School of Physics and Astronomy
Queen Mary University of London

Modeling and Analysis of Exoplanet Radial Velocity Data

Saashiv Valjee (190288688)

June 15, 2023

Supervisor: Dr Craig Agnor

SPA7033U Extended Independent Project
30 Credit Units

Submitted in partial fulfilment of the requirements for the degree of
MSci FT AstroPhysics from Queen Mary University of London

Contents

List of Figures iv

1 Introduction 3

2 Methods of Analysis 6

2.1 Obtaining a period . . . 6

2.1.1 Observational Analysis . . . 6

2.1.2 Lomb-Scargle periodogram . . . 8

2.1.3 Frequency Grids . . . 11

2.2 Fitting parameters . . . 12

2.2.1 Initial Fitting Parameters . . . 12

2.2.2 Circular Orbits . . . 13

2.2.3 Eccentric Orbits . . . 14

2.2.4 Chi-Squared . . . 15

2.2.5 The True Anomaly . . . 16

2.2.6 Refining the Initial Fitting Parameters . . . 17

2.2.7 Phase-Folded Diagrams . . . 18

2.3 MCMC . . . 21

3 Analysis 25

3.1 Posterior Slicing . . . 26

3.2 Distribution and Colour-maps . . . 26

3.3 Best Fits . . . 29

3.4 Graphing the orbit . . . 30

3.5 Residual Analysis . . . 32

4 Limit Testing and Experimenting 36

4.1 Observational Period vs Lomb scargle . . . 36

4.1.1 Observational Analysis . . . 37

4.1.2 Lomb-Scargle . . . 40

4.1.3 Comparison . . . 41

4.2 Eccentric Model vs Circular System 44

5 Discussion **47**

6 Conclusion **49**

Bibliography **50**

List of Figures

2.1 A closer look at System-00's [9] radial velocity data, with a focus on the region of higher data density. By examining smaller portions of the data with better sampling, we can obtain better-defined periodic signals through observational analysis. 7

2.2 Example of radial velocity data from system-05 that is difficult to analyze through observation. The data is irregularly sampled, making it challenging to discern periodicity. This is a common issue with radial velocity data, but can be addressed through the use of techniques such as the Lomb-Scargle periodogram, which is particularly well-suited for handling unevenly spaced data. 8

2.3 Raw radial velocity data (ms) v time, and corresponding Lomb-Scargle periodogram for system-03. The periodogram shows a clear peak, indicating a strong periodic signal in the data. This allows us to determine the orbital period of the system with high accuracy, despite the unevenly spaced nature of the data. 10

2.4 System-05: time (days) vs radial velocity (m/s). Despite having potentially accurate input parameters, the raw radial velocity data plotted against our model still appears to be random and we have no grasp over which input parameters could be improved or what each one controls with respect to the differences between the model and the data 18

2.5 Phase-folded diagram of the radial velocity data for System-00. The x-axis represents the orbital phase, while the y-axis represents the radial velocity in units of m/s. This diagram is produced using the refined initial input parameters. As we can see, the data points appear to be clustered around a sinusoidal curve, indicating a periodic signal. This is a good sign that our initial fitting parameters are accurate and that further analysis will yield promising results. 19

2.6 System-02: phase folded time (days) vs radial velocity (m/s).This phase folded diagram shows that while the model appears to be out of phase with the data, the model's curve still closely resembles the data. The phase difference between the two can be controlled by only one variable, meaning that while further refinement is needed, most of the variables in the model are accurately representing the system. . 20

2.7 Evolution of χ^2 per iteration for an MCMC algorithm with poor performance. The graph shows a lack of convergence in χ^2, with the value fluctuating greatly between 100 and 1000. This suggests that the choice of parameters for the MCMC process was not effective in minimizing χ^2, leading to a poor evolution of χ^2. 22

2.8 The graph depicts the evolution of χ^2 per iteration for an MCMC algorithm with poor performance. Although it seems to have converged to a relatively suitable level, the shape of the curve is actually a cause for concern as the MCMC algorithm is designed to explore parameter spaces and occasionally try out parameter sets that may lead to higher χ^2. In this case, the algorithm appears to have failed to do so, indicating that it may be stuck in a local minimum or may not be exploring the parameter space effectively. 23

2.9 Evolution of χ^2 per iteration for an MCMC algorithm with good performance. The graph displays a steady decrease in χ^2 per iteration with subtle increases, indicating the algorithm is exploring parameter spaces in a systematic way, leading to a convergence on a suitable set of fitting parameters. 24

3.1 This graph displays the posterior distribution obtained from the MCMC algorithm for periodicity and drift velocity, which samples from the probability distribution of the parameters given the data. The distribution appears to be normally distributed, indicating that the algorithm has converged to a good set of fitting parameters 27

3.2 This color map shows the correlation between the velocity semi-amplitude and the time of periastron passage, with the color representing the value of χ^2. Darker colors indicate a lower χ^2, indicating a better fit. We can observe a clear correlation between these two parameters, with the best-fit values located in the center of the color map. The shape of the correlation is approximately elliptical, and the range of values for the parameters appears to be well-explored by the MCMC method. 28
3.3 The color map displayed demonstrates the correlation between velocity semi-amplitude and time of periastron passage, where the color represents the corresponding chi-squared value. However, in this version of the color map, we include the burn-in period from the MCMC method, which causes a significant loss of extractable information due to the decrease in resolution caused by having to zoom out of the low chi-squared area in order to include the burn-in within the plot. This emphasizes the importance of properly selecting the portion of data to include in the color map to ensure that the information obtained is of high quality and resolution. 29
3.4 This phase folded diagram displays the models fit for the phase versus the radial velocity in ms^{-1}. The strong similarity between the models curve and the data trend indicates the model has accurately predicted the behaviour of the system. 30
3.5 The graph displays the orbit of system-00 visualized through the use of the orbital equations. The x and y coordinates represent the position of the star as it orbits around the center of mass of the system in terms of astronomical units. This graph allows us to visualize the shape of the orbit to gain a better understanding of the system's dynamics. 31
3.6 System-00 plots showing the residuals against radial velocity in m/s on the left and the corresponding power spectrum on the right with a phase-folded inset. The power spectrum reveals a potentially significant periodicity at a high frequency, which is confirmed by the phase-folded inset. 32
3.7 System-00s residual phase folded plots showing a circular model of the data on the left, and an eccentric model against the data on the right. The graphs themselves are the phase folded time in days against the radial velocity in m/s. 33

3.8 System-00 plots showing the residual χ^2 evolution during the MCMC for a circular (left) and eccentric (right) model 34

3.9 System-00 plots showing the residual phase folded circular (left) and eccentric (right) models against the data. time in days vs radial velocity in m/s 34

4.1 Phase folded time (days) vs radial velocity (ms) The phase folded diagram shown here was obtained using a period obtained via observational analysis. The resulting period did not provide a good fit to the data and thus the model appears scattered and does not align well with the recorded values. 37

4.2 The chi squared evolution per iteration graph shows poor performance of the MCMC method. The chi squared values do not seem to converge to a suitable level. This indicates that the MCMC method is not efficiently exploring the parameter space and potentially selecting random parameter sets. 38

4.3 The bar code-like appearance indicates that the chains are not exploring the parameter space thoroughly, resulting in poor performance. The lack of diversity in the periodicity values across the chains suggests that the algorithm is getting stuck in local minima and not exploring other regions of the parameter space. This is problematic for MCMC as the goal is to explore the entire parameter space to find the global minimum, rather than being stuck in local ones. 38

4.4 The Lomb-Scargle periodogram displays the power spectral density of the signal as a function of frequency. The peak in the periodogram indicates the most likely period of the signal. The inset phase folded diagram shows the signal folded at the most likely period 40

4.5 This graph shows a good evolution of the periodicity per step for the MCMC algorithm. The plot displays a consistent and stable periodicity over the posterior slice. This indicates that the MCMC algorithm has successfully converged to a good solution, and that the obtained periodicity is likely to be accurate. 41

4.6 Comparison of phase folded diagrams for System-00, showing the impact of initial period value on model performance. The left diagram, with initial period value given by observational analysis, shows poor fit to the data. The right diagram, with initial period value given by Lomb-Scargle analysis, shows a much improved fit after MCMC optimization. The red and blue lines represents the best-fit model after MCMC. 44
4.7 This phase folded diagram displays the radial velocity data of system-00, previously thought to be circular, but now fit with an eccentric orbit. 45
4.8 This plot shows the evolution of eccentricity per iteration of the MCMC algorithm. The model began with a relatively high eccentricity (0.495-0.5) but MCMC quickly stepped down to near 0 eccentricities. 45
4.9 This histogram shows the distribution of eccentricity values obtained from the MCMC algorithm. 46

Abstract

In this dissertation, I present a comprehensive investigation into the determination of periods and orbital fitting parameters of astronomical objects, with a focus on the radial velocity technique. Through a combination of observational analysis, Lomb-Scargle periodogram, and various orbit fitting techniques, the research aims to enhance our understanding of celestial phenomena by comparing the strengths and limitations of different methods for period determination and orbit modeling.

The study begins by examining the use of observational analysis and Lomb-Scargle periodogram for period determination, highlighting the differences and respective strengths and weaknesses of each method. Additionally, the complexities of fitting parameters for both circular and eccentric orbits are explored, utilizing techniques such as Chi-Squared and Phase-Folded Diagrams to refine initial fitting parameters.

To strengthen our analysis, we incorporate Markov Chain Monte Carlo (MCMC) techniques for a more robust and reliable means of orbit fitting. This allows for thorough exploration of the parameter space, posterior slicing, distribution and colour-map generation, and the identification of best fits.

The importance of employing a multi-faceted approach for period determination and orbit fitting is demonstrated, as well as the need to consider both circular and eccentric orbits when modeling celestial objects. This research contributes to the ongoing advancement of our understanding of the cosmos and provides valuable insights for future studies aiming to further enhance the accuracy and efficiency of modeling astronomical objects.

1 Introduction

Exoplanets are planets that orbit stars outside of our solar system and have been discovered using various methods, including radial velocity, transits, gravitational microlensing, astrometry, and direct imaging. Among these methods, the radial velocity technique stands out as a crucial tool in the search for exoplanets and will be the primary focus of this dissertation.

The discovery of 51 Pegasi b, the first exoplanet detected orbiting a main sequence star, was a significant breakthrough in the field of exoplanet research. Prior to this discovery, it was believed that planets could only form around cold, lower mass stars. However, 51 Pegasi b challenged this notion and validated the radial velocity technique, paving the way for further advancements in the field.

The radial velocity technique relies on measuring changes in a star's velocity caused by the gravitational pull of an orbiting planet. As a potential exoplanet orbits its host star, the star wobbles under the influence of its gravitational pull, causing the light received from the star to appear redshifted. By analyzing the shift in the star's spectrum, we can extract valuable information about the exoplanet's properties, such as its mass, distance from the star, and eccentricity. However, this method has limitations, such as its inability to directly detect an exoplanet's size or atmospheric properties.

The radial velocity technique is favored by many astronomers due to its ability to provide high precision measurements of exoplanet masses and orbits. This method is particularly effective at detecting more massive exoplanets and those that are closer to their host star, as described by the equation for gravitational pull:

$$F_g = \frac{Gm_1m_2}{r^2} \tag{1.1}$$

However, the radial velocity technique has its drawbacks, as low mass planets or those further away from their host star may not cause a significant shift in the star's spectrum, making it difficult to obtain useful data. To overcome these limitations, various analysis techniques and algorithms have been developed to maximize the information that can be extracted from radial velocity data.

The key components of the radial velocity technique involve a high-precision spectrometer, stable light source, and reference spectrum. By monitoring the changes in the star's light over time, we can detect the wobble caused by the exoplanet's gravitational pull. To utilize this technique, we require spectroscopic observations with high precision over a sufficiently long time span, at least the duration of one full orbit.

To analyze the data and obtain useful information, we must fit the data with an appropriate algorithm to optimize our fitting parameters, such as the exoplanet's period and characteristic variables. This fitting process enables us to accurately represent, predict, and visualize the orbit of an exoplanet. By employing the radial velocity technique and analyzing the resulting data, we can gain valuable insights into exoplanetary systems and their formation, as well as further our understanding of the universe as a whole.

The primary objective of this dissertation is to analyze a sample of exoplanets orbiting various types of stars using the radial velocity technique. Our focus will be on studying the exoplanets' orbital parameters and masses, and investigating how these parameters relate to the properties of their host stars. By analyzing the radial velocity data, we hope to gain insights into the formation and evolution of exoplanetary systems.

To achieve these goals, we will use a combination of observational data analysis, Lomb-Scargle periodograms, fitting algorithms, and Markov Chain Monte Carlo (MCMC) simulations. We will begin by obtaining the period of the exoplanet's orbit using the Lomb-Scargle periodogram, a common tool in the analysis of time-series data. Next, we will use fitting algorithms to optimize the parameters that describe the exoplanet's orbit, such as its eccentricity and true anomaly. Finally, we will use MCMC simulations to sample the posterior distribution of the fitting parameters and obtain a better understanding of the uncertainties associated with our measurements.

Overall, this dissertation aims to contribute to our understanding of exoplanetary systems and their formation, and to provide insights into the techniques and challenges associated with radial velocity data analysis.

2 Methods of Analysis

2.1 Obtaining a period

When analyzing data, the first step is to examine it thoroughly. In this study, we use time-series radial velocity data recorded in Julian days, a time measurement system commonly used in astronomy. While Julian days are useful for astronomical calculations, we convert the time recordings for all data points into days relative to the first recording to simplify our analysis. The orbital period is defined as the time taken to complete one full orbit around a center of mass, and is a critical parameter to obtain. There are also are multiple ways to do so. One common method is through observational analysis, which involves visually inspecting the data for periodic patterns. However, this approach can be subjective and may not be suitable for all data-sets. Another popular technique is the Lomb-Scargle periodogram, which uses Fourier analysis to identify periodic signals in irregularly sampled data. This method is particularly effective for data-sets with unevenly spaced observations, but it requires careful parameter tuning to obtain accurate results. Ultimately, the choice of method depends on the characteristics of the data-set and the research question at hand.

2.1.1 Observational Analysis

One of the primary ways to obtain an estimate of the orbital period of a system is through observational analysis. By graphically visualizing our data using tools like Matplotlib, we can often identify clear and distinct signs of periodicity in the signal. However, this process can become more challenging when the data spans over a long period of time, especially relative to the orbital period of the system.

