Parameter Estimation of Global 21cm Signal

Aryana Haghjoo

Department of Physics

McGill University

Montréal, Québec, Canada

August 2023

A thesis presented for the degree of Masters of Physics

Abstract

The global 21cm signal has been used to study the period between the end of the cosmic dark ages and the formation of the first stars and galaxies.

In this study, we explore the potential of this signal to reveal non-standard physics by means of providing a new path to test fundamental physical theories. The 21cm signal is sensitive to the density and temperature of neutral hydrogen in the early universe and the presence of the first stars and galaxies. Therefore, any deviation from the predictions of the standard cosmological model of this signal could indicate the presence of new physics beyond the standard model.

We adopt the Markov Chain Monte Carlo (MCMC) method combined with the Levenberg Marquardt (LM) algorithm to estimate the best-fit physical parameters (e.g., clumping factor, star formation efficiency) of the theoretical 21cm curves. We use the Accelerated Reionization Era Simulations (ARES) to generate models for the global 21-cm signal. Our method is flexible to the choice of parameters from ARES.

The knowledge of these best-fit parameters will help us to constrain future proposed

models and set theoretical limits for the precision of upcoming experiments to observe non-standard effects.

Also mention what fields of literature you will review.

Abrégé

How are you translating into French?

Le signal global de 21 cm a été utilisé pour étudier la période entre la fin de l'âge sombre cosmique et la formation des premières étoiles et galaxies.

Dans cette étude, nous explorons le potentiel de ce signal pour révéler la physique non standard en fournissant une nouvelle voie pour tester les théories physiques fondamentales. Le signal de 21 cm est sensible à la densité et à la température de l'hydrogène neutre dans l'univers primordial et à la présence des premières étoiles et galaxies. Par conséquent, tout écart par rapport aux prédictions du modèle cosmologique standard de ce signal pourrait indiquer la présence d'une nouvelle physique au-delà du modèle standard.

Nous adoptons la méthode Markov Chain Monte Carlo (MCMC) combinée à l'algorithme de Levenberg Marquardt (LM) pour estimer les paramètres physiques les mieux adaptés (par exemple, facteur d'agrégation, efficacité de formation d'étoiles) des courbes théoriques de 21 cm. Nous utilisons les simulations d'ère de réionisation accélérée (ARES) pour générer des modèles pour le signal global de 21 cm. Notre méthode est flexible au choix des paramètres d'ARES.

La connaissance de ces paramètres de meilleur ajustement nous aidera à contraindre les futurs modèles proposés et à fixer des limites théoriques pour la précision des expériences à venir pour observer les effets non standard.

Acknowledgements

Reminder: name of my friends, my roommates, something in Persian if allowed

I would like to express my gratitude to my supervisors, Jonathan Sievers and Oscar Hernandez, for their invaluable support throughout this project. I am also deeply grateful to my parents for their unwavering support, even from thousands of kilometers away. Additionally, I want to thank my new friends in Montreal for the wonderful experiences we shared.

Furthermore, I would like to acknowledge the significant role that my therapist at McGill Wellness Hub, Romeo Bidar, played in my graduate journey. His constant presence was a true blessing.

I must also mention the immense help that my two favorite Persian singers, Kamran and Hooman, provided through their Spotify playlist, which kept me motivated while debugging my code.

Lastly, I want to acknowledge my own perseverance and hard work, which enabled me to gain new experiences and insights.

To all the lifelong learners and doers: you are changing the world every day.

Contents

List of Figures

List of Tables

List of Acronyms

ARES	The Accelerated Reionization Era Simulations.
CMB	Cosmic Microwave Background.
DLS	Damped Least-Squares.
EDGES	Experiment to Detect the Global EoR Signature.
EoR	Epoch of Reionization.
IGM	Intergalactic Medium.
LM	Levenberg-Marquardt.
MCMC	Markov Chain Monte Carlo.
SARAS	Small Array for Research in Astrophysics of the South.
SFR	Star Formation Rate.

Chapter 1

Introduction

Reminder: The preface can be expanded, I should come back to writing this chapter after I am done with chapter 2 and 3. I guess it would help to have some addressing of those information here.

21cm cosmology in a relatively new opened window in the study of universe during its early states. It has been proven to be one of the only available probes of large-scale structure in the most distant reaches of the universe [1].

In this chapter we will talk bout the importance of this signal and its applications. Then, we will describe the motivations of this research to clarify the effects of non-standard physics on this signal. Although this signal can be studied in various forms (e.g. power spectrum, fluctuations, global signal, and etc.), we will only focus on the global 21cm signal.

Finally, in the last section of this chapter, we will briefly go over all the materials provided

in this thesis.

1.1 Background and motivation

Not very long after the big bang, before the formation of stars and galaxies, the universe was composed of huge clouds of gas (mostly from hydrogen and helium) and the remaining photons from the big bang called cosmic microwave background (CMB).

Neutral hydrogen has a special spin-flip transition in its ground state which results in absorption or emission of a radio signal with the wavelength of 21cm. When the CMB photons redshift through expansion of universe, there comes a period when they reach wavelengths close to that of hydrogen intrinsic bands. During this period, hydrogen molecules are able to absorb CMB photons (which later, in chapter 2, will be recalled as coupling of spin temperature T_s with kinetic temperature of the gas T_k). This mechanism happens roughly during the dark ages/epoch of reionization (EoR) and will leave the brightness temperature of 21cm line affected. This change can be detected through precise observation using radio telescopes. These observations are capable of giving us valuable information about the distribution of neutral hydrogen and evolution of cosmic structure over time [2].

The global 21cm signal is the average over the brightness temperature of 21cm line across the entire sky. It is a measure of overall state of intergalactic medium (IGM) which might present as absorption or emission in different redshift regions. This radiation is an

observational evidence for certain characteristics of the IGM in the early universe (e.g. temperature, density and reionization state). These properties are determined by the complex interplay between the cosmic radiation field, the formation and evolution of the first stars and galaxies, and the feedback processes that these sources exert on their surroundings [3].

Besides all the above-mentioned applications, the global 21cm signal is a **strong probe for non-standard physics** during the dark ages. It has the potential to shed a light on mysteries surrounding dark matter/dark energy, existence of cosmic strings, and even certain particle interaction (citation) [1]. This capacity of global 21cm signal is the main motivation of this research.

1.2 Research questions and objectives

Reminder: citations

The effects of non-standard physics in 21cm signal has gained lots of attraction in the recent literature (cite). ~~However, most of~~ the reported effects has been investigated only using analytical methods, regardless of the observational data.

We aim to fill this research gap by probing these effects using semi-analytical simulations of the global 21cm signal. With the recently released data at one's disposal (cite edges), we use certain fitting algorithms to compare them to theoretical simulated models (cite ares).

[1] These effects will be discussed more in 2.3

We first estimate the physical parameters of theses curves by only taking the standard physical mechanisms into account.

Need to include a foreground model first.

In future studies, we will upgrade our simulator to include arbitrary non-standard effects. We will again fit the observational data to the upgraded version of theoretical model to check if any of the best-fit physical parameters show patterns of change.

1.3 Overview of the thesis

Reminder: citations

This thesis consists of two major components: the literature review and the research.

The literature review chapters (~~which go half way through this document~~ mention chapter numbers), we first talk about the global 21cm signal and the its physical principals in chapter 2. In section 2.3, we explore the effects of non-standard physics on global 21cm signal mentioned in the literature. Finally, in section 2.4, we talk about the importance of estimating the parameters of this signal and how these estimations affects future proposed observations. Furthermore, we briefly talk about different methods used to serve this purpose and advantages of each one.

In chapter 3 we will briefly introduce all of the past, present and future observational project who aim to detect the global 21cm signal. From all these experiments, we will specifically talk about two of them in detail.

In the research half of this thesis, we will talk about the parameter estimation of global 21cm signal. The specific method used in this this study is described in chapter 4 thoroughly.

Subsequently, we apply this method to fit the chosen observational data. The results are demonstrated in chapter 5. Finally, in chapter 6, we interpret the outcome and talk the implications of these results for cosmology and astrophysics. Also, the limitations and the future prospective of the project is mentioned in this chapter.

Chapter 2

The Global 21cm Signal

2.1 Theoretical basis and physical principles of the Global 21cm Signal

2.2 Simulating the global 21cm signal

2.2.1 The Accelerated Reionization Era Simulations (ARES)

2.2.2 21cmFast

2.3 Effects of non-standard physics on the global 21cm signal

2.3.1 Cosmic Strings

2.3.2 Dark Matter

While neural networks have been shown to be effective, one drawback is the lack of clear physical interpretation of the resulting parameter values. In contrast, sampling algorithms like Markov chain Monte Carlo (MCMC), which preserve the basis vectors of the parameter space, can overcome this limitation. The parameters used in these methods are directly linked to the underlying physical mechanisms of the model.

2.4.1 Machine Learning

2.4.2 MCMC

Chapter 3

Observations of the Global 21cm Signal

List of experiments (reminder: rephrasing and citing each of them, things that I need for each experiment: dipole antenna, location, time frame of results, frequency range):

1. EDGES (Experiment to Detect the Global EoR Signature): This was the first experiment to report a detection of the global 21cm signal in 2018, by measuring the absorption of radio waves from distant sources by neutral hydrogen in the intergalactic medium.

2. SARAS (Small Array for Research in Astrophysics of the South): This is a radio telescope located in India that has been used to study the global 21cm signal since 2015.

3. SCI-HI (Small Radio Telescope for Cosmic Hydrogen Intensity Mapping): This is a low-frequency radio telescope located in the United States that was designed to detect the global 21cm signal during the epoch of reionization.

4. BIGHORNS (Broadband Instrument for Global HydrOgen ReioNisation Signal): This is a radio interferometer located in Australia that was designed to detect the global 21cm signal during the epoch of reionization.

5. HERA (Hydrogen Epoch of Reionization Array): This is a radio interferometer located in South Africa that is designed to detect the global 21cm signal during the epoch of reionization and cosmic dawn.

6. Tianlai: This is a radio interferometer located in China that is designed to detect the global 21cm signal during the epoch of reionization and cosmic dawn.

7. Murchison Widefield Array: This is a radio interferometer located in Australia that is used to study the early Universe, including the global 21cm signal.

8. LOFAR (Low-Frequency Array): This is a radio interferometer located in the Netherlands that is used to study a wide range of astrophysical phenomena, including the global 21cm signal.

9. PRIZM (Polarized Radiometer In Space) is another experiment designed to study the global 21cm signal. It is a small satellite mission developed by NASA's Goddard Space

Flight Center in collaboration with the Korean Astronomy and Space Science Institute (KASI).

The PRIZM mission aims to detect the global 21cm signal from the cosmic dawn era using a radiometer operating at a frequency range of 30-50 MHz. The satellite was launched in December 2018 as a secondary payload on a SpaceX Falcon 9 rocket and is in a low-Earth orbit.

PRIZM's scientific goals include measuring the temperature and polarization of the cosmic microwave background radiation, studying the reionization epoch, and detecting the global 21cm signal. It is expected to be the first space-based mission to detect the global 21cm signal and could provide crucial insights into the early Universe. These are just a few of the experiments designed to detect the global 21cm signal. Other experiments include the Canadian Hydrogen Intensity Mapping Experiment (CHIME), the Dark Ages Radio Explorer (DARE), and the Square Kilometer Array (SKA).

3.1 Experiment to Detect the Global EoR Signature (EDGES)

3.1.1 error bars and foreground removal

3.1.2 theoretical model and associated parameters

3.2 Small Array for Research in Astrophysics of the South (SARAS)

Chapter 4

Parameter Estimation Methods

We formerly discussed the parameter estimation and model fitting of global 21cm in 2.4. Various computational algorithms, including machine learning and neural networks, have been used to accomplish this task. While neural networks have been shown to be effective, certain drawbacks put sampling algorithms like MCMC on an advantaged position for global 21cm applications.

In this chapter, we describe the main fitting algorithm used in our study, which combines Levenberg-Marquardt and MCMC. We explain the reason behind utilzing both of these algorithms for this particular fitting challenge and provide a detailed explanation of the procedure to combine them. We also discuss two different tests used to measure the quality of the proposed fit.

In chapter 5, we present the results of applying this fitting method to the observed global

21cm curve[1] and its corresponding theoretical model[2].

4.1 Levenberg-Marquardt (LM)

The Levenberg-Marquardt (LM) algorithm, also known as the damped least-squares (DLS) method, is a fitting algorithm used for non-linear least-squares problems. This iterative algorithm is based on another gradient decent method referred to as "Newton's method".

LM is perfectly capable of fitting models with Gaussian-shaped likelihood spaces. However, it's abilities are limited when it comes to more complicated surfaces.

We continue this chapter by deriving the basic analytical definition of this method. In order to do so, we start by defining the matrix form of chi-square. Subsequently, we attempt to calculate the second order derivative of chi-square with respect to the parameters of the model. We will show that this calculation leads to defining the covariance matrix, which will be later used as our basis to generate correlated noise (4.3.1).

