Developing Statistical Inference Tools for Future Observations of the Cosmic Dawn and the Epoch of Reionization

M.Sc. Thesis

By

Himanshu Tiwari

Discipline of Astronomy, Astrophysics and Space Engineering

INDIAN INSTITUTE OF TECHNOLOGY INDORE

June, 2020

Developing Statistical Inference Tools for Future Observations of the Cosmic Dawn and the Epoch of Reionization

M.Sc. Thesis

Submitted in partial fulfillment of the
requirements for the awards of the degree
of
Master of Science

by

Himanshu Tiwari

Discipline of Astronomy, Astrophysics and Space Engineering

INDIAN INSTITUTE OF TECHNOLOGY INDORE

June, 2020

INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **DEVELOPING STATISTICAL INFERENCE TOOLS FOR FUTURE OBSERVATIONS OF THE COSMIC DAWN AND THE EPOCH OF REIONIZATION** in the partial fulfillment of the requirements for the award of the degree of **MASTER OF SCIENCE** and submitted in the **DISCIPLINE OF ASTRONOMY, ASTROPHYSICS AND SPACE ENGINEERING, Indian Institute of Technology Indore**, is an authentic record of my own work carried out during the time period from **JULY, 2019** to **JUNE, 2020** under the supervision of **Assistant Professor, Dr. Suman Majumdar.**

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

21/06/20

Signature of the student with date
Himanshu Tiwari

This is to certify that the above statement made by the candidate is correct to the best of my/our knowledge.

Suman Majumdar
21/06/2020

Signature of the Supervisor of M.Sc. thesis
Dr. Suman Majumdar

Himanshu Tiwari has successfully given his M.Sc. Oral Examination held on **12/06/2020.**

Suman Majumdar
Signature of the Supervisor of M.Sc. thesis
Dr. Suman Majumdar
Date: 21/06/2020

Suman Majumdar
Dr. Suman Majumdar
Convener, DPGC
Date: 22/06/2020

Subhendu Rakshit
Signature of PSPC Member #1
Prof. Subhendu Rakshit

Date: 21/06/2020

S. Das
Signature of PSPC Member #2
Dr. Saurabh Das

Date: 21/06/2020

ACKNOWLEDGEMENTS

I would like to thank DAASE PhD students **Madhurima Chaudury** and **Mohd Kamran** for helping me on various aspects of my project. I would also like to express my sincere gratitude to my supervisor **Dr. Suman Majumdar** for his constant guidance, invaluable feedback, motivation and support in every step of this project.

Abstract

The epoch of reionization (EoR) is one of the least known periods in the history of the Universe. The next-generation telescopes (e.g. the SKA, HERA) will have the capability to directly probe the distribution of neutral hydrogen from this era through the redshifted 21-cm line. Once such observations are successful in detecting the 21-cm signal from the EoR, one would then aim to constrain the astrophysical parameters of the EoR accurately by doing statistical analysis of the observed data. One way of achieving this is through Bayesian inference of various signal statistics. To draw inference in a Bayesian framework, one would need to compare the observed signal statistic with a model statistic while performing random walks in the multi-dimensional astrophysical parameter space. In case of the EoR 21-cm signal, this implies simulating the signal in a large cosmological volume with reasonably high precision for each step of the random walker in this parameter space. The conventional simulations of the EoR (radiative transfer or semi-numerical) takes a large amount of computer memory ($\sim$ 1 TB in RAM) and a significant amount of computation time to simulate the signal in a reasonable volume for one such set of parameters (i.e. one step of the random walker). The requirement of computing time will be proportionately higher when one has to take of the order of millions of random steps in the parameter space. To circumvent this problem, we have developed an artificial neural network (ANN) based EoR 21-cm signal statistics (presently for power spectrum and bispectrum) emulator **EmuPBk**[1][2], which is orders of magnitude faster than a semi-numerical simulation. The EmuPBk was trained over 1000 semi-numerically simulated (using Region-Yuga code) EoR 21-cm power spectra and bispectra. Our tests show that it has a reasonable capability of predicting the EoR 21-cm power spectra and bispectra. Further, using these emulated power spectra and bispectra of the signal, our MCMC analysis based Bayesian inference shows that one will be able to put tighter constraints on the reionization parameters using the bispectra compared to the power spectra. This is due to the fact that the 21-cm signal from the EoR is highly non-Gaussian and power spectra do not capture this non-Gaussianity. However, the bispectra is sensitive to such non-Gaussianities which are dependent on the time evolving topology of the signal.

[1] https://github.com/himmng/EmuPBk
[2] https://emupbk.readthedocs.io/en/latest/

Contents

List of Figures i

List of Tables iii

Acronyms iii

1 Introduction 1

2 21-cm transition in the context of EoR 6

2.1 H I Transition Line . . . 7

2.2 21-cm Brightness Temperature . . . 9

2.3 Statistical Observables of the 21-cm signal form the EoR . . . 11

2.3.1 Power Spectrum . . . 11

2.3.2 Bispectrum . . . 12

2.4 Probing EoR using Radio Telescopes . . . 13

2.5 EoR model Parameters . . . 14

2.5.1 Ionizing Efficiency (ζ) . . . 14

2.5.2 Mean Free Path of ionizing photons (R_{mfp}) . . . 14

2.5.3 Minimum Halo mass (Mh_{min}) . . . 14

2.6 Simulations of the EoR . . . 15

2.6.0.1 Radiative Transfer simulations . . . 15

2.6.0.2 Semi-numerical Simulations . . . 15

3 Methodology 17

3.1 Bayesian Inference . . . 19

3.2 Emulating the EoR 21-cm Statistics . . . 23

3.2.1 Structure of an ANN . . . 24

4 Results and Discussion 27

4.1 ANN Emulators . . . 27

4.1.1 21-cm Powerspectrum Emulator 27
4.1.1.1 Accuracy and Loss 28
4.1.2 21-cm Bispectrum Emulator 30
4.1.2.1 Bispectrum for $k_1 = 0.3\,\mathrm{Mpc}^{-1}$ 31
4.1.2.2 Accuracy, Loss and Predictions for Bispectrum with $k_1 = 0.3\,\mathrm{Mpc}^{-1}$ 32
4.1.2.3 Bispectrum for $k_1 = 1.5\,\mathrm{Mpc}^{-1}$ 32
4.1.2.4 Accuracy, Loss and Predictions for Bispectrum with $k_1 = 1.5\,\mathrm{Mpc}^{-1}$ 34
4.2 Parameter Estimation 34
4.2.1 Impact of EoR parameters on 21-cm statistics 34
4.2.2 Bayesian inference: parameter estimation using individual statistics 38
4.2.3 Joint Posterior Plots 40
4.3 Summary 54

5 Future Scopes **56**

List of Figures

2.1 Hydrogen Hyper-fine Transition 9
2.2 CMBR in the background for the intervening H I cloud 11

3.1 This flowchart shows the basic building blocks and workings of our statistical inference pipelines. 18
3.2 Working of CosmoHammer **Image credit: Cosmopic.** 21
3.3 A random walker sampling a 2D parameter space. 22
3.4 Basic multi-layered ANN structure. 24
3.5 Data-set used for training (upper-triangular region) and testing (lower-triangular region). 25
3.6 The working of a supervised ANN. 26

4.1 Powerspectrum simulation (points) and emulation (line). 28
4.2 Training Accuracy and Loss for the PS emulator. 29
4.3 PS test predictions from ANN for 18 test sets. 30
4.4 Bispectrum 31
4.5 Training Accuracy and Loss for bispectrum with $k_1 = 0.3\,\mathrm{Mpc}^{-1}$ 32
4.6 Test predictions using ANN emulator for bispectrum with $k_1 = 0.3\,\mathrm{Mpc}^{-1}$ 33
4.7 Training Accuracy and Loss in the bispectrum emulation for $k_1 = 1.5\,\mathrm{Mpc}^{-1}$ 34
4.8 Test predictions using ANN emulator for bispectrum with $k_1 = 1.5\,\mathrm{Mpc}^{-1}$ 35
4.9 Powerspectrum for Case-8. 36
4.10 Powerspectrum for Case-9 37
4.11 Parameter estimation using individual statistics : P(k) (grey), B(k_1=0.3) (yellow), B(k_1=1.5) (blue) for Case-1 (ζ = 20, R_{mfp} = 20 Mpc, Mh_{min} = 500). 38
4.12 Parameter estimation using individual statistics : P(k) (grey), B(k_1 = 0.3) (yellow), B(k_1 = 1.5) (blue) for Case-2 (ζ = 18, R_{mfp} = 45 Mpc, Mh_{min} = 100). 39
4.13 Joint Plot-1, true values (20, 20, 500) 41
4.14 Joint Plot-2, true values (18, 45, 100) 42
4.15 Joint Plot-3, true values (130, 30, 100) 43
4.16 Joint Plot-4, true values (140, 30, 100) 44

4.17 Joint Plot-5, true values (170, 30, 800) 45
4.18 Joint Plot-6, true values (45, 60, 700) 46
4.19 Joint Plot-7, true values (40, 30, 500) 47
4.20 Joint Plot-8, true values (30, 10, 10) 48
4.21 Joint Plot-9, true values (23.21, 20, 1000) 49
4.22 Joint Plot-10, true values (20, 30, 800) 50
4.23 Joint Plot-11, true values (130, 45, 200) 51
4.24 Joint Plot-12, true values (100, 10, 100) 52
4.25 Joint Plot-13, true values (55, 10, 100) 53
4.26 Joint Plot-14, true values (50, 60, 800) 54

List of Tables

4.1 This shows the 15 test sets of parameters for which we have tested our Bayesian inference pipeline. 36
4.2 Case-1 42
4.3 Case-2 42
4.4 Case-3 43
4.5 Case-4 44
4.6 Case-5 45
4.7 Case-6 46
4.8 Case-7 47
4.9 Case-8 48
4.10 Case-9 49
4.11 Case-10 50
4.12 Case-11 51
4.13 Case-12 52
4.14 Case-13 53
4.15 Case-14 54

Chapter 1

Introduction

The Universe has supposedly emerged out from the Big-bang and it is continuously evolving since then. In its very primitive stage, it was in the form of a plasma (soup of electrons, protons and radiation) which remained so till the expansion cooled the Universe enough to allow the very first atoms to be formed by the combination of these electrons and protons (Barkana and Loeb 2001; S. R. Furlanetto, Peng Oh, and Briggs 2006). It made the Universe neutral for the first time and also transparent to the radiation. This radiation can still be observed as the relic Cosmic Microwave Background Radiation (CMBR). This dissociation between matter and radiation led the Universe into a dark phase, when there were no sources of light were present. The observations of CMBR have successfully quantified the imprints of the very early perturbations in the matter density which further grew during this dark era and finally led to the bound structures that we see around us today (Jonathan R Pritchard and Loeb 2012; Choudhury 2003). The Cosmic Dawn (CD) was the time when these first structures (first stars and galaxies) emerged out from the gravitationally bound structures in the matter. These were the first sources of light which produced huge amount of X-rays and ionizing UV photons, and started heating and ionizing the neutral hydrogen (H I) at their vicinity and then gradually the whole inter-galactic medium (IGM) leading to the Epoch of Reionization (EoR). Due to the complex heating and ionization processes, the astrophysics during this era is more complicated than the cosmology. Thus understanding these astrophysical processes through a parametric model is key to answer many of the unresolved questions during this period. Some of the important unanswered questions from this era are – how did the Universe got heated and ionized during this era, what were the typical sources that produced X-ray and UV photons during this period, how long did this period last etc.

Although, we have a good understanding of the phase of the universe when matter-radiation decoupled (i.e. the time when CMBR was formed), we do not have much information about the periods afterwards, specifically the cosmic dawn and the epoch of reionization. A thorough observation of these phases will help us to better understand the process of structure formation. However, observing the first sources of lights from this era is a challenge. As most of the IGM was neutral at this time thus most of radiation from these sources were absorbed on their way before reaching even us. As an alternative technique, one could try to observe the IGM during this period, which was mostly hydrogen and helium, through the redshifted 21-cm signal originating due to the spin-flip transition in H I. However, observing this particular signal from the CD-EoR requires highly sensitive radio telescopes as the signal is very faint compared to the other radiations coming from the our own galaxy and extra glactic sources in the same frequency range. Detection of this signal is thus very difficult because of the presence of these radio foreground emission, man made radio frequency interference (RFI), and system noise of radio telescopes. Overcoming these obstacles require highly sophisticated techniques of foreground removal (Liu and Tegmark 2011; Murray, Trott, and Jordan 2017) and system noise suppression. The expected 21-cm signal is 4-5 orders of times weaker than the foregrounds (Ali, Bharadwaj, and Chengalur 2008; Bonaldi and Brown 2015). The first generation of radio interferometers such as the GMRT (Paciga et al. 2011), PAPER (Kolopanis et al. 2019), LOFAR (Mertens et al. 2020), MWA (Barry et al. 2019; Li et al. 2019) will not be able make images of this era via conventional means of radio astronomy due to the lack of sensitivity. This is why they are trying to detect this signal via Fourier statistics e.g. 21-cm Powerspectrum (PS). Till date, only the weak upper limits on the 21-cm PS (Parsons et al. 2014; A. H. Patil et al. 2017) has been obtained. The future generation telescopes such as the SKA

(Koopmans et al. 2015), HERA (Pober et al. 2014)) will not only be able to detect the 21-cm PS but will also be able to make first three dimensional tomographic images of the 21-cm signal from this era (G. Mellema et al. 2015). Once such a detection is achieved then the next tusk would be to interpret it and extract vital information about the the EoR from it. To achieve such interpretations from observations computer simulations of this era has been developed, where one can tune different parameters to get an idea about all possible observable statistics. These kinds of simulations have two categories: radiative transfer and semi-numerical. The radiative transfer simulations are very accurate in terms of their physics as they follow the path of each individual photons (e.g. Garrelt Mellema et al. 2006). However, they have the drawback of being memory intensive and computationally expensive. These types of simulations are thus limited in terms of exploring a huge and multidimensional parameter space such as that of the CD-EoR. The semi-numerical simulations (e.g. 21cmFast: Mesinger, S. Furlanetto, and Cen 2010, ReionYuga: Majumdar, Garrelt Mellema, et al. 2014a; Mondal, Bharadwaj, and Majumdar 2016) on the other hand are approximate in their treatment of the physics but are fast and computationally cheap, thus can be used to simulate the signal in large cosmological volumes and to some extent can be used to explore the parameter space as well. However, they are also somewhat limited)in their ability (in terms of speed) in exploring are large and complex parameter space such as the CD-EoR.

In many of these simulations and theoretical models, the EoR is described in terms of three main astrophysical parameters, they are - ζ - the ionizing efficiency of the ionizing sources (stars and galaxies), R_{mfp}- the mean free path of the ionizing photons and Mh_{min}- the minimum mass of the dark-matter halo which hosts these ionizing sources (Choudhury 2003; Mesinger, S. Furlanetto, and Cen 2010). So far various approachses have been considered for predicting these parameters given a successful observation of the signal, they include - Bayesian inference, Fisher-matrix analysis, Variational inference etc. (Hassan, Andrianomena, and Doughty 2020; Hortúa, Volpi, and Malagò 2020; Kern et al. 2017; Greig and Mesinger 2015; Ajinkya H. Patil et al. 2014; McQuinn et al. 2006). In this thesis our focus is constrain these parameters (ζ, R_{mfp}, Mh_{min}) using Bayesian Inference. Bayesian

Inference provides a unique way to probe the parameter space. It also gives insight to many other aspects (i.e.: the reliability to underlying model used for the inference, correlation between different parameters, probability density functions (PDF) of the constrained parameters). However, even semi-numerical simulations have some drawbacks when used in a Bayesian framework. As one need to simulate a large cosmological volume thus even a semi-numerical simulation takes a considerable amount of time to simulate an observable. On top of that the MCMC random walker needs to take thousands to millions of steps to make a reasonable posterior distribution (see Figure 2.1). Therefore, to circumvent this problem, new fast methods (such as Principal component analysis(PCA), machine learning) can be used to develop emulation based EoR models trained by these semi-numerical simulations in a roboust manner. Machine learning techniques more specifically Artificial Neural Networks is one such widely used methods (Claude J. Schmit and Jonathan R. Pritchard 2017; Chardin et al. 2019; Doussot, Eames, and Semelin 2019) to emulate such simulated signals.

In their paper, C J Schmit and J R Pritchard 2017 have directly compared the 21cmFast based simulated and ANN emulated 21-cm PS and also showed that their ANN emulator can be used to constrain the EoR parameters through Bayesian inference. However 21-cm PS has several limitations of its own. The EoR 21-cm signal is expected to be highly non-Gaussian (Majumdar, Jonathan R Pritchard, et al. 2018) and being a two point statistic PS will not be able to describe this non-Gaussian signal completely. Thus a higher order statistics is essential to describe this non-Gaussian signal. Here, we use 21-cm Bispectrum (BS), the Fourier equivalent of the three point correlation function, to characterize the EoR 21-cm signal Majumdar, Jonathan R Pritchard, et al. 2018. Following the approach of C J Schmit and J R Pritchard 2017 and extending it beyond the PS, we for the very first time attempt to develop an Artificial Neural Network (ANN) based 21-cm BS emulator and use this emulator to build a Bayesian inference pipeline for estimating the EoR parameters. We have used the ReionYuga simulations (Majumdar, Garrelt Mellema, et al. 2014a; Mondal, Bharadwaj, and Majumdar 2016) to train our ANNs. We then have tried to constrain the EoR parameters using both PS and BS on our mock observation data autonomously, and checked which one among these two statistics put tighter constraints on the parameters. Further we have converted this entire pipeline into a single python-pip package called EmuPBk and made it public. This package can be used further for more complex scenarios (e.g. adding system noises(SKA), improving the ANNs by adding more training samples, training ANNs over different EoR-models etc.) of parameter estimation.

This thesis is organized in the following manner: in the Chapter 2, we have discuss the basics of 21-cm physics that is relevant for the EoR 21-cm signal and how one can observe the EoR using Radio-Telescopes and different obstacles for these observing. We have also described briefly different 21-cm statistics from the EoR, which have a greater possibility of detection due to their high signal-to-noise ratio. Additionally, we discuss different EoR simulation techniques and different EoR model parameters. In Chapter 3, we have discussed the methodology that we have adopted in developing our statistical inference tools. This includes a short discussiong on the workings of Bayesian inference, why we have used emulation based EoR models instead of simulations? The Chapter 4, contains our results and related discussions. In the last chapter (Chapter 5), we have discussed further future scopes of this project.

Chapter 2

21-cm transition in the context of EoR

The observations of Cosmic Microwave Background Radiation (CMBR) (Collaboration et al. 2018) confirm that the Universe was hot and dense plasma of hydrogen and helium during its early stages ($z \sim 1100$). These observations also supports the big-bang model of cosmology, according to which, the Universe continuously expanded after the big-bang, thus got cooled later on. After the matter-radiation decopuling they evolved differently and thus also cooled differently. The electrons and protons and some neutrons (mostly from hydrogen and helium) recombined to form neutral hydrogen and helium during this period. This is why this period is known as the recombination era of the Universe. After this, the Universe went into a dark age, when there were no sources of photons (other than the CMBR). The small density perturbations in matter which has their imprints on the CMBR were of $\sim 10^{-5}$ in relative magnitude (S. R. Furlanetto, Peng Oh, and Briggs 2006; Jonathan R Pritchard and Loeb 2012) at this point. However, during this dark age, these perturbations grew over time and become the seed of the very first structures of the Universe. These first structures eventually (when enough baryons collapsed into them) become the hosts for the first luminous objects (first stars and galaxies). These first sources of lights produced huge amount of Ultra-Voilet (UV) and X-ray radiations, and started ionzing and heating the neutral intergalactic medium around them. This lead to a phase change in the state of the IGM from neutral to ionized. As the number of first stars and galaxies grew over time and also subsequent generations of stars and galaxies were formed, they led to the ionization of the whole Universe. This process is known as the Epoch of Reionization(Barkana and Loeb 2001; Loeb 2010).

The observations of line of sight (LoS) opacity of the free electrons to the CMBR, suggest that the mean neutral fraction ($\langle x_{HI} \rangle$) of the IGM changes from completely neutral to $\sim$ 0.9 around $z \leq 10$ (Collaboration et al. 2018). Also analysis of the data from the high-redshift quasar serveys (specifically the absence of the Gun-Peterson trough at low redshift quasar spectra) suggest that most of the IGM became ionized by $z \leq 6$ (Fan et al. 2006). As the numbers of luminous sources (stars and galaxies) increased, the reionization process also got boosted, which can also be inferred from the the observations of luminosity functions of Lyman-α emitters (Zheng et al. 2017). All these observations have put a somewhat weak constrain on the timeline of the EoR ($6 \leq z \leq 12$).