To address this challenge, we can zoom in on regions of the data that are better recorded and show a more pronounced periodic signal. This can help us to identify the most well-defined waves in the signal and use them to estimate the orbital period

of the system. For example:

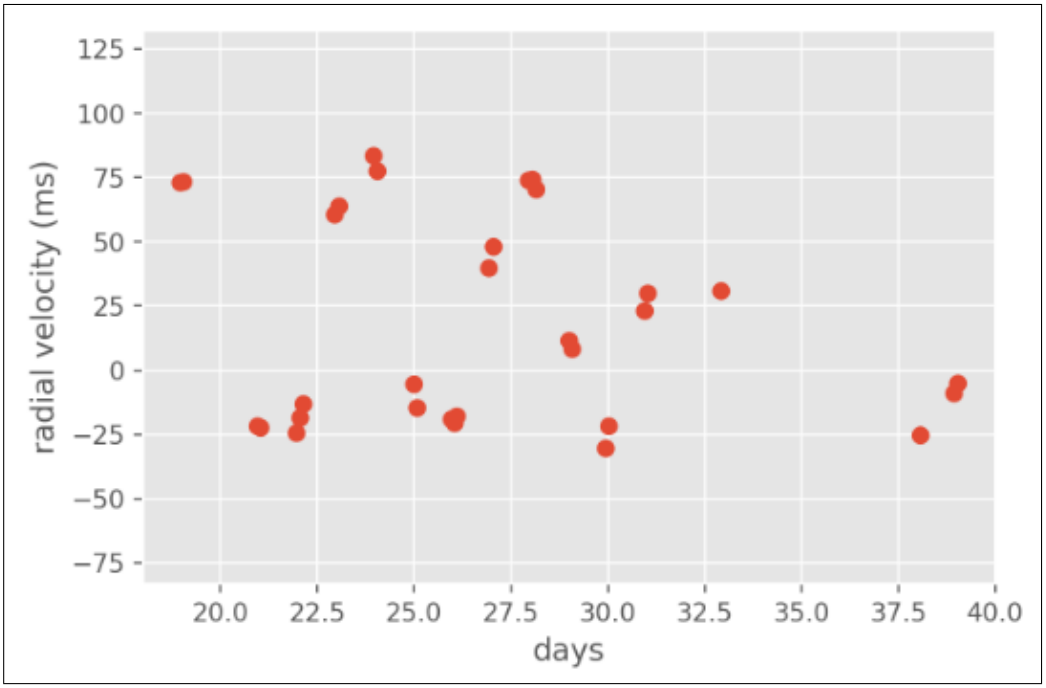

Figure 2.1: A closer look at System-00's [9] radial velocity data, with a focus on the region of higher data density. By examining smaller portions of the data with better sampling, we can obtain better-defined periodic signals through observational analysis.

in Figure 2.1, we can clearly see two well-defined waves in the region between the 22.5-day and 30.0-day marks. To calculate the orbital period from this data, we can take the duration of all the full cycles of the periodic signal we've identified and divide it by the number of cycles:

$$\frac{\Delta t}{n\lambda} \tag{2.1}$$

Where:

- $n\lambda$ represents how many portions/full waves there are
- Δt represents the total time all portions/full waves last for

In this case, we can calculate the period by dividing the duration of the two cycles we've identified by two. This gives us an estimate of the orbital period of roughly 3.75 days.

It's important to note that there are multiple ways to estimate the orbital period of a system, and this process can become more complex when dealing with data that

has noise or is influenced by other factors. Nonetheless, through careful observation and analysis of our data, we can often obtain valuable insights into the orbital characteristics of celestial objects.

2.1.2 Lomb-Scargle periodogram

Sometimes, obtaining accurate, consistent, and reliable radial velocity data can be challenging. Various sources of noise, irregular sampling, or unfavorable weather conditions could compromise the quality of the data. Radial velocity, in particular, is highly susceptible to stellar jitter, instrumental noise, and uneven sampling, which can result in messy data that makes it difficult to visually obtain an accurate estimation of the system's orbital period:

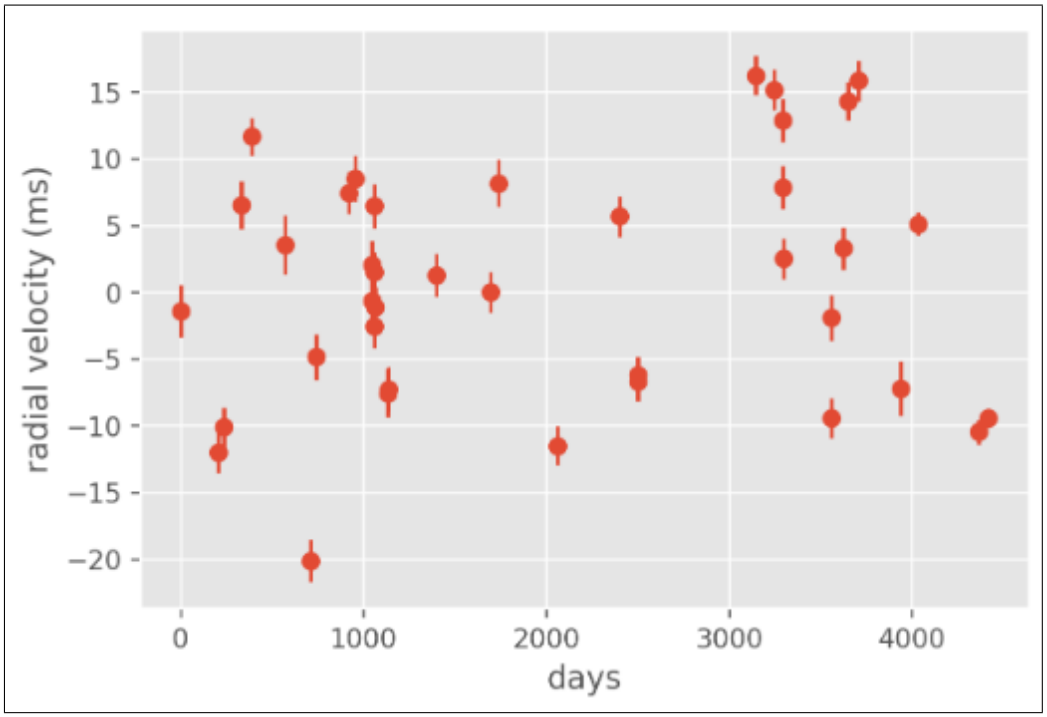

Figure 2.2: Example of radial velocity data from system-05 that is difficult to analyze through observation. The data is irregularly sampled, making it challenging to discern periodicity. This is a common issue with radial velocity data, but can be addressed through the use of techniques such as the Lomb-Scargle periodogram, which is particularly well-suited for handling unevenly spaced data.

While there are still semi-clear signs of periodicity, it will be much harder to observationally obtain a value for the systems orbital period. Here it is largely due to the uneven sampling, as we barely have a full clear cycle of readings. As such, we turn to the Lomb-Scargle Periodogram.

The Lomb-Scargle periodogram is a widely recognized method for estimating the power spectral density of periodic time series data. Introduced by Lomb in 1976 [10] and later developed by Scargle in 1982, the Lomb-Scargle periodogram fits a sine wave with a known frequency using the least squares approach. This process is repeated over a selected range of frequencies, which enables us to calculate the power of a fit and the known frequency, producing enough data to create a periodogram.

The Lomb-Scargle periodogram is particularly useful for obtaining accurate periodicity readings from radial velocity data and excels at dealing with unevenly spaced data [7]. This is advantageous because a large majority of the time, it is not possible to make radial velocity readings with even time intervals. To create a Lomb-Scargle periodogram, the data is subjected to a Fourier-like power spectrum that enables the visualization of the period of oscillation.

In addition to being particularly adept at handling unevenly spaced data, the Lomb-Scargle periodogram is far more efficient than alternative methods such as Fourier series analysis when dealing with larger amounts of data because it eliminates the need to compute the entire transform. Instead, it will only calculate the frequencies within the frequency grid. Its ability to provide a detailed, accurate analysis of periodicity in radial velocity data makes it an indispensable tool in the field of exoplanet research, helping to improve our understanding of celestial phenomena and expand our knowledge of the universe.

To implement the Lomb-Scarlge periodogram, initially, I had written my own code to calculate the power spectra using a formula [10]:

$$P_{LS}(f) = \frac{1}{2}\left[\frac{\left(\sum_n g_n \cos\left(2\pi f(t_n-\tau)\right)\right)^2}{\sum_n g_n \cos^2\left(2\pi f(t_n-\tau)\right)} + \frac{\left(\sum_n g_n \sin\left(2\pi f(t_n-\tau)\right)\right)^2}{\sum_n g_n \sin^2\left(2\pi f(t_n-\tau)\right)}\right] \tag{2.2}$$

$$\tau = \frac{1}{4\pi f}\tan^{-1}\left(\frac{\sum_n \sin\left(4\pi f t_n\right)}{\sum_n \cos\left(4\pi f t_n\right)}\right) \tag{2.3}$$

These equations represent the Lomb-Scargle periodogram, which is used to analyze unevenly sampled data to determine the periodicity of time series data. $P_{LS}(f)$ represents the power at frequency f, while τ is the phase offset that shifts the data to the center of the frequency range.

However, I later opted to use the Lomb-Scargle periodogram provided by AstroPy [6] for its flexibility and convenience. The data is fed into the method and the power spectra is computed, representing the periodic strength of a repeating signal at each frequency within the frequency grid. This allows us to easily visualize the data with subplots, displaying the original data and the periodogram. To obtain the period, we simply find the index of the maximum power using NumPy's argmax function, which returns the location of the largest value in an array. We apply this function to the y-axis to locate the highest signal, and since the two lists have the same length, the corresponding periodicity will have the same location within the x-axis array.

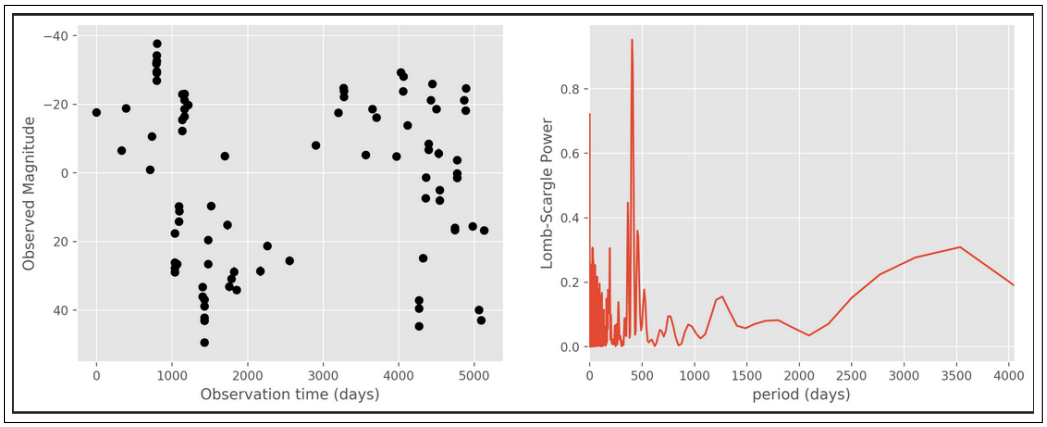

Figure 2.3: Raw radial velocity data (ms) v time, and corresponding Lomb-Scargle periodogram for system-03. The periodogram shows a clear peak, indicating a strong periodic signal in the data. This allows us to determine the orbital period of the system with high accuracy, despite the unevenly spaced nature of the data.

With the Lomb-Scargle periodogram, we were able to accurately determine the orbital period of data that was previously difficult to analyze through observation alone. The Lomb-Scargle periodogram is particularly beneficial in cases where the data is unevenly spaced, which is a common issue with radial velocity data. Additionally, this method is more efficient and suitable for analyzing large amounts of data when compared to Fourier series analysis. Overall, the Lomb-Scargle periodogram is a powerful tool for studying periodic signals in astronomical data, providing a means to improve our understanding of celestial objects and systems.

2.1.3 Frequency Grids

The Lomb-Scarlge method is a powerful tool for determining the periodicity of time series data, but its accuracy and reliability depend on the frequency grid and sampling rate used. Deviations from the optimal values can introduce inaccuracies and false positives in the power spectra. Therefore, it is essential to choose suitable values for these parameters when applying the Lomb-Scargle method to data.

One important concept to consider is the Nyquist frequency [10], which represents the highest frequency that a signal can be reconstructed from without introducing inaccuracies and false positives. The Nyquist frequency is half the sampling rate of a time series data-set. However, in the case of unevenly sampled data, the Nyquist frequency is not well defined, as there is no concrete sample rate. One workaround for this issue is to calculate the mean sample rate, but this modified version may not always provide a reliable estimate for the highest frequency that can be accurately analyzed.

When constructing the frequency grid for the Lomb-Scargle periodogram, it is important to avoid including frequencies that are too high or too low. This is because including these frequencies can lead to the creation of a noisy and unreliable power spectrum. At high frequencies, the power spectrum becomes noisy due to the limited time resolution of the data, leading to inaccurate results. Similarly, including frequencies that are too low can cause a large amount of aliasing, where the periodic signal is misrepresented due to inadequate sampling. Therefore, it is crucial to select a suitable frequency range based on the data, while avoiding these high and low frequencies.

To address this challenge, the Lomb-Scargle periodogram built into AstroPy offers an auto-power function that helps create a suitable frequency grid given the maximum and minimum frequency. This function reduces the sensitivity and impact of the input frequencies by creating a reliable and sensible frequency grid that fits the input values. As a result, the impact of unevenly sampled data is reduced, since the maximum frequencies can be further away from the data-set's optimal value.

In summary, understanding the Nyquist frequency, selecting a suitable frequency range, and utilizing the auto-power function in AstroPy are critical steps for ensuring the accuracy and reliability of the Lomb-Scargle method when dealing with unevenly

sampled data.

2.2 Fitting parameters

In order to extract more information about the observed system, we must estimate the other fitting parameters to use alongside the orbital periodicity previously obtained. To achieve this, we utilize a model based on Kepler's laws of planetary motion, which accurately describes the motion of a planet around a star under the influence of gravity.

In this section, we explore the fitting parameters of circular and eccentric models and introduce the χ^2 statistic to assess how well the model fits the data. We will also discuss the process of calculating the true anomaly and slightly refining the fitting parameters to make them suitable for use in the Markov Chain Monte Carlo (MCMC) method. Finally, we use phase folded diagrams to confirm that the fitting parameters are optimal and ready for further analysis.

2.2.1 Initial Fitting Parameters

To use our period obtained by the Lomb-Scargle periodogram, we must obtain the rest of the fitting parameters. These parameters include the eccentricity e and the orbital inclination ω (if the orbit is eccentric), the drift velocity V_0, the velocity semi-amplitude K, and the time of periastron passage T_P. The eccentricity refers to a dimensionless quantity that represents how much the orbit of a body deviates from a perfect circle. The orbital inclination is used to describe the angle between the plane of orbit and the observers reference plane. The drift velocity of a body is a combination of the objects true motion and the motion of the observer. The velocity semi-amplitude is used to describe the radial velocity variations of a star caused by an orbiting body. Achieving a certain level of precision for each of these parameters is crucial for the accuracy of our analysis.

Fortunately, we can use suitable ranges for each of these parameters to aid us in finding the values that satisfy our precision requirement:

- e must range from 0 to 1 to reflect the eccentricity of the orbit.
- ω must range from 0 to 2π to reflect the angle of inclination of the orbit.

- T_P must range between 0 and the period to reflect the time of periastron passage.
- V_0 can be estimated by averaging the radial velocity, which provides a baseline for the velocity measurements.
- K can be estimated by dividing the highest difference between any two radial velocity measurements by 2, providing an estimate for the velocity amplitude of the orbit.

By obtaining precise initial fitting parameters, we can improve the efficiency and accuracy of our methods later on in the analysis.

2.2.2 Circular Orbits

In radial velocity analysis, circular orbits are a simpler case to model than eccentric orbits. The radial velocity of a star in a circular orbit can be described using the following equation [3]:

$$V_r = V_0 + K \cos\left(\frac{2\pi(t - T_p)}{p}\right) \tag{2.4}$$

Where:

- V_r is radial velocity
- V_0 is average drift velocity of the star
- K is the velocity semi amplitude
- T_p is the time of periastron passage
- t is time
- P is period

This equation is derived from Kepler's laws, which describe the motion of a planet around a star under the influence of gravity. In a circular orbit, the planet's distance from the star remains constant, and the velocity of the planet is given by $v = 2\pi r/P$, where r is the distance from the star and P is the orbital period. The velocity of the planet can be related to the radial velocity of the star using the Doppler effect,

resulting in the equation above.

Compared to other models, we prefer this equation because it has a different set of fitting parameters, which makes the process of finding them less involved. The circular model assumes that the orbit of the planet around the star is perfectly circular, and can be useful when analyzing radial velocity data from systems where the eccentricity is low.