[1] released by the EDGES group
[2] generated using ARES

4.1.1 Covariance Matrix

The goal of gradient decent algorithms is to minimize the following sum called the "chi-square":

$$\chi^2 \equiv \sum \frac{(x_i - \mu_i)^2}{\sigma_i^2} \tag{4.1}$$

Where x_i is the observed data, μ_i is the expected value with respect to the theoretical model, and σ is the error associated with each data point. Is it often much easier to calculate the derivatives if we write the above expression in the language of linear algebra. μ can be defined as $\mu_i =< d_i >= A_i(m)$, where A is the model which is dependent on the set of parameters m (the dependency can be nonlinear). We can also define a diagonal noise matrix N, where $N_{i,i} = \sigma_i^2$. Substituting these two in equation 4.1 we get:

$$\chi^2 = (d - A(m))^T N^{-1} (d - A(m)) \tag{4.2}$$

Where d is the array containing the data points.

We continue by calculating the first two derivatives of the above expression, leading to the construction of the gradient decent method.

$$\frac{d\chi^2}{dm} = -(\frac{dA(m)}{dm})^T N^{-1}(d - A(m)) - (d - A(m))^T N^{-1} \frac{dA(m)}{dm} \tag{4.3}$$

Since we know that:

$$(N^{-1})^T = N^{-1} \tag{4.4}$$

$$[\frac{dA(m)}{dm} N^{-1}(d - A(m))]^T = (d - A(m))^T N^{-1} \frac{dA(m)}{dm} \tag{4.5}$$

Substituting in 4.3, we get:

$$\frac{d\chi^2}{dm} = -2(\frac{dA(m)}{dm})^T N^{-1}(d - A(m)) \tag{4.6}$$

$$\tag{4.7}$$

Thus, we can calculate the second derivative:

$$\frac{d^2\chi^2}{dm^2} = -2(\frac{d^2A(m)}{dm^2})^T N^{-1}(d - A(m)) - 2(\frac{dA(m)}{dm})^T N^{-1}(-\frac{d\chi^2}{dm}) \tag{4.8}$$

The first term can simply be neglected due to the following reasons:

1. The $(d - A(m))$ component, which is the residual comparing the model to data, can take both negative and positive values; Thus on average, it will be close to zero.

2. The best-fit model is expected to mimic the behavior of data closely. Therefore, we the residuals must have a small value: $(d - A(m)) \approx 0$

Finally, using the above mentioned logic, we are left with:

$$\frac{d^2\chi^2}{dm^2} = 2(\frac{dA(m)}{dm})^T N^{-1}(\frac{dA(m)}{dm}) \tag{4.9}$$

Which is the definition of the **Covariance Matrix**. The dimensions of this square matrix is the the same as that of m. The diagonal elements of this matrix are simply the inverse of variance on each parameter ($\sigma_{i,i}$), while the off-diagonal elements represent the inverse of covariance, measuring the dependency of a pair of parameters ($\sigma_{i,j}$). If calculated correctly, this matrix is always semi-positive definite[3].

4.1.2 Derivation of the Levenberg-Marquardt Algorithm

In LM method, on each iteration, the set of parameters m is replaced with $m + \delta m$. To find the δm, the function $\chi^2(m + \delta m)$ is approximated by its linearization:

$$\chi^2(m) = (d - A(m))^T N^{-1}(d - A(m)) \tag{4.10}$$

$$\chi^2(m + \delta m) = \chi^2(m) + \frac{d\chi^2}{dm}\delta m \tag{4.11}$$

[3]A positive definite matrix only posses positive eigenvalues. However, for a semi-positive definitive matrix, eigenvalues are non-negative.

Similar to the procedure done in section 4.1, we calculate the derivative of 4.11:

$$\frac{d\chi^2(m+\delta m)}{dm} = \frac{d}{dm}(\chi^2) + \frac{d}{dm}(\frac{d\chi^2}{dm}\delta m) \tag{4.12}$$

We already have the expression for the first order derivative of chi-square in 4.3. Therefore:

$$\frac{d\chi^2(m+\delta m)}{dm} = -2(\frac{dA(m)}{dm})^T N^{-1}(d - A(m)) + \frac{d^2\chi^2}{dm^2}\delta m + \frac{d\chi^2}{dm}\frac{d}{dm}(\delta m) \tag{4.13}$$

Where the last term equals zero since δm does not have any fundamental dependencies on m. Looking at the second term, it is simply inferred that we have already found the expression for second derivative of chi-square in 4.9. Thus, we are left with:

$$\frac{d\chi^2(m+\delta m)}{dm} = -2(\frac{dA(m)}{dm})^T N^{-1}(d - A(m)) + 2(\frac{dA(m)}{dm})^T N^{-1}(\frac{dA(m)}{dm}) \tag{4.14}$$

We define $d - A(m) \equiv r$, and $\frac{dA(m)}{dm} \equiv A'$, so the expression takes the following form:

$$A'^T N^{-1} A' \delta m = A'^T N^{-1} r \tag{4.15}$$

$$\delta m = (A'^T N^{-1} A')^{-1} A'^T N^{-1} r \tag{4.16}$$

Equation 4.16 represents the basis for **Newton's method**. However, as previously mentioned, this method suffers from convergence issues, especially on complicated

likelihood surfaces. To overcome this obstacle, we add a new term to the left-hand side of 4.16. This term includes a control parameter Λ that is updated on each iteration depending on the quality of fit.

$$(A'^T N^{-1} A' + \Lambda I)\delta m = A'^T N^{-1} r \tag{4.17}$$

$$\delta m = (A'^T N^{-1} A' + \Lambda I)^{-1} A'^T N^{-1} r \tag{4.18}$$

Now the basic idea is apparent: On each iteration, a set of parameters m will be replaced by $m + \delta m$ and the chi-square is calculated based on the perturbed parameters. Subsequently, the new chi-square is compared to its value in the last step. If we encounter a higher value, the Λ will be multiplied to a constant arbitrary number (> 1). Otherwise, it will be divided with another constant value (> 1). For practical purposes, if Λ takes a value lower than a constant small arbitrary number, we set it equal to zero. If the Λ is zero, and the chi-square is less than an arbitrary threshold value, we declare the **convergence** of the algorithm.

A flow chart summary of this method is shown in 4.1. The Python code for LM is also available in appendix 7.1.1.

4.2 Markov Chain Monte Carlo (MCMC)

Given that Levenberg-Marquardt (LM) is not always effective in finding the best-fit point for complex likelihood spaces, a more powerful algorithm such as Markov Chain Monte Carlo (MCMC) is necessary in many real-life applications.

MCMC is particularly effective in fitting non-Gaussian likelihood spaces and has a guaranteed convergence, regardless of the starting point in parameter space. However, the algorithm is computationally heavy due to its iterative process, which is based on the evaluation of chi-square on each step. Initially, an arbitrary point in parameter space is given to MCMC as its initial trial step and the associated chi-square is calculated. Then, a random point is drawn from a Gaussian distribution, where the mean is set at the last point in the chain[4]. Subsequently, the new chi-square is compared to the previous one from the last trial step. If the new chi-square is lower, the new trial point is accepted in the chain. If the new chi-square is higher, the trial point is accepted with a probability determined by a specific criterion.

The above-mentioned probability threshold is typically defined as:

$$Probability = e^{\frac{-1}{2}(\chi^2_{new} - \chi^2)} \tag{4.19}$$

Which again has Gaussian insights.

[4]In Python applications, the numpy.randn function is used to serve this purpose.

As a measure of MCMC performance, the acceptance ratio is used to determine the fraction of trial steps that end up getting accepted into the chain. An ideal MCMC would typically have an acceptance ratio of 25 percent. However, even with a lower acceptance ratio, the MCMC can still converge, but it will require more trial steps.

A visual summary of MCMC algorithm is shown in 4.2, and the Python code is available at 7.1.2.

4.2.1 Convergence Test

The MCMC algorithm is designed to explore different regions of the parameter space in order to reach convergence. Various methods have been developed to ensure that the MCMC has converged, one of which is to check the power spectrum.

The power spectrum represents the distribution of power at different frequencies in the MCMC chain. A converged MCMC chain must have the behaviour of a white noise, with power uniformly distributed among all frequencies. On the other hand, an unconverged chain will show more power at lower frequencies compared to higher ones. Therefore, the criterion for checking the convergence of an MCMC chain is the flatness of the power spectrum in low frequencies when plotted on a log-log scale. Figure 4.3 and 4.4 illustrates the difference between a converged and an unconverged chain.

4.3 Combination of MCMC and LM

As mentioned before, MCMC is a computationally heavy algorithm due to its iterative nature. If calculating the chi-square takes a long time on each step, the MCMC itself will have a rather long run time before reaching the converged state. Different methods have been proposed to deal with this issue and help the MCMC to converge faster, one of which is to use the insights from running LM.

We previously discussed that parameters of a model might be correlated (4.1.1). During a simple MCMC, we are drawing random samples from a gaussian distribution. This samples do not take the possible correlations into account. However, if we generate samples with such characteristics, the probability of them getting accepted into the actual chain is higher. Therefore, this approach (feeding the MCMC with a posterior distribution) will eventually assist the MCMC to explore more efficient regions of parameter space, and converge faster. Figure 4.5 illustrates this technique briefly.

Since the covariance matrix of these needed samples is already in hand, we are able to easily generate a set of correlated noise samples from that. This procedure is described in the following section.

4.3.1 Generating correlated noise

As discussed before, the off-diagonal elements of a covariance matrix correspond to the inverse of covariance between each pair of parameters. Thus, if we draw samples from the

inverse of covariance matrix (which needs to be calculated at the point of "best-fit"), we are essentially sampling from the multivariate normal distribution with the deviation values describing the uncertainties in the parameters[5].

The equivalent methods can be used to generate correlated noise: Cholesky and eigenvalue decomposition[6]. For practical reasons, we prefer to use the eigenvalue decomposition for 21cm applications. The procedure is a s follows: A matrix of normal gaussian-drawn random variables is constructed in the desired shape ($n \times m$, corresponding to the number of samples and number of parameters respectively). Then, it is multiplied by the eigenvalue matrix and scaled by the square root of the eigenvalues. The transpose of the product will give us the correlated samples.

To gain more precision, is it possible to use the eigenvalues decomposition of the normalized covariance matrix (where diagonal samples all equal unity). The normalization process is done by multiplying the covariance matrix with its own diagonal. Eventually, the drawn samples need to be scaled by the square root of the diagonal matrix.

The Python implementation of the above-mentioned procedure is given in 7.1.3.

[5]Figure 4.6, which shows the corner plots for a four-variable gaussian model, emphasizes the importance of sampling from a multivariate normal distribution and taking the correlations between the parameters into account.

[6]Normally, calculating the Cholesky decomposition takes a shorter amount of time.

4.4 Testing the Algorithm

The complicated algorithm described in the previous sections, does not always behave as expected. Thus, naturally, one seeks options to weigh the output. In sections 4.4.1 and 4.4.2, we introduce two methods to measure the overall quality of the model fitting.

4.4.1 The chi-square test

This method is more focused on inspecting the output of LM and drawn samples. A large number of samples are drawn from the covariance matrix and the corresponding chi-squares are calculated. Since the covariance matrix describes the uncertainties in the parameters, we expect that the average difference in the chi-square statistics for two different samples should be of order unity per each parameter. This comes from the fact that the chi-square statistic scales with the uncertainties in the data and model, which are typically of order 1.

Figure 4.7 shows an example of the distribution of difference between the chi-square values.

4.4.2 Chi-Square vs Parameters Plots

Another method to verify the results is to plot the chi-square values of drawn samples versus the each of the parameters. According to the definition of chi-square 4.2 (the non-linear dependency of chi-square on model parameters) we expect to observe a parabolic behaviour.

Figure 4.8 demonstrates chi-square vs parameters plots for the same model as 4.7.