However, these observations described above, most of which are indirect in nature, are limited in their capabilities to understand the EoR. The direct observation of EoR is one of the most challanging task, as there were very few luminous sources and the high optical depth of the medium prevented most of the photons produced during that time to reach us. The most promising way to observe the EoR is thus not the observations of luminous sources but observations of the radiation coming from the neutral Hydrogen H I from this era.

2.1 H I Transition Line

The gaseous matter in between the stars (interstellar medium) is a mix of gas (atoms and molecules), radiation and dust. Hydrogen is the most abundant amongst them, consisting of almost 75% of total baryonic budget of the Universe. Most of the hydrogen is in its atomic form. The number density of hydrogen varies place to place, from diffuse low dense ($n_{gas} \sim 10^6$ atoms/m^3) to high dense regions ($n_{gas} \sim 10^{13}$ atoms/m^3). The kinetic temperature of these H I regions can vary between $T \sim 80 - 6000\,K$. The protons and electrons in the hydrogen atom have spins associated with them, therefore they possesses magnetic moment. The total magnetic moment of the atom is the sum of the magnetic moments of both electron and proton.

$$\mathbf{F} = S + I \tag{2.1}$$

The magnetic moment is defined as:
For electron:

$$\mu_e = g_e \frac{\mu_B}{\hbar} S$$

where g_e = -2.00232 and μ_B is the bohr's magneton ($\mu_b = \frac{e\hbar}{2m_e} = 9.274 \times 10^{-24}$ J/T).
Similarly for proton:

$$\mu_p = g_p \frac{\mu_N}{\hbar} I$$

where g_p = +5.586 and μ_N is the nuclear magneton ($\mu_N = \frac{e\hbar}{2m_p} = 5.051 \times 10^{-27}$ J/T)

The direction of magnetic moment of an electron having angular momentum along $\hat{z}$ direction will be along -$\hat{z}$ direction, and for proton, it will be along the same direction. In the triplet state, the magnetic moment of electron and proton are in the opposite direction, and for the singlet state they are in the same direction. The magnetic moment of proton is found to be ~ 650 times lower than electron's magnetic moment. The difference in these energy states arise due to difference in the magnetic interaction energy. The magnetic interaction energy is defined as:

$$E_{pot} = -\mu_e . B_P$$

Taking proton size to be finite with radius R_p.

$$B_p = \frac{\mu_0}{2\pi R_p^3} \mu_p$$

where, R_p is the radius of the proton, $\mu_0 = 4\pi \times 10^{-7}\ TmA^{-1}$.

$$E_{pot} = -\frac{4}{3} \frac{R_P^3}{a_0^3} \mu_e \cdot B_p$$

where, a_0 is the bohr radius.

$$E_{pot} = -\frac{4}{3}\frac{\mu_0}{2\pi}\frac{g_e g_p \mu_e \mu_N}{a_0^3}\left(\frac{\mathbf{S}\cdot\mathbf{I}}{\hbar^2}\right)$$

$$\mathbf{S}\cdot\mathbf{I} = \frac{1}{2}\hbar^2[F(F+1) - S(S+1) - I(I+1)] \tag{2.2}$$

- For $F = 0$ in above equation results $\mathbf{S}\cdot\mathbf{I}$ = -$\frac{3}{4}\hbar^2$.
- For $F = 1$ in above equation results $\mathbf{S}\cdot\mathbf{I}$ = -$\frac{1}{4}\hbar^2$.

This creates the energy difference which is:

$$\Delta E_{pot} = |E_{pot}|_{F=1} - |E_{pot}|_{F=0} = \frac{4}{3}\frac{\mu_0}{2\pi}\frac{g_e g_p \mu_e \mu_N}{a_0^3}\left[\frac{1}{4}\hbar^2 - \left(-\frac{3}{4}\hbar^2\right)\right] \tag{2.3}$$

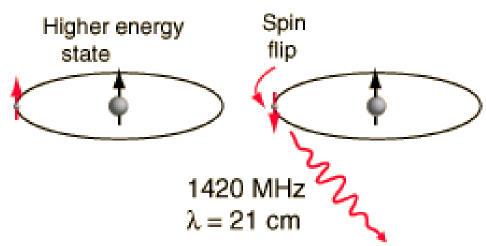

Figure 2.1: Hydrogen Hyper-fine Transition

$$\Delta E_{pot} = 9.42762 \times 10^{-25}\left[\frac{1}{4} - \left(-\frac{3}{4}\right)\right]$$

The frequency correspond to this difference is:

$$\nu = \Delta E_{pot}/h \approx 1420MHz$$

This hyperfine transition line corresponds to 21-cm wavelength. This transition occurs when electron spin flips from the triplet state to a singlet state (Bradt 2008).

2.2 21-cm Brightness Temperature

The excitation temperature corresponding to this hyperfine line is known as the spin temperature T_s. This represents the relative population between these two states mentioned above and are defined as (Morales and Wyithe 2010):

$$n_1/n_0 = (g_1/g_0)exp(-T_*/T_s) \tag{2.4}$$

Where, subscript 0 & 1 correspond to 1S singlet and 1S triplet level, (g_1/g_0) = 3, $T_* \equiv hc/k\lambda_{21cm}$ = 0.068 K.

The Universe is non-homogeneous and anisotropic on the small length scales, due to which we might have spatial variations in spin temperature which conveys information about the astrophysical processes.

We will only be able to detect the 21-cm signal if there is some deviation of spin temperature form its background temperature. The three processes that can make the spin flip transition of H I are:

- Absorption/Emission of the 21-cm signal from the radio background (CMBR).
- Collisions of hydrogen atoms with other hydrogen atoms or electrons, that can also excite/deexcite hydrogen for 21cm absorption/emission.
- Due to the presence of Ly-α sources, which can cause a spin flip via intermediate excited state (Wouthuysen-Field effect).

The spin-temperature thus takes the form:

$$T_S^{-1} = \frac{T_\gamma^{-1} + x_c T_c^{-1} + x_\alpha T_\alpha^{-1}}{1 + x_c + x_\alpha} \tag{2.5}$$

Where, x_c and x_α are the coupling coefficients (due to the atomic collision and Ly-α photons; see e.g. Jonathan R Pritchard and Loeb 2012; Carilli 2015). The CMBR acts as a background radio source for the intervening H I clouds, which have a tendency to absorb the CMBR. Thus, the signal depends upon the radiative-transfer (RT) through gas along the LoS. The basic RT equation for the transmission of this radiation is governed by the equation for $\partial I_\nu/\partial \nu$ (ignoring the scattering along LoS):

$$\frac{dI_\nu}{ds} = -\alpha_\nu I_\nu + j_\nu \tag{2.6}$$

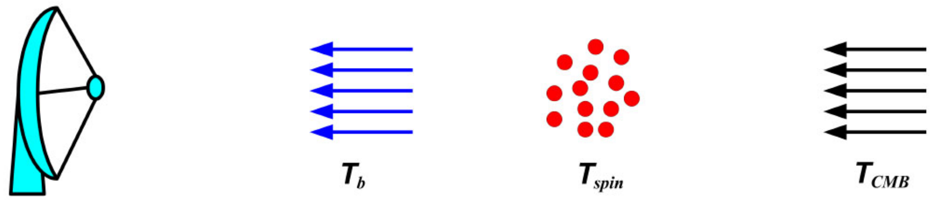

Figure 2.2: CMBR in the background for the intervening H Icloud

The absorption and emission coefficients are $-\alpha_\nu$, j_ν respectively. This lead to the differential brightness temptrature, following (Jonathan R Pritchard and Loeb 2012):

$$\delta T_b = \frac{T_S - T_R}{1+z}(1 - e^{-\tau_\nu}) \tag{2.7}$$

$$\approx \frac{T_S - T_R}{1+z}\tau \tag{2.8}$$

$$\delta T_b \approx 27 x_{\mathrm{H\,I}}(1+\delta_b)\left(\frac{\Omega_b h^2}{0.023}\right)\left(\frac{0.15.(1+z)}{\Omega_m h^2.10}\right)^{1/2}\left(\frac{T_S - T_R}{T_S}\right)\left[\frac{\partial_r v_r}{(1+z)H(z)}\right]\ \mathrm{mK} \tag{2.9}$$

where $x_{\mathrm{H\,I}}$ is the neutral fraction of hydrogen, δ_b is the fractional overdensity in baryons and the final term arises form the peculiar velocity gradient $\partial_r v_r$ along the LoS (Jonathan R Pritchard and Loeb 2012).

2.3 Statistical Observables of the 21-cm signal form the EoR

2.3.1 Power Spectrum

The brightness temperature of the 21-cm signal can have a spatial variation at same redshifts, because of the spin temperatue T_S changes as local hydrogen density, presence of the Ly-α sources changes and also due to the difference in the state of ionization of hydrogen at different points. These spatial features are dependent on the early structures of the Universe. To observe these features power spectrum statistics is used, which is the Fourier transform of the 2-point correlation function, a measure of the amplitude of the fluctuations in the signal at different spatial scales. Before the ionization, the 21-cm powerspectrum effectively follows the matter powerspectrum, and for the latter stages of the reionization it follows the ionisation field power spectrum (see e.g. Jonathan R Pritchard and Loeb 2012). The powerspectrum in general for the 21-cm field can be defined as:

$$\delta_{21}(x) = (\delta T_b(x) - \delta\bar{T}_b)/\delta\bar{T}_b \tag{2.10}$$

$$\langle \delta_{21}(k_1)\delta_{21}^*(k_2)\rangle = (2\pi)^3 \delta_D(k_1 - k_2) P_{21}(k_1) \tag{2.11}$$

where, δ_D is the dirac-delta and $\langle\rangle$ denotes the ensemble average. It is convenient to use dimensionless powerspectrum for our analysis, which is expressed as:

$$\Delta_{21}^2(k) = \frac{k^3}{2\pi^2} P_{21}(k) \tag{2.12}$$

2.3.2 Bispectrum

The powerspectrum in principle can represent all information about the 21-cm signal only if, the fluctuations present in the signal are Gaussian random in nature. However, the signal is expected to be highly non-Gaussian due to the presence of non-random distribution of ionized regions and their evolution with time and the inherent non-Gaussianity in the initial perturbation form the inflation period. This non-Gaussianity in the signal is also expected to evolve with the progressing state of reionization, resulting in the correlation in the signal between different Fourier modes (k), which can not be probed by the powerspectrum (Mondal, Bharadwaj, and Majumdar 2015; Majumdar, Jonathan R Pritchard, et al. 2018). Thus one would need to use some higher order statistics which can extract the non-Gaussian information from the 21-cm signal. The bispectrum, which is a Fourier transform of the 3-point correlation function, is one such higher order statistics and is defined as:

$$\langle \delta_{21}(k_1)\delta_{21}(k_2)\delta_{21}(k_3)\rangle = (2\pi)^3 \delta_D(k_1 + k_2 + k_3) B(k_1, k_2, k_3) \tag{2.13}$$

(Majumdar, Jonathan R Pritchard, et al. 2018). The bispectrum provides a scale-dependent measure of the non-Gaussianity in the 21-cm signal and thus contain additional astrophysical information over powerspectrum. The studies of non-Gaussianity will be most useful for learning about astrophysics in more detail and for constraining the astrophysical parameters more precisely.

2.4 Probing EoR using Radio Telescopes

The 21-cm signal from the EoR is redshifted towards the meter-wave band of the spectrum due to the expansion of the Universe. This signal can be detected using radio telescopes. However, observation of the 21-cm signal from the EoR is an extremly challanging due to the presence of different obstacles, e.g.:

- **Galactic Foreground:** The galactic emission is orders of magnitude higher than that of signal from EoR.
- **System Noise:** System Noises also (e.g.: Radio Telescope noise) makes it difficult to observe the signal.
- **RFI:** Most of our daily communications uses radio channels (i.e.: Television Broadcasts, Mobile network signals). These sources interfere with the 21-cm signal.

From all these sources the 21-cm signal gets contaminated severely. So, one requires a good modeling of the foreground, system noise, and preferred locations of the radio telescopes to be away from human population for a successful detection (Ali, Bharadwaj, and Chengalur 2008; Bonaldi and Brown 2015).

Using multiple radio-dishes instead of a single dish, is a great alternative to probe the EoR. The collection of these dishes acts as a signal interferrometer, which is sensitive to the spatial fluctuations in the 21-cm signal across the sky (Jonathan R Pritchard and Loeb 2012). Typical First generation radio telescopes e.g. GMRT (Paciga et al. 2011), PAPER (Kolopanis et al. 2019), LOFAR (Mertens et al. 2020), MWA (Barry et al. 2019; Li et al. 2019), etc. due to their lack of sensitivity, are unable make images of this signal but is trying to detect it through Fourier statistics such as the powerspectrum. However, next-generation telescopes such as the SKA (Koopmans et al. 2015), HERA (Pober et al. 2014) will have large collecting area, which will be able to make 21-cm images of the EoR.

2.5 EoR model Parameters

The brightness temperature eq. (2.9) depends on cosmology as well as the astrophysics. For, higher redshifts, before the reionization begins, the neutral fraction $x_{HI} \approx 1$ results in the brightness temperature being proportional to the matter density fluctuations which probes the cosmology. For later redshifts, when the reionization starts, the brightness temperature becomes dependent on both the density fluctuations and neutral fraction. The neutral fraction fluctuations and evolution is dependent on the astrophysics. In our simulations the parameters that affect the reionization process are:

2.5.1 Ionizing Efficiency (ζ)

The ionizing efficiency of high redshift sources of photons is the combination of several other degenerate parameters, define as:

$$\zeta = A_{He} f_{\star} f_{esc} N_{ion}$$

where A_{He} is a correction factor for the presence helium, $f_{\star}$ is star formation efficiency, f_{esc} is the escape fraction and N_{ion} is number of photons produced per baryons. More ionizing efficiency will speed up the ionization process.

2.5.2 Mean Free Path of ionizing photons (R_{mfp})

During the formation of first structures, there were dense clouds of neutral hydrogen where the recombination (electron-proton) rate was much higher than rest of the IGM. These regions effectively absorb all ionizing radiation which affect the size of ionized bubble regions during reionization. So, the R_{mfp} is defined as the mean distance travelled by those ionizing photon before getting absorbed by such regions.

2.5.3 Minimum Halo mass (Mh_{min})

The minimum mass of the dark matter halos which host the luminous sources (stars and galaxies etc) to form and produce the ionizing radiation, which will ionize their surroundings. Typical star formation happens after the stability between in-falling gravity and thermal pressure of gas (H_2 primarily), gas cools fast and instability happen in the system results the star formation, which leads to dissociation of H_2 gas, and further ionize the dissociated neutral hydrogen.

2.6 Simulations of the EoR

The 21-cm signal contains a wealth of information about the cosmology and astrophysical processes that takes place during the EoR. In order to learn about the underlying physical processes resulting the re-ionisation, simulations with dynamical ranges spanning orders of magnitude are required. In principle these simulations should be able to deal with the physics between Mpc to sub-kpc scales. There exists different approaches of doing these simulations:

2.6.0.1 Radiative Transfer simulations

Radiative Transfer simulations are physically more accurate as they simulate the underlying astrophysical processes more accurately by solving radiative transfer equations along the line-of-sight of each individual photons. This makes them computationally more intensive thus they are very expensive to generate a single realization of ionization and brightness temperature maps.

2.6.0.2 Semi-numerical Simulations

Semi-numerical simulations are developed by considering some approximations over the radiative transfer methods which mainly involves photon counting at different smoothing length scales. These approximations makes them fast and also gives more freedom to make:

- Simulations of large cosmological volumes ($\sim$ Gpc^3) that are comparable to the survey volumes of future radio interferometers like the SKA.
- Allows fast exploration of the reionization parameter space.

There are several semi-numerical simulations that are openly available, e.g. Reion-Yuga (Majumdar, Garrelt Mellema, et al. 2014a; Mondal, Bharadwaj, and Majumdar 2015), 21cmFast (Mesinger, S. Furlanetto, and Cen 2010), etc. **21cmFast** is one of the fast and most used semi-numerical codes, however, it doesnot use N-body simulations to generate the matter density (thus do not properly simulate the non-linearities in the matter distribution) and also does not include the effect of redshift-space-distortion properly in the simulated 21-cm signal. However, **Reion-Yuga** uses a particle-mesh N-body simulation to simulate the matter distribution and includes the effect of redshift space distortion properly. Which makes this code more accurate than 21cmFast (Majumdar, Garrelt Mellema, et al. 2014a). It has been observed that redshift space distortions can alter the signal significantly upto $100-200\%$ at large length scales (Mao et al. 2012).

Chapter 3

Methodology

In this chapter, we try to explain the methods that we have adopted for the development of our statistical inference pipeline. It is apprent from the discussions in Section 2.6 that there are significant amount of limitations in use of simulations for Bayesian parameter estimation. The use of simulations will increase the time for statistical inference by many folds. Both the radiative transfer and semi-numerical simulations are expensive enough that they cannot be used to take millions of random steps in a multidimensional parameter space. One alternative for speeding up the inference process is to use well trained emulations instead of simulations. The first generations of radio-telescopes have successfully put the upper-limits in the PS of the signal and in the upcoming years next-generation telescope (e.g. SKA) comes with the promise of detecting the 21-cm PS and BS but also making the first 3D-tomographic images from this era. Once this signal is detected one would require a roboust inference mechanism to constrain the EoR model parameters from these observations. One such inference mechanism is the Bayesian inference and we use this to build our statistical inference tool box for this project. The flowchart in Figure 3 shows the basic building blocks and working principle of our Bayesian inference pipeline, which we discuss in details in this chapter.

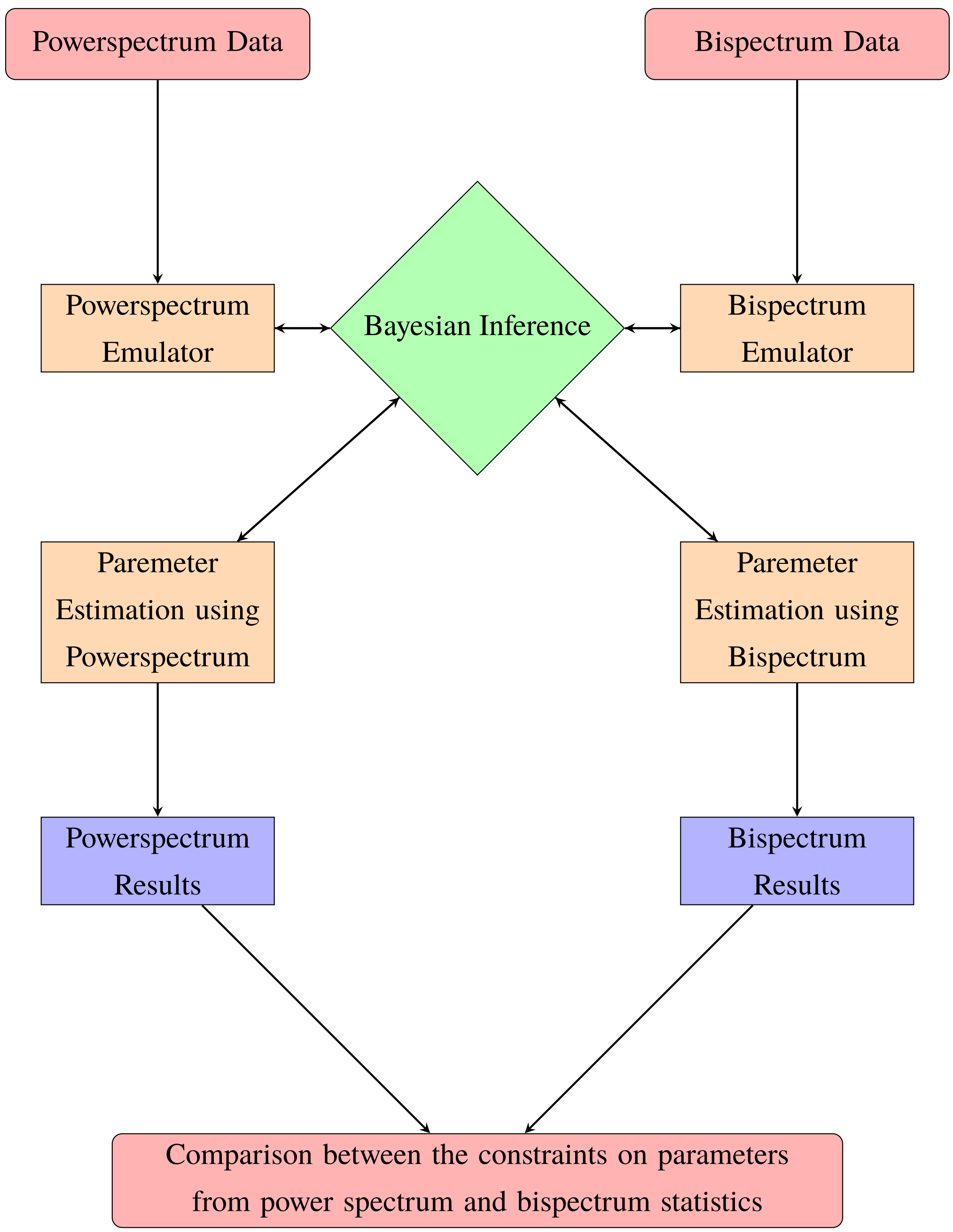

Figure 3.1: This flowchart shows the basic building blocks and workings of our statistical inference pipelines.