2.2.3 Eccentric Orbits

When modeling an eccentric orbit, the radial velocity of the star can be expressed as [3]:

$$V_r = V_0 + K\cos(f + \omega) + \cos(\omega) \tag{2.5}$$

Where:

- f is the true anomaly
- e is the eccentricity
- ω is angle of inclination

This equation takes into account the eccentricity and angle of inclination of the orbit, represented by e and ω respectively. The true anomaly f represents the angle between the periastron (point of closest approach) and the current position of the star in its orbit. The velocity semi-amplitude K represents the amplitude of the radial velocity variations, and V_0 is the average drift velocity of the star.

The equation for eccentric orbits, Equation 2.5, is derived from Kepler's laws of planetary motion. The equation includes additional parameters, such as eccentricity and angle of inclination, to account for the non-circular shape of the orbit. By incorporating these parameters, the model can more accurately describe the motion of a planet around a star under the influence of gravity.

By using this equation, we can accurately model the radial velocity of a star in an eccentric orbit, and estimate the parameters that describe the orbit's shape and motion. These parameters can then be used to make further predictions and analyses of the system.

2.2.4 Chi-Squared

The χ^2 value is an important tool in fitting a model to data, as it provides a quantifiable measure of the accuracy of the fit. In astronomy, we can use the χ^2 equation to test different combinations of initial fitting parameters and find the combination that results in the lowest χ^2 value.

However, the standard χ^2 equation does not account for variations in the quality of the data. To address this, we use a "weighted" χ^2 equation, which assigns a weight to each data point based on its estimated error. This weighted equation is given by:

$$\chi^2 = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{y_{fit}(x_i) - y_i}{\sigma_i} \right)^2 \tag{2.6}$$

Here, y_{fit} is the value of the model function evaluated at x_i with the current set of fitting parameters, y_i is the observed value at x_i, σ_i is the estimated error associated with the observed value, and the sum is taken over all N data points.

The weighted χ^2 equation takes into account the quality of each data point, giving more weight to the more precise measurements. This is particularly useful in cases where the error associated with each data point is not equal, which is often the case in astronomy.

A smaller weighted χ^2 value for our fit indicates that the fit is closer to the model, and it is important to note that the weighted χ^2 value is also normalized to the number of data points. Normalizing the χ^2 value ensures that the value is found consistently between different sets of data, and reduces the impact of outliers.

By analyzing the confidence intervals on the model parameters, we can gauge the reliability of our model. The fitting parameters, including the eccentricity e and orbital inclination ω (for eccentric orbits), the drift velocity V_0, the velocity semi-amplitude K, and the time of periastron passage T_P, must be determined with a certain level of precision to ensure the accuracy and reliability of our model. The range for each of these values must be carefully selected to find the values that satisfy the precision requirement.

2.2.5 The True Anomaly

The equation (2.4) shows that obtaining a time series radial velocity array is straightforward - by putting in the appropriate parameters and times, we can obtain the radial velocities. However, the situation is more complicated for the eccentric orbit model, as we can see from (2.5). In this case, there is no direct time variable, and instead, we cycle through a variable known as the true anomaly to obtain a list of radial velocity values. To calculate the true anomaly [3] at a given time, we first need to find the mean anomaly. This is given by:

$$\frac{2\pi(t - T_p)}{P} \tag{2.7}$$

The next step in calculating the true anomaly is to solve Kepler's equation [5]:

$$E - esin(E) = M \tag{2.8}$$

Where:

- E is the eccentric anomaly
- M is the mean anomaly

One obstacle we face is that the equation for the eccentric anomaly, (2.8), is a transcendental equation, which cannot be solved analytically. To obtain a numerical solution, we will employ the Newton-Raphson method [1], a commonly used technique for solving tricky equations. This method involves providing an initial guess for the root of the equation, evaluating the function and its derivative at this guess, and using the formula:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_0)} \tag{2.9}$$

to update the guess and iteratively converge to the root. The Newton-Raphson method is favored for its relatively quick convergence and versatility in solving a wide range of nonlinear equations.

After we have solved the Kepler equation and have a value for our eccentric anomaly, we can finally find the true anomaly using:

$$\cos(f) = \frac{\cos(E) - e}{1 - e\cos(E)} \tag{2.10}$$

$$\sin(f) = \frac{\sqrt{1-e^2}\sin(E)}{1-e\cos(E)} \tag{2.11}$$

$$f = \tan^{-1}\left(\frac{\sin(f)}{\cos(f)}\right) \tag{2.12}$$

Arctan2 is a preferred function over atan because it can handle the edge cases and singularities where atan fails. For example, if we use atan to calculate the angle of a point (x, y), we may get incorrect results when $x = 0$, resulting in division by zero. On the other hand, arctan2 is designed to handle these cases gracefully, by taking both x and y as input and determining the correct quadrant of the angle.

This is especially important in astronomy, where we often deal with trigonometric functions and angles in a range of $[0, 2\pi]$. For instance, when computing the true anomaly using $\arctan 2$, we can get the correct value in the range of $[0, 2\pi]$ even when y is negative or zero, or when x is negative and y is positive. This robustness and convenience make arctan2 a much better choice for calculating angles in astronomical applications.

2.2.6 Refining the Initial Fitting Parameters

To obtain a suitable set of initial fitting parameters, we must first determine the range of the parameters we cannot currently calculate and calculate the rest. Then, we can create a set of arrays or lists representing the potential range of values for each parameter. Using nested for-loops, we can iterate over each value in each list and calculate the value of χ^2 using (2.6). This method is similar to a "brute force" approach, where we check every possible combination of input parameters against how well they fit against our observational data. We aim to minimize χ^2 by finding the combination of parameters with the closest fit.

This method is useful because it can work for many different parameters at once, even in high input parameter dimensionality. However, increasing either the number of values contained in the lists for each parameter or the number of parameters themselves can drastically increase the time this method takes due to its exponential nature.

2.2.7 Phase-Folded Diagrams

Once we have a preliminary set of fitting parameters, it's important to ensure their accuracy before proceeding with more fine-tuning. Although we can visualize our model plotted against the raw data, it can be difficult to determine whether the model is an accurate representation of the data and which parameters may need adjusting. For example:

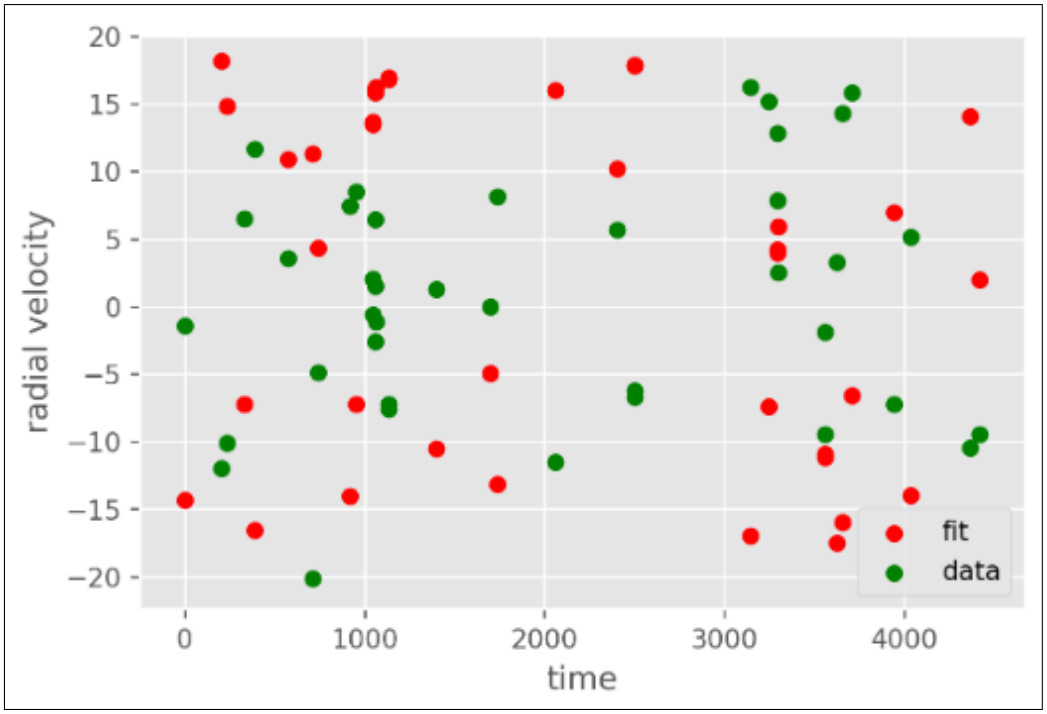

Figure 2.4: System-05: time (days) vs radial velocity (m/s). Despite having potentially accurate input parameters, the raw radial velocity data plotted against our model still appears to be random and we have no grasp over which input parameters could be improved or what each one controls with respect to the differences between the model and the data

By simply plotting our model against the raw data, it can be difficult to determine which parameter(s) may be causing the fit to not align with the data or which parameter(s) are controlling specific aspects of the data. To overcome this issue, we can plot Phase-Folded diagrams which provide a graphical representation of data commonly found in radial velocity data analysis.

A phase-folded diagram is a common way to represent data in radial velocity analysis. This graphical representation is achieved by plotting the data as a function of a repeating cycle, such as an orbital period signal. To create the phase-folded data, we must map the data's times relative to the system's last complete cycle. This can be done using the modulo operation, as shown in the following code:

$$tfolded = x\%Pguess$$

After folding the data, we can view our phase-folded diagram by plotting the mapped time against the corresponding radial velocity readings. The resulting diagram should show the x-axis representing portions of any elapsed cycle, along with the corresponding radial velocity reading.

By graphing a phase folded diagram of our model on top of a folded version of the original data, we can assess the accuracy of our fitting parameters. A good model will closely follow the trend of the original data, while a bad model won't. A well-suited set of initial parameters can produce phase folded diagrams that accurately match the data trend, as shown in the following example:

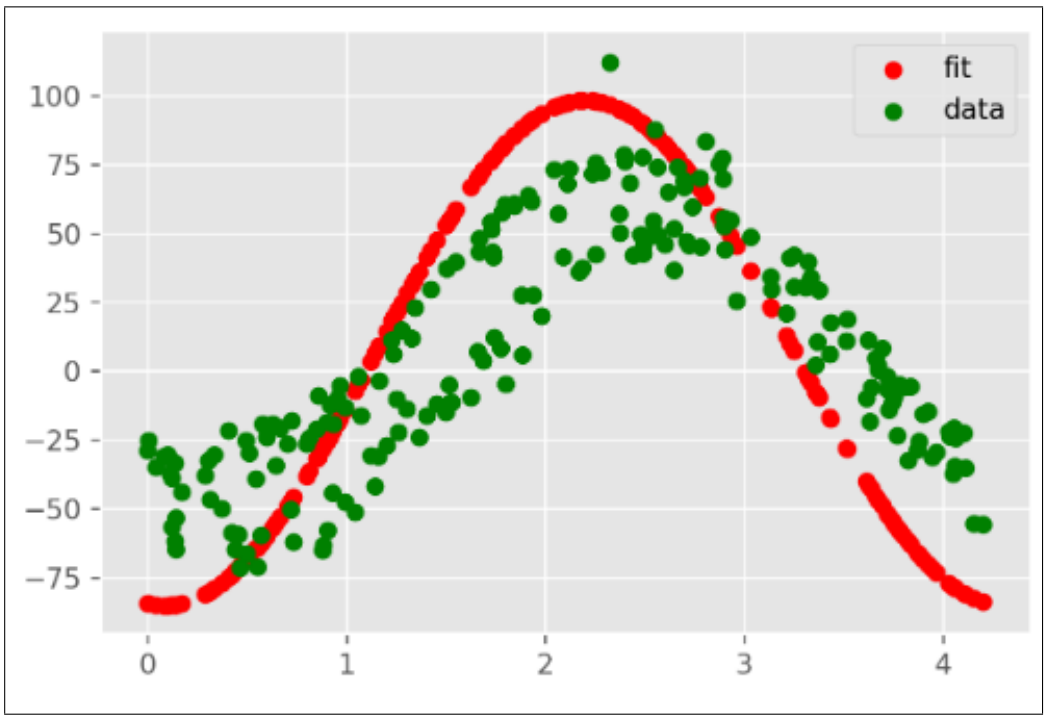

Figure 2.5: Phase-folded diagram of the radial velocity data for System-00. The x-axis represents the orbital phase, while the y-axis represents the radial velocity in units of m/s. This diagram is produced using the refined initial input parameters. As we can see, the data points appear to be clustered around a sinusoidal curve, indicating a periodic signal. This is a good sign that our initial fitting parameters are accurate and that further analysis will yield promising results.

The model closely follows the data, meaning all the parameters used to create our fit are likely very close to the actual values representing the system, and that they're ready to be refined further. Phase folded diagrams can sometimes give off a deceptive portrayal of the model accuracy. Take this diagram for example:

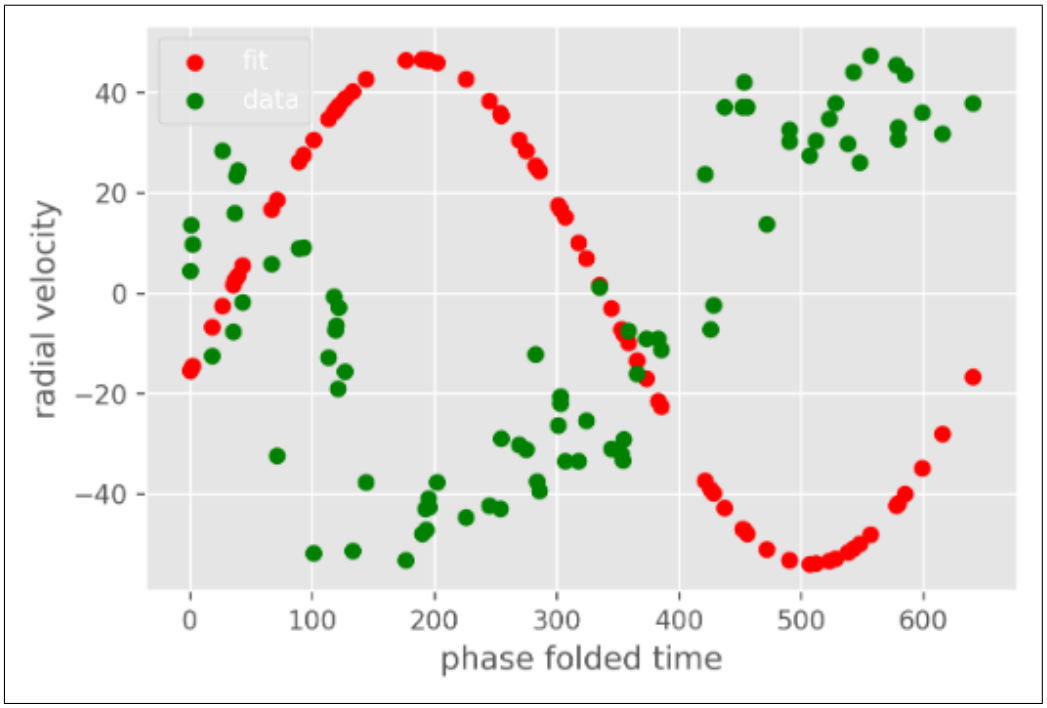

Figure 2.6: System-02: phase folded time (days) vs radial velocity (m/s).This phase folded diagram shows that while the model appears to be out of phase with the data, the model's curve still closely resembles the data. The phase difference between the two can be controlled by only one variable, meaning that while further refinement is needed, most of the variables in the model are accurately representing the system.

