Planetary Systems

Pages 1–11
2023

Radial-velocity data and derivation of planetary physical properties

Telmo Monteiro (202308183)
MSc. in Astronomy and Astrophysics
Physics and Astronomy Department, Faculty of Sciences of University of Porto
Course: Planetary Systems (AST4007)
Course director: Nuno C. Santos

Keywords: Markov Chain Monte Carlo methods, Radial Velocity, Exoplanets, Orbital parameters

ABSTRACT

The aim of this work was to present the results of the resolution of three interconnected exercises related to the analysis of radial velocity data. First, the concept of radial velocity signals (RV) is introduced. Then, the Markov Chain Monte Carlo algorithm is presented, as well as the Generalised Lomb-Scargle periodogram, though shortly. These methods were used to fit Keplerian functions to the two data sets provided. Two other methods used but not formally presented are the *curve_fit* function from *scipy* and the RadVel tool.

Exercise 1 consisted in the fitting of a circular orbit, obtaining a mass of $0.8760^{+0.0210}_{-0.0299}$ $[M_J]$ and a semi-major axis a of $0.1636^{+0.0003}_{-0.0003}$ AU, for the MCMC method, which is deemed the best method. Considering these two results, the planet seems to consist in a Hot Jupiter. The objective of exercise 2 was to draw the shape of the Keplerian function for 4 different sets of parameters, which layed down the work for exercise 3. Finally, exercise 3 consisted in the fitting of an eccentric orbit ($e \approx 0.4$), obtaining a mass of $1.7235^{+0.0298}_{-0.0278}$ $[M_J]$ and a semi-major axis a of $1.6553^{+0.0027}_{-0.0027}$ AU, for the MCMC method. The planet in this exercise would consist in a giant gas planet.

CONTENTS

1 INTRODUCTION

Exoplanets can be detected with various observational techniques. Among them, radial velocity (RV) has the key advantages of revealing the architecture of planetary systems and measuring planetary mass and orbital eccentricities (Hara and Ford [2023]).

The radial velocity of a star can be measured due to the Doppler effect. If a source emits a photon with wavelength λ_0 and velocity $\mathbf{v}$ of modulus v relative to an observer, then the wavelength of the photon received is

$$\lambda = \lambda_0 \frac{1 + \frac{1}{c}\mathbf{k}\cdot\mathbf{v}}{\sqrt{1 - \frac{v^2}{c^2}}}, \tag{1}$$

where $\cdot$ is the scalar product and $\mathbf{k}$ is the unit vector from the observer to the source. Stellar spectra contains thousands of absorption lines, which implies that photons are only absorbed in short intervals of wavelengths, in the upper parts of stellar atmospheres. Duo to the star's motion, the Doppler effect enters the scene and changes the apparent wavelength of the spectral lines (Hara and Ford [2023]).

Acquiring the spectrum of a star multiple times allows the measuring of changes in the apparent wavelength of each spectral line as function of time. This amounts to measuring $\mathrm{RV}(t) \equiv \mathbf{k}\cdot\mathbf{v}(t)$ in equation 1 (Hara and Ford [2023]). When a star is orbited by a planet, the gravitational pull from the orbiting companion implies that the star will orbit the center of mass of the star-planet system. As a consequence, this motion is observable as variations in the radial velocity (the velocity in the direction of the line-of-sight) of the star as a function of time. The Keplerian signal,RV signal as a function of time as a result of the orbit of a planet, is given by

$$f(t; K, P, e, \omega, M_0) = K\left[e\cos\omega + \cos(\omega + \nu(t; e, P, M_0))\right], \tag{2}$$