3.1 Bayesian Inference

Bayesian inference works with the fundamental principle of Bayes' theorem and Marcov's Chain Monte Carlo (MCMC) algorithm. This essentially predicts the posterior probability from a prior, likelihood and the evidence, which connected through the Bayes theorem as:

$$Pr(\theta|D, M) = \frac{Pr(D|\theta, M)\pi(\theta|M)}{Pr(D|M)} \tag{3.1}$$

where θ is a vector which contains the parameters, D is the observed data (given), M is model. The different probabilities mentioned above are labeled as:

- $Pr(\theta|D, M)$ is called **Posterior**.
- $\pi(\theta|M)$ is called **Prior**.
- $Pr(D|\theta, M)$ is the **Likelihood**.
- $Pr(D|M)$ is our **Normalization constant or Evidence**.

In our context the eq. 3.1 implies that given a successful detection of the 21-cm signal and assuming a reionization model to be true, we would like to know what is the posterior distribution of the model parameters. As discussed earlier our reionization model depends on three parameters:

$$M(\zeta, R_{mfp}, Mh_{min})$$

In scenarios when we do not have any apriori information about the parameters we assume as flat prior for each of them. This implies while sampling the parameter space, our random walker will not assume any predefined distribution around the randomly chosen parameter points. Additionally if we assume our model to be true then we can also ignore the estimation of the evidence as it just acts as an normalization constant. So our problem then boils down to estimating the likelihood alone. In the standard MCMCs (MH) algorithms the likelihood is generally assumed to be a multivariate-Gaussian of the form:

$$L \equiv C * exp(\frac{-(M(\theta_i) - D_{ob})^2}{Cov}) \tag{3.2}$$

Where, $\theta_i = set(\zeta, R_{mfp}, Mh_{min})$ is random step in the parameter space, D_{ob} and Cov are the data and covariance-matrix respectively from the observation. It is more reasonable to use the log of likelihood function for the convenience of estimation:

$$\log L \propto -0.5 \times ((M(\theta_i) - D_{ob})^T Cov^{-1} (M(\theta_i) - D_{ob})) \tag{3.3}$$

The MCMC computes the likelihood of parameters for a given data set and a given model using Bayes rule. As the posterior is proportional to the likelihood thus it essentially provides us the constraints on the parameters. As this exercise provides us the entire posterior distribution, thus one can go even beyond the simple constraints on the parameters and can even tell us about if there is any correlation between different parameters. The basic MCMC mechanism works in the following manner:

- Start over large numbers of iteration.
- for each iteration : **generate** a random point in parameter space(θ_i) (random walk).
- **calculate** the value of the model function at that point in the parameter space.
- **calculate** the **likelihood** or **log-liklihood**.

calculation of log-likelihood is preferred in those cases, where we just want the distribution of the likelihood over the given parameter values.

There are many **python based** MCMC libraries available e.g. EMCEE (Foreman-Mackey et al. 2013), CosmoHammer (Akeret et al. 2013), PyHMC, PyMC3 etc., which quickly computes the likelihood for a large numbers of random walkers at simultaneously using parallel computation for each MCMC chain. For the MCMC sampling in this work, we have used CosmoHammer which has been used for constraining the cosmological parameters from the WMAP data by Akeret et al. 2013. CosmoHammer uses EMCEE under its hood.

CosmoHammer chain has three basic components:

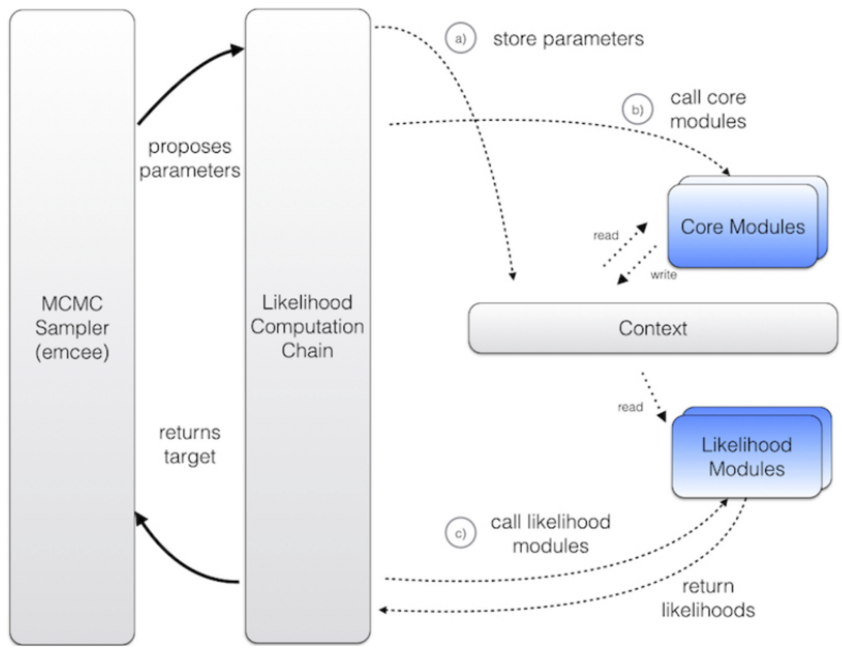

Figure 3.2: Working of CosmoHammer **Image credit: Cosmopic.**

- **Context:** The context is a dictionary for storing information created during the evaluation of the likelihood. It at least contains the parameter values of the current position proposed by the sampler.

- **Core Module:** The Core Modules can be used to calculate information which is needed for the evaluation of the likelihood. The information can then stored in the context.

- **Likelihood Module:** The Likelihood Modules use the information in the context to calculate the likelihood of the proposed position and return the log-likelihood to the chain.

The Likelihood Computation Chain $\Longrightarrow$ first stores the proposed parameters in the **context** $\Longrightarrow$ then moves on and invokes all available **Core Module** before $\Longrightarrow$ then calls the **Likelihood Module**.
The resulting log-likelihood values are gathered, summed, and returned to the sampler.
CosmoHammer is **parallelized**, so one can use several random walkers simultaneously for the sampling of the parameter space.

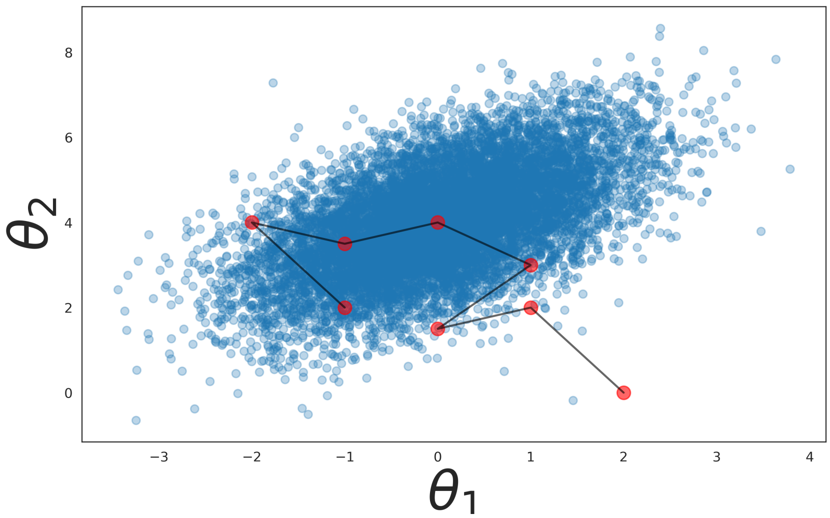

Figure 3.3: A random walker sampling a 2D parameter space.

As shown in Figure 3.3 the random walker needs to take thousands of steps to probe the parameter space properly. It calculates the likelihood at each of those steps, only after which we get a considerable posterior distribution. To estimate the likelihood at each of these steps one would need to estimate the specific statistic for a given model. For our problem we need to simulate the 21-cm signal in a cosmologically relevant volume and then estimate the specific statistics and compare it with the observed statistics. Simulation part of this flowchart is the computationally time consuming and heaviest segment of this process. Previously, Greig and Mesinger 2015; Greig and Mesinger 2017 had used the 21-cm PS from semi-numerical simulation 21cmFAST for evaluating the log-likelihood. As discussed earlier, though it is a faster way to explore the parameter space but not accurate enough to simulate various physical signatures or estimate higher order statistics e.g. 21-cm BS with reasonable accuracy at different length scales.

As an alternative, one can use different emulation techniques e.g. using Artificial Neural Networks(ANN), Principle component analysis(PCA) etc., to emulate the signal rather than simulating it. Agarwal et al. 2014 had developed PkANN, that emulates the non-linear matter power spectrum through ANN. Kern et al. 2017 had used Gaussian processes and PCA for the emulation of Cosmic Dawn 21-cm power spectrum. C. J. Schmit and J. R. Pritchard 2018 had used ANN emulated 21-cm PS signal for EoR parameter estimation. For building their ANN based emulator they have used semi-numerical simulation 21cmFast.

3.2 Emulating the EoR 21-cm Statistics

Keeping this motivation in mind that emulated models of the EoR can speed up the time consuming statistical inference, we have tried to develop Artificial Neural Network(ANN) based models of the EoR 21-cm statistics.

Artificial Neural Networks are one of many applications of Machine learning where the machine learns by itself without being explicitly programmed. It mimics the working of a human brain. The ANN is a collection of individual artificial neurons, which are nothing but some well defined mathematical functions. These artificial neurons are simply connected to each other (like the neuron cells in our brain). There exists different kinds of ANN architectures e.g.: Artificial Neural Networks(ANN)- used as both classification and regression solver, Convolutional Neural Networks(CNN)- generally used on image data, and classification, Recurrent Neural Networks(RNN)- are used in pattern recognition, language processing etc. These different ANNs were then trained over the data, accordingly. Here, our approach is the supervised machine learning, where we train the ANN model before doing any predictions out of it. The basic working of a single artificial neuron is the following:

- Let us have the data in the form of some (x_i, y_i). The neuron tries to fit a functional relation between the data points, using the simple equation:

  $$z_i = W_i.x_i + b_i$$

  where W_i are the weights and b_i are the biases.

- Then z_i is inserted in some kind of activation function, e.g. Relu, Sigmoid, Elu , tanh etx.

  $$\hat{y}_i = act(z_i)$$

- The difference (loss) between the $\hat{y}_i$, y_i is:

$$\delta y_i = \frac{\sqrt{\hat{y}_i{}^2 \sim y_i^2}}{N} \tag{3.4}$$

where N are the number of epochs. The forward and back propagation algorithms (e.g. gradient decent) are then used to minimise the loss function. The whole idea behind this is to adjust the weights and biases in such a way, that minimises the loss.

3.2.1 Structure of an ANN

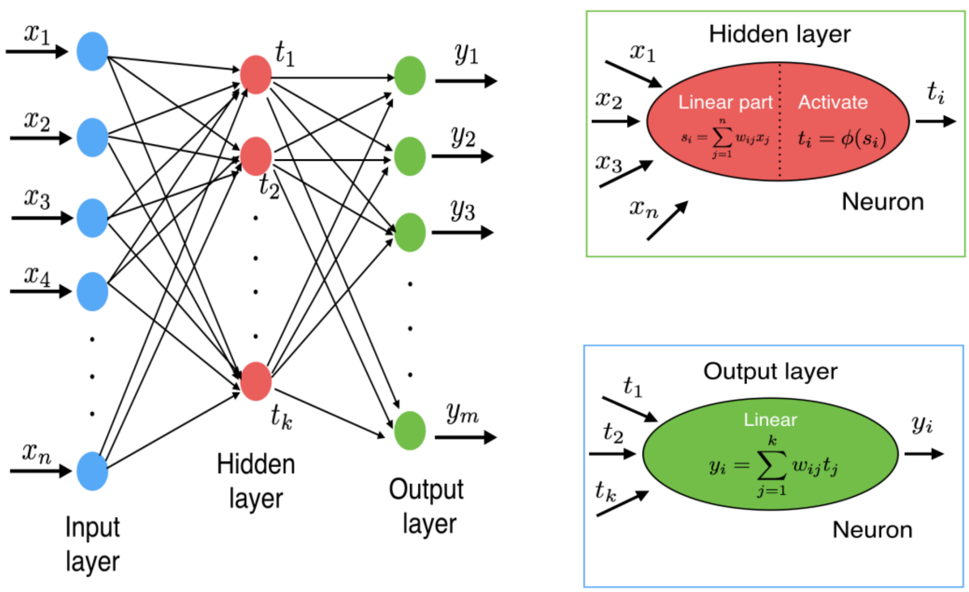

Figure 3.4: Basic multi-layered ANN structure.

Emulating the EoR 21-cm statistics is a kind of a regression problem. So, before making any predictions the ANN should be well trained, which requires huge amount of training data, i.e. simulated 21-cm statistics. The more data we put into the network and the more paramter space is covered, the better will be the accuracy of the ANN when it predicts.

Our training simulated data-sets come from the EoR simulation **Reion-Yuga** (Majumdar, Garrelt Mellema, et al. 2014b; Mondal, Bharadwaj, and Majumdar 2016). We estimate both the power spectrum and bispectrum from this simulated training set.

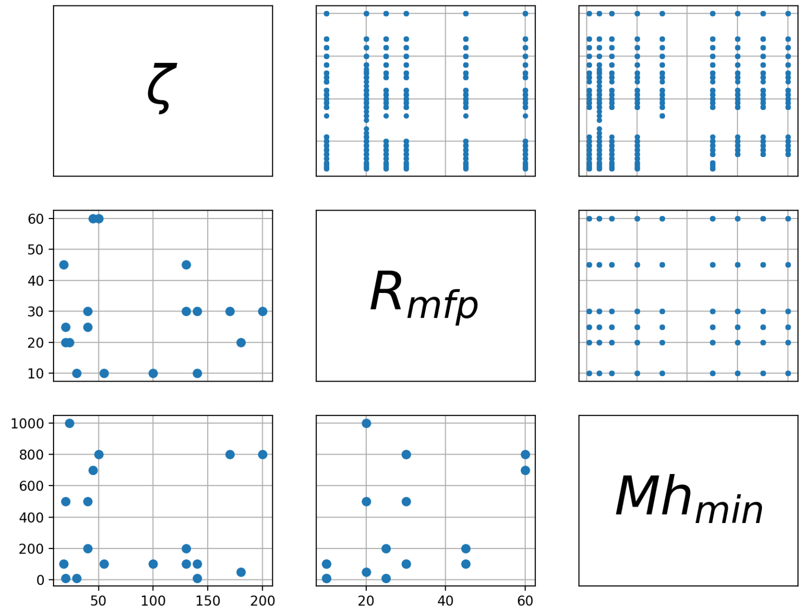

Figure 3.5: Data-set used for training (upper-triangular region) and testing (lower-triangular region).

The data-set consists of 1058 points in the parameter space (see Figure 3.5) out of which we have used **1000** samples for **training** purpose, **40** samples for **validation** and **18** as our **test-set**. The simulations data has the box size of **215 Mpc**, at a single redshift **z = 9.210**. The reason for choosing this redshift is that the 21-cm signal is expected to have its highest level of fluctuations around this redshift. We followed a simple Multi-layered ANN model as our structure of ANN. The basic working of the ANN is as follows:

- ANN uses **training data** i.e. **for given Parameters** $\Longrightarrow$ **what is Powerspectrum/Bispectrum ?**

- After completion of training, it should be able to predict i.e. **given the parameters** $\Longrightarrow$ **what will be the Powerspectrum/Bispectrum?**

Out of several different Python based open source machine learning libraries

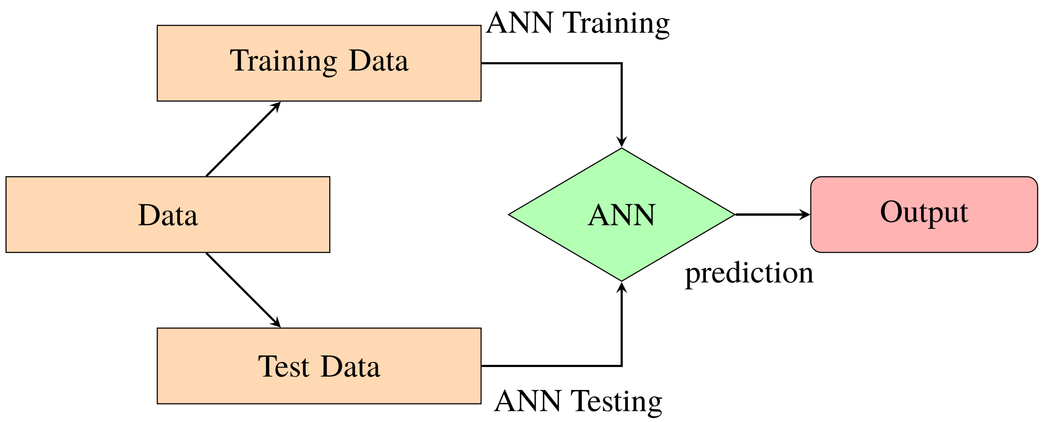

Figure 3.6: The working of a supervised ANN.

available, we have used **Tensorflow[1] and keras[2]** to build the ANN models.

[1] https://www.tensorflow.org/
[2] https://keras.io/

Chapter 4

Results and Discussion

In this chapter we discuss our major results. Our results have two segments. In the first segment, we discuss about the performance of our ANN based emulator in emulating the power spectrum and bispectrum. In the second segment we discuss how well our Bayesian inference pipeline works in constraining the EoR parameters.

4.1 ANN Emulators

4.1.1 21-cm Powerspectrum Emulator

We have used **1000** simulated normalized 21-cm PS over seven k-bins to train our ANN based PS emulator. So the emulator effectively has ten parameters (three EoR parameters (dependent) + seven k-bins : free parameters, Figure 4.1) on which the PS depends. As all PS data-set are on the same k-bins, it is convenient to order the data for training. Thus the emulator is designed such that it can predict the 21-cm PS over all of these seven k-bins when supplied with just three EoR parameter values.

The PS emulator has following structure:

- The input layer consists of **3 neurons** according to the 3 given parameters with **elu** as the activation function.
- first hidden layer has **48 neurons**.
- second hidden layer has **24 neurons**.

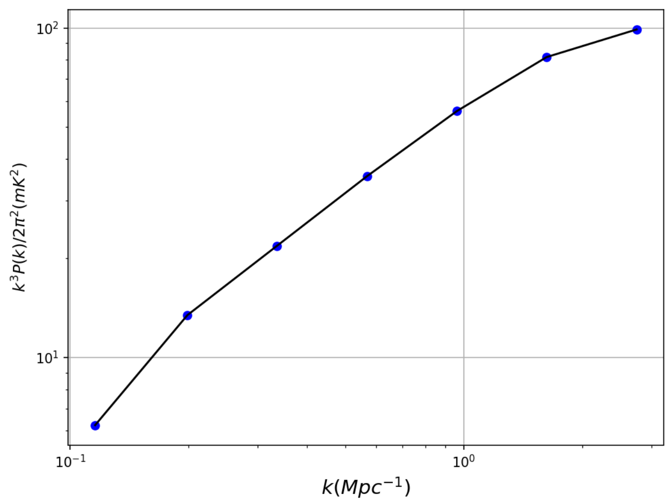

Figure 4.1: Powerspectrum simulation (points) and emulation (line).

- The third hidden layer has **14 neurons**.
- The output layer has **7 neurons** (correspond to seven desired Powerspectrum values over seven k-bins), with **linear** activation function and **0.0 dropout** rate. For all layers, **elu** activation function has been used.
- The **Adam** optimizer is used with the learning rate of 0.001.
- The loss function is **Mean Squared Error**, with 1000 epochs and 10 batch size.

4.1.1.1 Accuracy and Loss

Figure 4.1 shows the simulated and emulated powerspectrum simultaneously. The difference between simulation and emulation is non-differentiable with eye. As discussed in Section 3.2 and Chapter 3, the back-end working rule of Artificial Neural Network is to minimise the loss function which improves accuracy of the predictions. The Figure 4.2 shows the accuracy and loss of the emulation process. We can clearly see from these plots that, our model has achieved $\sim 98\%$ accuracy for about 1000 training epochs, similarly the loss is also significantly low.