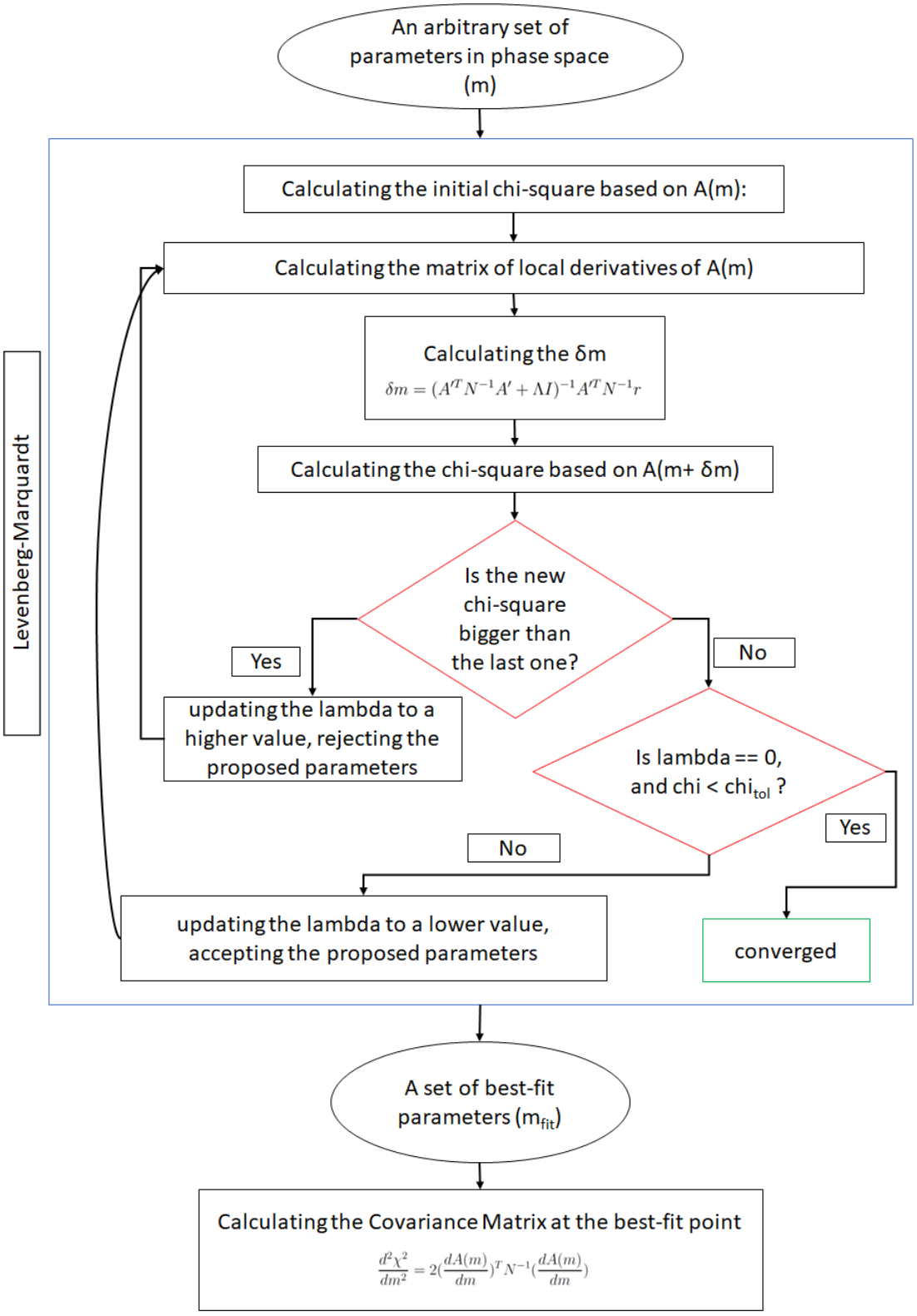

Figure 4.1: Flow chart of Levenberg-Marquardt algorithm

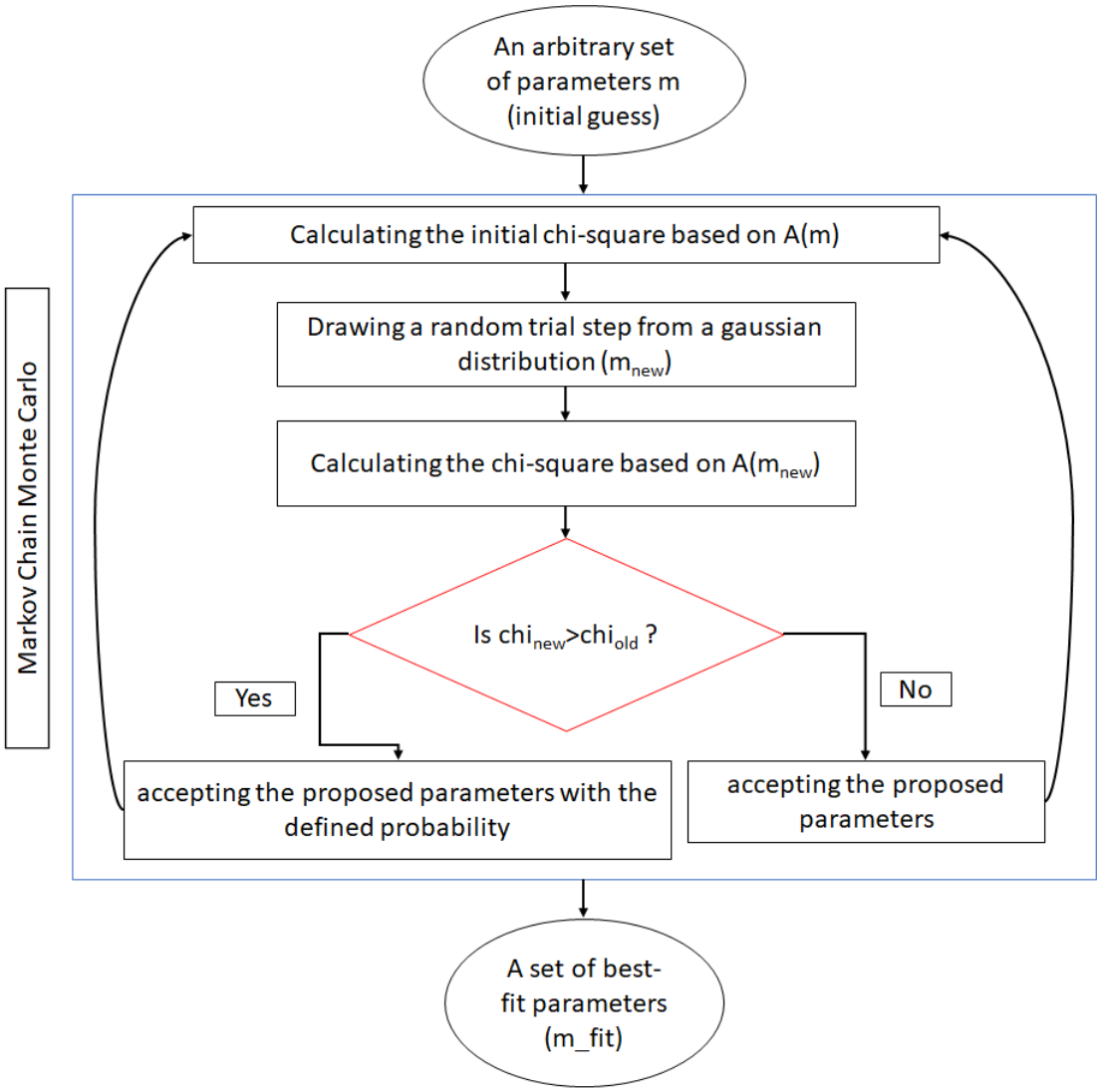

Figure 4.2: Flow chart of MCMC algorithm

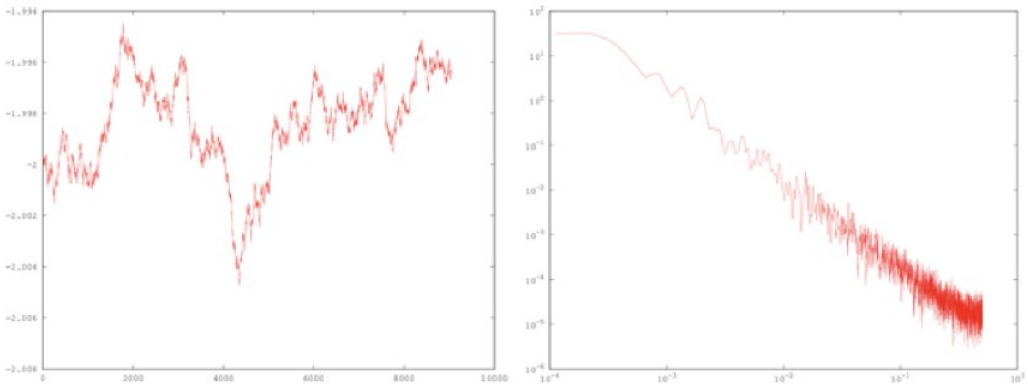

Figure 4.3: An unconverged MCMC chain (left panel) and its power spectrum (right panel): The power tend to increase in lower frequencies, the chain itself does not indicate the behaviour of white noise, plot from Jonathan Sievers

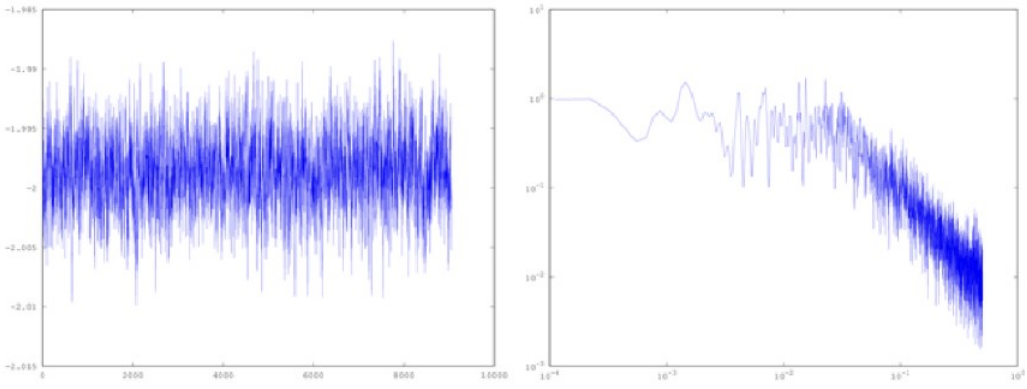

Figure 4.4: A converged MCMC chain (left panel) and its power spectrum (right panel): Note the flatness on low frequencies, plot from Jonathan Sievers