where K is the velocity amplitude, P is the orbital period, e is the orbital eccentricity, ω is the argument of periastron and M_0 is the mean anomaly describing the motion of the star. The true anomaly, ν, is the angle that parameterizes the position of the star on its orbit (Murray and Correia [2010]). The eccentricity of the ellipse is confined to [0, 1[for a bound system undergoing periodic motion. Orbits with $e \approx 0$ are nearly circular and can be well approximated by the leading terms of a series expansion in the mean anomaly. The signal amplitude K depends on the mass of the star and the planet, as well as on the orbital eccentricity, according to

$$K = \left(\frac{2\pi G}{P}\right)^{\frac{1}{3}} \frac{m \sin i}{(m+M)^{2/3}} \frac{1}{\sqrt{1-e^2}}, \quad (3)$$

where M is the stellar mass, m is the planet mass, G is the gravitational constant and i is the angle between the plane in which the planet orbits and the plane perpendicular to the line of sight. Each observed star has a proper motion due to the motion of the star (and the Sun) around the galactic center. On the timescale of RV observations, 1 to 20 years, this usually appears as constant, in some cases as a linear trend that must be added to equation 2. Furthermore, most Sun-like stars host multiple planet and, in principle, the planet-planet interactions cause deviations from the Keplerian model. Nevertheless, for the vast majority of planetary systems, the motion of the star can be well approximated over the timescale of RV surveys as the linear superposition of the Keplerian orbit due to each planet. This way, a model for a time series of RV measurements of a star hosting n planets with parameters $(K_j, P_j, e_j, \omega_j, M_{0j})_{j=1,\ldots,n}$ is

$$\mathrm{RV}(t) = c_0 + c_1 t + \sum_{j=1}^{n} f(t; K_j, P_j, e_j, \omega_j, M_{0j}), \quad (4)$$

as according to Hara and Ford [2023]. Substituting equation 2 in equation 4, assuming that the star only has one planet orbiting it and calling the star's average velocity c_0 as γ:

$$\mathrm{RV}(t) = K\left[e\cos\omega + \cos(\omega + \nu(t; e, P, M_0))\right] + \gamma. \quad (5)$$

Let's finally define the true anomaly ν. Firstly, the mean anomaly, M or M_0, is defined as

$$M = \frac{2\pi}{P}(t - T_0), \quad (6)$$

where t is a given moment, T is the moment when its velocity is at zero phase (or time of periastron passage) and P is the period of the orbit. The mean anomaly is also defined with Kepler's equation as

$$M = E - e\sin E, \quad (7)$$

where E is the eccentric anomaly (Murray and Correia [2010]). The relation between the true anomaly and the eccentric anomaly is defined by Vallado [2013] (for the full deduction of the equation, see this book):

$$\tan\frac{\nu}{2} = \sqrt{\frac{1+e}{1-e}}\tan\frac{E}{2}. \quad (8)$$

The Kepler's Equation is transcendental, so it needs to be solved iteratively, using for example a Newton-Raphson method.

Once all the orbits parameters are obtained, one can then compute the minimum value for the mass of the planet $M_{\mathrm{pl}}\sin i$ or $m\sin i$ by solving equation 3:

$$\frac{M_{\mathrm{pl}}^3 \sin^3 i}{(M_{\mathrm{pl}} + M_{\mathrm{st}})^2} = 1.036 \times 10^{-7} K^3 (1-e^2)^{3/2} P \ [M_\odot], \quad (9)$$

where P is the period of the signal (in days), K is the semi-amplitude of the radial velocity signal of the star in km/s, and M_{pl} and M_{st} are the planetary and stellar masses, respectively, in solar masses.

The semi-major axis of the orbit a can also be derived using Kepler's third law:

$$P^2 = \frac{a^3}{M_{\mathrm{pl}} + M_{\mathrm{st}}}, \quad (10)$$

where the masses of the planet M_{pl} and of the star M_{st} are in solar masses, P in years and the semi-major axis a in AU.