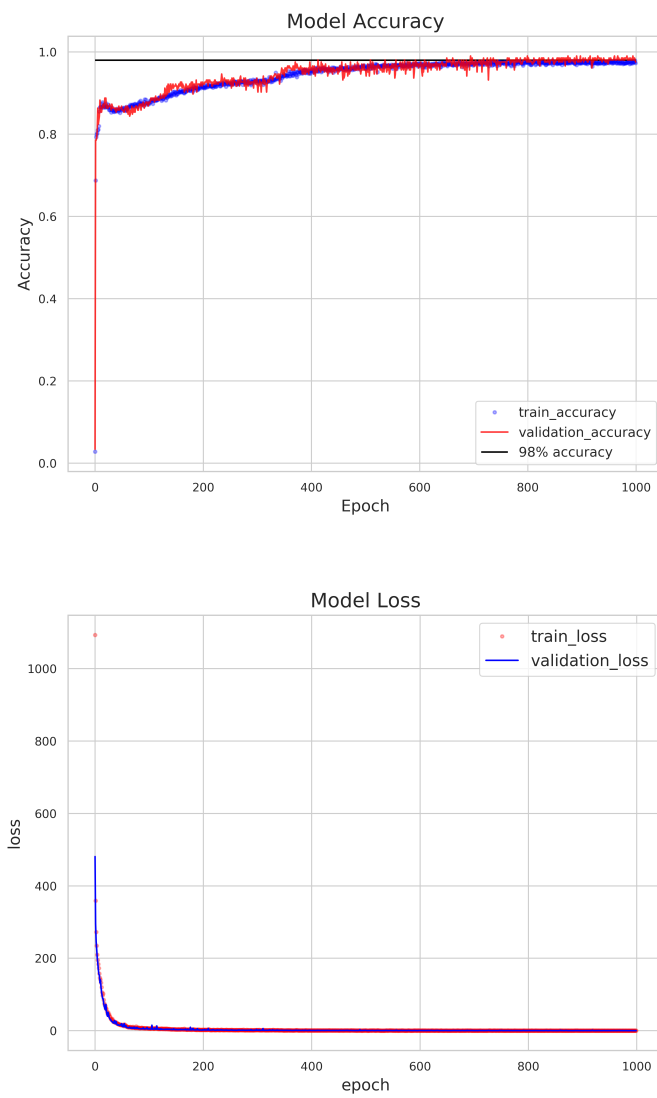

Figure 4.2: Training Accuracy and Loss for the PS emulator.

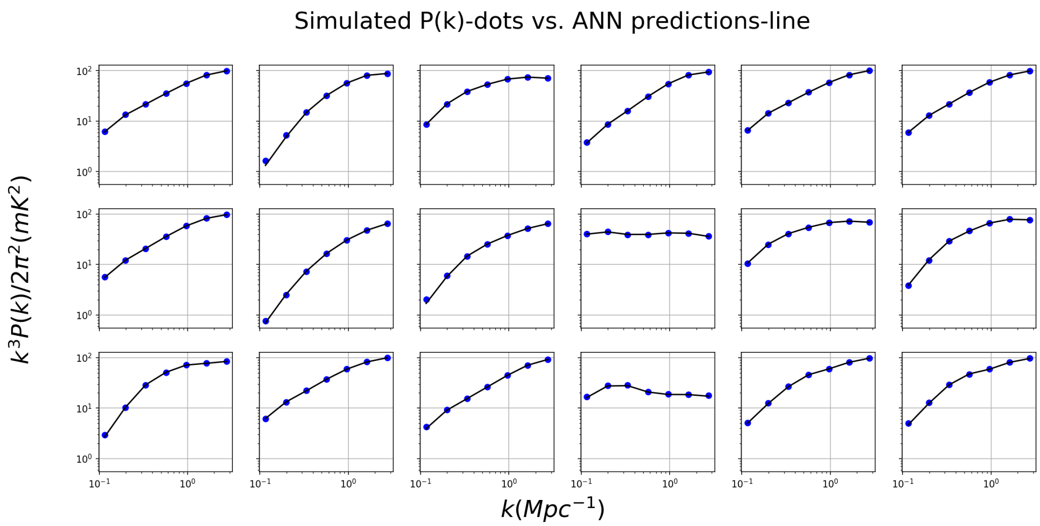

Figure 4.3: PS test predictions from ANN for 18 test sets.

Figure 4.3 shows the predictions of PS from our 18 test-sets. It is evident that predictions are as good as the simulations.

4.1.2 21-cm Bispectrum Emulator

Bispectrum for binned triangle configurations can be defined as:

$$\hat{B}(\vec{k}_1, \vec{k}_2, \vec{k}_3) = \frac{1}{NV} \sum_{(\vec{k}_1+\vec{k}_2+\vec{k}_3=0)\in n} \Delta_b(k_1)\Delta_b(k_2)\Delta_b(k_3) \tag{4.1}$$

The k-triangle shape for BS estimation is defined via two paramters:

- $k_2/k_1 = n$
- $cos(\theta) = \frac{\vec{k}_1.\vec{k}_2}{k_1.k_2}$

We have used 21-cm Bispectrum with the following configurations:

- we have only used two different k_1 values ($0.3\,\mathrm{Mpc}^{-1}$ and $1.5\,\mathrm{Mpc}^{-1}$).
- k_2/k_1 ranges from $0.50 \rightarrow 1.00$ with the steps of 0.05.
- $cos(\theta)$ ranges from $0.50 \rightarrow 0.99$ with the step size 0.01.
- for a given value of k_1, we have 50×11 (free) + 3 EoR parameters (dependent).

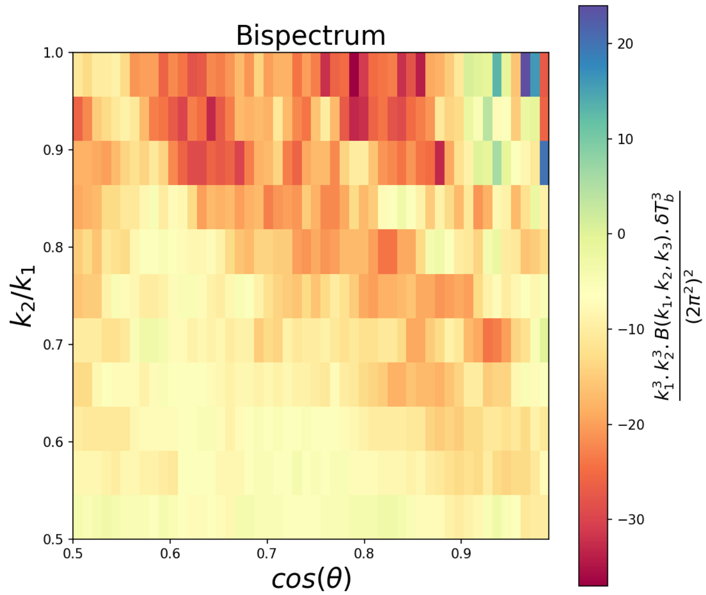

Figure 4.4: Bispectrum

Figure 4.4 shows the BS at single k_1 and for k_2/k_1-$cos(\theta)$ parameter space. Since all the dependent parameters are fixed, we have only used 3 EoR-parameters and 50×11 BS values at a single k_1 mode for the training of the ANN emulator. It is convenient to plot Bispecturm for only the unique triangles, which follows following conditions:

- $$k_1 \geq k_2 \geq k_3$$

- $$k_2/k_1 \times cos(\theta) \geq 0.5$$

4.1.2.1 Bispectrum for $k_1 = 0.3\,\mathrm{Mpc}^{-1}$

The bispectrum ANN emulator has the following structure:

- The input layer consists of **3 neurons** as we use 3 parameters.

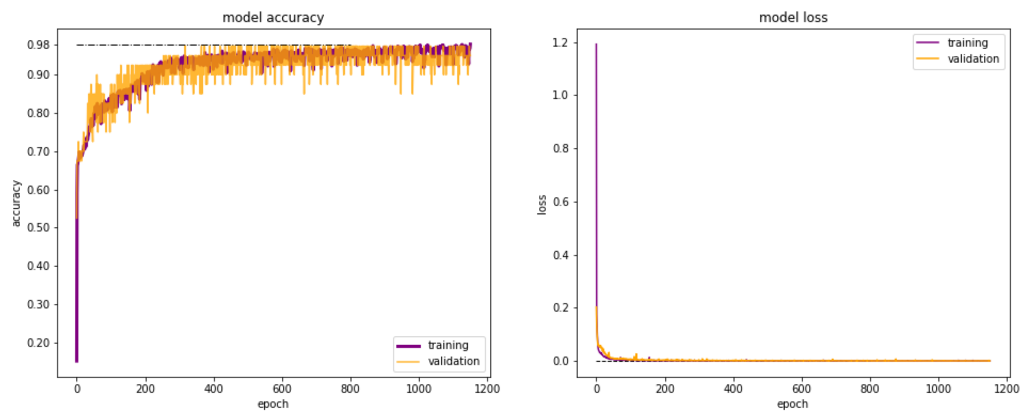

Figure 4.5: Training Accuracy and Loss for bispectrum with $k_1 = 0.3\,\text{Mpc}^{-1}$

- We have total 6 hidden layers on which first, second, third, fourth, fifth, sixth hidden layers have **80, 320, 460, 560, 260, 100 neurons**.
- The output layer has **550 neurons** (corresponding to 550 desired bispectrum values), with **linear** activation function and **0.0 dropout** rate.
- The **Adam** optimizer is used with the learning rate of 0.0001, and **Mean Squared Error** as the loss function, with 1200 epochs and 10 batch size. For all layers except output layer, **elu** activation function has been used.

4.1.2.2 Accuracy, Loss and Predictions for Bispectrum with $k_1 = 0.3\,\text{Mpc}^{-1}$

For both the training set and the validation set, accuracy certainly reached around 98% for about 1200 training epochs, and the loss also reached close to zero.

4.1.2.3 Bispectrum for $k_1 = 1.5\,\text{Mpc}^{-1}$

The base structure for the bispectrum with k_1 values $0.3\,\text{Mpc}^{-1}$and $1.5\,\text{Mpc}^{-1}$ are the same. For both the training sets and the validation sets, accuracy certainly reached around 98% for less than 1000 training epochs, and the loss also reached close to zero. ANN predictions of bispectrum, shows a great resemblance with the simulated bispectrum.

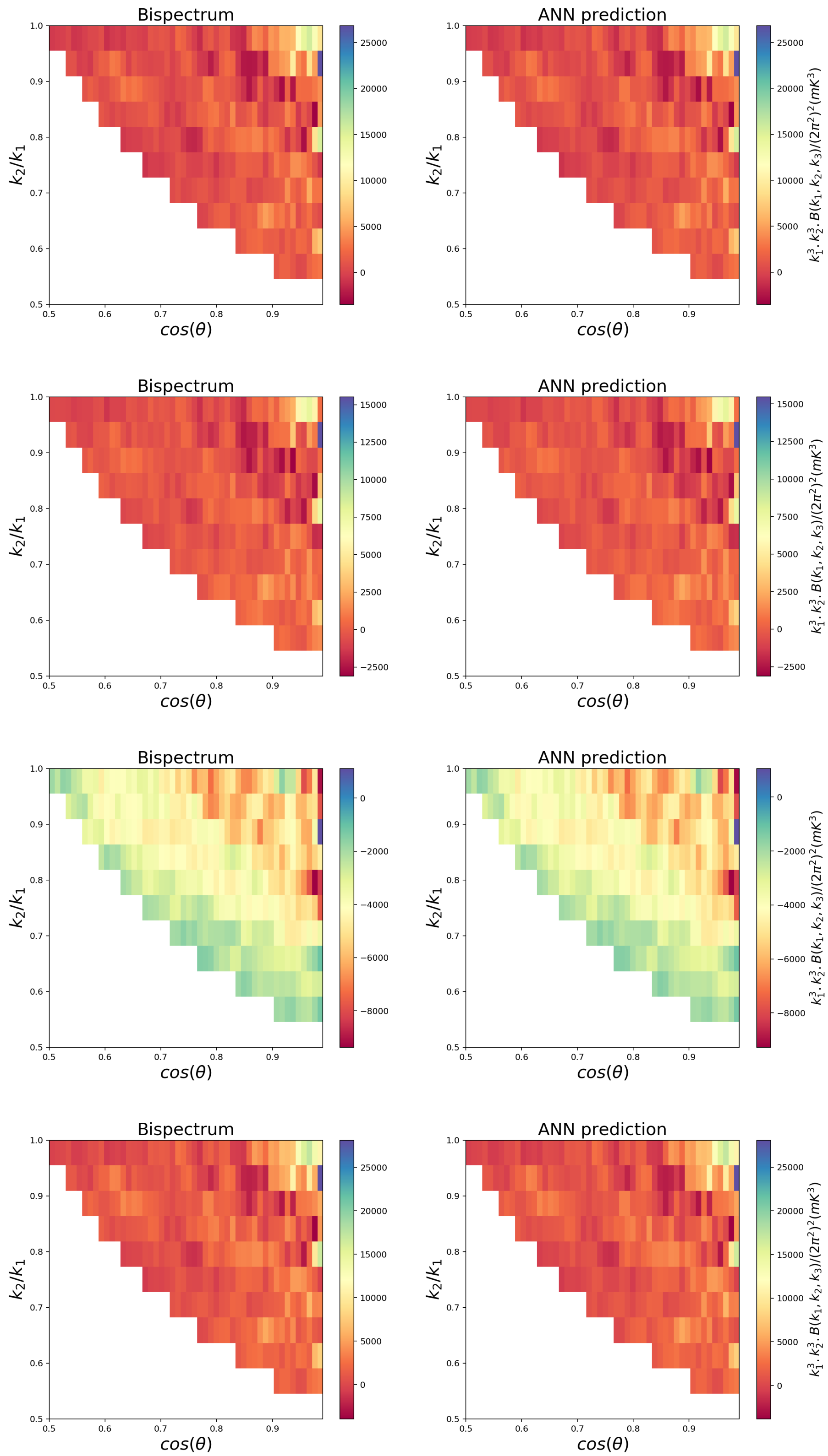

Figure 4.6: Test predictions using ANN emulator for bispectrum with $k_1 = 0.3\,\mathrm{Mpc}^{-1}$

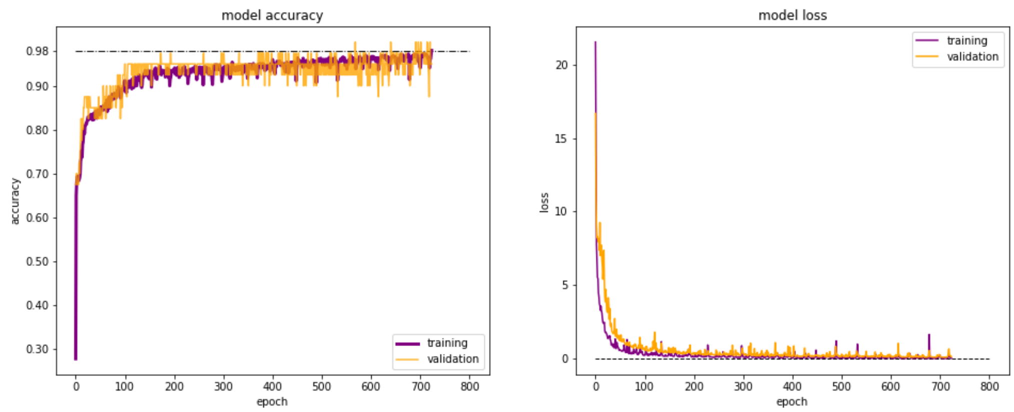

Figure 4.7: Training Accuracy and Loss in the bispectrum emulation for $k_1 = 1.5\,\mathrm{Mpc}^{-1}$

4.1.2.4 Accuracy, Loss and Predictions for Bispectrum with $k_1 = 1.5\,\mathrm{Mpc}^{-1}$

4.2 Parameter Estimation

4.2.1 Impact of EoR parameters on 21-cm statistics

The global ionization state of the IGM is dependent on the EoR parameters. Among the parameters under consideration, ζ implies efficiency of the ionizing sources. R_{mfp} implies the maximum distance travelled by these ionizing photons, before getting absorbed by the IGM. This also in a way tells about the size of the ionized regions. Mh_{min} implies the minimum mass of the halo that hosts the ionizing sources. The typical size of the ionized regions directly depends on the halo mass as the photon production rate is also proportional to the halo mass.

We choose the test sets of parameters in such a way that it covers a significant amount of the parameter space. Table 4.2 shows the test sets of parameters that we have chosen to test our Bayesian inference pipeline. From this table we choose few extreme sets of parameters and discuss their imapct on one of the 21-cm statistic, the powerspectrum. The same can be done for the bispectrum, however, we limit our discussion here to powerspectrum alone.

We start our discussion with **Case-8**. Figure 4.9 shows the power spectrum for this set of parameter values ($\zeta = 30$, $R_{mfp} = 10.$, $Mh_{min} = 10$). The

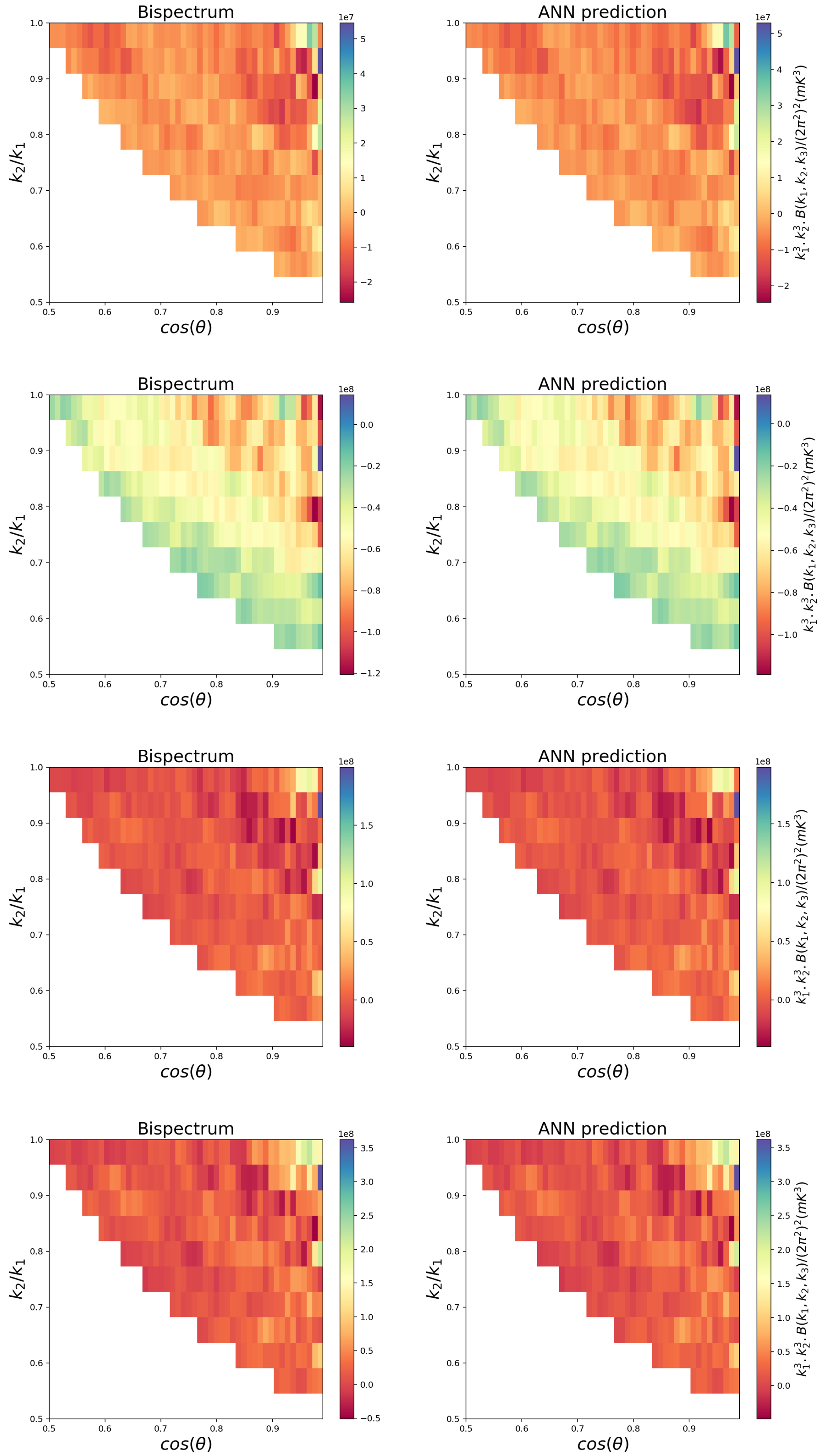

Figure 4.8: Test predictions using ANN emulator for bispectrum with $k_1 = 1.5\,\mathrm{Mpc}^{-1}$

True Values	ζ	$R_{mfp}(Mpc)$	$Mh_{min}(10^8 M_\odot)$
Case-1	20.	20.	500.
Case-2	18.	45.	100.
Case-3	130.	30.	100.
Case-4	140.	30.	100.
Case-5	170.	30.	800.
Case-6	45.	60.	700.
Case-7	40.	30.	500.
Case-8	30.	10.	10.
Case-9	23.21	20.	1000.
Case-10	200.	30.	800.
Case-11	130.	45.	200.
Case-13	100.	10.	100.
Case-14	55.	10.	100.
Case-15	50.	60.	800.