The phase folded diagram in Figure 2.6 may seem misleading as it appears that the data and the fit are completely out of phase, but it still provides valuable information. Even though one parameter may be inaccurate, we can tell that the fit still closely resembles the data. Experimentation can help us determine which parameter affects which aspect of the phase folded diagrams, making them a powerful tool for analyzing data. Each parameter controls a specific characteristic of our data. For example, the period controls the defined periodicity of the recorded data. An inaccurate period will make the data look random, while an accurate one will create a wave-like appearance. The semi-amplitude velocity controls the amplitude of the fit. A good fit will have matching amplitudes with the data. The time of periastron passage can be determined to be accurate or not depending on whether the fit and the data appear in or out of phase. Finally, the drift velocity can also affect the alignment between the fit and the data if not properly calculated. A good and complete set of fitting parameters will create phase folded diagrams similar to Figure 2.5.

2.3 MCMC

Obtaining a precise set of fitting parameters is crucial to our analysis, and we need a flexible and reliable method to achieve this. Many before us have used the MCMC (Markov Chain Monte Carlo) method [4][2], which is a powerful and versatile tool that has been used in various fields. For our analysis, the MCMC method's ability to efficiently explore high dimensionality spaces is invaluable.

To use the MCMC method, we start by providing a set of initial guesses for each input parameter. Using our fitting equation (either (2.5) for an eccentric orbit or (2.4) for a circular one), we create a list of radial velocity values within a given time range and then calculate χ^2. Next, we randomly alter each input parameter by slightly increasing or decreasing it by a specific "step size", while keeping track of the value that each parameter takes.

It is important to remember that some variables have restricted ranges, as outlined in 2.2.1. For example, eccentricity can only range from 0 to 1, the time of periastron passage ranges from 0 to the time of one full period, and the angle of inclination ranges from 0 to 2π. Ignoring these constraints could break the method.

Another important factor is that the scale of our input parameters varies greatly, so we cannot alter them all by the same step size. To address this, we alter each input parameter based on a fraction of its own value. Specifically, we create a random number between (using 100's as examples, the scale of which the step-range is cut by can be fine tuned by us) one one-hundredth and negative one one-hundredth of each input parameter, add or subtract it (depending on the sign of the random number), and then recalculate χ^2. The scalar variable that controls the range of random numbers is specific to each input parameter and needs to be fine-tuned.

We can compare different sets of input parameters to determine which one returns a better model that closely represents our observational data. However, we do not simply choose the set with a lower χ^2. Instead, we use a piece-wise probability function to decide whether to accept or reject a new set of parameters. The function is given by:

$$P_{\mathrm{a}} = \begin{cases} 1 & \text{if} \quad F_{i+1} < F_i \\ \exp\{-\beta(F_{i+1} - F_i)\} & \text{if} \quad F_{i+1} \geq F_i \end{cases} \tag{2.13}$$

Where:

- F_i is the function at variable set i
- F_{i+1} is the function at variable set $i + 1$
- β is our acceptance parameter

Here, F_i and F_{i+1} denote χ^2 at two different sets of input parameters, while β is the acceptance parameter. To utilize this probability function, we generate a random number between 0 and 1 and compare it to the second case of the function. This simulates the probability of accepting a new set of parameters.

It's important to fine-tune the value of the acceptance parameter to achieve good performance of the MCMC method. We can plot the development of χ^2 with each iteration to assess whether the MCMC is able to converge on an absolute minimum with sufficient confidence:

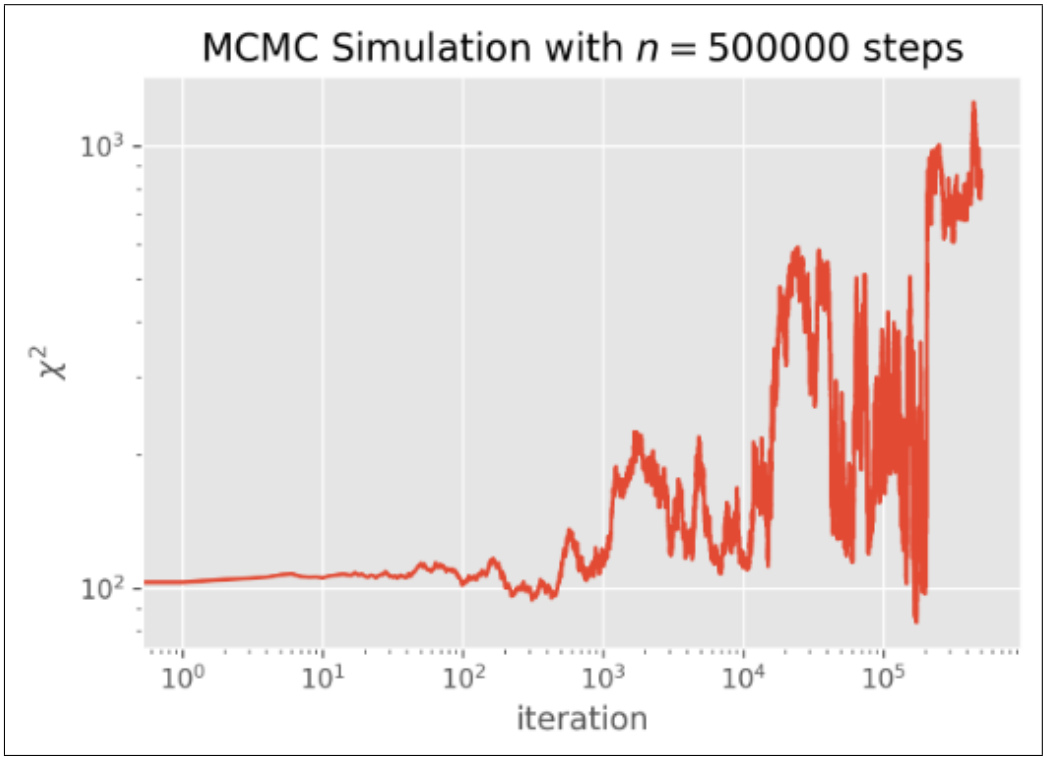

Figure 2.7: Evolution of χ^2 per iteration for an MCMC algorithm with poor performance. The graph shows a lack of convergence in χ^2, with the value fluctuating greatly between 100 and 1000. This suggests that the choice of parameters for the MCMC process was not effective in minimizing χ^2, leading to a poor evolution of χ^2.

If we observe poor convergence of the MCMC as with Figure 2.7, it may indicate that the acceptance parameter was set too low. In (2.13), we can see that a low acceptance parameter will increase the probability of accepting a parameter set that does not necessarily lead to a lower χ^2. However, the effect of the acceptance parameter on the probability of accepting a new parameter set may vary depending on how probability is simulated.

Take this evolution of χ^2:

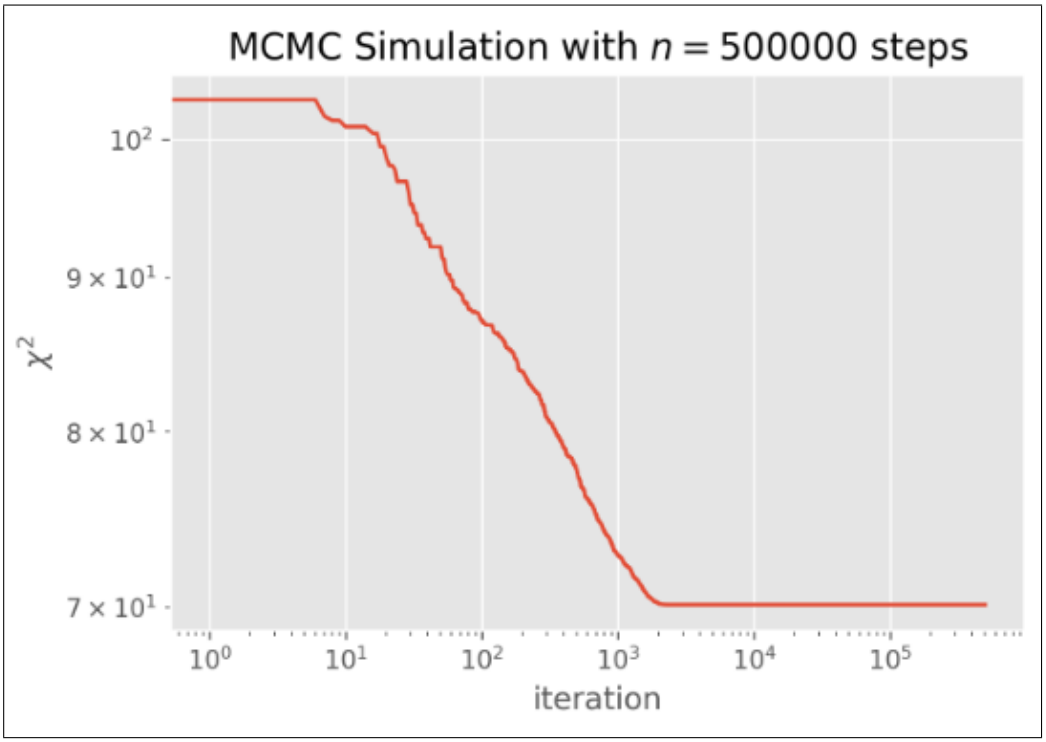

Figure 2.8: The graph depicts the evolution of χ^2 per iteration for an MCMC algorithm with poor performance. Although it seems to have converged to a relatively suitable level, the shape of the curve is actually a cause for concern as the MCMC algorithm is designed to explore parameter spaces and occasionally try out parameter sets that may lead to higher χ^2. In this case, the algorithm appears to have failed to do so, indicating that it may be stuck in a local minimum or may not be exploring the parameter space effectively.

In contrast to Figure 2.7, we can have more confidence in the convergence of χ^2 in Figure 2.8. However, this does not necessarily indicate a good MCMC model. The MCMC is desirable for its adaptability, versatility, and ability to escape local minima in search of the absolute minimum using the piece-wise function (2.13), specifically through the probability section. The idea is that at times, we may discard lower χ^2 parameter sets to avoid becoming trapped in local minima and better explore the parameter space. Therefore, we anticipate seeing small but noticeable changes

in χ^2 as the MCMC progresses. In Figure 2.8, we see no increase in χ^2, indicating that the acceptance parameter may be too high - leading to the secondary function having an extremely low probability for accepting higher χ^2 parameter sets - ending with the MCMC not adequately exploring the parameter space. While the model shows signs of convergence, we cannot say for sure whether it has converged at a local or absolute minimum of χ^2.

One final example, of what a good χ^2 evolution looks like:

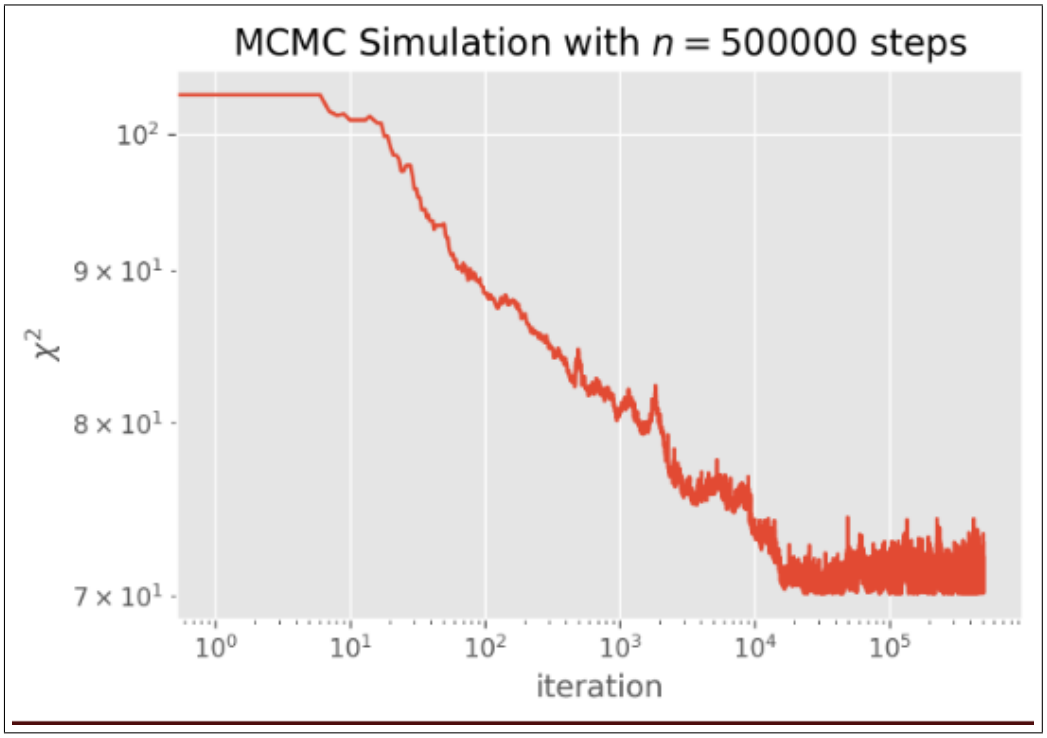

Figure 2.9: Evolution of χ^2 per iteration for an MCMC algorithm with good performance. The graph displays a steady decrease in χ^2 per iteration with subtle increases, indicating the algorithm is exploring parameter spaces in a systematic way, leading to a convergence on a suitable set of fitting parameters.

This is a favorable development of χ^2, unlike what is observed in Figures 2.7 and 2.8. The fact that χ^2 did not decrease further after a considerable number of iterations indicates that it has converged. Furthermore, the increases and reversals in the values of χ^2 suggest that the MCMC method has explored the space around the minima.

3 Analysis

In this section, we will explore the analysis of our radial velocity data using various techniques. We will start by looking at posterior slicing and colour-maps of parameter values obtained from the Markov Chain Monte Carlo (MCMC) method. These visualizations will help us understand the distribution of parameter values and identify correlations between them.

We will then move on to finding the best-fit parameters for our model and use phase-folded diagrams to assess the accuracy of the fit. We will discuss how to identify the best-fit parameters and what to look for in phase-folded diagrams to ensure that our fit accurately represents the data.

From there, we will explore methods of graphically visualizing the orbits given our best-fit parameters. We will use these visualizations to gain a better understanding of the dynamics of the system and how the different parameters affect the shape and orientation of the orbit.

Finally, we consider the possibility of signals residing within the residuals of our model and data. We use our introduced analysis techniques on the residual data to learn more about the systems.

Overall, this section will provide an in-depth analysis of our radial velocity data, using a combination of statistical techniques and visualizations to understand the underlying physical processes.

3.1 Posterior Slicing

The MCMC method is a powerful tool for obtaining precise fitting parameters for our model. However, not all data collected during the process is useful to us.

The burn-in period, which for our 500,000 iterations, lasts from the beginning of the chain until around the $10^4 - 10^5$ mark. It represents the method still trying to converge on a parameter set and should be discarded. After this period, the MCMC method reaches a stationary distribution, indicating that χ^2 has reached its minimum value(s). At this point, the method tries to accept higher χ^2 parameter sets to escape its minima and explore other parameter spaces. However, it eventually realizes that the current area is its absolute minima and returns to it. The posterior distribution obtained from the stationary area contains all the useful data required for analysis. To extract the relevant data, we slice the arrays representing each parameter's steps at the 10^5 mark, where the method is safely in its minima zone. With the useful data at hand, we can now visualize the distributions and explore potential relationships between parameters to gain a deeper understanding of the data.