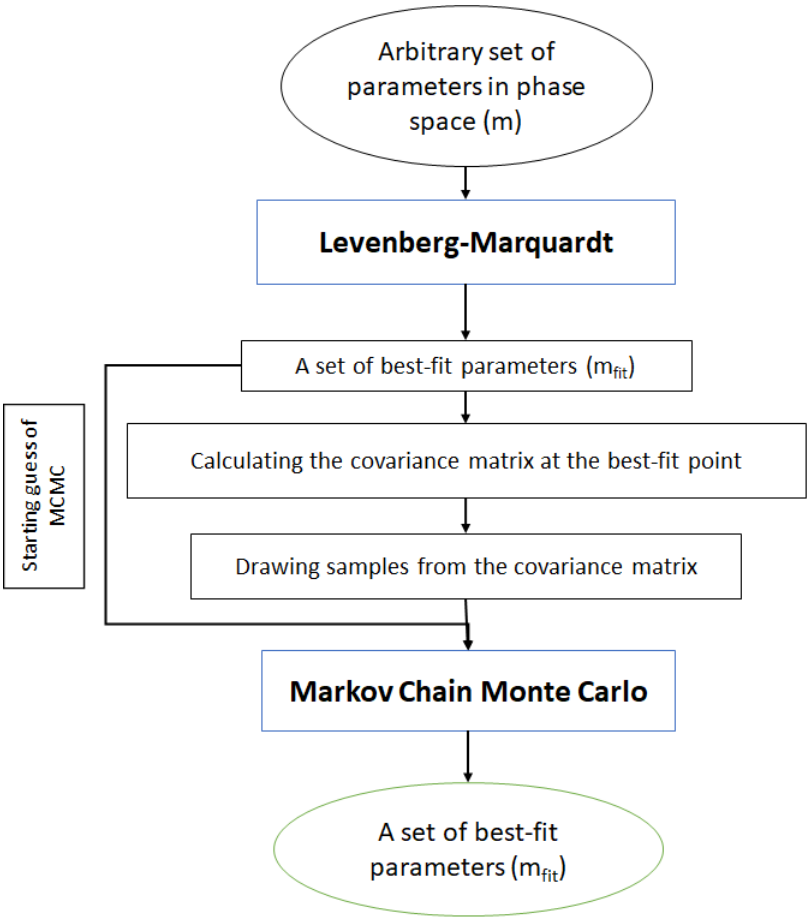

Figure 4.5: Flow chart of the procedure to combine MCMC and LM

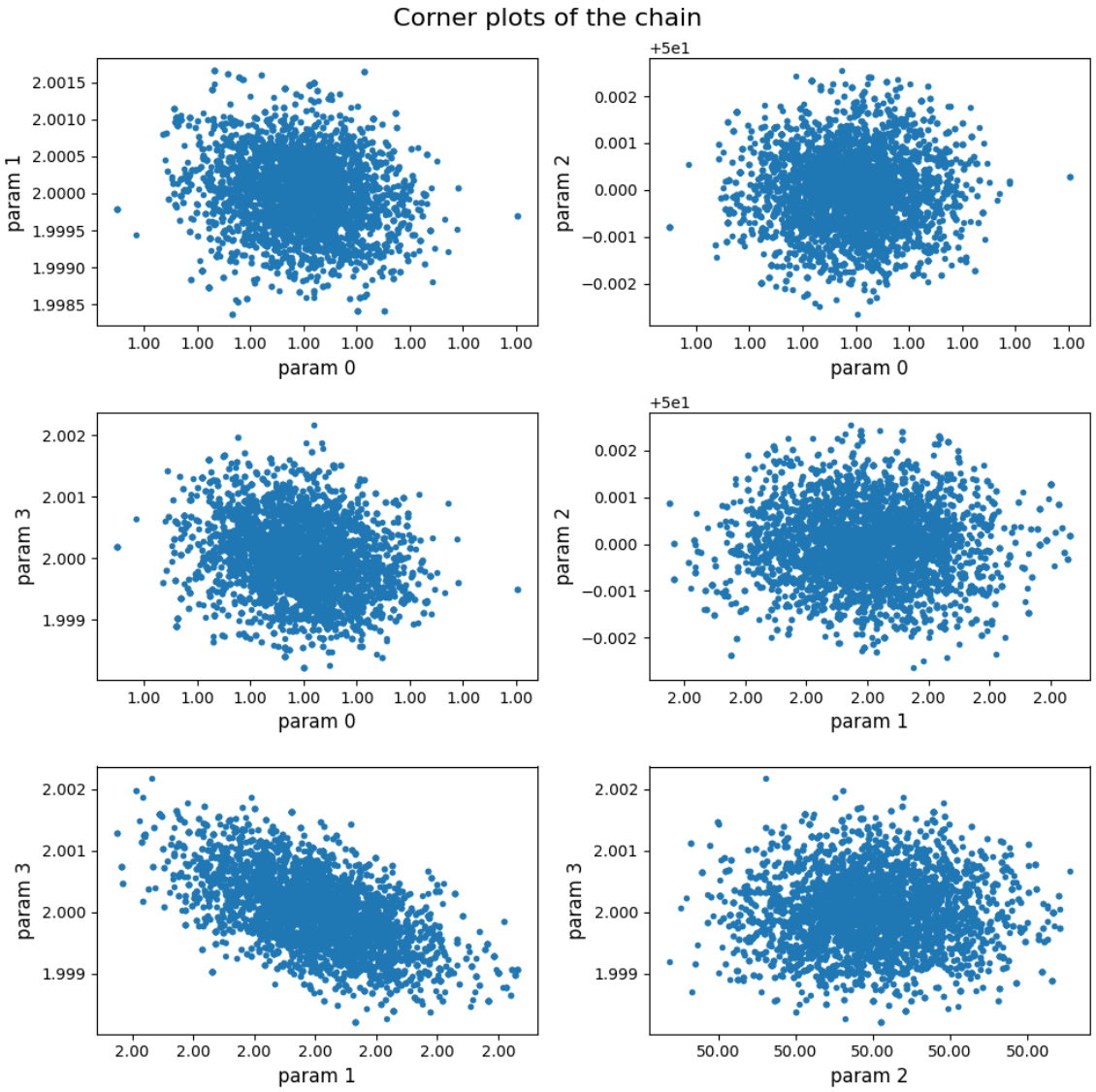

Figure 4.6: Corner plots (parameter vs parameter) of a typical MCMC chain imposed on a four-variable gaussian model. The lower right panel clearly shows patterns of correlation between a pair of parameters. Sampling from a multivariate normal distribution based on the covariance matrix of these samples helps taking the correlations between these parameters into account.

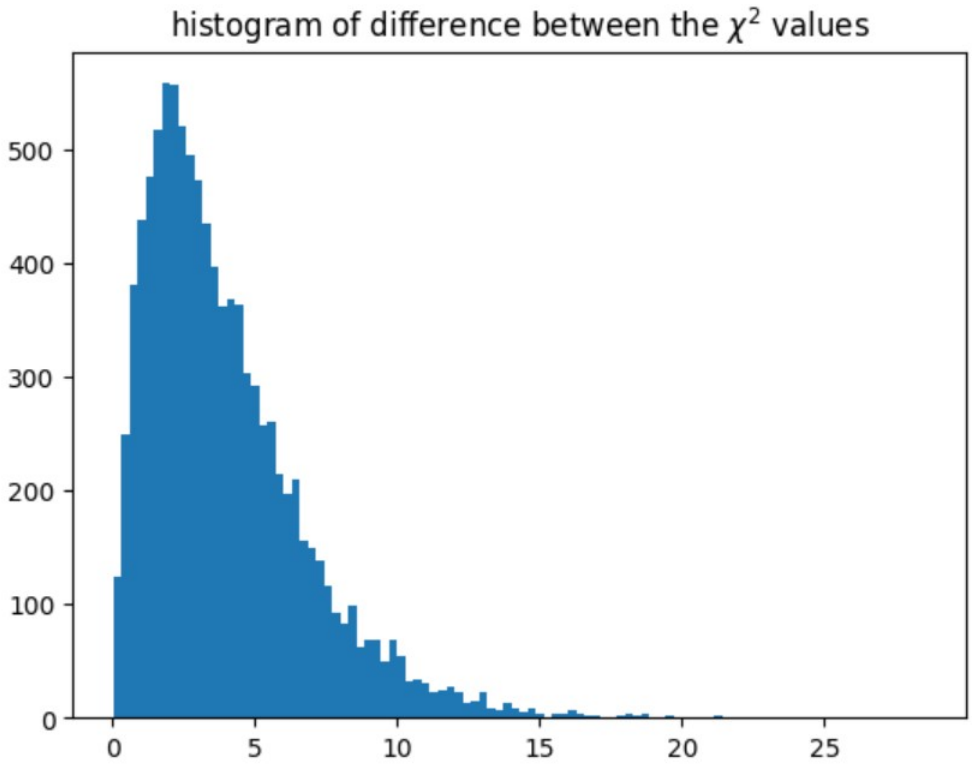

Figure 4.7: Histogram illustrating the distribution of the values of difference between the chi-squares of drawn samples for a four variable model. The average is 3.97 and the standard deviation is 2.83, in agreement with our expectations. The best-fit parameters is considered as a "good fit".

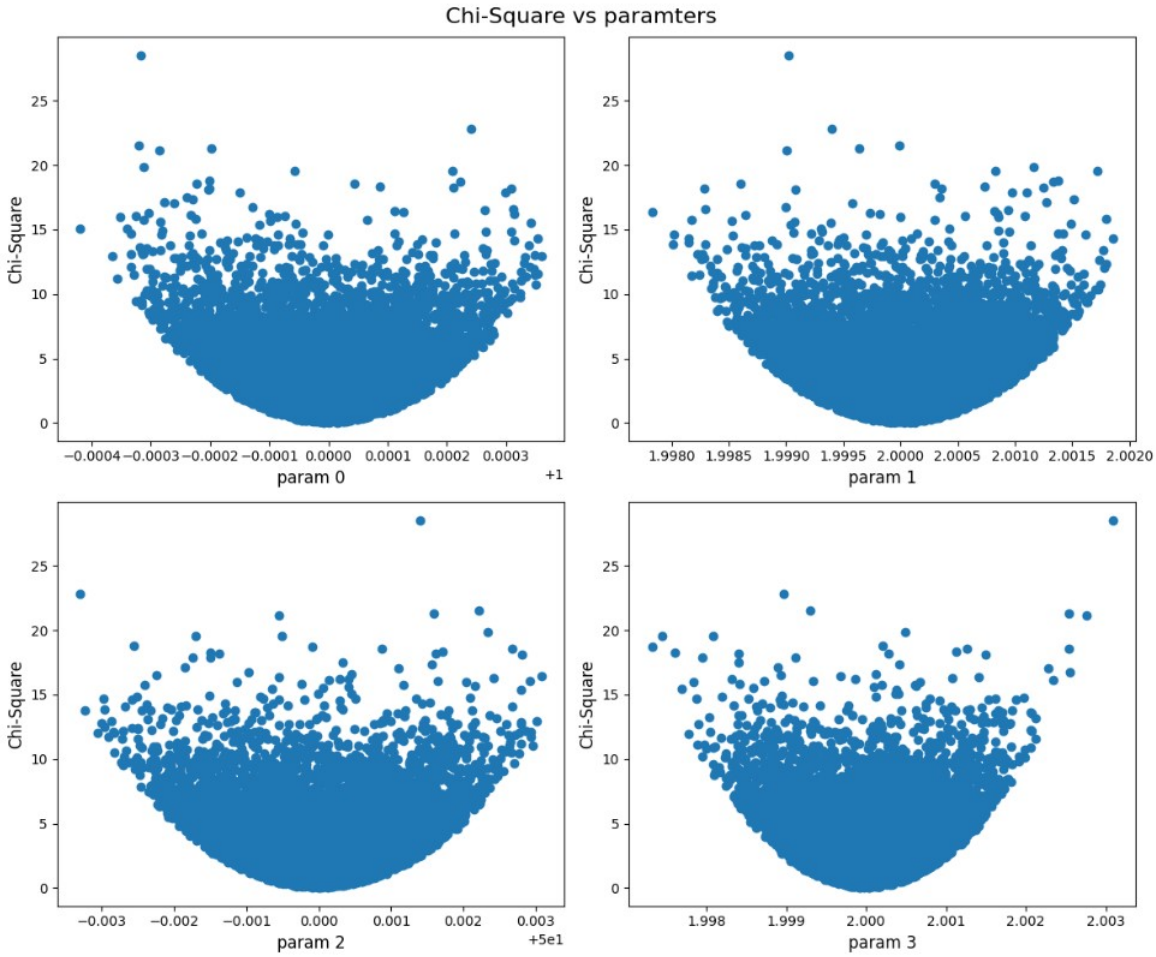

Figure 4.8: Plots of the values of chi-square of drawn samples versus the values of parameters for a four variable model. The parabolic behaviour is obvious.

Chapter 5

Results and Analysis

Reminder: citations for physical explanation of parameters

Chapter 4 carefully described a specific method to fit an experimental data set to its corresponding theoretical model. In this chapter, we are going to impose this method to actual global 21cm curves. We chose to use ARES as our simulator to generate these curves.

As discussed in chapter 2, the physical model of 21cm curve is dependent on a large number of parameters. However, this model can be effectively described using the following parameters:

1. N_{lw}: Number of photons emitted in the Lyman-Werner band ($11.2 - 13.6eV$) per baryon of star formation (This parameter is referred to as *pop_rad_yield_0_* in ARES documentation)

2. N_{ion}: Mean number of ionizing photons produced per baryon of star formation (This

parameter is referred to as *pop_rad_yield_2_* in ARES documentation)

3. f_{esc}: Fraction of ionizing photons that escape their host galaxy into the IGM

4. f_X: High-redshift normalization factor in the relation between X-ray luminosity and SFR

Figure 5.1 shows the effect of changing these parameters on an actual 21cm curve.

We take the above-mentioned physical parameters as fitting parameters and we try to find their best-fit values and corresponding error-bars.

5.1 Parameter Estimation of an ARES generated curve

Before using our developed script to fit actual data, we use it to fit a known ARES curve as a verification test. We generate an ARES curve with a fixed set of parameters and we use these curve as the imaginary "data". If the algorithm returns the same parameter values (with in the error-bar range), we can make sure that it is working correctly.

We begin by using the LM and we find inverse of covariance matrix for our chosen combination of parameters, which we will later use to generate our correlated samples.

Parameters:

$$pop\_rad\_yield\_0\_ : 10^4, \quad pop\_rad\_yield\_2\_ : 10^3, \quad f_{esc} : 0.1, \quad f_X : 0.1 \tag{5.1}$$

Parameter / Value	*pop_rad_yield_0_*	*pop_rad_yield_2_*	f_{esc}	f_X
Mean	9.99999993×10^3	1.00002933×10^3	9.99962873×10^2	1.00000029×10^1
Standard Deviation	4.08174690×10^2	4.03138658×10^0	3.74407952×10^4	3.72581029×10^6

Table 5.1: Mean and standard deviation of samples drawn from the covariance matrix 5.2

Inverse of covariance matrix for this set of parameters:

$$\begin{pmatrix} 1.64 \times 10^{-3} & 1.96 \times 10^{-2} & -4.39 \times 10^{-7} & -5.37 \times 10^{-8} \\ 1.96 \times 10^{-2} & 1.60 \times 10^{1} & -1.44 \times 10^{-3} & -5.94 \times 10^{-6} \\ -4.39 \times 10^{-7} & -1.44 \times 10^{-3} & 1.38 \times 10^{-7} & 2.15 \times 10^{-10} \\ -5.37 \times 10^{-8} & -5.94 \times 10^{-6} & 2.15 \times 10^{-10} & 1.38 \times 10^{-11} \end{pmatrix} \tag{5.2}$$

Now, we are prepared to draws our samples based on method described in 4.3.1. Figure 5.10 illustrates the distribution of samples for the above-mentioned four parameters. As a sanity check, we calculate the mean and standard deviation of these distribution. We expect the mean to be close to values of 5.1. Table 5.1 represents the values of mean and standard deviation, which is in agreement with the expectations.

In section 4.4, we introduced two methods to check the algorithm and the quality of fit. Performing the chi-square test on our samples showed that the average difference between the chi-square of drawn samples is 49.20, which we expected to be 4. Although it is approximately one order of magnitude bigger that the expectations, it can be still considered a good fit.

Parameter / Value	*pop_rad_yield_0_*	*pop_rad_yield_2_*	f_{esc}	f_X
True values	1×10^4	1×10^3	0.1	0.1
Fitted values	9.99998275×10^3	9.99419824×10^2	$1.00025031 \times 10^{-1}$	$1.00001169 \times 10^{-1}$
Error-bar of fit	$3.54171104 \times 10^{-2}$	3.86126208×10^0	$3.57822431 \times 10^{-4}$	$3.39295495 \times 10^{-6}$
Fitting Error	0.00013093%	0.02645403 %	0.01527014%	0.00032365%

Table 5.2: Results of fitting a known ARES curve with MCMC chain

We proceed by performing the second test, which is plotting the values of ch-square of samples versus the parameter values. The results are shown in 5.11. In the hypothetical ideal case, We expected to see a parabolic behaviour. Figure 5.5 shows the zoomed version of results. We can infer a parabolic behaviour in low values of chi-square.

