc-RybLightman-ch1.pdf.

Santos, M. G. et al. (2009). "Fast Large Volume Simulations of the 21 cm Signal from the Reionization and pre-Reionization Epochs". In: 000.June, p. 14. DOI: 10.1111/j.1365-2966.2010.16898.x. arXiv: 0911.2219. URL: http://arxiv.org/abs/0911.2219.

Santos, Mario G et al. (2008). "Cosmic Reionization and the 21 cm Signal: Comparison between an Analytical Model and a Simulation - Abstract - The Astrophysical Journal - IOPscience". In: *The Astrophysical Journal* 689.1, pp. 1–16. arXiv: arXiv:0708.2424v2. URL: http://stacks.iop.org/0004-637X/689/i=1/a=1{\%}5Cnpapers2://publication/doi/10.1086/592487.

Sheth, R. K. and G. Tormen (1999). "Large-scale bias and the peak background split". In: *Monthly Notices of the Royal Astronomical Society*. ISSN: 0035-8711. DOI: 10.1046/j.1365-8711.1999.02692.x. arXiv: 9901122 [astro-ph].

Skilling, John (2004). "Nested Sampling". In: *AIP Conference Proceedings* 735, pp. 395–405. ISSN: 0094243X. DOI: 10.1063/1.1835238. URL: http://link.aip.org/link/?APC/735/395/1{\&}Agg=doi{\%}5Cnpapers2://publication/doi/10.1063/1.1835238.

So, Geoffrey C. et al. (2014). "Fully Coupled Simulation of Cosmic Reionization. II. Recombinations, Clumping Factors, and the Photon Budget for Reionization". In: *The Astrophysical Journal* 789, p. 149. ISSN: 0004-637X. DOI: 10.1088/0004-637X/

789/2/149. arXiv: arXiv:1311.2152v1. URL: http://adsabs.harvard.edu/abs/2014ApJ...789..149S{\%}5Cnhttp://adsabs.harvard.edu/cgi-bin/nph-data{_}query?bibcode=2014ApJ...789..149S{\&}link{_}type=ARTICLE.

Sobolev, Viktor Viktorovich (1960). *Moving envelopes of stars*. URL: http://adsabs.harvard.edu/abs/1960mes..book.....Sï¿¡{\"{U}}.

Springel, Volker, C Frenk, and S White (2006). "The large-scale structure of the Universe". In: *Nature*, pp. 1–34. ISSN: 0028-0836. DOI: 10.1038/nature04805. arXiv: 0604561 [astro-ph]. URL: http://www.nature.com/nature/journal/v440/n7088/abs/nature04805.html.

Tingay, S J et al. (2013). "The Murchison Widefield Array: The Square Kilometre Array Precursor at Low Radio Frequencies". In: *Publications of the Astronomical Society of Australia* 30, p. 7. URL: http://adsabs.harvard.edu/cgi-bin/nph-data{_}query?bibcode=2013PASA...30....7T{\&}link{_}type=ABSTRACT{\%}5Cnpapers2://publication/doi/10.1017/pasa.2012.007.

Trotta, Roberto (2008). "Bayes in the sky: Bayesian inference and model selection in cosmology". In: *Contemporary Physics* 49.2, pp. 71–104. ISSN: 0010-7514. DOI: 10.1080/00107510802066753. arXiv: 0803.4089. URL: http://arxiv.org/abs/0803.4089.

Van de Hulst, HC (1945). "The Origin of Radio Waves from Space". In: *Ned. Tijdschr. Natuurk* 11, pp. 201–21.

Weinberg, S (2008). *Cosmology*, p. 320. ISBN: 978-0-19-852682-7.

White, Richard L. et al. (2003). "Probing the Ionization State of the Universe at z >6". In: *The Astronomical Journal* 126.1, pp. 1–14. ISSN: 0004-6256. DOI: 10.1086/375547. URL: http://stacks.iop.org/1538-3881/126/i=1/a=1.

Willott, Chris J. et al. (2010). "The Canada-France High-z Quasar Survey: nine new quasars and the luminosity function at redshift 6". In: *The Astronomical Journal* 139.3, pp. 906–918. ISSN: 0004-6256. DOI: 10.1088/0004-6256/139/3/906. arXiv: 0912.0281. URL: http://arxiv.org/abs/0912.0281.

Wilman, R. J. et al. (2008). "A semi-empirical simulation of the extragalactic radio continuum sky for next generation radio telescopes". In: *Monthly Notices of the Royal Astronomical Society* 388.3, pp. 1335–1348. ISSN: 00358711. DOI: 10.1111/j.1365-2966.2008.13486.x. URL: http://mnras.oxfordjournals.org/cgi/content/long/388/3/1335.

Wouthuysen, S. A (1952). "On the excitation mechanism of the 21-cm (radio-frequency) interstellar hydrogen emission line". In: *The Astronomical Journal* 57, pp. 31–32.

Xu, Yidong et al. (2009). "THE 21 cm FOREST AS A PROBE OF THE REIONIZATION AND THE TEMPERATURE OF THE INTERGALACTIC MEDIUM". In: *The Astrophysical Journal* 704.2, pp. 1396–1404. ISSN: 0004-637X. DOI: 10.1088/0004-637X/704/2/1396. URL: http://stacks.iop.org/0004-637X/704/i=2/a=1396?key=crossref.8927f385d9a19ccf92f62b09867397cb.

Yoshisato, Ayako et al. (2005). "Why is the Zel'dovich Approximation so Accurate?" In: pp. 1–16. arXiv: 0510107 [astro-ph]. URL: http://arxiv.org/abs/astro-ph/0510107v2.

Zahn, Oliver et al. (2007). "Simulations and Analytic Calculations of Bubble Growth during Hydrogen Reionization". In: *The Astrophysical Journal* 654.1, pp. 12–26. ISSN: 0004-637X. DOI: 10.1086/509597. arXiv: 0604177 [astro-ph]. URL: http://arxiv.org/abs/astro-ph/0604177{\%}5Cnhttp://stacks.iop.org/0004-637X/654/i=1/a=12.

Zaroubi, Saleem (2012). "The Epoch of Reionization". In: *ArXiv e-prints*, pp. 1–59. arXiv: 1206.0267. URL: http://arxiv.org/abs/1206.0267.

Zel'Dovich, YA B (1970). "Gravitational instability: An approximate theory for large density perturbations." In: *Astronomy and astrophysics* 5, pp. 84–89.

Zhou, Jie et al. (2013). "Semi-numerical simulation of reionization with semi-analytical modeling of galaxy formation". In: *Research in Astronomy and Astrophysics* 13.4, pp. 373–386. ISSN: 1674-4527. DOI: 10.1088/1674-4527/13/4/001. arXiv: 1212.6099. URL: http://arxiv.org/abs/1212.6099.