3.2 Distribution and Colour-maps

Histograms are a useful tool for visualizing the spread of MCMC-generated data for each model parameter. Using Matplotlib's histogram function and the posterior set of each parameter list, we can visualize the distribution of favored parameter values and the range of values explored by the MCMC algorithm. The Gaussian distribution is expected in our model because the MCMC method uses conditional probability to determine which parameter values to accept or reject based on the current χ^2 value. As the probability of accepting a parameter set leading to a lower χ^2 value is higher than accepting a set leading to a higher χ^2, this bias towards lower χ^2 values results in a more probable distribution of values around the minimum χ^2 value, leading to Gaussian-shaped histograms:

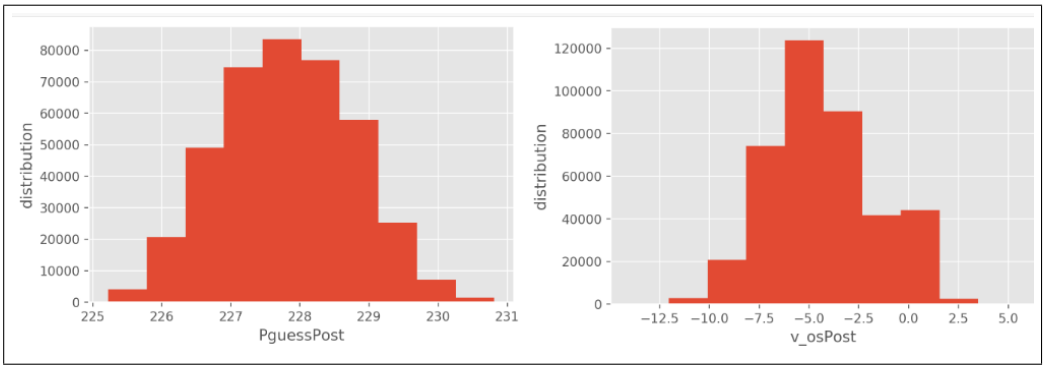

Figure 3.1: This graph displays the posterior distribution obtained from the MCMC algorithm for periodicity and drift velocity, which samples from the probability distribution of the parameters given the data. The distribution appears to be normally distributed, indicating that the algorithm has converged to a good set of fitting parameters

Color-maps are a commonly used visualization tool in data analysis. They provide a way to represent data in the form of colors that can help to identify patterns and trends. Each color in the map corresponds to a range of values for a particular variable. By plotting a color-map of our data, we can quickly identify areas where certain values are more frequent, and see how these areas are distributed across the data set.

Color-maps can be especially useful for identifying correlations between variables. By plotting two variables on the x and y axis and using color to represent a third variable, we can identify areas where there are strong correlations between the variables. These areas will be characterized by similar colors, indicating that the value of the third variable is consistent across a range of values for the other two variables.

Color-maps can also be used to visualize the step-wise evolution of parameters during an MCMC algorithm. By using color to represent the value of the χ^2, we can identify regions in which sets of parameter combinations yield smaller χ^2 values. We can also observe how the color changes as we move further away from this region, allowing us to identify potential correlations between parameters.

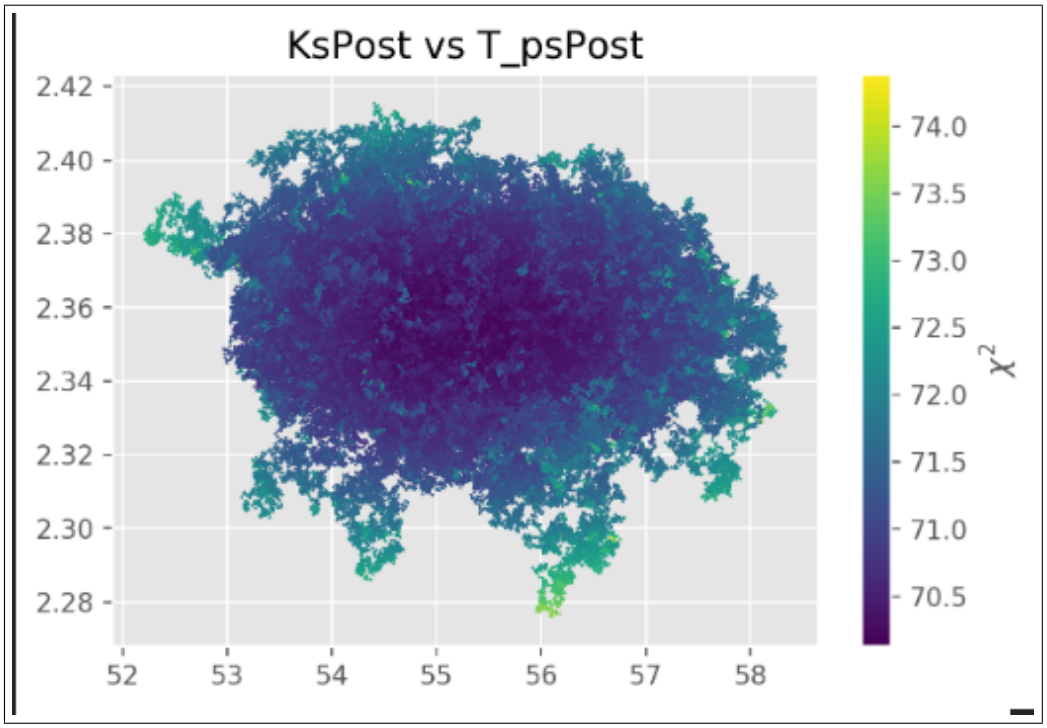

Figure 3.2: This color map shows the correlation between the velocity semi-amplitude and the time of periastron passage, with the color representing the value of χ^2. Darker colors indicate a lower χ^2, indicating a better fit. We can observe a clear correlation between these two parameters, with the best-fit values located in the center of the color map. The shape of the correlation is approximately elliptical, and the range of values for the parameters appears to be well-explored by the MCMC method.

The color-map depicted in figure 3.2 is a typical representation of what we expect to see. Color-maps allow us to visualize parameter combinations resulting in lower χ^2 values. The center of the parameter value range in the color-map represents the lowest χ^2 values. This is an expected outcome because we anticipate that the MCMC method has explored thoroughly the space around the parameter combinations with the lowest χ^2 values. The MCMC method should have returned to the lowest combination after exploring different combinations of higher and lower parameter values to determine if χ^2 could be further decreased.

It is important to ensure that the slicing is done correctly to obtain accurate diagrams. An example of an inaccurate diagram due to a poor cutoff point of 10^3 would appear like:

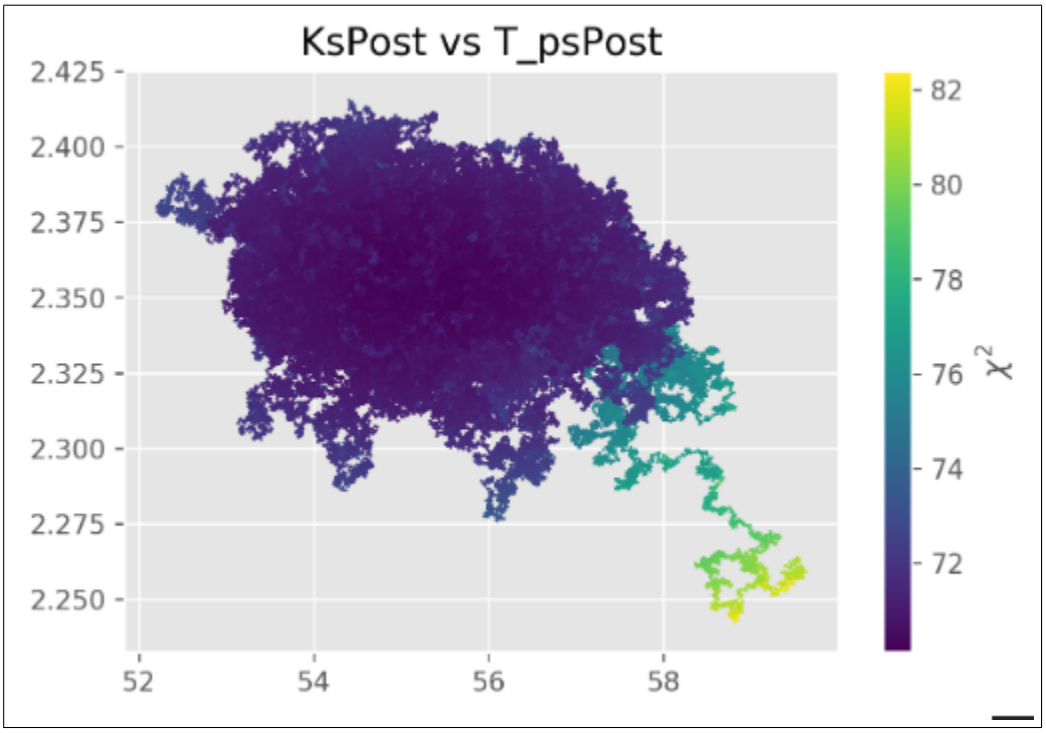

Figure 3.3: The color map displayed demonstrates the correlation between velocity semi-amplitude and time of periastron passage, where the color represents the corresponding chi-squared value. However, in this version of the color map, we include the burn-in period from the MCMC method, which causes a significant loss of extractable information due to the decrease in resolution caused by having to zoom out of the low chi-squared area in order to include the burn-in within the plot. This emphasizes the importance of properly selecting the portion of data to include in the color map to ensure that the information obtained is of high quality and resolution.

By zooming out the graph to plot the burn-in portion of the data, we lose a lot of detail in the correlation between the two parameters.

3.3 Best Fits

We can create a more precise model that accurately represents the observational data by utilizing the best-fit parameters obtained from the MCMC method. To assess the quality of the model, we generate best-fit curves for the phase-folded data. To estimate the accuracy of the model, we plot the upper and lower quartiles, allowing us to observe the precision of the model's fit. By comparing this plot to the previous phase-folded diagram, we can visualize the improved accuracy of the new model in recreating the orbit.

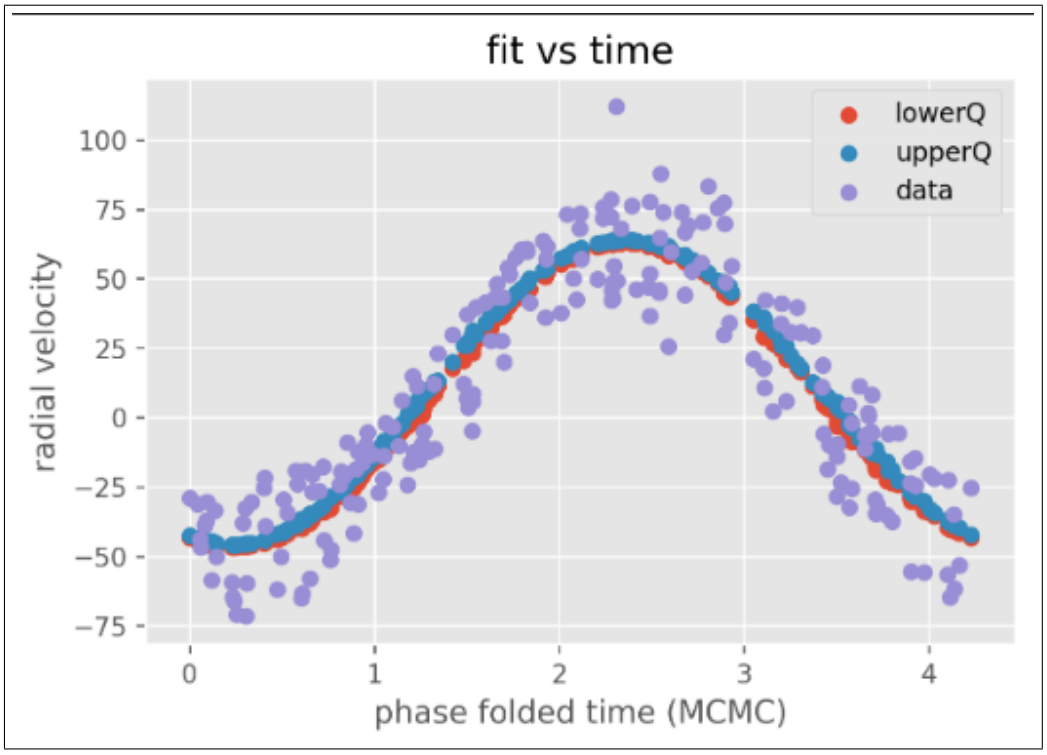

Figure 3.4: This phase folded diagram displays the models fit for the phase versus the radial velocity in ms^{-1}. The strong similarity between the models curve and the data trend indicates the model has accurately predicted the behaviour of the system.

Our analysis has produced an excellent result. The original phase folded diagram, Figure 2.5, has been significantly improved through the refinement of parameters using our methods of analysis. The model now closely represents the observational data, indicating the effectiveness of our approach.

3.4 Graphing the orbit

After obtaining a reliable model, the next step is to visualize the orbit using the eccentricity, semi-major axis, and true anomaly [8]. The equation we use to begin this process is:

$$r = \frac{a(1-e^2)}{1+ecos(f)} \tag{3.1}$$

This gives us the distance of the exoplanet from the orbits centre, it's "Radial Distance". Using this, we can find coordinates of the exoplanet:

$$x = rcos(f) \tag{3.2}$$

$$y = rsin(f) \tag{3.3}$$

We can obtain the x and y coordinates of our orbit by using the equation. This method is applicable even for non-eccentric orbits since the vanishing of e from 3.1 does not cause any mathematical errors. If e was the sole variable in the numerator, it would cause an error due to division by 0. By applying this method to one of our data sets, we were able to generate a visual representation of the orbit:

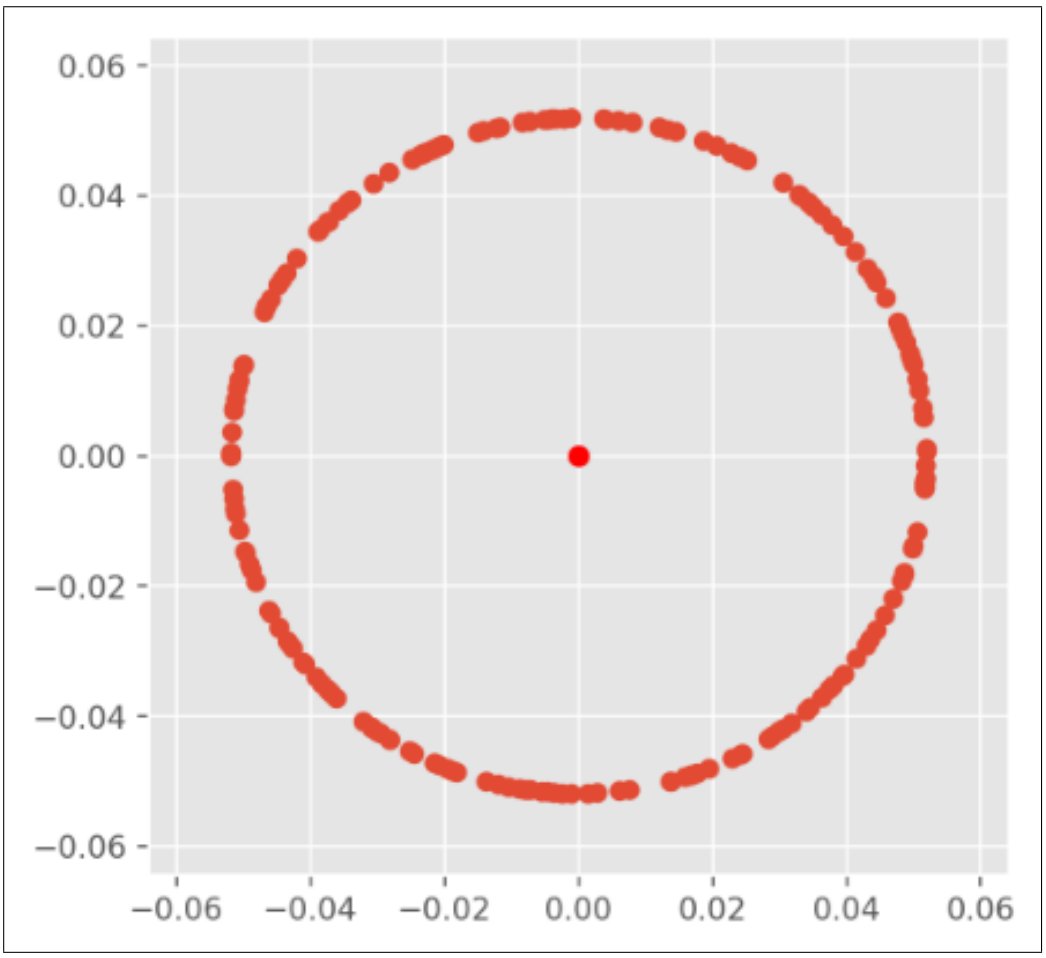

Figure 3.5: The graph displays the orbit of system-00 visualized through the use of the orbital equations. The x and y coordinates represent the position of the star as it orbits around the center of mass of the system in terms of astronomical units. This graph allows us to visualize the shape of the orbit to gain a better understanding of the system's dynamics.