2 MARKOV CHAIN MONTE CARLO (MCMC)

The Markov Chain Monte Carlo (MCMC) methods consists in a class of algorithms for sampling from a probability distribution, being mostly used for calculating numerical approximations of multi-dimensional integrals (Wikipedia contributors [2023]). They are extremely useful for statistical inference, or in other words, to fit models to data, which is the case in this work (Pasha [2020], Hogg and Foreman-Mackey [2018]). It is possible to obtain a sample of the desired distribution by establishing a Markov chain that has that distribution as its equilibrium distribution and recording states from the chain.

A Markov chain or Markov process is a stochastic (or random process) model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event (Maltby, Henry and others [2023], Gagniuc [2017]). In a simple way, it is a process for which predictions can be made regarding future outcomes based solely on its present state and, most importantly, such predictions are as good as the ones that could be made knowing the process's full history (Oksendal [2000]).

Despite there being optimizers to find the best parameters of a model that describes a data set, MCMC shines as it is able to sample from the posterior distribution around the optimum values, generatively modelling the data. What this implies is that it allows to obtain better and more robust uncertainties for the parameters of the model, helping understanding multi-modalities or covariances in the data, as well as marginalizing out nuisance parameters that are needed to include to obtain accurate results. The fundamental process of running an MCMC is then to compare models generated by a set of parameters against data and to sample from the set of parameters that produces the models that well-fit the data (Pasha [2020], Hogg and Foreman-Mackey [2018]).

The MCMC process is inherently Bayesian, which means that it requires the user to impose priors in the model parameters. The Posterior Probability, $\mathrm{P}(\theta|\mathrm{D})$, the probability of our model given our data, is then calculated applying the Bayes theorem

$$\mathrm{P}(\theta|\mathrm{D}) = \frac{\mathrm{P}(\mathrm{D}|\theta)\mathrm{P}(\theta)}{\mathrm{P}(\mathrm{D})}, \quad (11)$$

where $\mathrm{P}(\mathrm{D}|\theta)$ is the likelihood (probability of the data given the model), $\mathrm{P}(\theta)$ is the prior (probability of the model) and $\mathrm{P}(\mathrm{D})$ is the evidence (probability of the data). The MCMC estimates (samples) the posterior distribution, integrating numerically the right handed side of the equation, for a given expectation value. If one wants the expectation value of θ (the parameters of the model), one computes

$$\mathrm{E}[\theta] = \int \theta \mathrm{p}(\theta) d\theta.$$

Briefly, the process running the MCMC is:

- Establish a function that outputs a model given a set of input parameters;
- Establish an ensemble of walkers, defined by a vector θ that contains the parameters. In the case of a circular orbit, the vector would contain the parameters K, T, P and γ;
- A way of imagining it is a grid of possible values of θ within the prior ranges declared;
- Each "walker" will explore the parameter space, by taking a step to a new value of θ and generating the corresponding model. Then it is compared to the data, through a χ^2-type check:

$$\text{“Likeliness”} = -\frac{1}{2}\sum\left(\frac{y_{\text{data}} - y_{\text{model}}}{y_{\text{data,err}}}\right)^2$$

- The ratio of the Likeliness of the new model with the data is compared with the current model. In an iterative process, the algorithm checks if the new location produces a better match to the data and, if so, the walker moves there and repeats the process. If the new location is worse, it goes back to the previous position and tries a new direction. An important quantity to be defined is the acceptance ratio, that makes sure that walkers don't get all trapped in individual peaks of high probability;
- Eventually, the walkers all begin climbing towards the regions of highest "likeness" between the models generated from the data.

Another name for the process is "production run" and the result of said process is a posterior distribution or chain: each walker contains a record of all the vectors θ accepted and the corresponding likelihood. As the MCMC consists in a sampler, it cannot tell which model θ is the "best", but it returns a representation of the user ability to constrain the parameters in the models, via the spread of those models, extracted from the posterior distribution (Pasha [2020], Hogg and Foreman-Mackey [2018]).