Now that we have our samples, we proceed to run the MCMC chain with them. The overall results of the fit and the final resulting curve are shown in 5.2 and **??** respectively. Figure 5.13 indicates the white-noise behaviour of parameters throughout the chain which is consistent with our anticipation for a converged chain. We confirm the convergence by looking at the power spectrum in Figure 5.14. Flat behaviour in low frequencies is apparent.

5.2 Choice of ”multiplication factor”

5.3 Parameter estimates of EDGES data and uncertainties

We proceed by doing the same procedure as section 5.1. However, this time we use the EDGES data (with half of the actual amplitude) as our observational data in MCMC.

5.3.1 error bar calculation

5.4 Comparison with previous studies and observations

I will write this section after the literature review

Figure 5.15 and 5.16 present the trend of values of chi-square throughout the chain and corner plots of parameters.

Figure 5.15 and 5.16 present the trend of values of chi-square throughout the chain and corner plots of parameters.

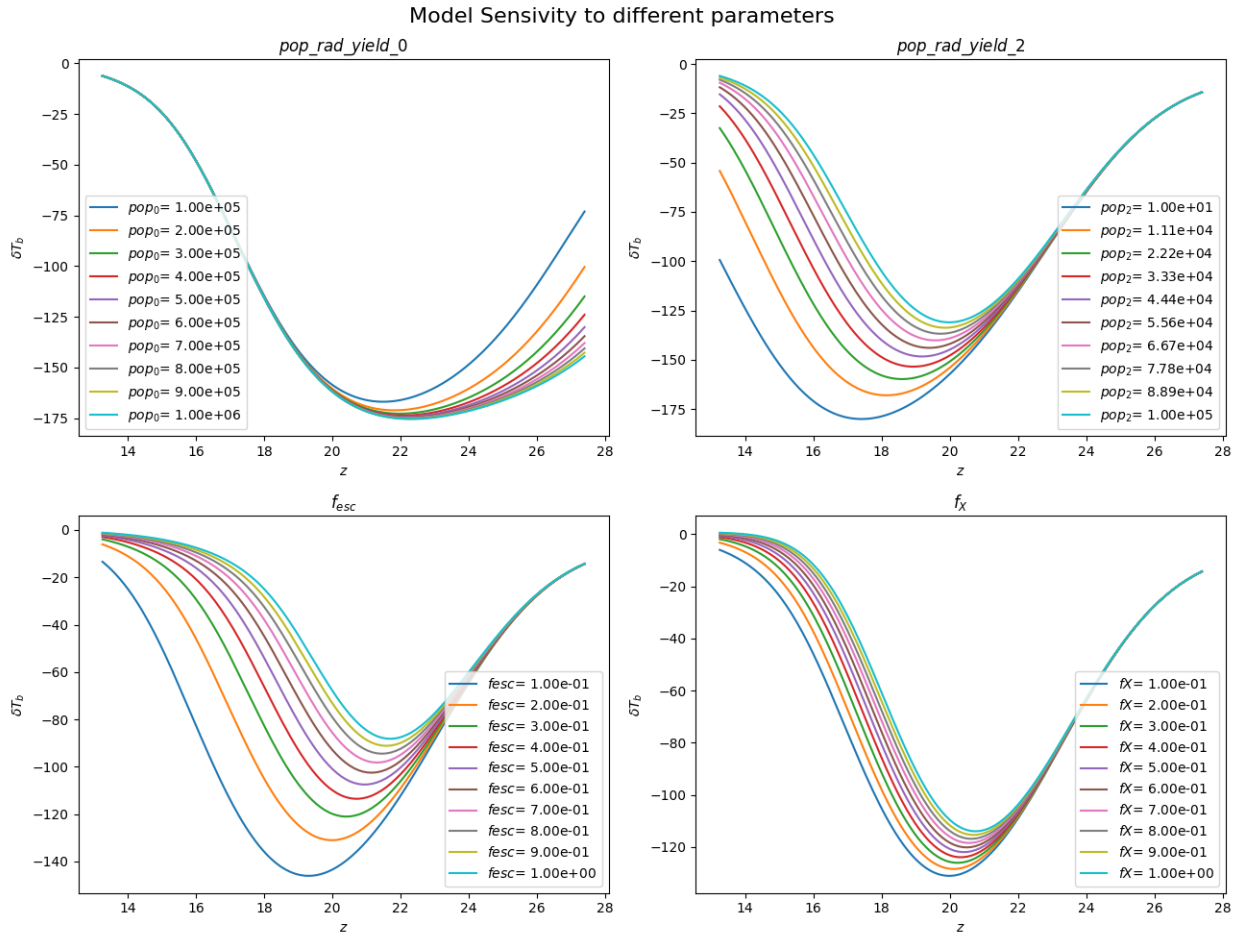

Figure 5.1: Behaviour of ARES-generated global 21cm models with respect to four parameters: *pop_rad_yield_0_*, *pop_rad_yield_2_*, f_{esc}, and f_X. In each panel, the corresponding parameter posses 10 different values and all other parameters are kept at default values by ARES

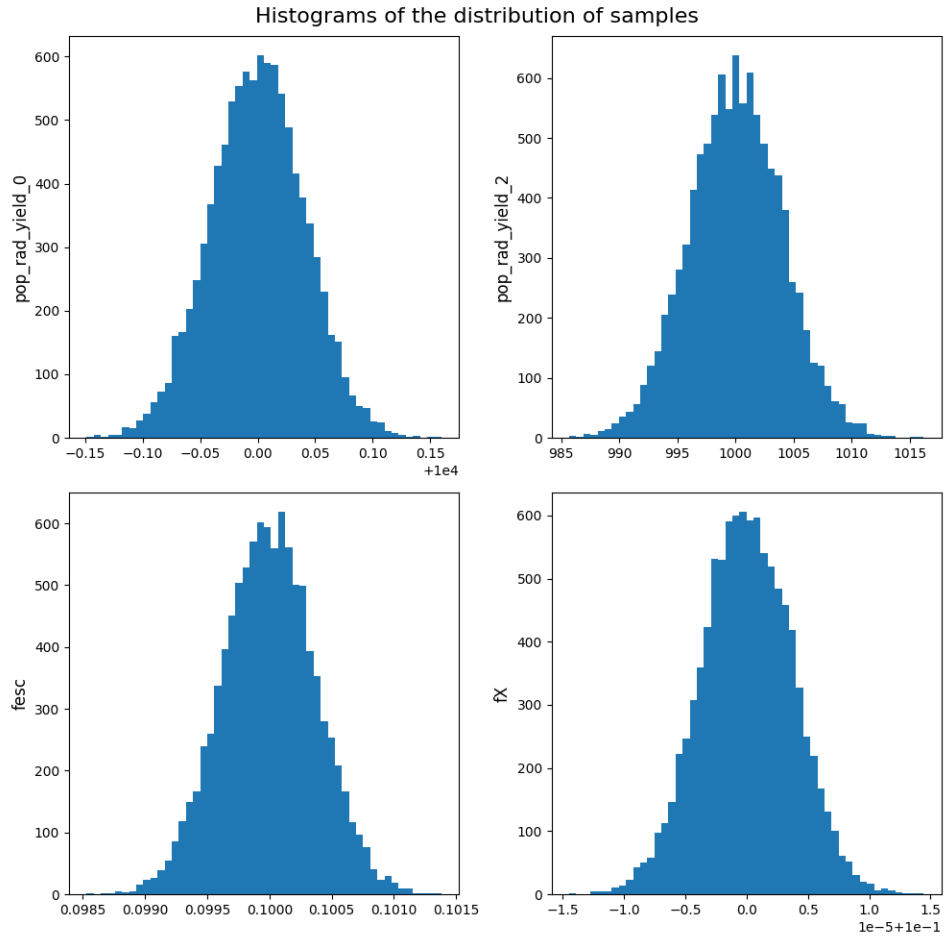

Figure 5.2: Histogram of distribution of samples before feeding to MCMC, The mean of these distributions are consistent with values given in 5.1.

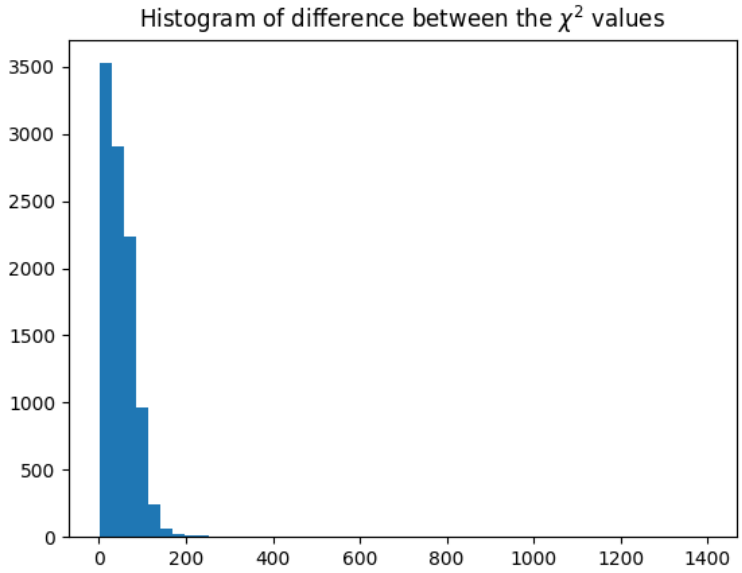

Figure 5.3: Histogram of Chi-Square of drawn samples, the mean is 49.20 with standard deviation of 41.98.

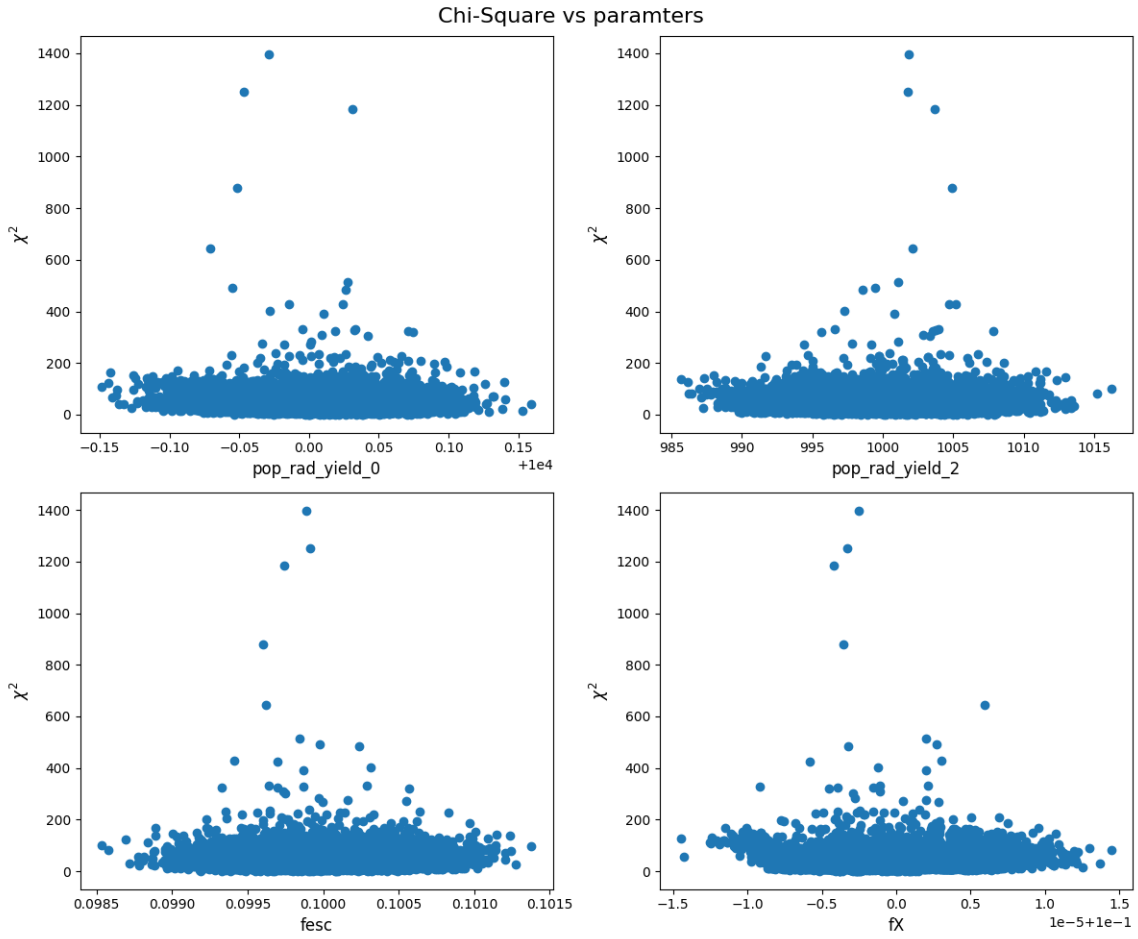

Figure 5.4: Chi-Square of Drawn samples versus parameter values for a known ARES curve with four parameters, The parabolic behaviour is not obvious in these plots.