In the same way, it is also possible to visualise eccentric orbits.

3.5 Residual Analysis

After obtaining a best-fit model for a system, we can analyze the residuals by subtracting the model's predictions from the observed radial velocity data. The resulting residuals represent the difference between the model and the observed data, and they contain valuable information about any remaining systematic errors or discrepancies in the model. One possible approach for analysing the residuals is to treat them as their own system and perform our introduced analysis techniques such as Lomb-Scargle periodograms and MCMC fitting. By doing so, we can investigate whether any periodicities or correlations exist in the residuals that were not captured by the original model.

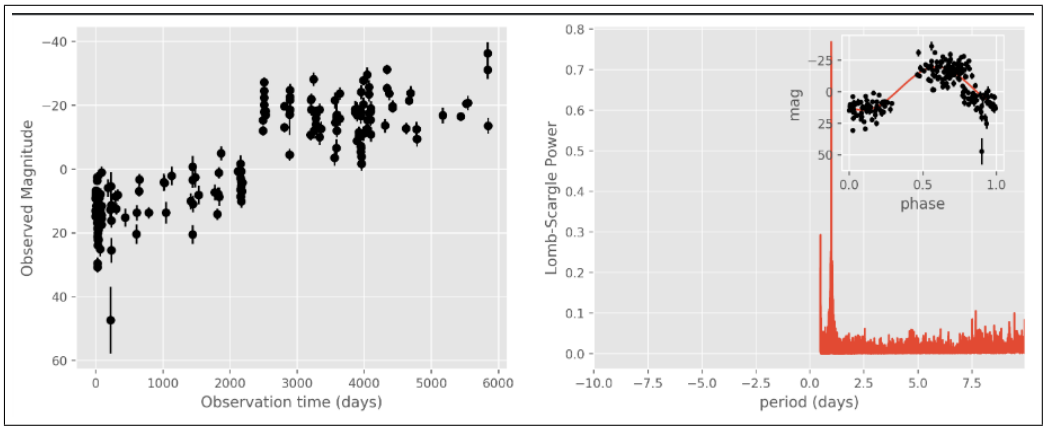

Figure 3.6: System-00 plots showing the residuals against radial velocity in m/s on the left and the corresponding power spectrum on the right with a phase-folded inset. The power spectrum reveals a potentially significant periodicity at a high frequency, which is confirmed by the phase-folded inset.

When working with high frequency data, it is important to ensure that the signals within the power spectrum from the residuals are not false positives. This can happen when extremely high frequencies are allowed as potential periodicities, where an entire period could be fit between each recording. To prevent this, we can check if a given frequency is a significant periodicity in the residuals using a Lomb-Scargle periodogram. This helps us confirm whether the observed signal is a genuine periodicity or simply a result of noise or other factors. By setting a threshold for peak significance (relative to a chosen percentile of the power values in the entire periodogram), we can ensure that only truly significant periodicities are identified, allowing us to confidently analyze the data.

Once a true periodic signal has been detected within the residuals, it can be assumed that there is another body in the system that the previous model failed

to detect. By treating the residuals as a new system, we can apply the analysis techniques discussed in Section 2 to find a new set of parameters that describe the potentially hidden secondary body. Since the phase folded diagram does not clearly indicate the eccentricity of the secondary body, both circular and eccentric models are attempted. Although the eccentric model would be more comprehensive, the circular model is approximately 10-12 times faster due to less computational stress during the MCMC method and serves as a baseline. If the eccentric model yields an extremely eccentric orbit, the analysis will continue with the eccentric model. If the eccentric model yields a low eccentricity orbit, the orbit will be assumed to be circular and the analysis will continue with the circular model. The selection of the model to continue with will likely be based on both the quality of fit and χ^2. To begin with, a brute force method is applied to arrays of potential parameter values to find the best sets for both the circular and eccentric model, resulting in varying phase folded diagrams:

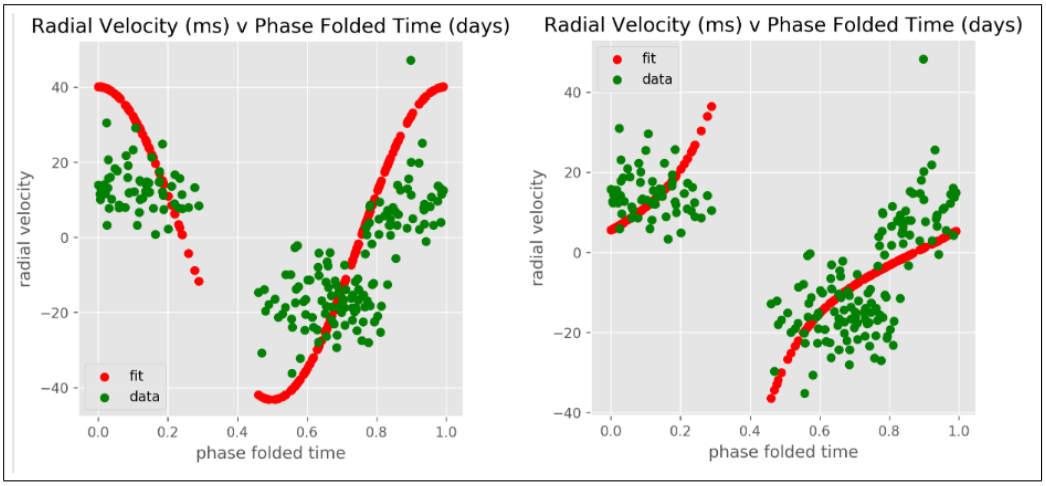

Figure 3.7: System-00s residual phase folded plots showing a circular model of the data on the left, and an eccentric model against the data on the right. The graphs themselves are the phase folded time in days against the radial velocity in m/s.

The fact that the eccentric fit resembles a tan function indicates that the eccentricity of the system is relatively high. This is because a highly eccentric orbit causes the object to move much faster when it is closer to the primary object, and much slower when it is further away. This results in a curved, non-linear shape that can resemble a tan function when phase folded.

Continuing on from here, we may have a semi accurate set of initial fitting parameters for the respective models. With these and the right MCMC parameters, we can use the MCMC algorithm to lower χ^2 in the hopes of refining our systems

parameter values in order to determine whether there are more bodies in system.

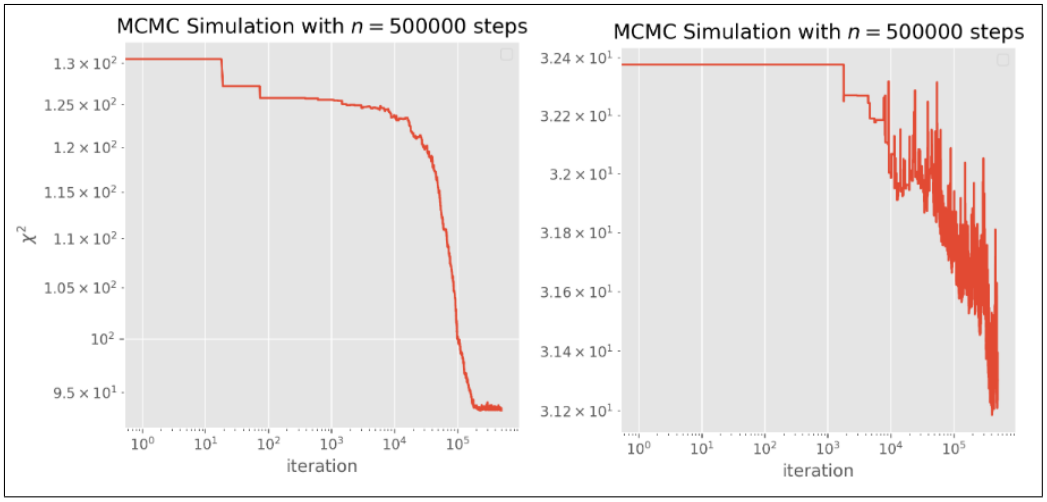

Figure 3.8: System-00 plots showing the residual χ^2 evolution during the MCMC for a circular (left) and eccentric (right) model

The evolution for χ^2 with an eccentric model has both started and converged on values that are much lower then the circular model. From here, we can assume that the secondary body within the orbit is more likely to be on an eccentric orbit as opposed to a circular orbit. Finishing the analysis using both models to observe the phase folded diagrams with the refined parameter sets we see:

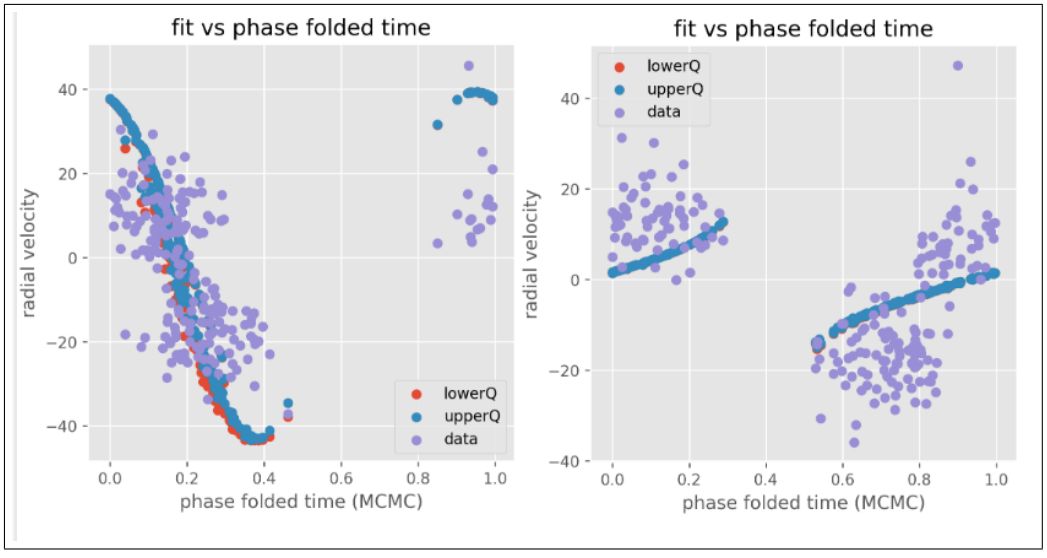

Figure 3.9: System-00 plots showing the residual phase folded circular (left) and eccentric (right) models against the data. time in days vs radial velocity in m/s

The nature of both phase folded graphs put the existence of the secondary body into question. After analysis, we still don't have enough information and cannot

confirm or deny whether more bodies within system-00 exist as the models are not representative enough of an orbiting body.

4 Limit Testing and Experimenting

In this chapter, we will explore the limitations of the analysis methods we have employed. By pushing the methods to their limits, we can gain insight into how they behave in different scenarios and understand their strengths and weaknesses. This experimentation will allow us to identify situations where one method may be more suitable than another, and provide a deeper understanding of the methods and their alternatives. Additionally, we will investigate the effects of surpassing the limits to gain a better understanding of the consequences of such actions.

4.1 Observational Period vs Lomb scargle

We employed two primary methods, namely observational analysis and the Lomb-Scargle periodogram, to obtain initial estimates for our model parameters with some level of accuracy. However, each method has its own strengths, weaknesses, and limitations.

To assess the effectiveness of the Lomb-Scargle periodogram in improving our analysis, we repeated our analysis of the radial velocity data of a star twice, while varying only the period supplied to the MCMC method. Furthermore, we may adjust the range cutter for the period as required because using a range cutter value that is optimal for one method on an initial parameter set supplied by a different method may not reflect the MCMC method's ability to converge on the lowest χ^2 parameter set.

It is worth noting that while we vary the methodology employed to determine the period, the primary focus of our stress test is on the MCMC method's ability to converge on the correct period from different numerical ranges.

To begin the comparison of the performance of the MCMC method with respect to the different period obtaining methods, we will use data sets that have short intervals between recordings of the radial velocity of the system. This ensures well defined periodic signals for observational analysis, allowing for a fair comparison. Our analysis is focused on system-00, which was introduced earlier (see 2.1). We investigate how starting values for periodicity from observational analysis affect the final result obtained by the MCMC method. We use the evolution of the period per iteration to determine good values for the range cutter and β. We ensure that the MCMC method varies the periodicity both above and below the converged value, accepts values outside of the convergence value, and pushes further out, all of which eventually result in returning to the converged-upon value.

4.1.1 Observational Analysis

Using observational analysis, we are given a periodicity of: 3.75 days (2.1.1). With this value, we can check our phase folded diagram to asses the precision of the period as well as the rest of the fitting parameters:

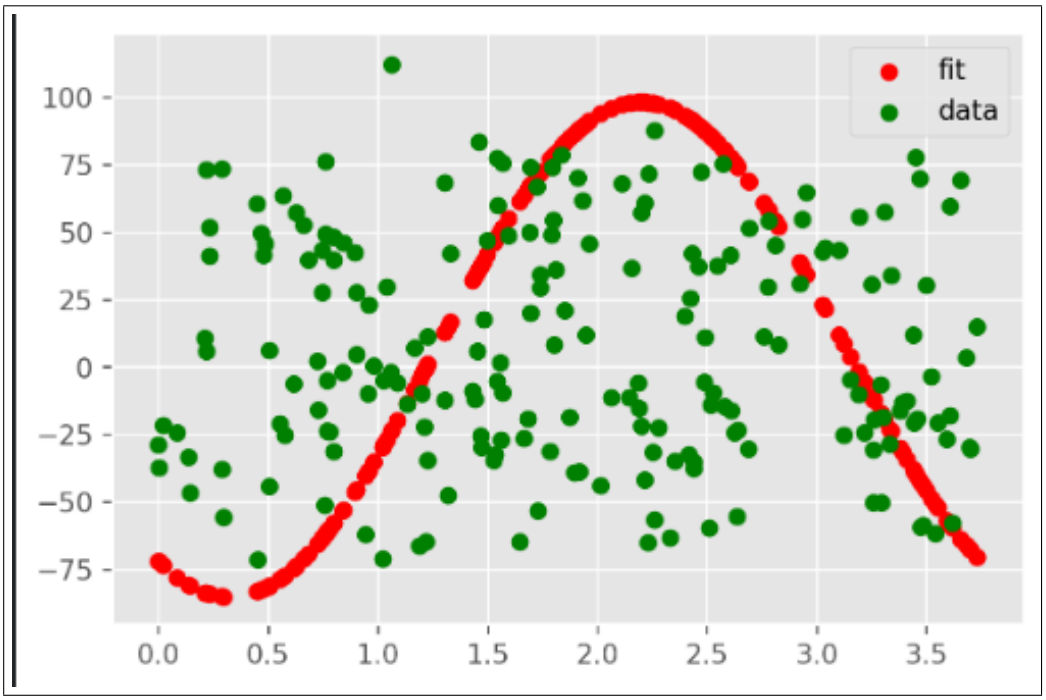

Figure 4.1: Phase folded time (days) vs radial velocity (ms) The phase folded diagram shown here was obtained using a period obtained via observational analysis. The resulting period did not provide a good fit to the data and thus the model appears scattered and does not align well with the recorded values.