3 GENERALISED LOMB-SCARGLE (GLS) PERIODOGRAM

The Lomb-Scargle periodogram (Scargle [1982]) is a widely used tool in period searches and frequency analysis of time series composed by unequally spaced data. It is equivalent to fitting sine waves of the form $y = a\cos(\omega t) + b\sin(\omega t)$. While standard fitting procedures require the solution of a set of linear equations for each sampled frequency, the Lomb-Scargle method provides an analytic solution and is therefore both convenient to use and efficient.

In Zechmeister and Kürster [2009], an analytic solution for the generalisation to a full sine wave fit is given, including an offset and weights (χ^2 fitting). Compared to the Lomb-Scargle periodogram, the generalisation is superior as it provides more accurate frequencies, is less susceptible to aliasing, and gives a much better determination of the spectral intensity.

An algorithm that implements this generalisation is presented in Zechmeister and Kürster [2009], that can be accessed in a webpage[1].

[1] https://pyastronomy.readthedocs.io/en/latest/pyTimingDoc/pyPeriodDoc/gls.html#the-generalized-lomb-scargle-periodogram-gls

4 EXERCISE 1

In the case of a circular orbit, $e = 0$, which simplifies equations 7 and 8 to $M = E = \nu$, so equation 4 turns to

$$\text{RV}(t) = K\sin\left(\frac{2\pi(t-T)}{P}\right) + \gamma, \tag{12}$$

turning the cosine into a sine, as the argument of the periastron ω is not defined in a circular orbit. As one can see, the equation for RV is now much simpler.

4.1 Orbital parameters

To find the orbital parameters K, P, T and γ, the first method employed was the function *curve_fit* from the *scipy* package (Virtanen et al. [2020]). As input, the initial parameters were guessed by visual inspection of the RV plot and by roughly estimating (for example, for the T and γ parameters I took the mean of the time series and RV, respectively).

Then, the obtained parameters from the *curve_fit* are used as the initial parameters for the Markov-Chain Monte Carlo (MCMC) algorithm, which code was adapted from Pasha [2020]. There is no defined optimal number of walkers and steps to use, so I tried some and saw that the final results didn't change much, so I adopted a final number that is big enough so it's credible but small enough to be computationally feasible: 500 walkers and 500 steps. The final model adopted was the one with the highest likelihood. Figure 1 shows a corner plot for the parameters obtained and figure 2 shows the three σ posterior spread of the fit.

The third method used was the Generalised Lomb-Scargle (GLS) periodogram. The code was adapted from a webpage[2]. I computed the GLS periodogram for the sinusoidal signal of this planet, which returned the parameters of the function: in the statistical output, the "best sine period" is P, the "Amplitude" is K, "Phase (T0)" is T and "Offset" is γ. Figure 3 shows the power as a function of the frequency of the signal, along with False-Alarm-Probability (FAP) levels at 1%, 5% and 10% to help distinguish between significant and spurious peaks.

Finally, the fourth method used was an open-source Python package for modeling Keplerian orbits in radial velocity (RV) timeseries, RadVel (Fulton et al. [2018]). RadVel provides a convenient framework to fit RVs using maximum a posteriori optimization and to compute robust confidence intervals by sampling the posterior probability density via Markov Chain Monte Carlo (MCMC). RadVel allows users to float or fix parameters, impose priors, and perform Bayesian model comparison. The code was adapted from a tutorial from the package[3], allowing only the parameters K, P, T and γ to vary and setting the others to zero. I wasn't able to retrieve the uncertainties for the parameters, as there was an error related to the package that I am not qualified to correct.

Table 1 shows the orbital parameters K, P, T and γ for the four methods, and the respective uncertainties. Figure 4 shows the fitted models for each algorithm and one can see that the models obtained from *Curve_fit*, MCMC and GLS overlap, while the one obtained from RadVel is displaced. This occurs due to the difference in around 1000 days in the T parameter.