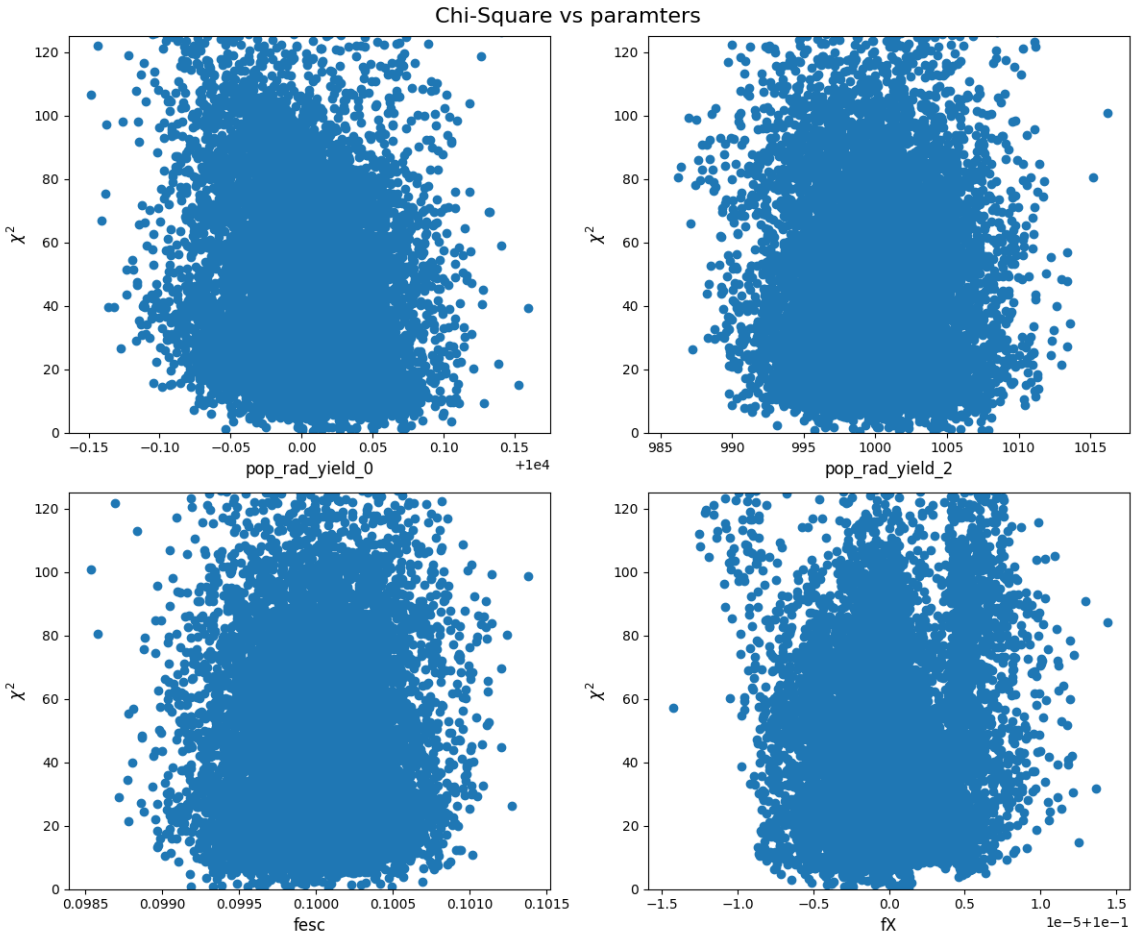

Figure 5.5: Zoomed version of figure 5.12, A weak parabolic behaviour is observed for low values of χ^2.

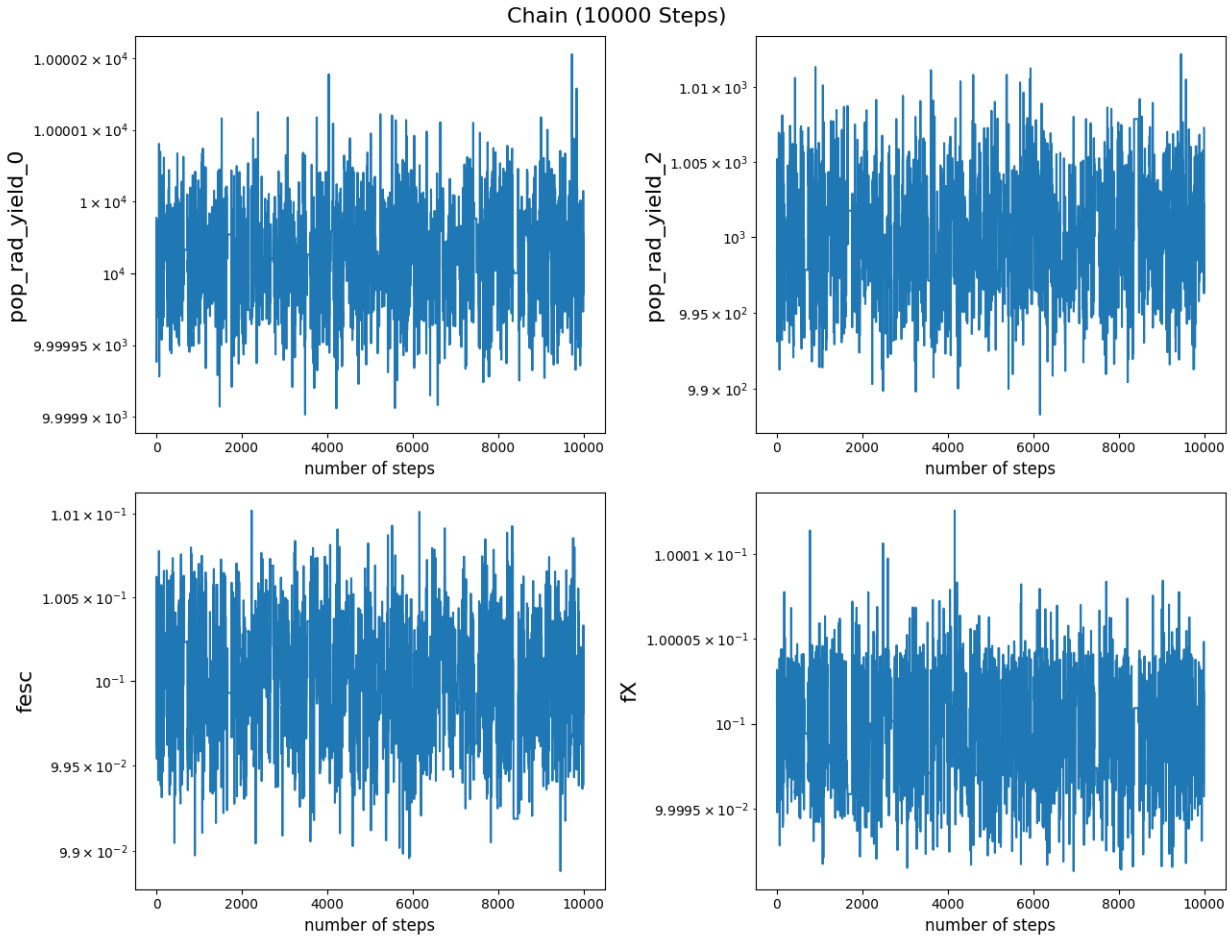

Figure 5.6: Trend of parameters in the MCMC chain, White noise behaviour is considered as a sign of convergence. The acceptance ratio of the chain is 22.12%.

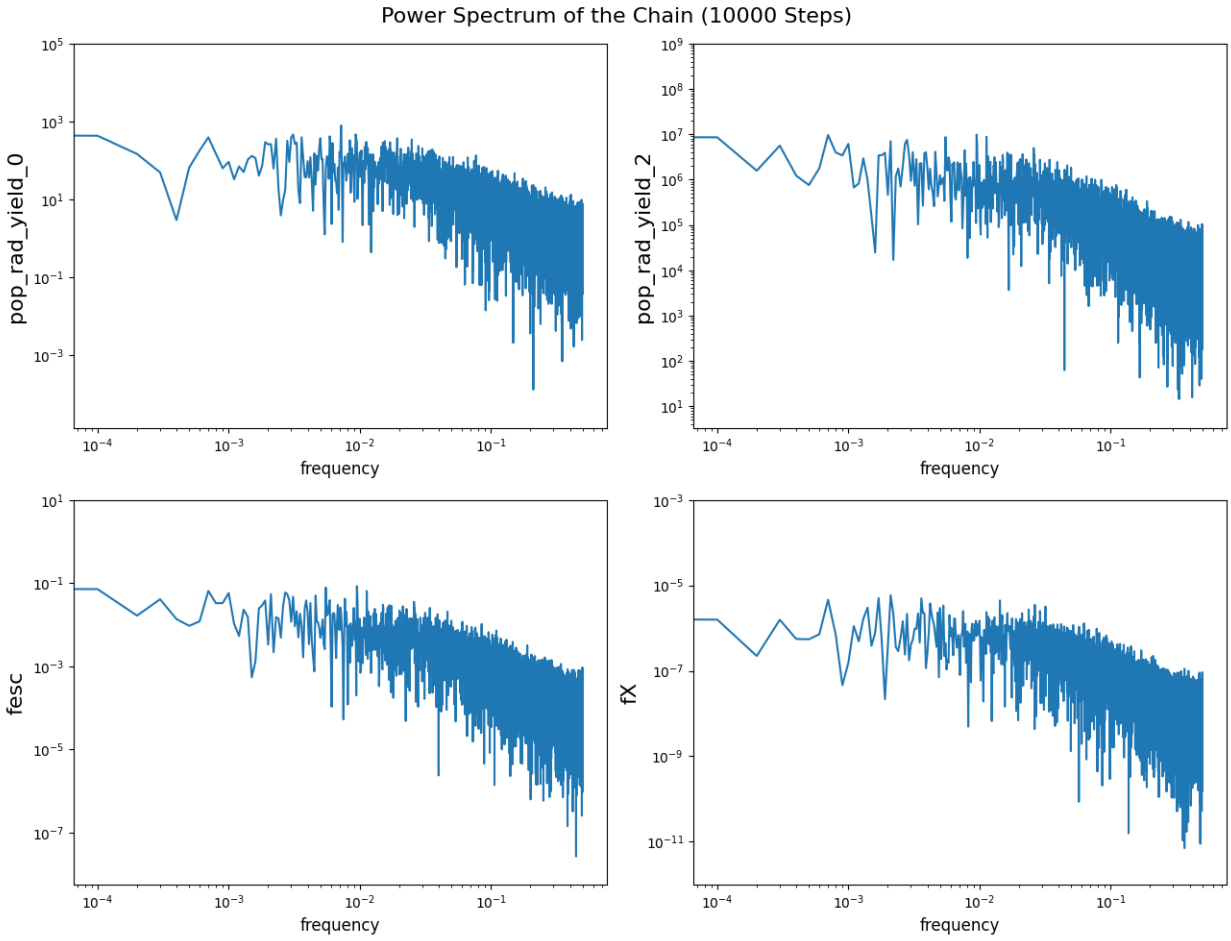

Figure 5.7: Power spectrum of the chain, Flat behaviour in low frequencies is in agreement with results of 5.13.

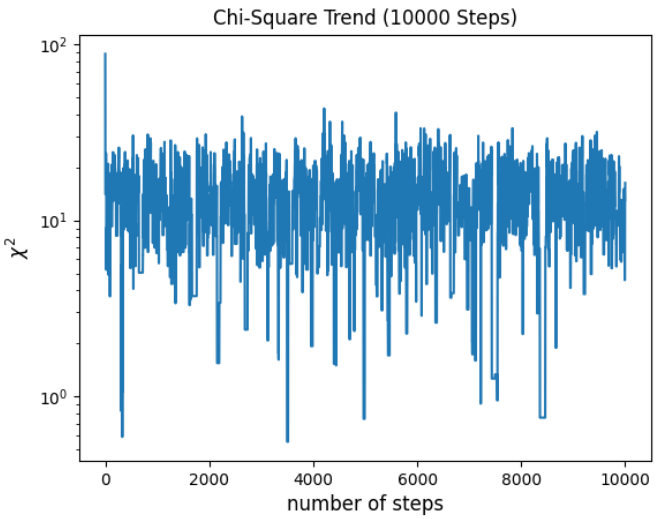

Figure 5.8: Trend of chi-square in the MCMC chain, The chain seems to have reached an equilibrium state.