The phase folded diagram indicates that the model's prediction has scattered the

recordings, which can be attributed to an inaccurate period. The key challenge now is to see if the MCMC method can utilize this inaccurate period to converge on a value that better represents the system. We can evaluate the performance of the MCMC algorithm by examining the evolution of χ^2 versus each iteration. In addition, since we are also interested in how the algorithm handles the period, we can plot the evolution of periodicity values for each iteration:

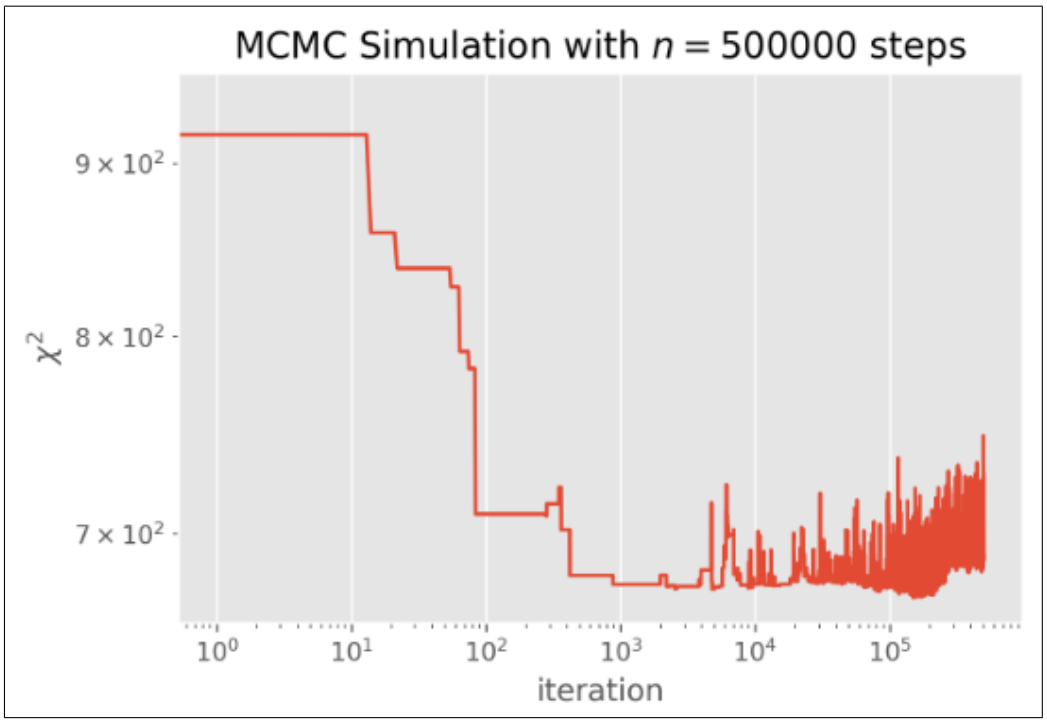

Figure 4.2: The chi squared evolution per iteration graph shows poor performance of the MCMC method. The chi squared values do not seem to converge to a suitable level. This indicates that the MCMC method is not efficiently exploring the parameter space and potentially selecting random parameter sets.

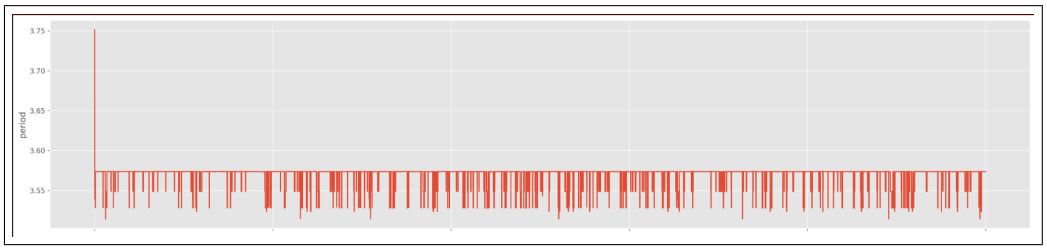

Figure 4.3: The bar code-like appearance indicates that the chains are not exploring the parameter space thoroughly, resulting in poor performance. The lack of diversity in the periodicity values across the chains suggests that the algorithm is getting stuck in local minima and not exploring other regions of the parameter space. This is problematic for MCMC as the goal is to explore the entire parameter space to find the global minimum, rather than being stuck in local ones.

It is immediately apparent that the parameters for the MCMC algorithm have not been properly fine-tuned. However, in this case, have fine-tuned parameters will not be enough to solve the issue since the initial periodicity is too far away from the true value. Due to the high dimensionality of the search space, there are many local minima, and the fine-tuning parameters are used to help escape them and converge on the nearby absolute minima. However, in this case, the absolute minima was too far away in terms of the parameter values required to reach it. As a result, the MCMC got trapped in a deeper local minima, which is a common occurrence due to the large number of potential parameter combinations. With each combination, the MCMC introduces false minima that happen to have lower χ^2 values than their neighboring combinations. This underscores the importance of having a strong set of initial parameters.

Therefore, having a good set of initial parameters becomes even more crucial when dealing with a higher-dimensional search space. A good starting set of parameters can help the algorithm to escape from false minima and sink into the nearby absolute minima. By exploring a wider range of the parameter space, the algorithm has a better chance of finding the global minimum and producing accurate results.

4.1.2 Lomb-Scargle

Using the Lomb-scargle periodogram on this data set, are given a periodicity of 4.23059 days. Alongside this value, we also see a periodogram of the power spectra aswell as a phase folded fit of the data against just the period:

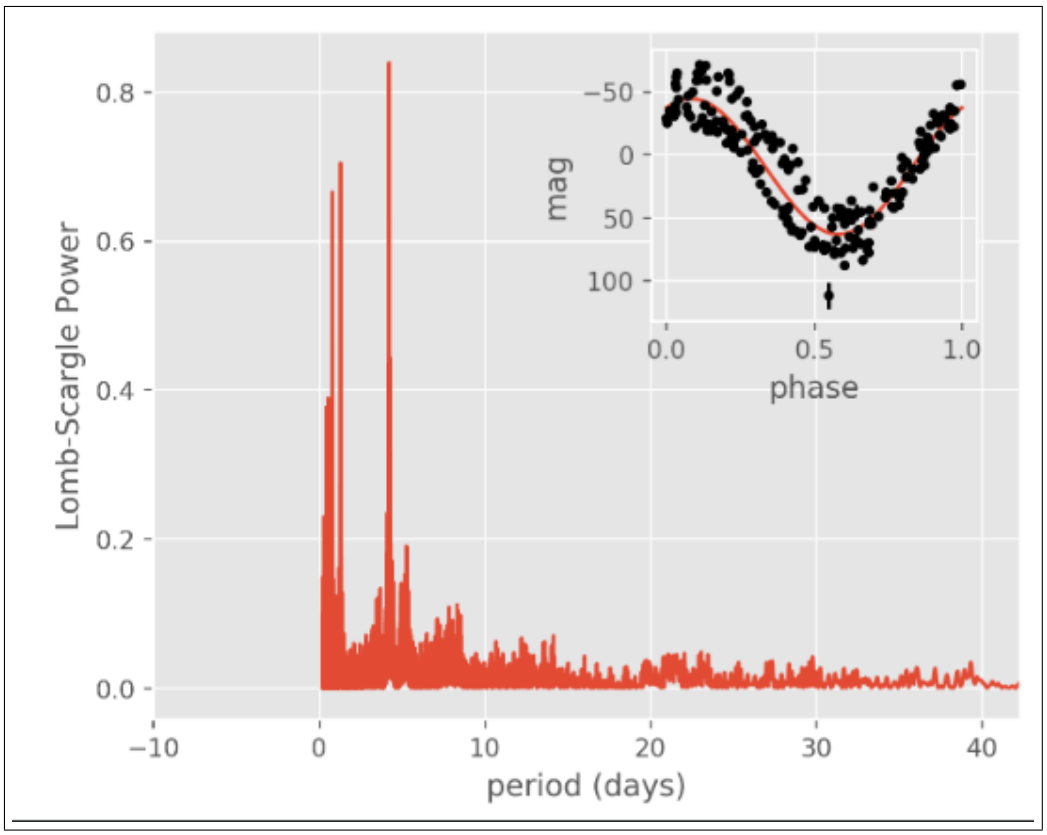

Figure 4.4: The Lomb-Scargle periodogram displays the power spectral density of the signal as a function of frequency. The peak in the periodogram indicates the most likely period of the signal. The inset phase folded diagram shows the signal folded at the most likely period

There is a prominent signal present at the 4 day period mark in the radial velocity data. Utilizing this information, along with the other fitting parameters, and assuming a circular model, we are able to produce a phase folded diagram of the data against a full model of the radial velocity. As shown in Figure 2.5, this initial set of fitting parameters appears to be very promising. We can use these values as a starting point for the MCMC method and evaluate how well the algorithm performs by plotting the change in χ^2 per iteration, as well as the evolution of the periodicity per iteration. These plots are shown in Figure 2.9 and:

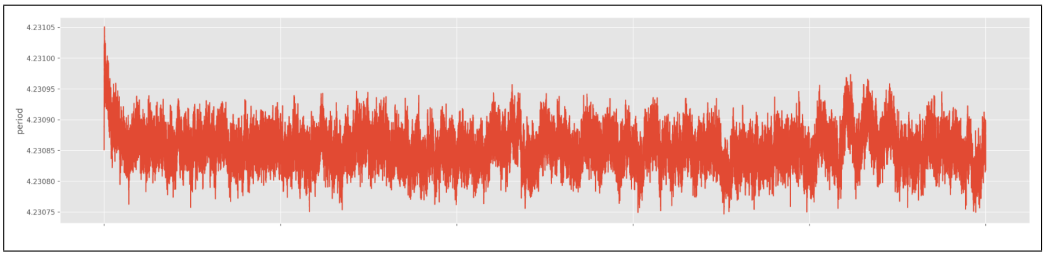

Figure 4.5: This graph shows a good evolution of the periodicity per step for the MCMC algorithm. The plot displays a consistent and stable periodicity over the posterior slice. This indicates that the MCMC algorithm has successfully converged to a good solution, and that the obtained periodicity is likely to be accurate.

By analyzing the output of both the MCMC and the Lomb-Scargle periodogram, we can conclude that they agree on the true period of the system being around the 4.230 range. This is supported by the fact that both methods produce similar periodicity values and a relatively low χ^2. Therefore, we can have a high degree of confidence that this value accurately represents our system.

4.1.3 Comparison

One of the main challenges in studying stars is determining their period accurately. There are several methods that astronomers use to determine the period, including observational analysis and the Lomb-Scargle periodogram. In this section, we will compare the results obtained from these two methods by using the MCMC algorithm to refine the periodicity estimate. We will present a table of the periodicities obtained from the MCMC algorithm for multiple systems, where the starting value is given by either observational analysis or the Lomb-Scargle periodogram. By comparing these values alongside their respective χ^2's., we can gain insights into the strengths and weaknesses of each method and their respective standings.

System	Pre MCMC (days)	Post MCMC (days)	Pre χ^2	Post χ^2	Quality
00-(LS)	4.23059	4.23090	428.32484	71.81191	Good
00-(OA)	3.75	3.57363	1597.97229	702.31707	Bad
01-(LS)	228.45286	228.45247	5.24232	3.03575	Good
01-(OA)	225	227.27515	7.65343	3.00260	Good
02-(LS)	634.20631	637.75742	173.86121	122.49011	Good
02-(OA)	500	496.94216	1270.46819	636.90074	Bad
03-(LS)	405.65301	406.42249	173.86121	25.90478	Good
03-(OA)	380	366.35086	1282.93496	331.56988	Bad
04-(LS)	405.65301	406.42249	173.86121	25.90478	Good
04-(OA)	380	366.35086	1282.93496	331.56988	Bad
05-(LS)	23.04934	10.44454	458.21129	458.06283	Good
05-(OA)	91.13450	36.91966	500	500.22142	Good
06-(LS)	17.05293	10.44454	17.05155	11.94776	Good
06-(OA)	82.38812	80.16846	100	99.55944	bad

System	semi-major axis (AU)	$M_p \sin(I)$ (Kg)
00-(LS)	0.05187	6.93528e+25
00-(OA)	0.04785	7.994562+25
01-(LS)	0.77031	6.60139e+27
01-(OA)	0.76269	6.15534e+27
02-(LS)	1.65703	4.17955e+27
02-(OA)	1.41343	2.00875e+27
03-(LS)	1.22013	6.64930e+25
03-(OA)	1.16401	1.96836e+25
04-(LS)	1.22013	6.64930e+25
04-(OA)	1.16401	1.96836e+25
05-(LS)	1.25298	1.840407+25
05-(OA)	1.16401	1.96836e+25
06-(LS)	0.12188	3.60667+24
06-(OA)	0.39511	2.18294e+24

Here, we are recording (for each system with varying methods: Lomb-Scargle and Observational Analysis) the periodicity of each system before and after MCMC, χ^2 before and after MCMC, the quality of the phase folded fits, the semi major axis of

the system in AU and the proxy mass of the exoplanet in Kg using:

$$G(M_* + M_p) = a^3 \left(\frac{2\pi}{P}\right)^2 \tag{4.1}$$

and

$$M_p \sin I = K(\frac{2\pi G}{P})^{-1/3}\sqrt{1-e^2}(M_\odot \frac{M_*}{M_\odot})^{2/3} \tag{4.2}$$

We use a proxy mass because the inclination of the orbit is often unknown, meaning the closest value we have that can represent the mass is $M_p \sin(I)$.

The table shows the results of applying the MCMC algorithm with different initial periodicities obtained from Lomb-Scargle periodogram (LS) and observational analysis (OA). The Pre MCMC column represents the initial periodicity, while Post MCMC column shows the periodicity obtained after running the MCMC algorithm. We can see a significant difference in the final periodicity as well as the χ^2 values for the two methods. The Pre χ^2 column shows the initial χ^2 value, and the Post χ^2 column shows the χ^2 value after the MCMC algorithm. The quality of the phase folded diagram after the MCMC algorithm is also shown, with "Good" referring to a well-folded diagram and "Bad" referring to a poorly-folded diagram. That allows us to represent whether the analysis settings used provided us with a model that represents the data.

As anticipated, we have confirmed a significant disparity in the final periodicity and χ^2 when using the Lomb-Scargle periodogram versus observational analysis. In order to better comprehend the impact of these methods, we can plot our post analysis phase folded diagrams side by side, which will enable us to better visualize and understand the magnitude of the differences.

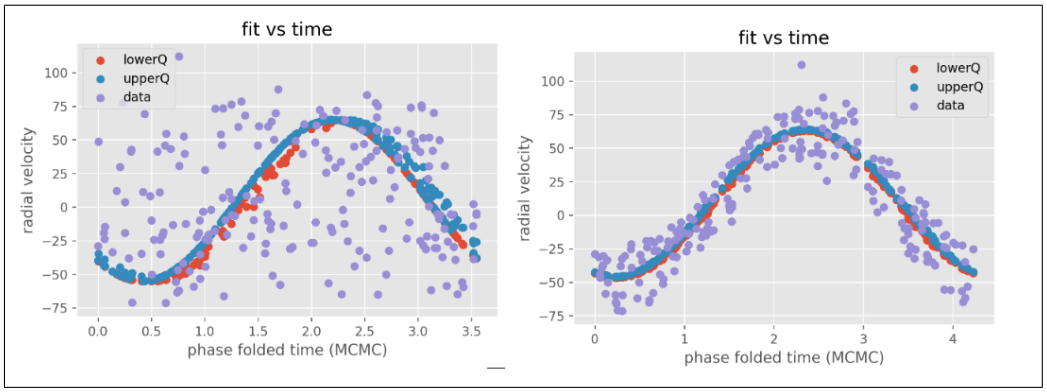

Figure 4.6: Comparison of phase folded diagrams for System-00, showing the impact of initial period value on model performance. The left diagram, with initial period value given by observational analysis, shows poor fit to the data. The right diagram, with initial period value given by Lomb-Scargle analysis, shows a much improved fit after MCMC optimization. The red and blue lines represents the best-fit model after MCMC.