[2] https://pyastronomy.readthedocs.io/en/latest/pyTimingDoc/pyPeriodDoc/gls.html#example-adding-an-error-column-and-fap-levels

[3] https://radvel.readthedocs.io/en/latest/tutorials/K2-24_Fitting+MCMC.html

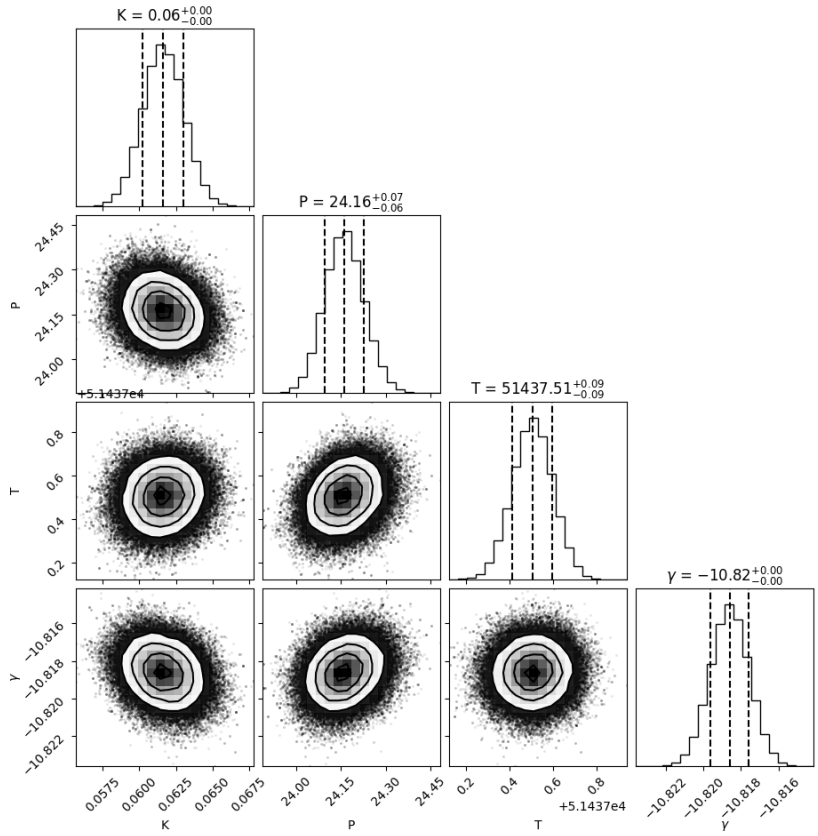

Figure 1: Corner plot of the parameters obtained with MCMC. It shows 2D Gaussian functions between the parameters and their uncertainties.

Table 1: Orbital parameters K, P, T and γ and respective uncertainties, obtained with *Curve_fit*, MCMC, GLS and RadVel.

Parameter	***Curve_fit***	**MCMC**	**GLS**	**RadVel**	**Unit**
K	$0.061645^{+0.002113}_{-0.002113}$	$0.061615^{+0.001401}_{-0.001389}$	$0.061684^{+0.001991}_{-0.001991}$	0.061647	km/s
P	$24.159835^{+0.097374}_{-0.097374}$	$24.160618^{+0.066352}_{-0.064175}$	$24.150426^{+0.132991}_{-0.132991}$	24.1584	days
T	$51437.507539^{+0.138729}_{-0.138729}$	$51437.505189^{+0.091535}_{-0.092516}$	$51365.052944^{+0.124094}_{-0.124094}$	52627.3	Julian days
γ	$-10.818611^{+0.001529}_{-0.001529}$	$-10.818604^{+0.001018}_{-0.001022}$	$-10.818647^{+0.001408}_{-0.001408}$	−10.8186	km/s

4.2 Mass of planet and semi-major axis of the orbit

Assuming that $M_{\rm pl} << M_{\rm st}$ and that the stellar mass is one solar mass, the derivation for the minimum mass function is also simplified to

$$M_{\rm pl} \sin i = \left(1.036 \times 10^{-7} K^3 P\right)^{1/3} \ [M_\odot]. \tag{13}$$

To convert the mass to Jupiter or Earth masses, I considered the mass of the Sun to be 1.9885×10^{30} kg[4], the mass of Jupiter 1898.125×10^{24} kg and the mass of Earth 5.97217×10^{24} kg[5].