Table 4.1: This shows the 15 test sets of parameters for which we have tested our Bayesian inference pipeline.

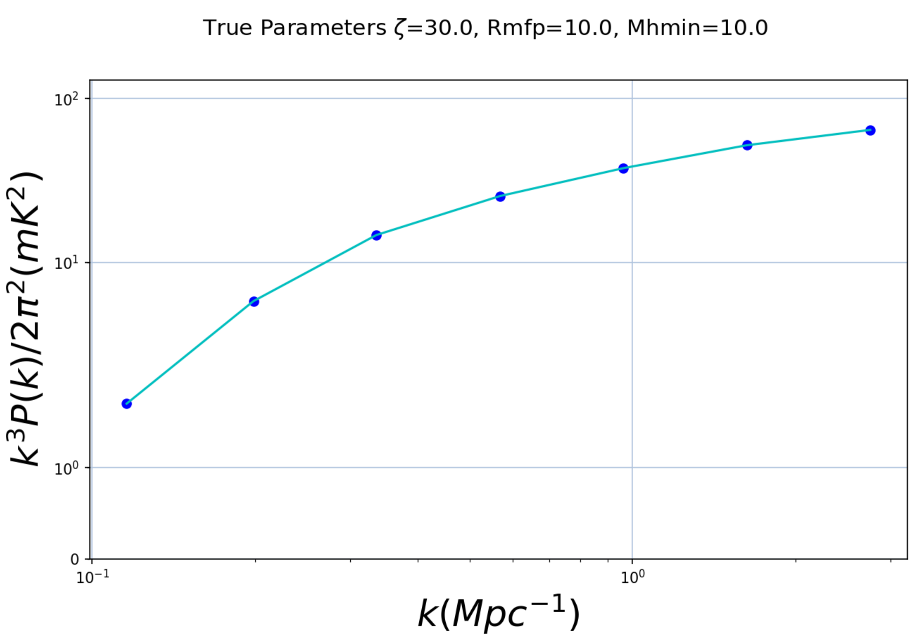

Figure 4.9: Powerspectrum for Case-8.

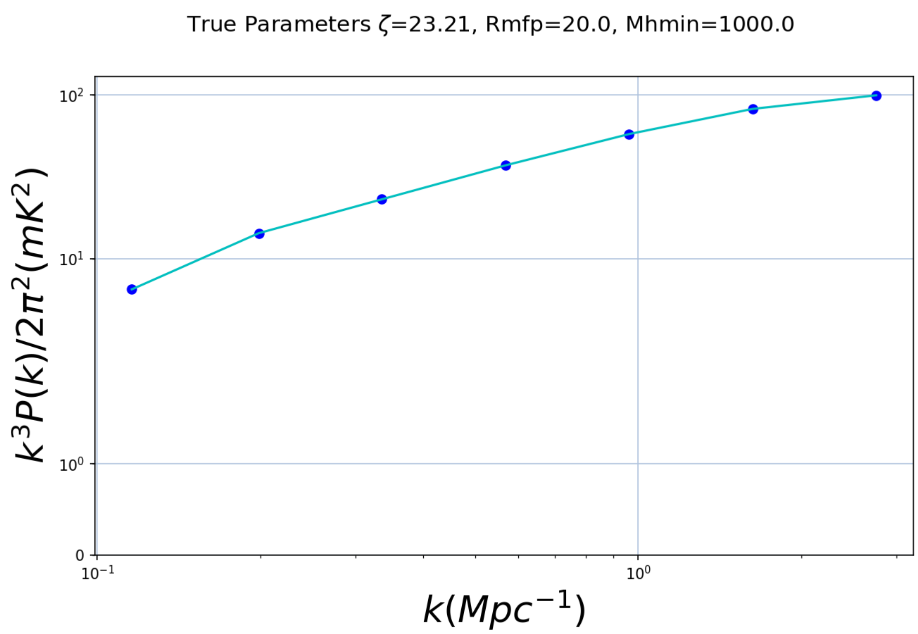

Figure 4.10: Powerspectrum for Case-9

Mh_{min} implies that the smallest halo mass is of the order of $10 \times 10^9 M_\odot$, which is on the lower-end of halo mass function. This means that almost the entire population of halo at any given time contributes to the EoR. As the low mass halos are more numerous than the high mass halos, one would expect to have alarge number of small ionized regions along with few large ionized regions by this redshift (given that $\zeta = 30$ and $R_{mfp} = 10$ Mpc also has somewhat fiducial values). Thus the signature of this set of parameters will be a significant amplitude in the powerspectrum at large k bins, whereas a relatively small amplitude in the powerspectrum at small k modes.

Next we take up **Case-9**. Figure 4.10 shows the powerspectrum for Case-9 which has $\zeta = 23.21$, $R_{mfp} = 20$ Mpc, $Mh_{min} = 1000$. This set of parameters lead to an end of reionization by $z \sim 6$. This also provides us a Thompson-scattering optical depth of CMBR photons that is in agreement with the estimates from the PLANCK data. Here we have $Mh_{min} = 1000 \times 10^9 M_\odot$, which implies only the high mass end of the halo mass function is contributing as ionizing sources. As they are less numerous, they lead to a less number of ionized regions by this time. However, as they are more massive the number of ionizing photons produced by them is quite high, which leads to larger ionized regions. This increases the fluctuations in the large length scales compared to Case-8, which implies more power at small k modes.

We find that in general, ζ and Mh_{min} can significantly alter the ionization state of IGM. However, the sensitivity of R_{mfp} can be seen only in a very small region of the parameter space. If we choose high ζ and low-moderate Mh_{min}, which will lead to large number of small mass halos with high ionizing efficiency, in this region of the parameter space the ionization state of the IGM will be sensitive to R_{mfp}.

4.2.2 Bayesian inference: parameter estimation using individual statistics

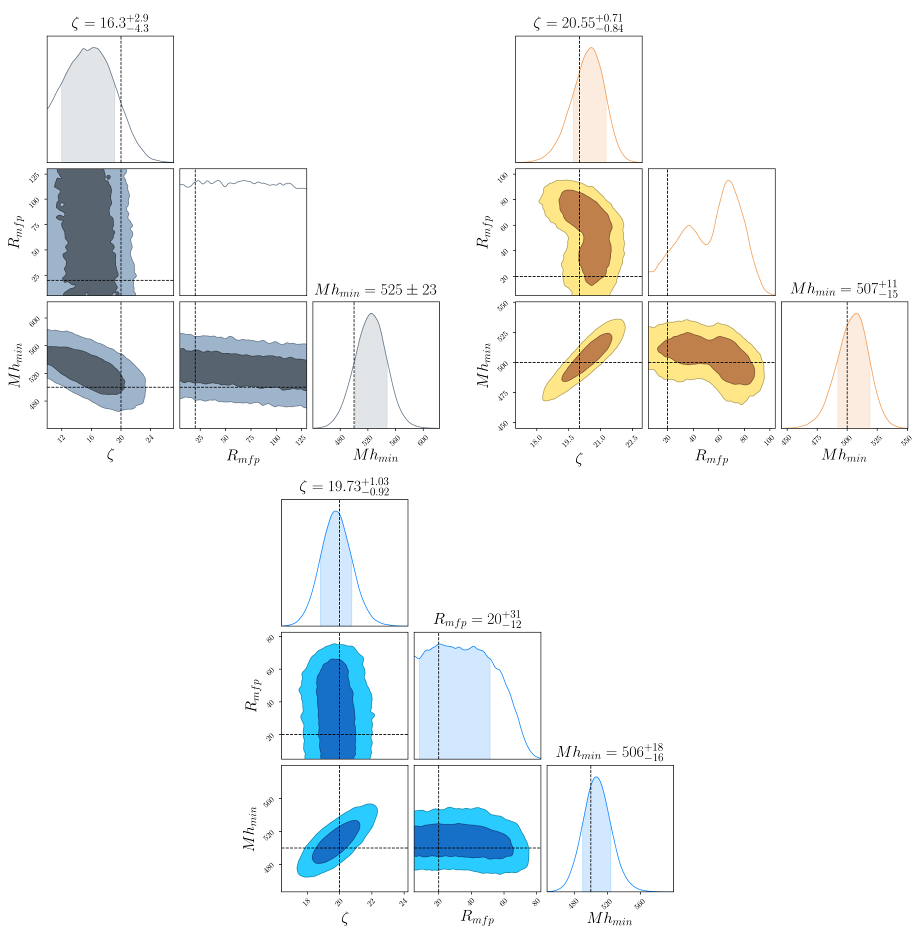

Figure 4.11: Parameter estimation using individual statistics : P(k) (grey), B(k_1=0.3) (yellow), B(k_1=1.5) (blue) for Case-1 (ζ = 20, R_{mfp} = 20 Mpc, Mh_{min} = 500).

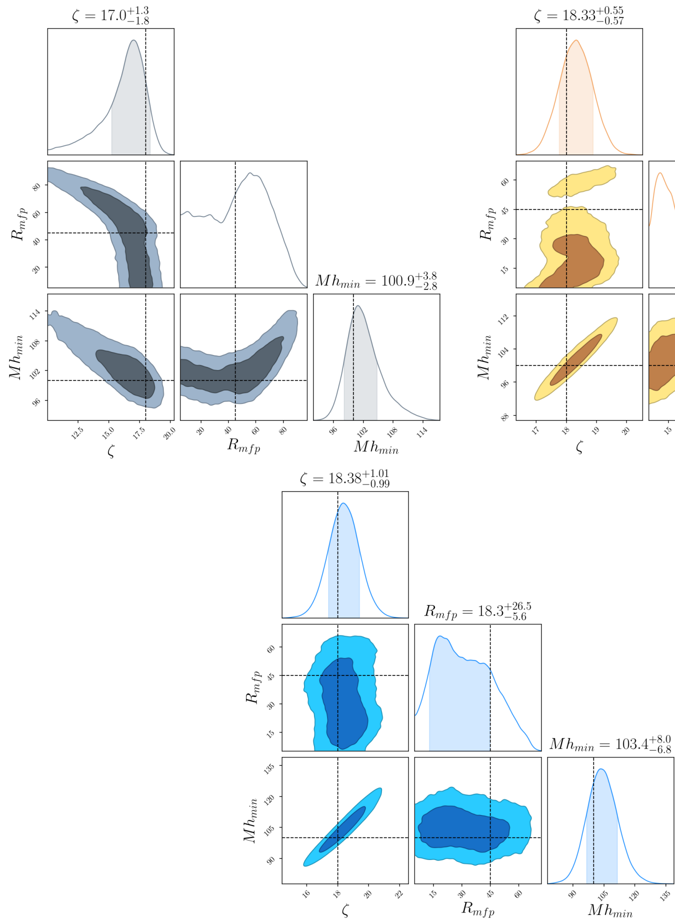

Figure 4.12: Parameter estimation using individual statistics : P(k) (grey), B(k_1 = 0.3) (yellow), B(k_1 = 1.5) (blue) for Case-2 (ζ = 18, R_{mfp} = 45 Mpc, Mh_{min} = 100).

Figure 4.11 and 4.12 show the estimated posterior distributions for the three parameters from our Bayesian inference pipeline for three individual statistics (powerspectrum and bispectra for k_1 = 0.3 and 1.5 Mpc^{-1}) for the test Case-1 and 2 respectively. These plots gives us a clear idea about which parameter is sensitive to which statistics. The intersecting points of the dashed show the true values of the parameters. It is clear from these plots that the posterior is not a simple multidimensional Gaussian in any of the cases. The powerspectrum puts a rather weaker constraints on the parameters in both cases compared to the bispectra. Among the bispectra fro two k_1 modes considered here, $B(k_1 = 0.3)$ puts a tighter constrain on the parameters compared to $B(k_1 = 1.5)$. However, they have significantly different shapes in their posterior, which implies they probe different non-Gaussian characteristics. We also find that powerspectrum is not sensitive to the R_{mfp} in this regime, whereas $B(k_1 = 1.5)$ manages to put somewhat weaker constraints on the R_{mfp} in Case 1 and 2.

4.2.3 Joint Posterior Plots

Though Figure 4.11 and 4.12 provides us some insight about how different statistics constrain different parameters (through the shape of their individual posteriors), still own would gain more insight when all of these posteriors from different statistics would be plotted together. In its true sense just this kind of joint posterior plots cannot be treated as joint parameter estimation. However, they provide us some sort of pictorial insight to what a joint parameter estimation would lead to. We next show the joint posterior plots for all test cases in Figures 4.13-4.26.

We next divide our test cases based on to what extent the parameters have been constrained for them. First we focus on Cases 1, 6, and 7 (Figures 4.13 , 4.18, 4.19 respectively). In all of these cases the parameters ζ and Mh_{min} are both well constrained by all the statistics and their mean/median are very close to the true values of the parameters. The other important point to note is that the error elipse for the powerspectrum is differently alligned compared to the error ellipses of the bispectra. Additionally, the powerspectra has a larger uncertainty for its estimated parameters compared to the bispectra. We also observe that the bispectrum for larger $k_1 = 1.5\,\mathrm{Mpc}^{-1}$ has put somewhat weaker constraint even around R_{mfp} parameter.

For the test Cases 3, 4, 13 (Figures 4.15 , 4.16, 4.26), we find some of the best obtained constraints on the parameters ζ and Mh_{min}. Interestingly, here the error ellipses for all statistics are alligned with each other. The marginalized posterior for ζ and Mh_{min} also appears to be well behaved Gaussian. On top of these the bispectrum for $k_1 = 0.3\,\mathrm{Mpc}^{-1}$ provides a somewhat weak constrain on R_{mfp}.

In Cases 5 and 10 (Figures 4.17 and 4.22) we find that all statistics provide a good constrain on ζ and Mh_{min}. In these two cases both bispectra are able to set weaker constrains on the R_{mfp} as well.

The Case 8 (Figure 4.20), is an example of MCMC sampling being restricted by the boundary of the parameter space. Due to our simulation resolution limits we were unable to go below a certain value of Mh_{min}. This has defined our lower limit of Mh_{min} range for the emulator as well. This has lead to a trimmed posterior.

For all other test cases to a significant degree the parameters ζ and Mh_{min} reasonably constrained.

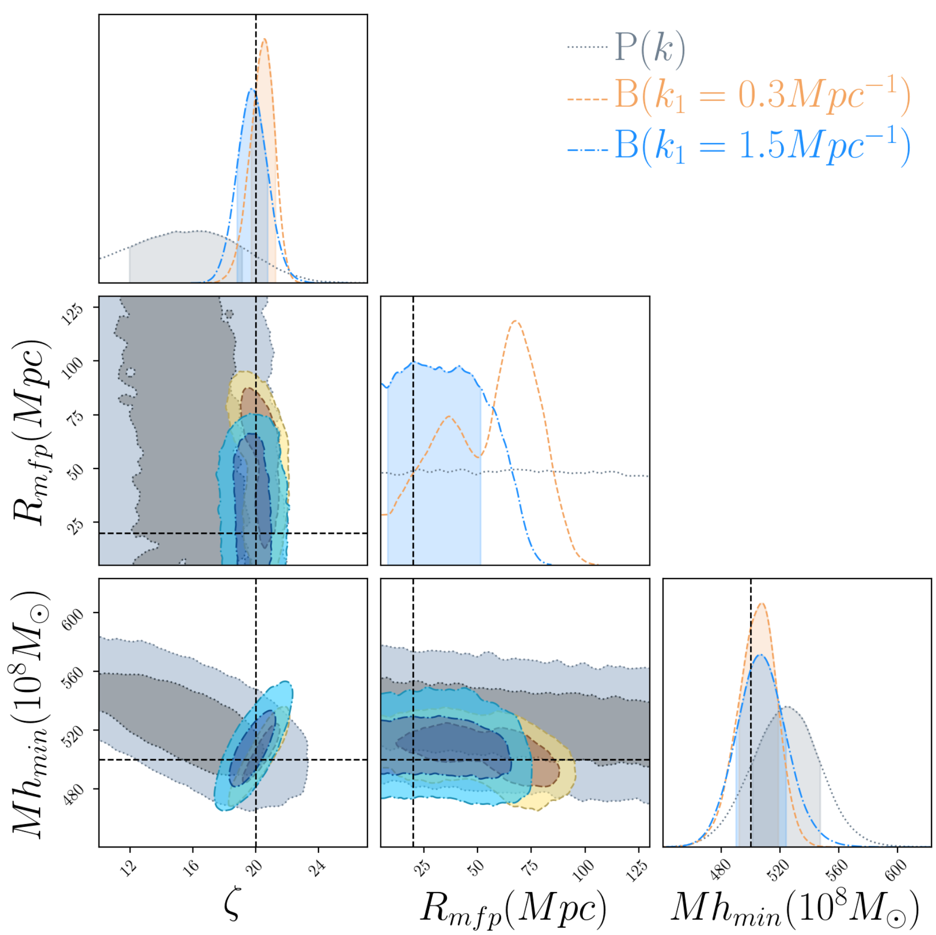

Figure 4.13: Joint Plot-1, true values (20, 20, 500)

Statistics	ζ	$R_{mfp}(Mpc)$	$Mh_{min}(10^8 M\odot)$
$P(k)$	$16.3^{+2.9}_{-4.3}$	-	525 ± 23
$B(k_1 = 0.3Mpc^{-1})$	$20.55^{+0.71}_{-0.84}$	-	507^{+11}_{-15}
$B(k_1 = 1.5Mpc^{-1})$	$19.73^{+1.03}_{-0.92}$	20^{+31}_{-12}	506^{+18}_{-16}

Table 4.2: Case-1

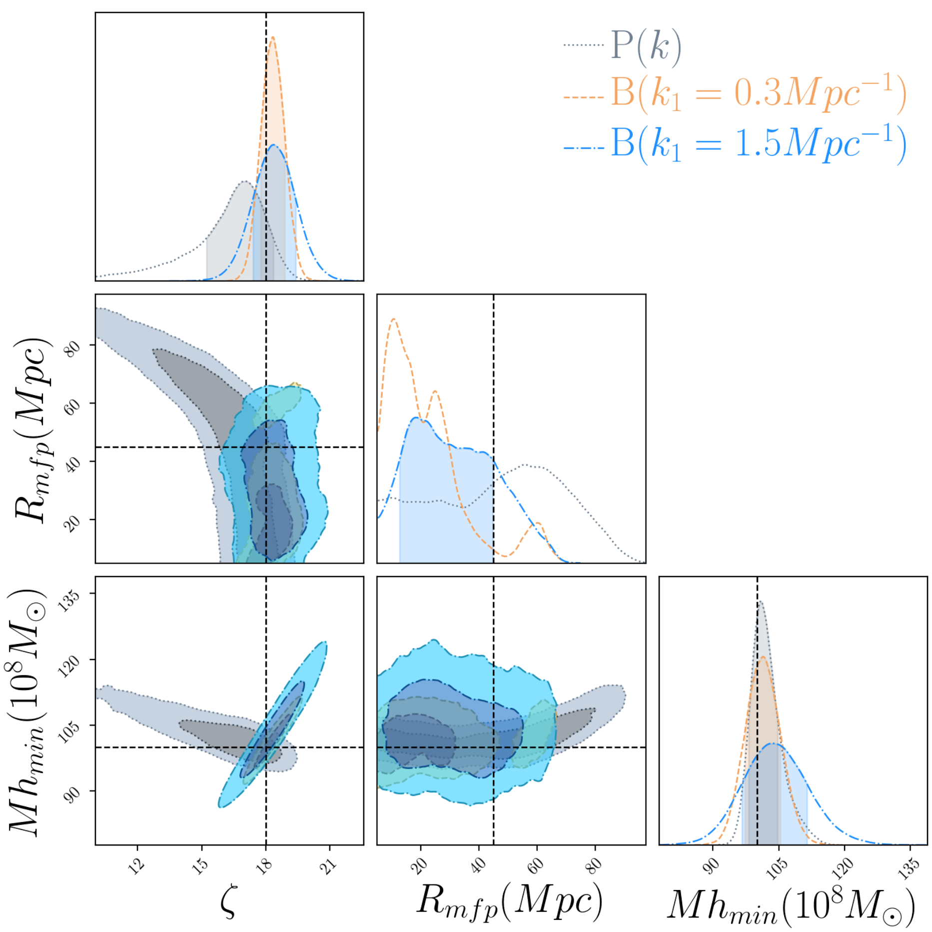

Figure 4.14: Joint Plot-2, true values (18, 45, 100)