Based on the table, we can observe that the value of χ^2 before the MCMC algorithm plays a crucial role in determining the accuracy of the resulting parameters. The difference in the periodicities obtained using different methods is relatively small, but the variation in the value of χ^2 is significant. When the pre-MCMC χ^2 value is small enough, the MCMC algorithm can converge on an accurate set of parameters, irrespective of the method used for obtaining the period. However, in most cases, the Lomb-Scargle method provides a more accurate value of the system's periodicity, resulting in a smaller pre-MCMC χ^2 value. For systems with evenly recorded data or a high recording frequency, the choice of method is irrelevant as the periodicity value obtained would meet the MCMC convergence χ^2 threshold (an example being system-01). Although the threshold value is not clear, we can estimate it to be within the range of 400-500. Meeting this requirement would help the MCMC reduce χ^2 to around 150-50, resulting in obtaining an accurate set of input parameters for modeling the system. Failure to meet this threshold could result in the MCMC converging on a local minima.

4.2 Eccentric Model vs Circular System

To investigate the impact of the eccentricity on our analysis, we have made a new assumption on one of our previously introduced data sets. The objective is to observe how the output of our methods change when we consider an eccentric orbit. We aim to see if the MCMC algorithm can converge to an eccentricity close to 0 and realize that it is working with a non-eccentric orbit. To achieve this, we experimented with a small test set of values for the initial parameters with an inclination angle and

eccentricity in mind (as described in section 2.2.6). Using these new parameters, we generated a phase folded diagram of our new model:

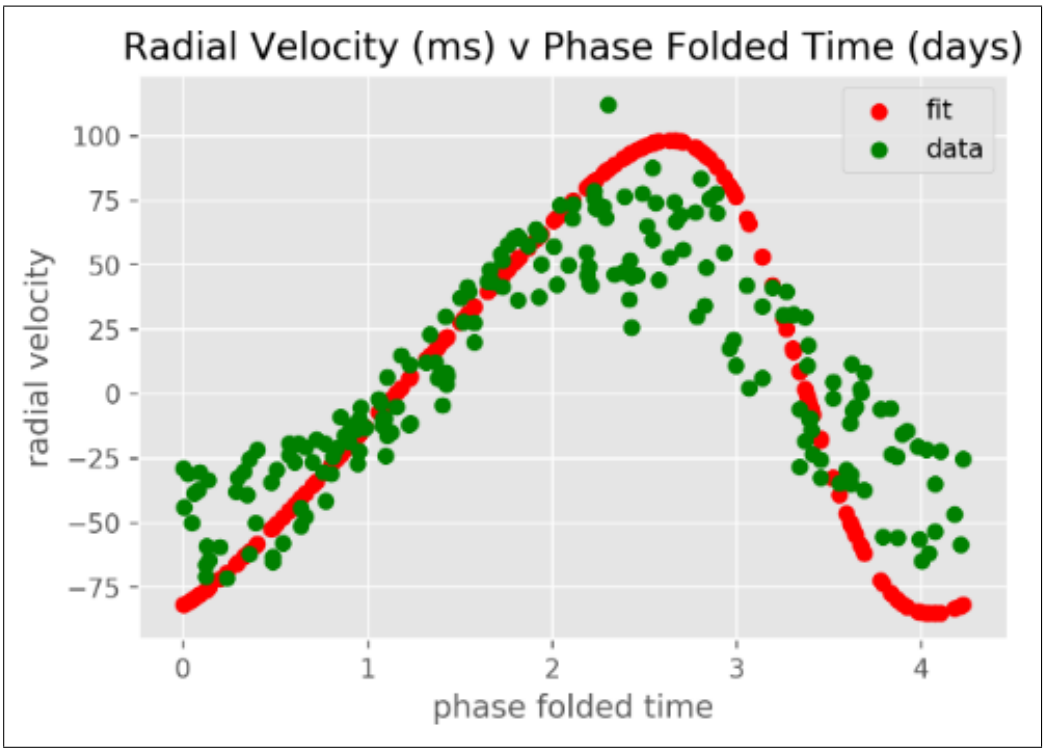

Figure 4.7: This phase folded diagram displays the radial velocity data of system-00, previously thought to be circular, but now fit with an eccentric orbit.

After refining the initial fitting parameters through brute force, the algorithm has determined that a non-negligible eccentricity of 0.495 provides the best fit for the model. With this new set of parameters, which already closely mirrors the system, we can re-run the MCMC method and analyze how it affects the fitting parameters and their respective values. By using the exact same MCMC parameters that led to the previous best fit for the data-set, we can observe the evolution of eccentricity during the MCMC process:

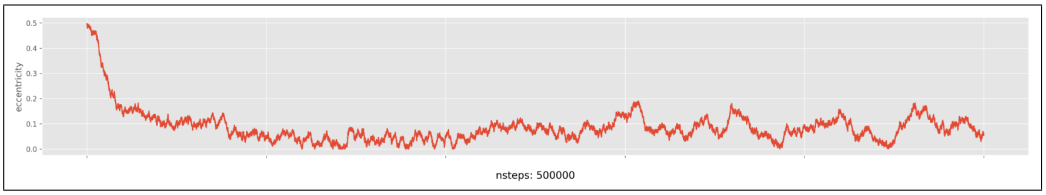

Figure 4.8: This plot shows the evolution of eccentricity per iteration of the MCMC algorithm. The model began with a relatively high eccentricity (0.495-0.5) but MCMC quickly stepped down to near 0 eccentricities.

The model clearly favours eccentricities between 0 and 0.1, using a histogram, we

can inspect this distribution further:

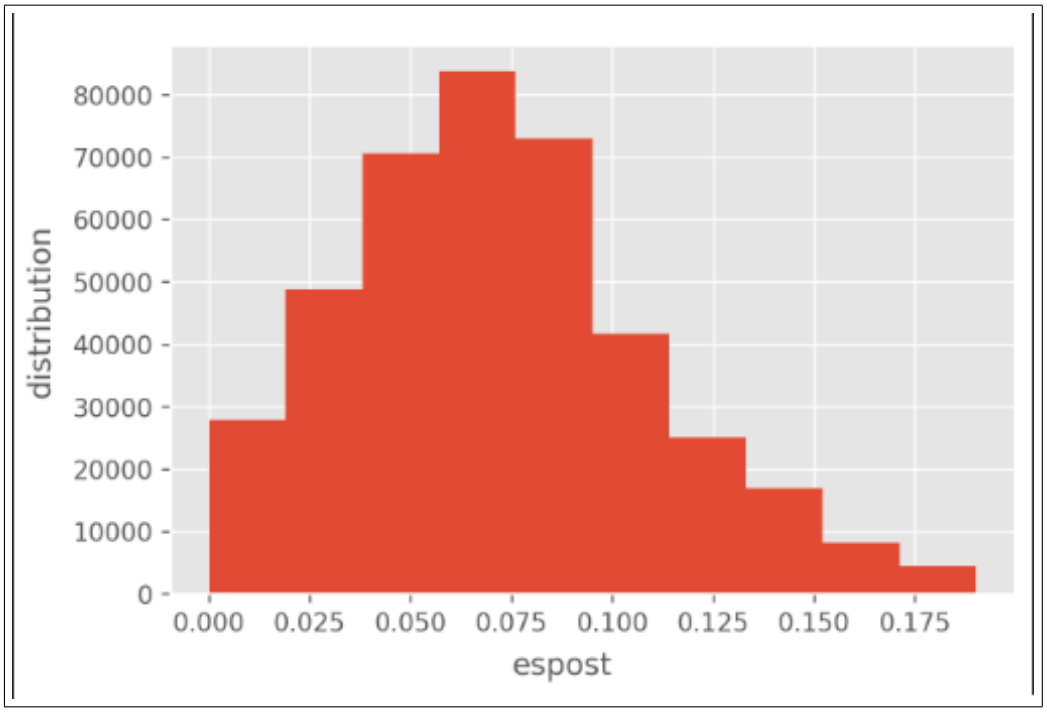

Figure 4.9: This histogram shows the distribution of eccentricity values obtained from the MCMC algorithm.

We observe that during the MCMC process, the majority of the eccentricity values held during the process were between 0.05 to 0.08. There are a couple of reasons that could explain this. Firstly, it's possible that the orbit actually has a non-zero eccentricity that was hard to detect using a circular model because it's already very close to being circular. This small difference may not be easily noticeable in the phase folded diagrams. Secondly, the MCMC algorithm may not have been able to stay at 0 due to its random exploration of the parameter space. As it tried to explore the space around 0, because it could never go lower, it was forced to increase, leading to an increase in the weighting on the space just above 0 in the histogram. One approach to deal with this is to vary the scale of the eccentricity step size. Decreasing the step-size could lead to the "best" eccentricity displaying as much lower, indicating that the true eccentricity is likely much closer to 0. Another way to compare the circular and eccentric models is to calculate the χ^2. A small but non-negligible difference between the two models could suggest that a small non-zero eccentricity better represents the system than a 0 eccentricity. Finally, the MCMC algorithm could be permitted to explore χ^2 for eccentricities below 0. Assuming a normal behaviour of χ^2, a normal distribution would be present.

5 Discussion

One of the key findings within the analysis is the importance of residual analysis. By treating the residuals as a new system and performing analysis on them, we've created the chance to detect additional planets that had not detected by the initial model. This highlights the potential of using residual analysis as a complementary tool for exoplanet detection. Furthermore, experimentation with: data manipulation techniques, improved refining techniques, or more suitable MCMC parameters could have resulted in a more conclusive result on whether more bodies exist within system-00.

Our investigation also highlights the significance of selecting appropriate MCMC parameters. Our results indicate that even a single parameter that is not appropriately set can completely derail the algorithm. Additionally, we observed that MCMC parameters can differ significantly even when performing similar tasks or analyzing similar systems. Therefore, careful fine-tuning of these parameters is crucial to ensure an accurate and detailed analysis.

System-00's exoplanet closely resemble 51 Pegasi b [11], which is a well-studied exoplanet. Through a comparison of our findings with the recorded and accepted characteristics of 51 Pegasi b, we were able to identify that the exoplanet 51 Pegasi b has an extremely low eccentricity of 0.01, indicating that its orbit is close to circular. This is confirmed by our phase folded diagram of the circular model against the data, as shown in Figure 4.6 which tells us that the exoplanet within system-00 can be accurately modelled using a circular fitting equation. Our analysis also revealed that the predicted eccentricity using an eccentric model was in the range of 0.05 to 0.08, likely due to the step size. Further experimentation with a smaller step size may yield a range closer to the observed eccentricity. Additionally, our calculated semi-major axis of 0.05187AU is in agreement with the recorded and accepted value of 0.0527AU. Finally, our analysis found a periodicity of 4.23090 days, which closely matches the recorded value of 4.2 days for 51 Pegasi b. These findings demonstrate the ability of our analysis techniques to accurately identify and characterize exo-

planets, even in complex systems.

An alternative approach to adjusting the parameters for the MCMC algorithm is to change each parameter separately instead of adjusting all of them by a random float within the step-size restriction. By doing this, the χ^2 value would be calculated for each single parameter change instead of all parameters. This method may potentially slow down the convergence process due to calculating χ^2 each time a single parameter changes as opposed to a change in all of the parameters. With an eccentric model, this delay in convergence would have a drastic effect, due to the increase in complexity the true anomaly adds to calculating χ^2. However, it could have a stabilizing effect on the entire process by giving more control over which parameter accepts or declines the step. The piece-wise function would be able to accept or reject changes for each parameter separately, providing more control over staying within the correct parameter space/minima. This approach also has the potential to make the process less volatile compared to the current method of changing all parameters simultaneously. While the current approach has been addressed through the use of MCMC parameters and "Range Cutters", it is worth exploring this alternative approach.

It should be noted that while some of the calculated proxy masses for the exoplanets are consistent with their recorded values, many are not. However, given that the semi-major axis matches the recorded values, it is probable that the period does as well. As such, it is possible that the issue lies with the function responsible for computing these values, or the other parameters involved.

6 Conclusion

We have explored various methods for obtaining the period of an astronomical object, comparing observational analysis and Lomb-Scargle periodogram techniques. We have delved into fitting parameters for circular and eccentric orbits, establishing initial fitting parameters and refined them using Chi-Squared and Phase-Folded Diagrams. The application of MCMC allowed us to generate distribution and colourmaps by performing posterior slicing, as well as identify the best fits for the orbits and their characteristics, including analysis of residuals for multi-body systems. Finally, we conducted limit testing and experimenting, examining the differences between observational periods and Lomb-Scargle methods, as well as comparing eccentric and circular orbits.

We have demonstrated the importance of employing a combination of methods to accurately determine the period and model the orbit of astronomical objects. Through the comparison of different techniques, we have highlighted the strengths and limitations of each method, emphasizing the need for a holistic approach to understanding and predicting the behavior of celestial bodies. Additionally, this research has shed light on the significance of considering both eccentric and circular orbits when modeling orbits, as assumptions about the nature of an orbit can lead to discrepancies and inaccuracies.

In conclusion, the methodologies and findings presented in this dissertation contribute to a better understanding of the intricacies involved in modeling celestial orbits and offer a comprehensive approach to period determination and orbit fitting with the aid of computer programming.

Bibliography

[1] *Newton Raphson Method.* Brilliant.org, https://brilliant.org/wiki/newton-raphson-method/, April 2023.

[2] J. Brownlee. *A Gentle Introduction to Markov Chain Monte Carlo for Probability.* , https://machinelearningmastery.com/markov-chain-monte-carlo-for-probability/, Nov 2019.

[3] C.D.Murray and S.F.Dermott. *Solar System Dynamics.* , Sept 2006.

[4] A. D. Christophe Andrieu, Nando De Freitas and M. I. Jordan. *An Introduction to MCMC for Machine Learning.* Kluwer Academics, 2003.

[5] J. M. Danby and T. M. Burkardt. *The solution of Kepler's equation. I". Celestial Mechanics.* , 1983.

[6] A. developers. *Lomb-Scargle periodogram.* , https://docs.astropy.org/en/stable/timeseries/lombscargle.html, April 2013.

[7] N. R. Lomb. *Least-Squares Frequency Analysis of Unequally Spaced Data* . Astrophysics and Space Science, Volume 39, Issue 2, https://ui.adsabs.harvard.edu/abs/1976Ap%26SS..39..447L/abstract, February 1976.

[8] C. Lovis and D. A. Fischer. *Radial Velocity.* , .

[9] M. Mayor and D. Queloz. *A Jupiter-mass companion to a solar-type star.* Nature, Nov 1995.

[10] J. T.Vanderplas. *Understanding the Lomb–Scargle Periodogram.* The American Astronomical Society, May 2018.

[11] K. Walbolt. *51 Pegasi b.* Exoplanet Exploration Program and the Jet Propulsion Laboratory for NASA's Astrophysics Division, https://exoplanets.nasa.gov/exoplanet-catalog/7001/51-pegasi-b/, Sept .