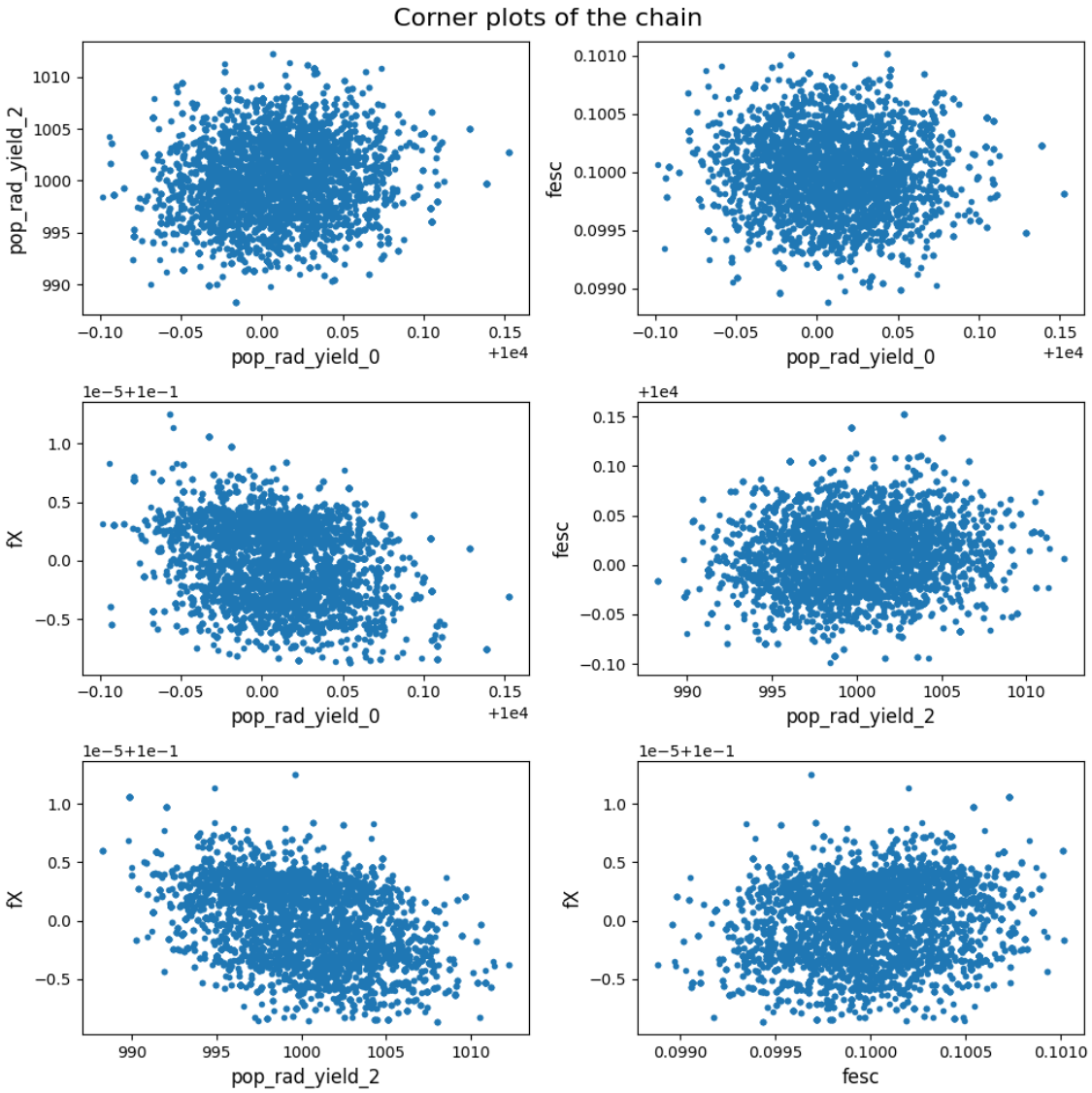

Figure 5.9: Corner plots of the chain indicating the correlation between the parameters specially in the lower left panel

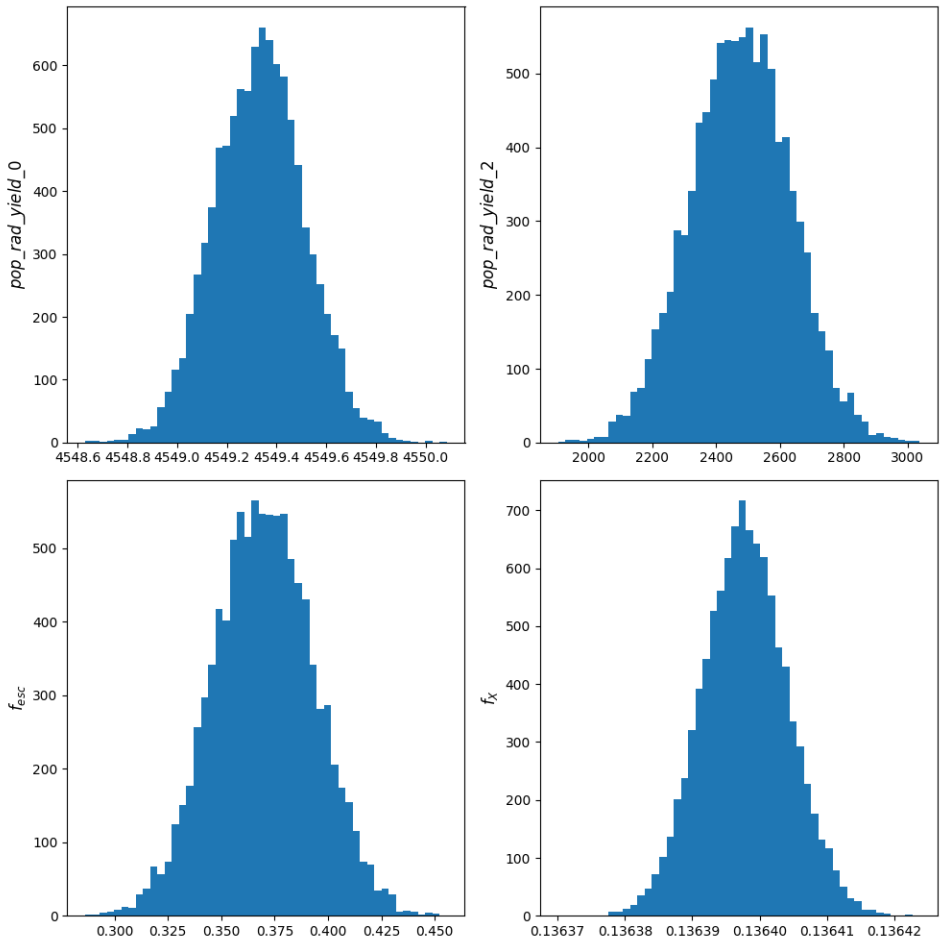

Figure 5.10: Histogram of distribution of samples before feeding to MCMC, The mean of these distributions are consistent with values given in 5.1.

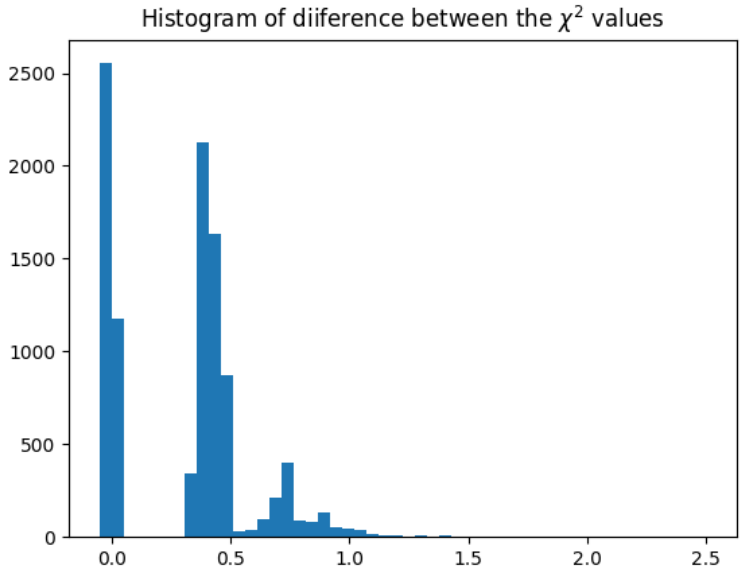

Figure 5.11: Histogram of Chi-Square of drawn samples, the mean is 49.20 with standard deviation of 41.98.

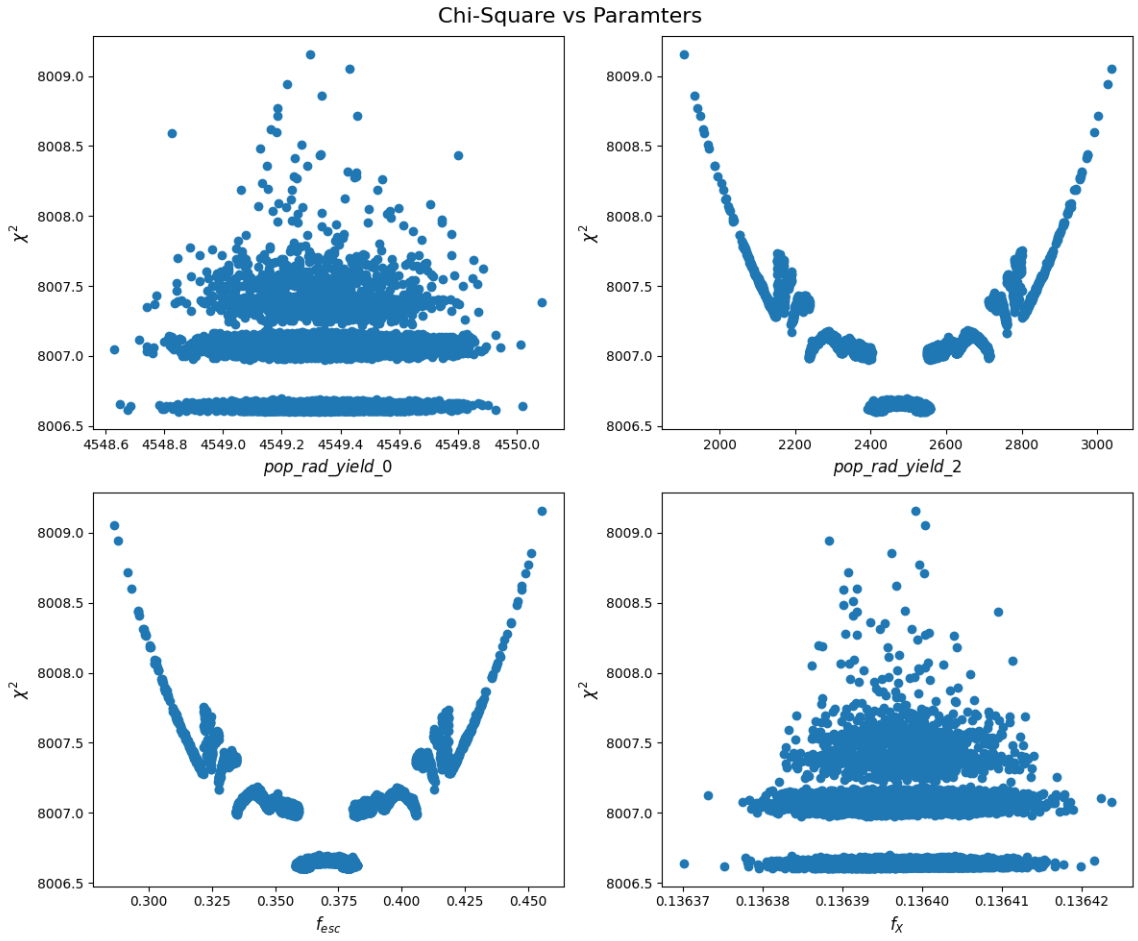

Figure 5.12: Chi-Square of Drawn samples versus parameter values for a known ARES curve with four parameters, The parabolic behaviour is not obvious in these plots.