Statistics	ζ	$R_{mfp}(Mpc)$	$Mh_{min}(10^8 M\odot)$
$P(k)$	$17.0^{+1.3}_{-1.8}$	-	$100.9^{+3.8}_{-2.8}$
$B(k_1 = 0.3Mpc^{-1})$	$18.33^{+0.55}_{-0.57}$	-	$101.4^{+4.0}_{-4.1}$
$B(k_1 = 1.5Mpc^{-1})$	$18.38^{+1.01}_{-0.99}$	$18.3^{+26.5}_{-5.6}$	$103.4^{+8.0}_{-6.8}$

Table 4.3: Case-2

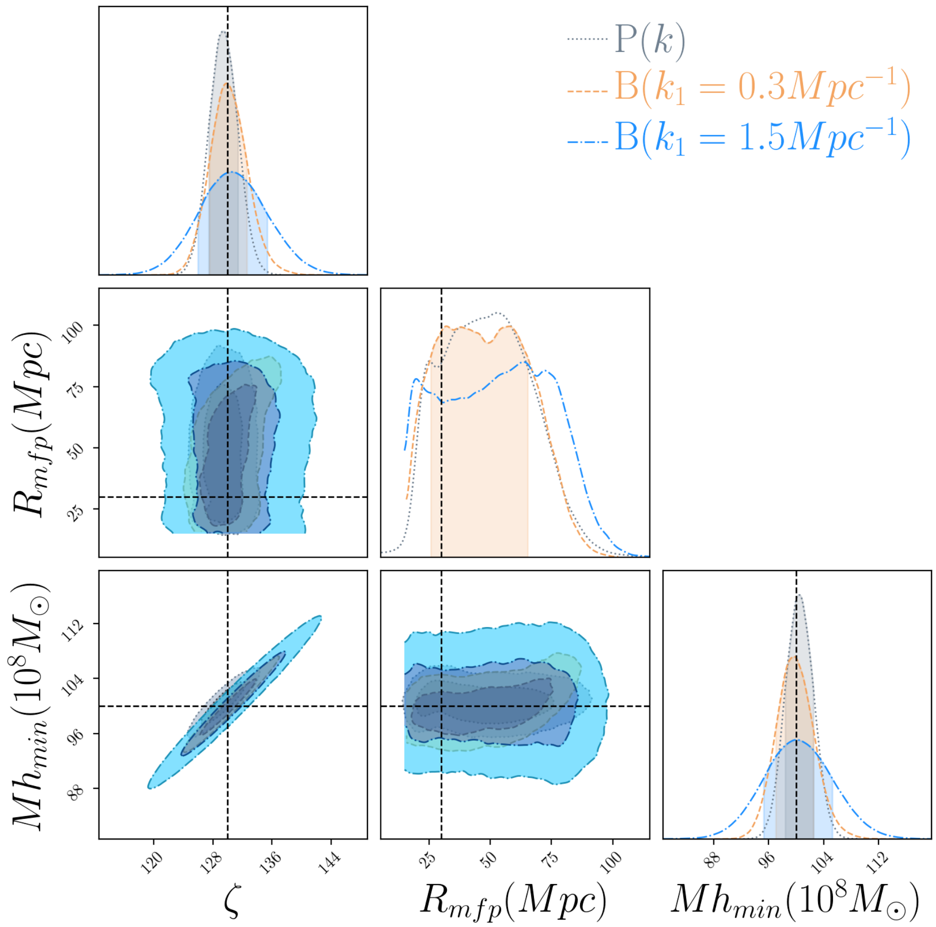

Figure 4.15: Joint Plot-3, true values (130, 30, 100)

Statistics	ζ	$R_{mfp}(Mpc)$	$Mh_{min}(10^8 M\odot)$
$P(k)$	$129.5^{+1.9}_{-2.0}$	-	$100.5^{+2.1}_{-2.0}$
$B(k_1 = 0.3Mpc^{-1})$	$129.9^{+2.7}_{-2.3}$	$32.0^{+33.2}_{-6.3}$	$99.7^{+2.9}_{-2.6}$
$B(k_1 = 1.5Mpc^{-1})$	130.6 ± 4.7	-	$100.2^{+5.1}_{-4.9}$

Table 4.4: Case-3

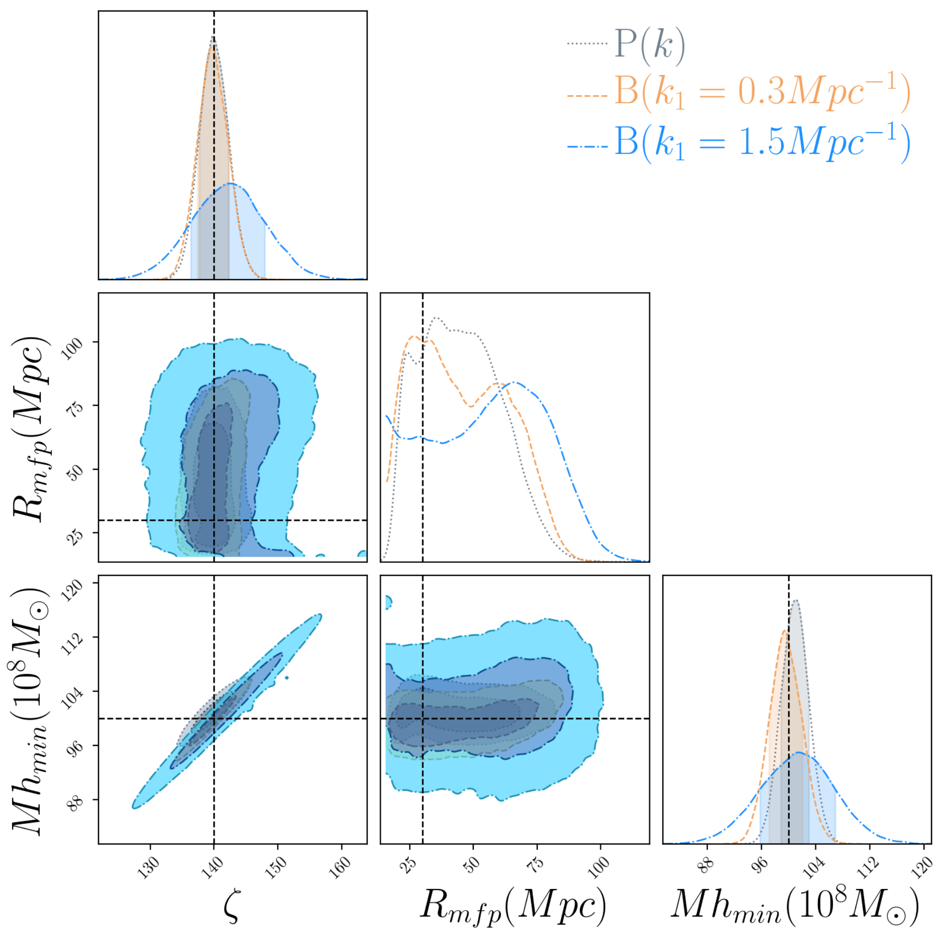

Figure 4.16: Joint Plot-4, true values (140, 30, 100)

Statistics	ζ	$R_{mfp}(Mpc)$	$Mh_{min}(10^8 M\odot)$
$P(k)$	$139.9^{+2.4}_{-2.2}$	-	$101.1^{+1.9}_{-2.2}$
$B(k_1 = 0.3Mpc^{-1})$	$139.7^{+2.6}_{-2.3}$	-	$99.4^{+2.6}_{-2.3}$
$B(k_1 = 1.5Mpc^{-1})$	$142.6^{+5.4}_{-6.2}$	-	$101.5^{+5.4}_{-5.7}$

Table 4.5: Case-4

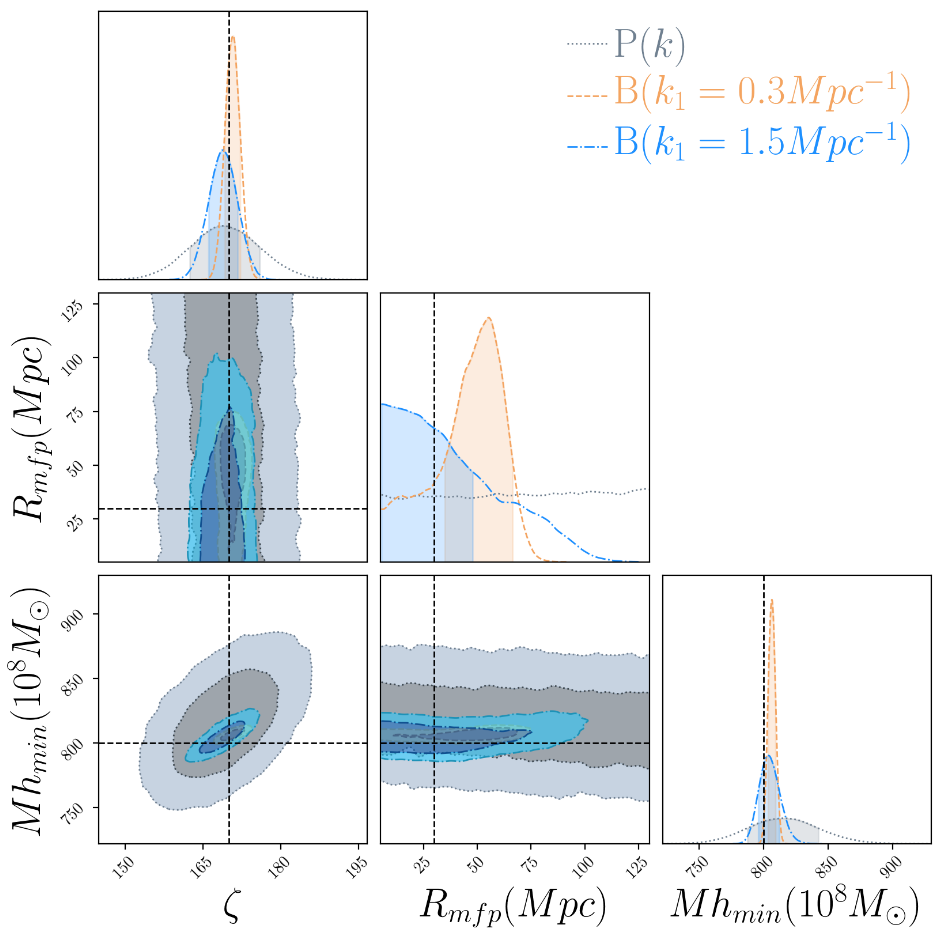

Figure 4.17: Joint Plot-5, true values (170, 30, 800)

Statistics	ζ	$R_{mfp}(Mpc)$	$Mh_{min}(10^8 M\odot)$
$P(k)$	$168.9^{+7.1}_{-6.4}$	-	815^{+27}_{-28}
$B(k_1 = 0.3Mpc^{-1})$	$170.8^{+1.4}_{-1.6}$	56^{+11}_{-21}	$806.4^{+2.8}_{-3.0}$
$B(k_1 = 1.5Mpc^{-1})$	$168.9^{+2.9}_{-2.8}$	5^{+42}_{-0}	$803.7^{+8.4}_{-7.9}$

Table 4.6: Case-5

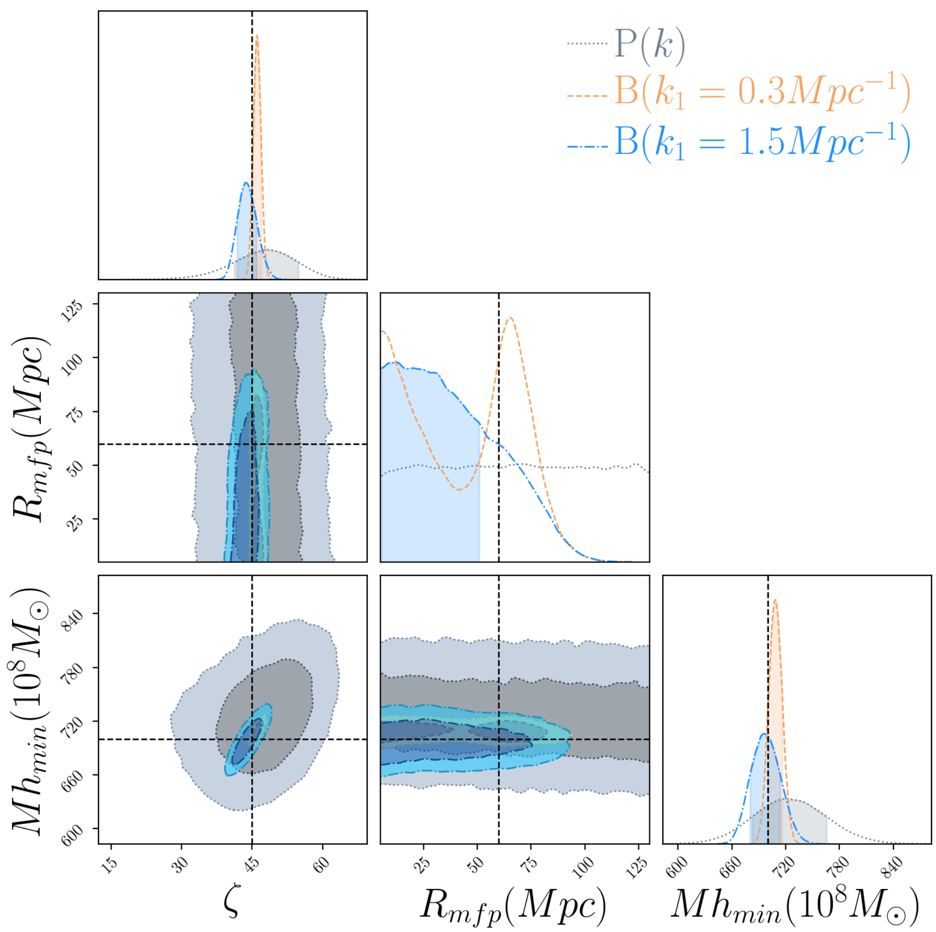

Figure 4.18: Joint Plot-6, true values (45, 60, 700)

Statistics	ζ	$R_{mfp}(Mpc)$	$Mh_{min}(10^8 M\odot)$
$P(k)$	$48.2^{+6.6}_{-7.0}$	-	720^{+45}_{-37}
$B(k_1 = 0.3Mpc^{-1})$	$46.05^{+0.84}_{-0.81}$	-	$708.4^{+7.8}_{-7.4}$
$B(k_1 = 1.5Mpc^{-1})$	$43.7^{+2.2}_{-1.9}$	$12.6^{+38.3}_{-7.3}$	697^{+17}_{-16}

Table 4.7: Case-6

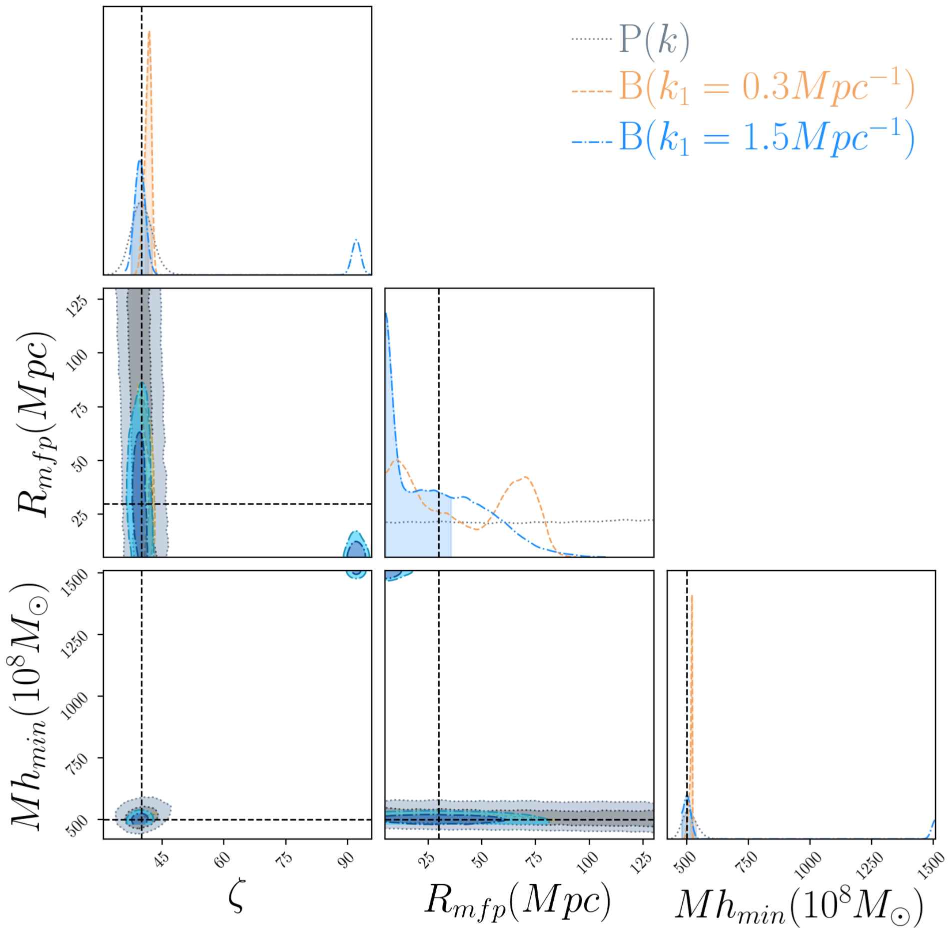

Figure 4.19: Joint Plot-7, true values (40, 30, 500)

Statistics	ζ	$R_{mfp}(Mpc)$	$Mh_{min}(10^8 M\odot)$
$P(k)$	$39.9^{+2.7}_{-2.5}$	-	507^{+30}_{-28}
$B(k_1 = 0.3Mpc^{-1})$	$41.86^{+0.67}_{-0.96}$	-	$521.5^{+2.5}_{-6.1}$
$B(k_1 = 1.5Mpc^{-1})$	$39.5^{+2.1}_{-1.9}$	5^{+30}_{-0}	500^{+38}_{-20}

Table 4.8: Case-7

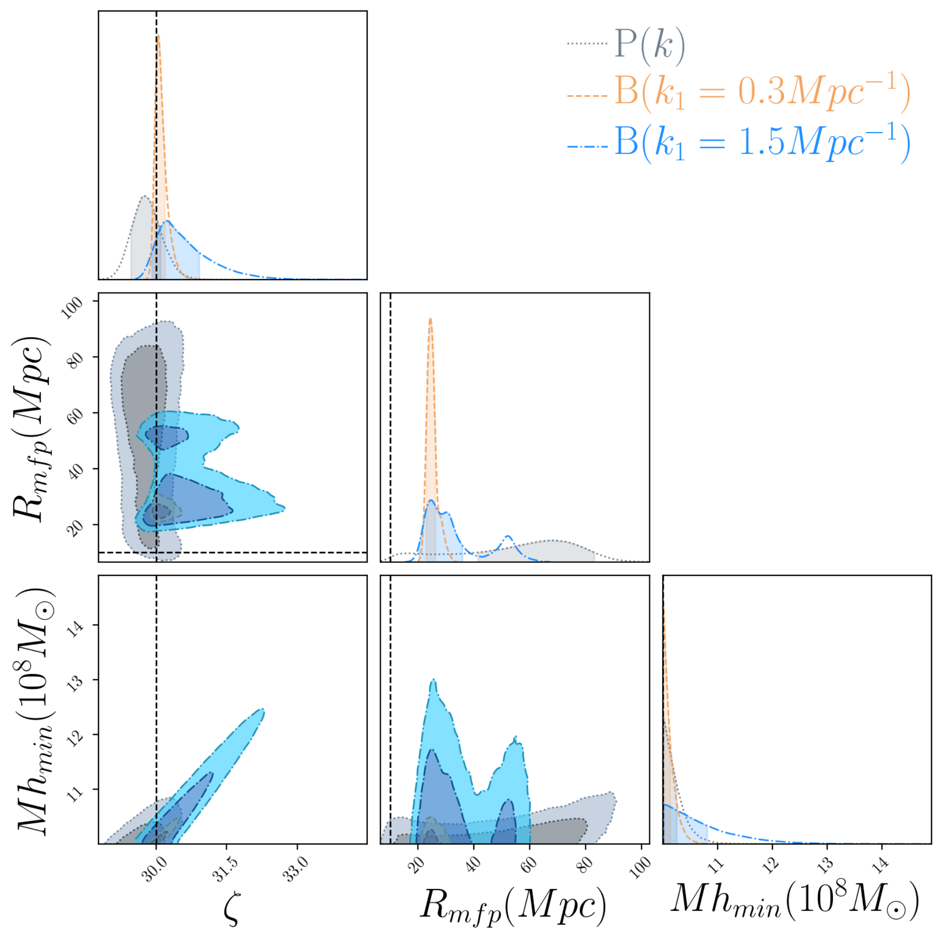

Figure 4.20: Joint Plot-8, true values (30, 10, 10)

Statistics	ζ	$R_{mfp}(Mpc)$	$Mh_{min}(10^8 M\odot)$
$P(k)$	$17.0^{+1.3}_{-1.8}$	-	$100.9^{+3.8}_{-2.8}$
$B(k_1 = 0.3Mpc^{-1})$	$18.33^{+0.55}_{-0.57}$	-	$101.4^{+4.0}_{-4.1}$
$B(k_1 = 1.5Mpc^{-1})$	$18.38^{+1.01}_{-0.99}$	$18.3^{+26.5}_{-5.6}$	$103.4^{+8.0}_{-6.8}$

Table 4.9: Case-8

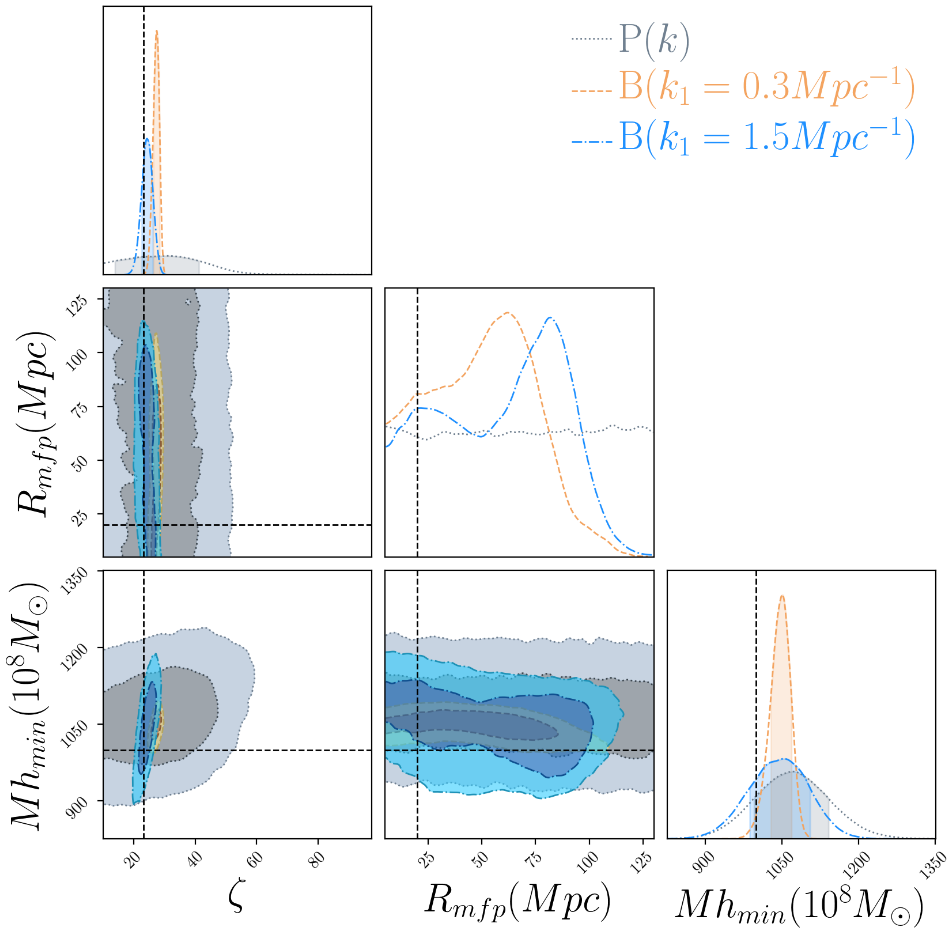

Figure 4.21: Joint Plot-9, true values (23.21, 20, 1000)

Statistics	ζ	$R_{mfp}(Mpc)$	$Mh_{min}(10^8 M\odot)$
$P(k)$	31^{+10}_{-17}	-	1073^{+68}_{-78}
$B(k_1 = 0.3Mpc^{-1})$	$27.47^{+0.99}_{-1.08}$	-	1051^{+18}_{-22}
$B(k_1 = 1.5Mpc^{-1})$	$24.3^{+1.9}_{-1.7}$	-	1054^{+52}_{-66}

Table 4.10: Case-9

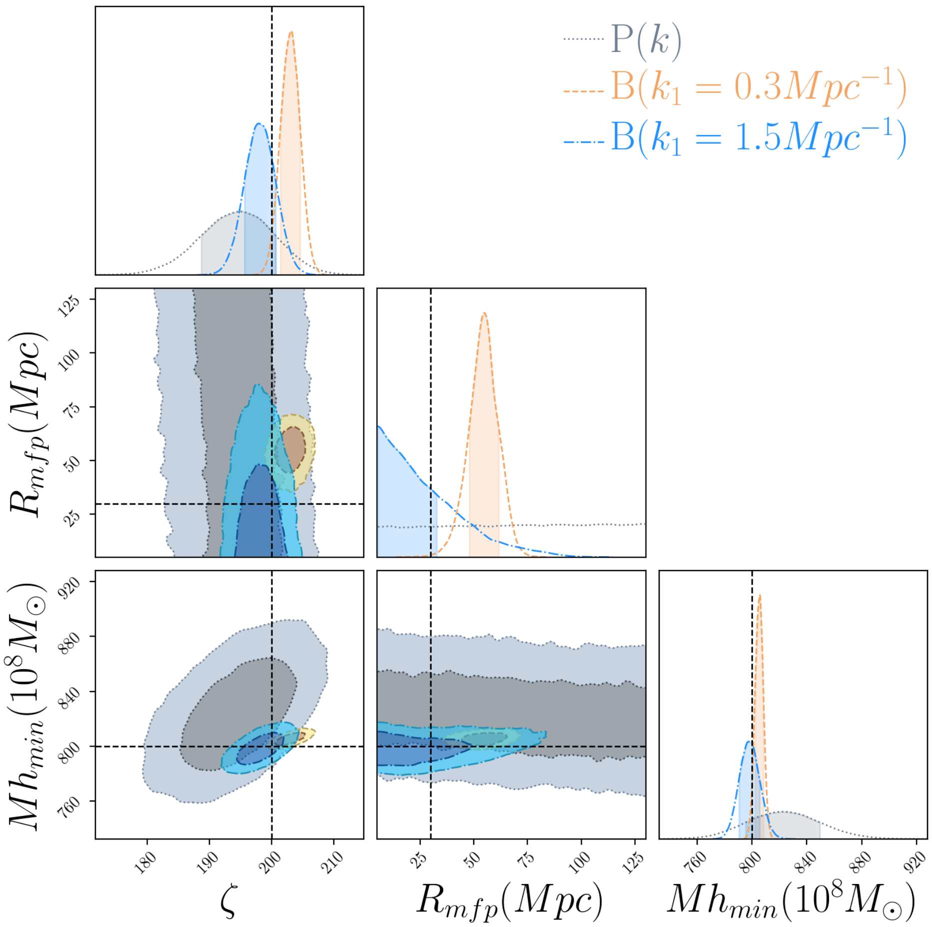

Figure 4.22: Joint Plot-10, true values (20, 30, 800)

Statistics	ζ	$R_{mfp}(Mpc)$	$Mh_{min}(10^8 M\odot)$
$P(k)$	$195.0^{+5.7}_{-6.3}$	-	822 ± 27
$B(k_1 = 0.3Mpc^{-1})$	$203.2^{+1.4}_{-1.8}$	$55.0^{+6.6}_{-7.1}$	$805.8^{+2.5}_{-3.5}$
$B(k_1 = 1.5Mpc^{-1})$	$197.9^{+2.8}_{-2.2}$	5^{+27}_{-0}	$797.4^{+8.2}_{-6.9}$

Table 4.11: Case-10

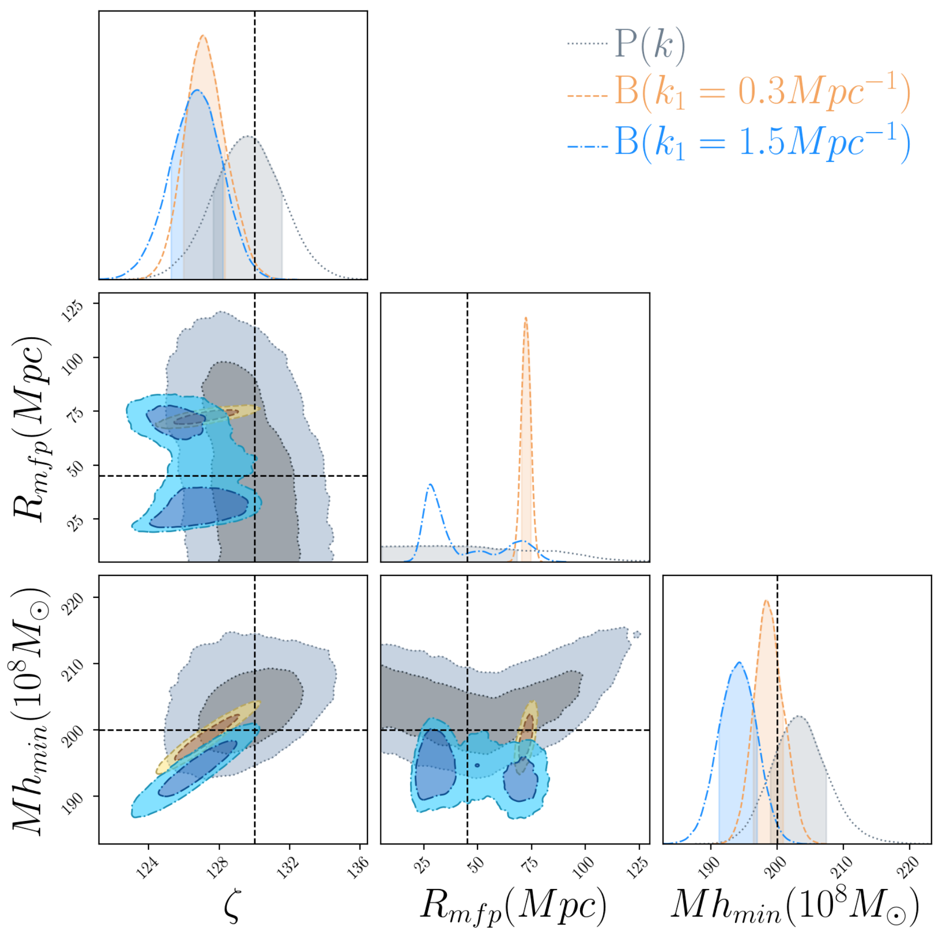

Figure 4.23: Joint Plot-11, true values (130, 45, 200)

Statistics	ζ	$R_{mfp}(Mpc)$	$Mh_{min}(10^8 M\odot)$
$P(k)$	$129.6^{+2.0}_{-1.9}$	32^{+37}_{-26}	$203.3^{+4.1}_{-4.3}$
$B(k_1 = 0.3Mpc^{-1})$	$127.1^{+1.2}_{-1.1}$	$72.4^{+2.2}_{-2.1}$	$198.4^{+2.5}_{-2.0}$
$B(k_1 = 1.5Mpc^{-1})$	$126.8^{+1.4}_{-1.5}$	-	$194.4^{+2.6}_{-3.1}$

Table 4.12: Case-11

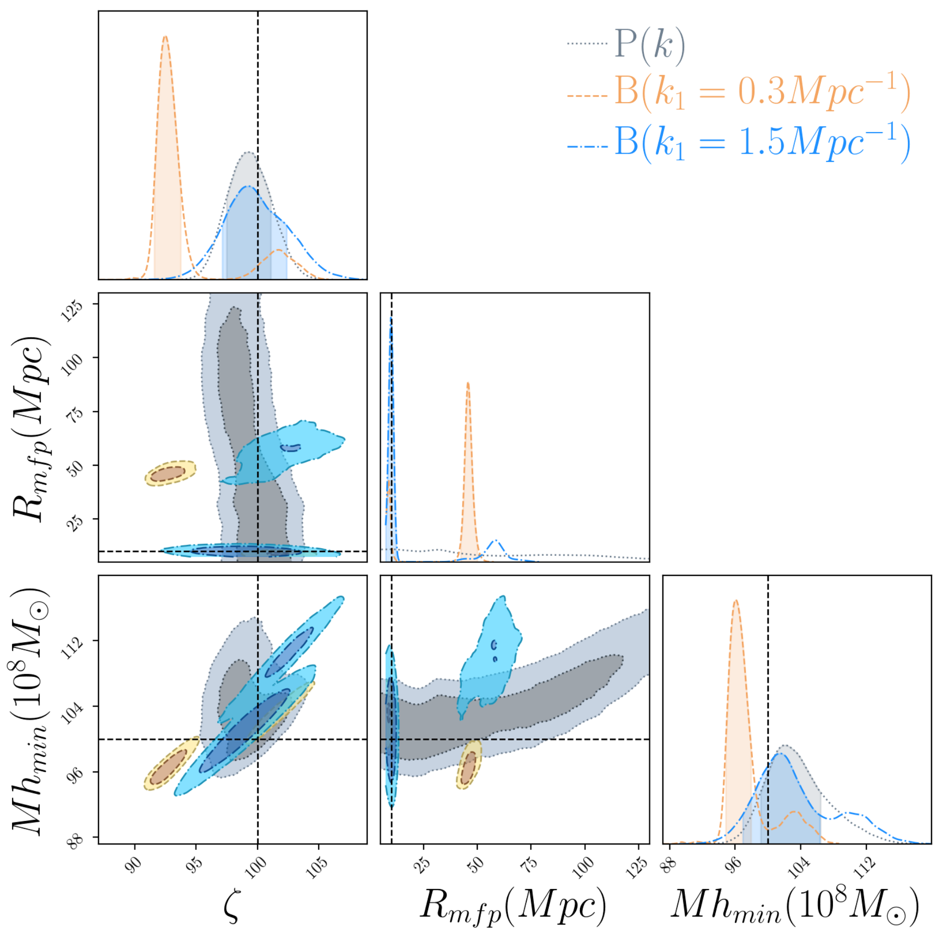

Figure 4.24: Joint Plot-12, true values (100, 10, 100)

Statistics	ζ	$R_{mfp}(Mpc)$	$Mh_{min}(10^8 M\odot)$
$P(k)$	$99.2^{+1.9}_{-1.7}$	-	$102.1^{+4.4}_{-3.0}$
$B(k_1 = 0.3Mpc^{-1})$	$92.48^{+1.25}_{-0.90}$	$45.7^{+4.3}_{-3.4}$	$96.1^{+1.8}_{-1.2}$
$B(k_1 = 1.5Mpc^{-1})$	$99.2^{+3.1}_{-2.1}$	$9.8^{+2.5}_{-2.4}$	$101.4^{+4.9}_{-4.5}$

Table 4.13: Case-12

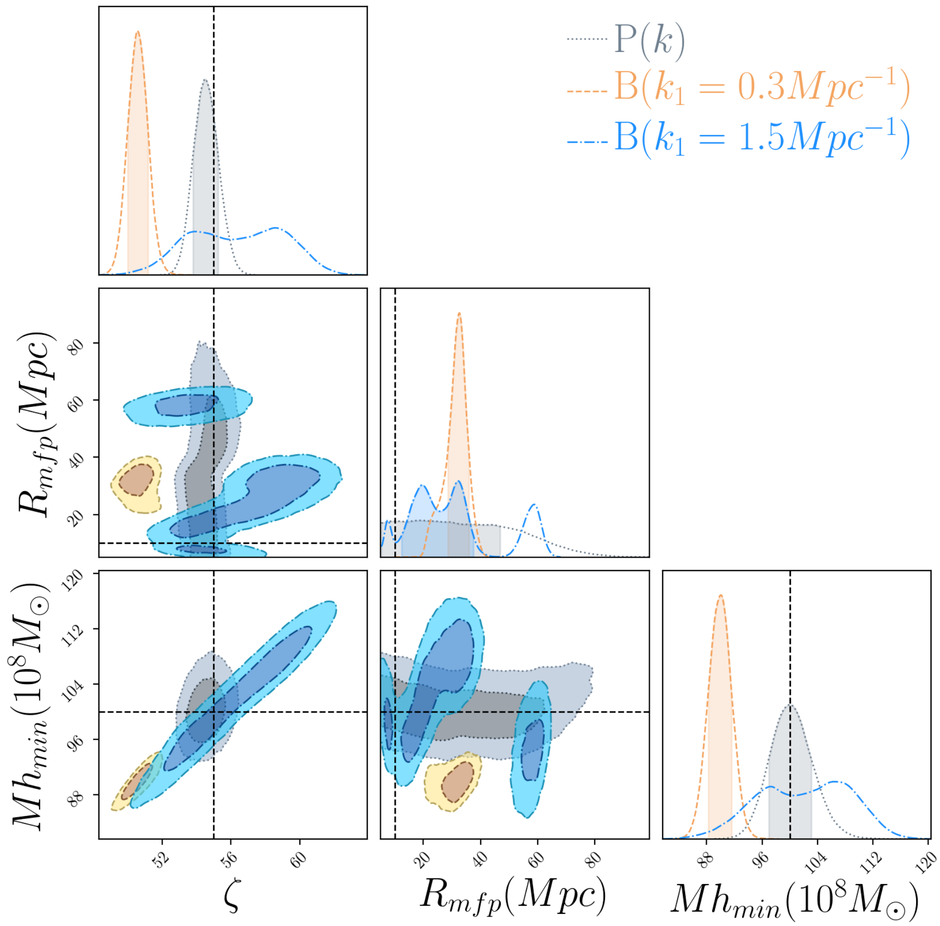

Figure 4.25: Joint Plot-13, true values (55, 10, 100)

Statistics	ζ	$R_{mfp}(Mpc)$	$Mh_{min}(10^8 M\odot)$
$P(k)$	$54.49^{+0.77}_{-0.69}$	19^{+28}_{-14}	$100.0^{+3.1}_{-3.0}$
$B(k_1 = 0.3Mpc^{-1})$	$50.57^{+0.61}_{-0.57}$	$32.6^{+3.3}_{-4.0}$	$90.0^{+1.5}_{-1.8}$
$B(k_1 = 1.5Mpc^{-1})$	-	$32.1^{+5.4}_{-19.8}$	-

Table 4.14: Case-13

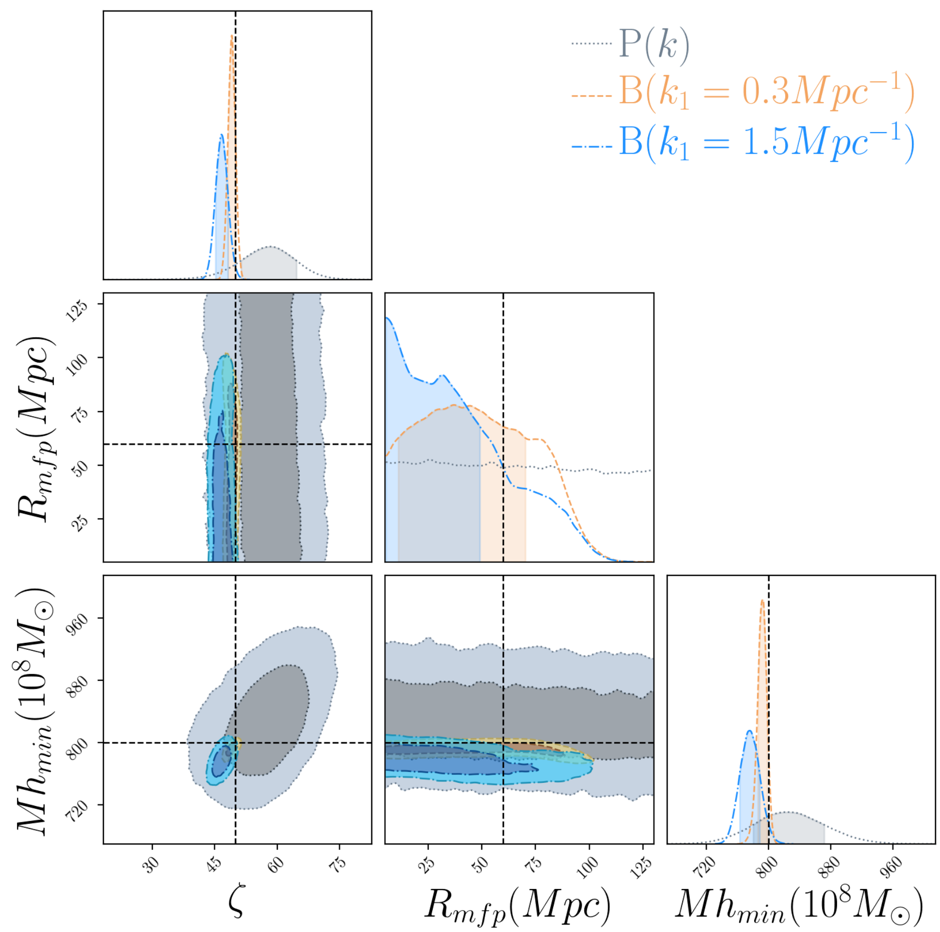

Figure 4.26: Joint Plot-14, true values (50, 60, 800)

Statistics	ζ	$R_{mfp}(Mpc)$	$Mh_{min}(10^8 M\odot)$
$P(k)$	$58.3^{+6.3}_{-7.0}$	-	824^{+48}_{-43}
$B(k_1 = 0.3Mpc^{-1})$	$49.13^{+0.90}_{-0.92}$	37^{+33}_{-26}	$792.8^{+5.4}_{-6.6}$
$B(k_1 = 1.5Mpc^{-1})$	46.7 ± 1.5	5^{+44}_{-0}	776 ± 13

Table 4.15: Case-14

4.3 Summary

The outcome this thesis can be summarized in terms of the following few points:

- In this thesis, for the first time, we have successfully demonstrated that it is possible to build a ANN based emulator for the 21-cm bispectrum from the EoR.