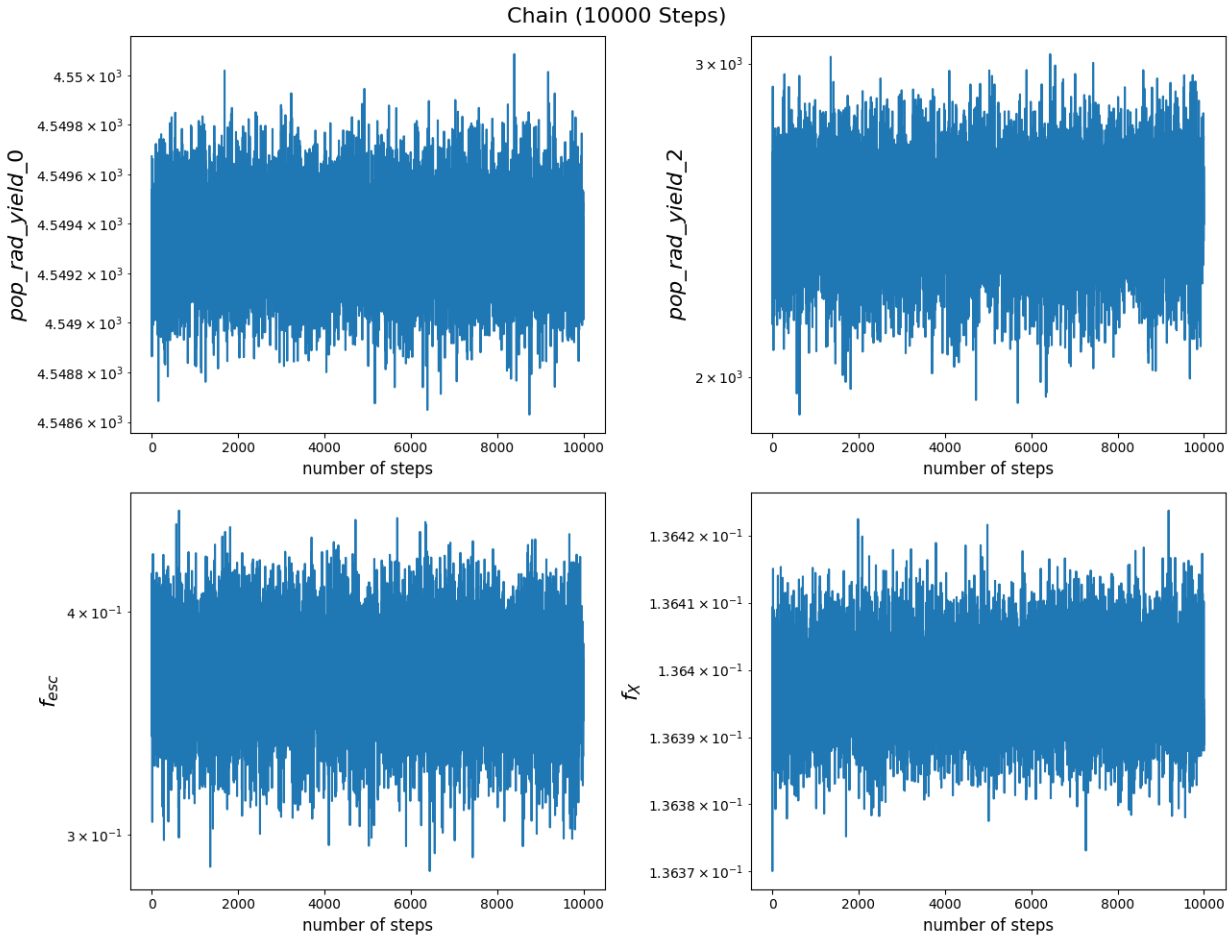

Figure 5.13: Trend of parameters in the MCMC chain, White noise behaviour is considered as a sign of convergence. The acceptance ratio of the chain is 22.12%.

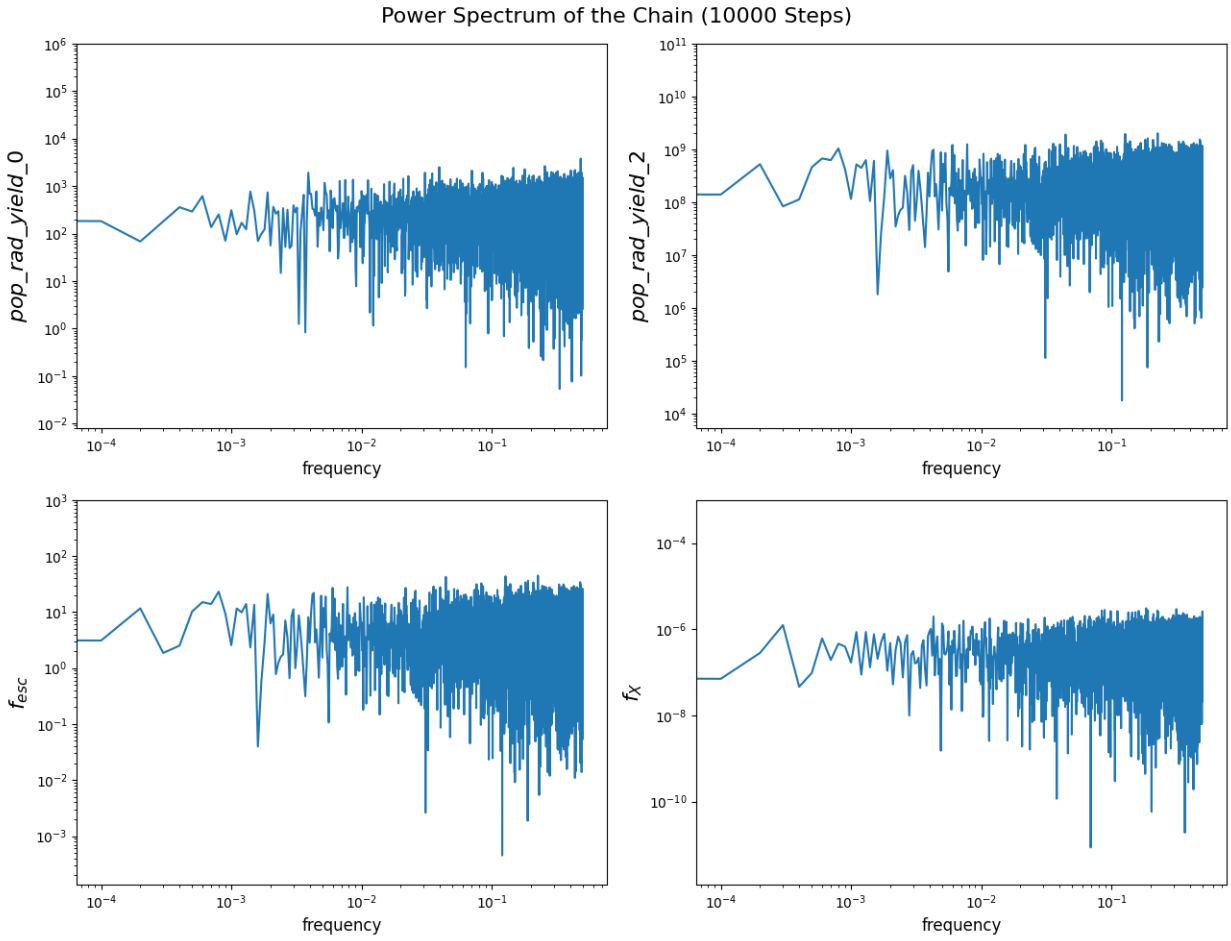

Figure 5.14: Power spectrum of the chain, Flat behaviour in low frequencies is in agreement with results of 5.13.

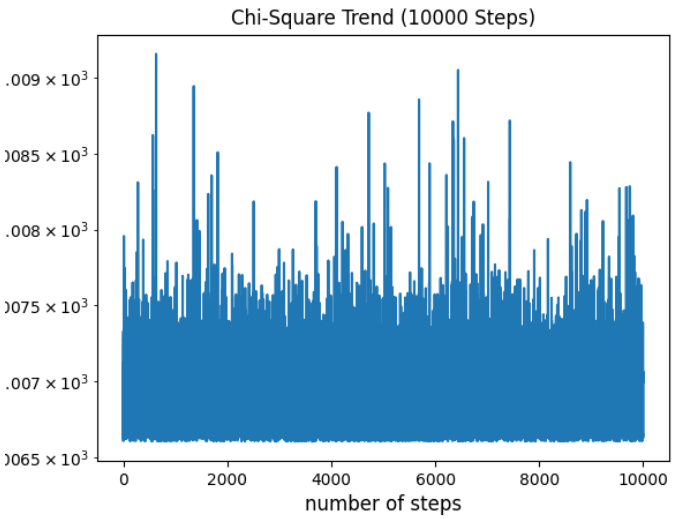

Figure 5.15: Trend of chi-square in the MCMC chain, The chain seems to have reached an equilibrium state.

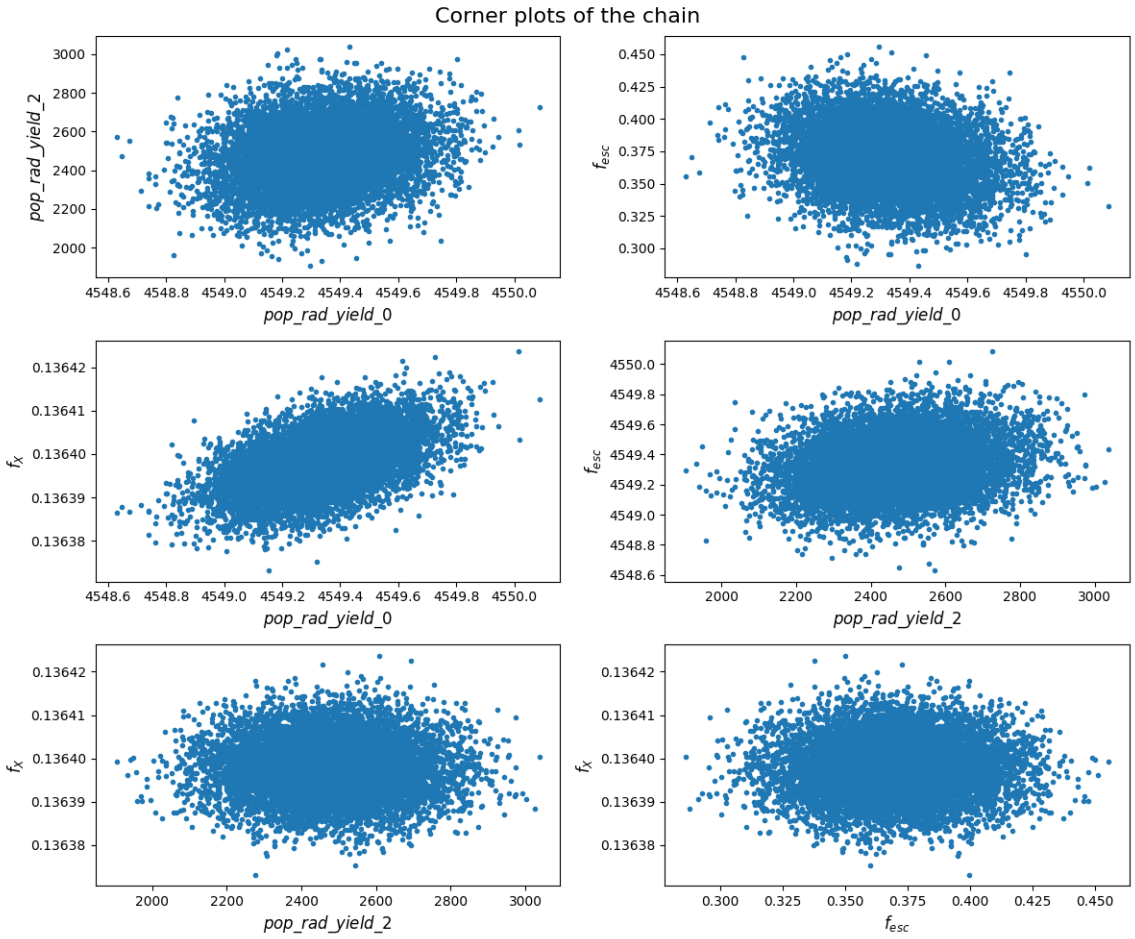

Figure 5.16: Corner plots of the chain indicating the correlation between the parameters specially in the lower left panel

Chapter 6

Discussion and Conclusion

6.1 Interpretation of the results

6.2 Summary of the main findings

6.3 Contributions and significance of the research

6.4 Limitations and future work

Chapter 7

Appendices

7.1 Code snippets and scripts

7.1.1 Levenberg-Marquardt

7.1.2 Markov Chain Monte Carlo

7.1.3 drawing samples from covariance matrix

Listing 7.1: Python example

```
def draw_samples(covariance_matrix, nset):
"""
covariance_matrix: covariance matrix
```

```
    nset:the number of samples

    returns: a matrix of samples

    This function calculates a series of correlated samples based on the pres

    covariance matrix and the number of samples.

    The shape of the output is (nset, m) where m comes from the shape of

    covariance matrix and it typically shows the number of parameters in the

    """

    #normalizing the covariance matrix

    D = np.diag(np.diag(covariance_matrix)) #diagonal matrix of covariance m

    D_sqrt = np.sqrt(D)

    D_inv_sqrt = np.linalg.pinv(D_sqrt)

    #normalized covariance matrix

    covariance_matrix_normalized = D_inv_sqrt @ covariance_matrix @ D_inv_sqr


    e,v = np.linalg.eigh(covariance_matrix_normalized)

    e[e<0]=0 #omitting any negative eigenvalues due to roundoff

    n = len(e)


    #make gaussian random variables

    g=np.random.randn(n, nset)
```

```
#scaling by the square root of the eigenvalues
rte=np.sqrt(e)
for i in range(nset):
    g[:,i]=g[:,i]*rte

#calculating the samples
samples = (v@g).T
#denormalizing the samples
samples_denormalized = samples @ D_sqrt
return samples_denormalized
```

Bibliography

[1] T. Minoda, S. Saga, T. Takahashi, H. Tashiro, D. Yamauchi, S. Yokoyama, and S. Yoshiura, "Probing the primordial universe with 21 cm line from cosmic dawn/epoch of reionization," *Publications of the Astronomical Society of Japan*, 2022.

[2] S. R. Furlanetto, S. P. Oh, and F. H. Briggs, "Cosmology at low frequencies: The 21 cm transition and the high-redshift universe," *Physics reports*, vol. 433, no. 4-6, pp. 181–301, 2006.

[3] J. R. Pritchard and A. Loeb, "21 cm cosmology in the 21st century," *Reports on Progress in Physics*, vol. 75, no. 8, p. 086901, 2012.