- We have tested our 21-cm bispectrum emulator for a wide range of parameter values and found its outcome to be quite robust so that it can be safely used in the Bayesian inference exercise for the EoR parameter estimation.

- Through our comparative Bayesian parameter estimation exercise we have shown that the bispectrum provides a much tighter constraints than the powerspectrum for the EoR parameters ζ and Mh_{min}.

- We have further shown that the parameter R_{mfp}, which is quite insensitive to the powerspectrum based parameter estimation and earlier have been reported to be a dependent parameter (Binnie and J R Pritchard 2019), can be constrained weakly by the bispectrum in some regions of the parameter space.

- Through overlapping posterior plots obtained from different statistics we have shown that one can put a much tighter constraints on the EoR parameters if simultaneouly independent signal statistics such as powerspectrum and bispectrum are used.

- The efficiency of bispectrum in putting tighter constraints on the EoR parameters can be ascribed to the fact that being a higher order statistic bispectrum contains more information than the powerspectrum.

Chapter 5

Future Scopes

In this project we have considered a simpler three parameter model of the EoR to build our ANN based signal emulator and the Bayesian inference pipeline. We have shown that it is possible to emulate the 21-cm EoR bispectrum reliably enough that it can be used in the Bayesian inference exercise. This project is a stepping stone in the direction of developing statistical inference tools in the relm of BIG DATA driven astronomy with next-generation telescopes. There are many directions in which this project can be progressed further. We note down a few of them below:

- Our entire analysis uses sample variance in the place of the error covariance matrix for the Bayesian inference. In a realistic scenario, the error covariance matrix will contain noise uncertainty in the measurement of the target statistics along with its sample variance. We are presently working on to take into account the system noise contribution to the error covariance matrix for an experiment like the SKA-low using the formalism of (Shaw, Bharadwaj, and Mondal 2019).

- Here we have considered an idealistic scenario where there is no foreground in the data. In case of imperfect foreground subtraction, the residual foreground in the data will impact the parameter estimation. We plan to explore the impact this realistic scenario in our follow up work.

- The training data set that we have used to train our emulator is sampled from an almost uniform grid from the parameter space. This is not the ideal way to building a training set for an emulator. We plan to use the latin-hypercube sampling of the parameter space for building up our training set (Claude J. Schmit and Jonathan R. Pritchard 2017). This is expected to make the training of the ANN far more robust.

- We have considered our EoR 21-cm signal to come from a single redshift. In a realistic observation the signal will come from a range of redshifts. This will make the constraints on the parameters much tighter.

- One can extend this work further to emulate the signal from the Cosmic Dawn where the number of parameter involved in the model is more than the EoR.

- Here for the Bayesian parameter estimation we have used only one model of the reionization. However, in a realistic scenario one should consider all possible models of the EoR. One can then first perform a Bayesian model selection exercise, followed by a parameter estimation for the preferred model.

- Here we have done independent parameter estimation using each of the 21-cm statistics. However, this parameter estimation would have been even more robust if it was done jointly using all available statistics together.

Bibliography

Agarwal, Shankar et al. (Feb. 2014). "pkann – II. A non-linear matter power spectrum interpolator developed using artificial neural networks". In: *Monthly Notices of the Royal Astronomical Society* 439.2, pp. 2102–2121. ISSN: 1365-2966. DOI: 10.1093/mnras/stu090. URL: http://dx.doi.org/10.1093/mnras/stu090.

Akeret, Joël et al. (Aug. 2013). "CosmoHammer: Cosmological parameter estimation with the MCMC Hammer". In: *Astronomy and Computing* 2, pp. 27–39. ISSN: 2213-1337. DOI: 10.1016/j.ascom.2013.06.003. URL: http://dx.doi.org/10.1016/j.ascom.2013.06.003.

Ali, Sk. Saiyad, Somnath Bharadwaj, and Jayaram N. Chengalur (Mar. 2008). "Foregrounds for redshifted 21-cm studies of reionization: Giant Meter Wave Radio Telescope 153-MHz observations". In: *Monthly Notices of the Royal Astronomical Society* 385.4, pp. 2166–2174. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2008.12984.x. eprint: https://academic.oup.com/mnras/article-pdf/385/4/2166/18235111/mnras0385-2166.pdf. URL: https://doi.org/10.1111/j.1365-2966.2008.12984.x.

Barkana, Rennan and Abraham Loeb (July 2001). "In the beginning: the first sources of light and the reionization of the universe". In: *Physics Reports* 349.2, pp. 125–238. ISSN: 0370-1573. DOI: 10.1016/s0370-1573(01)00019-9. URL: http://dx.doi.org/10.1016/S0370-1573(01)00019-9.

Barry, N. et al. (Oct. 2019). "Improving the Epoch of Reionization Power Spectrum Results from Murchison Widefield Array Season 1 Observations". In: *The Astrophysical Journal* 884.1, p. 1. DOI: 10.3847/1538-4357/ab40a8. URL: https://doi.org/10.3847%2F1538-4357%2Fab40a8.

Binnie, T and J R Pritchard (May 2019). "Bayesian model selection with future 21cm observations of the epoch of reionization". In: *Monthly Notices of the Royal Astronomical Society* 487.1, pp. 1160–1177. ISSN: 0035-8711. DOI: 10.1093/mnras/stz1297. eprint: https://academic.oup.com/mnras/article-pdf/487/1/1160/28753391/stz1297.pdf. URL: https://doi.org/10.1093/mnras/stz1297.

Bonaldi, Anna and Michael L Brown (Jan. 2015). "Foreground removal for SKA". In: *Royal Astronomical Society. Monthly Notices* 447.2. arXiv: 1409.5300, pp. 1973–1983. ISSN: 1365-2966. DOI: {10.1093/mnras/stu2601}.

Bradt, Hale (2008). *Astrophysics Processes*.

Carilli, Chris (May 2015). “Square Kilometre Array key science: a progressive retrospective”. In: p. 171. DOI: 10.22323/1.215.0171.

Chardin, Jonathan et al. (Sept. 2019). “A deep learning model to emulate simulations of cosmic reionization”. In: *Monthly Notices of the Royal Astronomical Society* 490.1, pp. 1055–1065. ISSN: 1365-2966. DOI: 10.1093/mnras/stz2605. URL: http://dx.doi.org/10.1093/mnras/stz2605.

Choudhury, T. Roy (2003). *Physics of Structure Formation in the Universe*. arXiv: astro-ph/0305033 [astro-ph].

Collaboration, Planck et al. (2018). *Planck 2018 results. VI. Cosmological parameters*. arXiv: 1807.06209 [astro-ph.CO].

Doussot, Aristide, Evan Eames, and Benoit Semelin (Sept. 2019). “Improved supervised learning methods for EoR parameters reconstruction”. In: *Monthly Notices of the Royal Astronomical Society* 490.1, pp. 371–384. ISSN: 1365-2966. DOI: 10.1093/mnras/stz2429. URL: http://dx.doi.org/10.1093/mnras/stz2429.

Fan, Xiaohui et al. (June 2006). “Constraining the Evolution of the Ionizing Background and the Epoch of Reionization with $z \sim 6$ Quasars. II. A Sample of 19 Quasars”. In: *The Astronomical Journal* 132.1, pp. 117–136. ISSN: 1538-3881. DOI: 10.1086/504836. URL: http://dx.doi.org/10.1086/504836.

Foreman-Mackey, Daniel et al. (Mar. 2013). “emcee: The MCMC Hammer”. In: *Publications of the Astronomical Society of the Pacific* 125.925, pp. 306–312. ISSN: 1538-3873. DOI: 10.1086/670067. URL: http://dx.doi.org/10.1086/670067.

Furlanetto, Steven R., S. Peng Oh, and Frank H. Briggs (Oct. 2006). “Cosmology at low frequencies: The 21cm transition and the high-redshift Universe”. In: *Physics Reports* 433.4-6, pp. 181–301. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2006.08.002. URL: http://dx.doi.org/10.1016/j.physrep.2006.08.002.

Greig, Bradley and Andrei Mesinger (Apr. 2015). “21CMMC: an MCMC analysis tool enabling astrophysical parameter studies of the cosmic 21-cm signal”. In: *Monthly Notices of the Royal Astronomical Society* 449.4, pp. 4246–4263. ISSN: 1365-2966. DOI: 10.1093/mnras/stv571. URL: http://dx.doi.org/10.1093/mnras/stv571.

— (Oct. 2017). “Simultaneously constraining the astrophysics of reionisation and the epoch of heating with 21CMMC”. In: *Proceedings of the International Astronomical Union* 12.S333, pp. 18–21. ISSN: 1743-9221. DOI: 10.1017/s1743921317011103. URL: http://dx.doi.org/10.1017/S1743921317011103.

Hassan, Sultan, Sambatra Andrianomena, and Caitlin Doughty (May 2020). “Constraining the astrophysics and cosmology from 21cm tomography using deep learning with the SKA”. In: *Monthly Notices of the Royal Astronomical Society* 494.4, pp. 5761–5774. ISSN: 1365-2966. DOI: 10.1093/mnras/staa1151. URL: http://dx.doi.org/10.1093/mnras/staa1151.

Hortúa, Héctor J., Riccardo Volpi, and Luigi Malagò (2020). *Parameters Estimation from the 21 cm signal using Variational Inference*. arXiv: 2005.02299 [astro-ph.CO].

Kern, Nicholas S. et al. (Oct. 2017). “Emulating Simulations of Cosmic Dawn for 21 cm Power Spectrum Constraints on Cosmology, Reionization, and X-Ray Heating”. In: *The Astrophysical Journal* 848.1, p. 23. ISSN: 1538-4357. DOI: 10.3847/1538-4357/aa8bb4. URL: http://dx.doi.org/10.3847/1538-4357/aa8bb4.

Kolopanis, Matthew et al. (Sept. 2019). “A Simplified, Lossless Reanalysis of PAPER-64”. In: *The Astrophysical Journal* 883.2, p. 133. DOI: 10.3847/1538-4357/ab3e3a. URL: https://doi.org/10.3847%2F1538-4357%2Fab3e3.

Koopmans, L. V. E. et al. (2015). *The Cosmic Dawn and Epoch of Reionization with the Square Kilometre Array*. arXiv: 1505.07568 [astro-ph.CO].

Li, W. et al. (Dec. 2019). “First Season MWA Phase II Epoch of Reionization Power Spectrum Results at Redshift 7”. In: *The Astrophysical Journal* 887.2, p. 141. DOI: 10.3847/1538-4357/ab55e4. URL: https://doi.org/10.3847%2F1538-4357%2Fab55e4.

Liu, Adrian and Max Tegmark (May 2011). “A method for 21-cm power spectrum estimation in the presence of foregrounds”. In: *Physical Review D* 83.10. ISSN: 1550-2368. DOI: 10.1103/physrevd.83.103006. URL: http://dx.doi.org/10.1103/PhysRevD.83.103006.

Loeb, Abraham (2010). *How Did the First Stars and Galaxies Form?*

Majumdar, Suman, Garrelt Mellema, et al. (Oct. 2014a). “On the use of seminumerical simulations in predicting the 21-cm signal from the epoch of reionization”. In: *MNRAS* 443.4, pp. 2843–2861. DOI: 10.1093/mnras/stu1342. arXiv: 1403.0941 [astro-ph.CO].

— (Aug. 2014b). “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: 0035-8711. DOI: 10.1093/mnras/stu1342. URL: http://dx.doi.org/10.1093/mnras/stu1342.

Majumdar, Suman, Jonathan R Pritchard, et al. (Feb. 2018). “Quantifying the non-Gaussianity in the EoR 21-cm signal through bispectrum”. In: *Monthly Notices of the Royal Astronomical Society* 476.3, pp. 4007–4024. ISSN: 1365-2966. DOI: 10.1093/mnras/sty535. URL: http://dx.doi.org/10.1093/mnras/sty535.

Mao, Yi et al. (Apr. 2012). “Redshift-space distortion of the 21-cm background from the epoch of reionization - I. Methodology re-examined”. In: *Monthly Notices of the Royal Astronomical Society* 422.2, pp. 926–954. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2012.20471.x. URL: http://dx.doi.org/10.1111/j.1365-2966.2012.20471.x.

McQuinn, Matthew et al. (Dec. 2006). “Cosmological Parameter Estimation Using 21 cm Radiation from the Epoch of Reionization”. In: *The Astrophysical Journal* 653.2, pp. 815–834. ISSN: 1538-4357. DOI: 10.1086/505167. URL: http://dx.doi.org/10.1086/505167.

Mellema, G. et al. (Apr. 2015). “HI tomographic imaging of the Cosmic Dawn and Epoch of Reionization with SKA”. In: *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, p. 10. arXiv: 1501.04203 [astro-ph.CO].

Mellema, Garrelt et al. (Mar. 2006). “C^2-ray: A new method for photon-conserving transport of ionizing radiation”. In: *New Astronomy* 11.5, pp. 374–395. DOI: 10.1016/j.newast.2005.09.004. arXiv: astro-ph/0508416 [astro-ph].

Mertens, F G et al. (Feb. 2020). “Improved upper limits on the 21 cm signal power spectrum of neutral hydrogen at $z \approx 9.1$ from LOFAR”. In: *Monthly Notices of the Royal Astronomical Society* 493.2, pp. 1662–1685. ISSN: 0035-8711. DOI: 10.1093/mnras/staa327. eprint: https://academic.oup.com/mnras/article-pdf/493/2/1662/32666766/staa327.pdf. URL: https://doi.org/10.1093/mnras/staa327.

Mesinger, Andrei, Steven Furlanetto, and Renyue Cen (Nov. 2010). “21cmfast: a fast, seminumerical simulation of the high-redshift 21-cm signal”. In: *Monthly Notices of the Royal Astronomical Society* 411.2, pp. 955–972. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2010.17731.x. URL: http://dx.doi.org/10.1111/j.1365-2966.2010.17731.x.

Mondal, Rajesh, Somnath Bharadwaj, and Suman Majumdar (Dec. 2015). “Statistics of the epoch of reionization 21-cm signal – I. Power spectrum error-covariance”. In: *Monthly Notices of the Royal Astronomical Society* 456.2, pp. 1936–1947. ISSN: 1365-2966. DOI: 10.1093/mnras/stv2772. URL: http://dx.doi.org/10.1093/mnras/stv2772.

— (Oct. 2016). “Statistics of the epoch of reionization (EoR) 21-cm signal – II. The evolution of the power-spectrum error-covariance”. In: *Monthly Notices of the Royal Astronomical Society* 464.3, pp. 2992–3004. ISSN: 0035-8711. DOI: 10.1093/mnras/stw2599. eprint: https://academic.oup.com/mnras/article-pdf/464/3/2992/18519471/stw2599.pdf. URL: https://doi.org/10.1093/mnras/stw2599.

Morales, Miguel F. and J. Stuart B. Wyithe (Aug. 2010). “Reionization and Cosmology with 21-cm Fluctuations”. In: *Annual Review of Astronomy and Astrophysics* 48.1, pp. 127–171. ISSN: 1545-4282. DOI: 10.1146/annurev-astro-081309-130936. URL: http://dx.doi.org/10.1146/annurev-astro-081309-130936.

Murray, S. G., C. M. Trott, and C. H. Jordan (Aug. 2017). “An Improved Statistical Point-source Foreground Model for the Epoch of Reionization”. In: *The Astrophysical Journal* 845.1, p. 7. ISSN: 1538-4357. DOI: 10.3847/1538-4357/aa7d0a. URL: http://dx.doi.org/10.3847/1538-4357/aa7d0a.

Paciga, Gregory et al. (Mar. 2011). “The GMRT Epoch of Reionization experiment: A new upper limit on the neutral hydrogen power spectrum at $z \approx 8.6$”. In: *Monthly Notices of the Royal Astronomical Society* 413, pp. 1174–1183. DOI: 10.1111/j.1365-2966.2011.18208.x.

Parsons, Aaron R. et al. (June 2014). “New limits on 21 cm epoch of reionization from paper-32 consistent with an x-ray heated intergalactic medium at z = 7.7”. English

(US). In: *Astrophysical Journal* 788.2. ISSN: 0004-637X. DOI: 10.1088/0004-637X/788/2/106.

Patil, A. H. et al. (Mar. 2017). “Upper Limits on the 21 cm Epoch of Reionization Power Spectrum from One Night with LOFAR”. In: *The Astrophysical Journal* 838.1, p. 65. ISSN: 1538-4357. DOI: 10.3847/1538-4357/aa63e7. URL: http://dx.doi.org/10.3847/1538-4357/aa63e7.

Patil, Ajinkya H. et al. (July 2014). “Constraining the epoch of reionization with the variance statistic: simulations of the LOFAR case”. In: *Monthly Notices of the Royal Astronomical Society* 443.2, pp. 1113–1124. ISSN: 0035-8711. DOI: 10.1093/mnras/stu1178. URL: http://dx.doi.org/10.1093/mnras/stu1178.

Pober, Jonathan C. et al. (Jan. 2014). “WHAT NEXT-GENERATION 21 cm POWER SPECTRUM MEASUREMENTS CAN TEACH US ABOUT THE EPOCH OF REIONIZATION”. In: *The Astrophysical Journal* 782.2, p. 66. DOI: 10.1088/0004-637x/782/2/66. URL: https://doi.org/10.1088%2F0004-637x%2F782%2F2%2F66.

Pritchard, Jonathan R and Abraham Loeb (July 2012). “21 cm cosmology in the 21st century”. In: *Reports on Progress in Physics* 75.8, p. 086901. ISSN: 1361-6633. DOI: 10.1088/0034-4885/75/8/086901. URL: http://dx.doi.org/10.1088/0034-4885/75/8/086901.

Schmit, C J and J R Pritchard (Dec. 2017). “Emulation of reionization simulations for Bayesian inference of astrophysics parameters using neural networks”. In: *Monthly Notices of the Royal Astronomical Society* 475.1, pp. 1213–1223. ISSN: 0035-8711. DOI: 10.1093/mnras/stx3292. eprint: https://academic.oup.com/mnras/article-pdf/475/1/1213/23565018/stx3292.pdf. URL: https://doi.org/10.1093/mnras/stx3292.

— (Mar. 2018). “Emulation of reionization simulations for Bayesian inference of astrophysics parameters using neural networks”. In: *MNRAS* 475.1, pp. 1213–1223. DOI: 10.1093/mnras/stx3292. arXiv: 1708.00011 [astro-ph.CO].

Schmit, Claude J. and Jonathan R. Pritchard (2017). “Neural Network Emulation of Reionization Simulations”. In: *Proceedings of the International Astronomical Union* 12.S333, pp. 43–46. DOI: 10.1017/S174392131700984X.

Shaw, Abinash Kumar, Somnath Bharadwaj, and Rajesh Mondal (June 2019). “The impact of non-Gaussianity on the error covariance for observations of the Epoch of Reionization (EoR) 21-cm power spectrum”. In: *Monthly Notices of the Royal Astronomical Society*. ISSN: 1365-2966. DOI: 10.1093/mnras/stz1561. URL: http://dx.doi.org/10.1093/mnras/stz1561.

Zheng, Zhen-Ya et al. (June 2017). “First Results from the Lyman Alpha Galaxies in the Epoch of Reionization (LAGER) Survey: Cosmological Reionization at z ∼ 7”. In: *The Astrophysical Journal* 842.2, p. L22. DOI: 10.3847/2041-8213/aa794f. URL: https://doi.org/10.3847%2F2041-8213%2Faa794f.