To compute the semi-major axis of the orbit we only have to isolate a in equation 10, giving

$$a = \left[P^2 (M_{\rm pl} + 1)\right]^{1/3}. \tag{14}$$

Table 2 shows the resultant minimum mass estimate for each method applied (*curve_fit*, MCMC, GLS and RadVel) in solar, Jupiter and Earth masses, as well as the semi-major axis of the orbit, in AU. As one can see, the spread around the results is small: all the methods return a value for the minimum mass and orbital semi-major axis within the uncertainty interval of each other.

Seeing that the minimum mass of the planet is around 0.88 Jupiter masses and the semi-major axis of the orbit is approximately 0.16 astronomical units, one can conclude that the planet consists in a giant gas planet, more specifically a Hot Jupiter, due to its proximity to the host star.

[4] `https://nssdc.gsfc.nasa.gov/planetary/factsheet/sunfact.html`

[5] `https://ssd.jpl.nasa.gov/planets/phys_par.html`

Table 2: Minimum mass estimate of the planet $M_{pl} \sin i$ and semi-major axis of the orbit a estimate according to each method applied. The minimum mass is shown in solar, Jupiter and Earth masses.

Algorithm	**Mass** $[M_\odot]$	**Mass** $[M_\mathrm{J}]$	**Mass** $[M_\oplus]$	**a** [AU]
Curve_fit	$0.0008370^{+0.00002985}_{-0.00002978}$	$0.8768^{+0.0313}_{-0.0312}$	$278.6815^{+9.9401}_{-9.9154}$	$0.1636^{+0.0004}_{-0.0004}$
MCMC	$0.0008362^{+0.00002006}_{-0.00001943}$	$0.8760^{+0.0210}_{-0.0204}$	$278.4222^{+6.6776}_{-6.4681}$	$0.1636^{+0.0003}_{-0.0003}$
GLS	$0.0008374^{+0.00002861}_{-0.00002852}$	$0.8776^{+0.0300}_{-0.0299}$	$278.8223^{+9.5270}_{-9.4959}$	$0.1636^{+0.0006}_{-0.0006}$
RadVel	0.0008370	0.8768	278.6862	0.1636

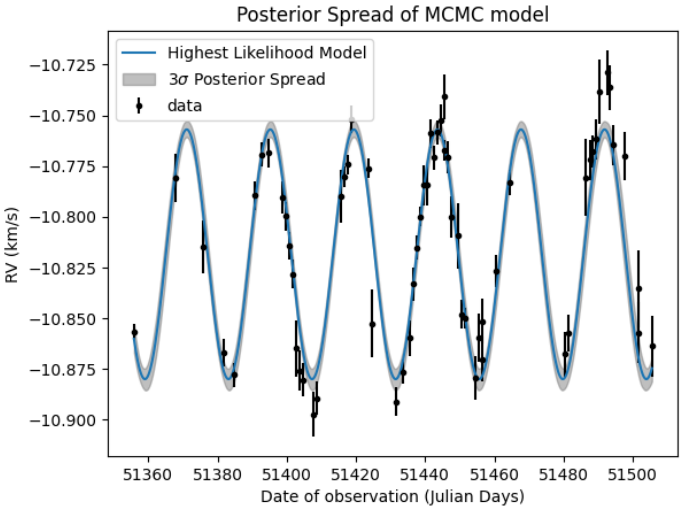

Figure 2: Three σ posterior spread of the fitted RV model with the highest likelihood.

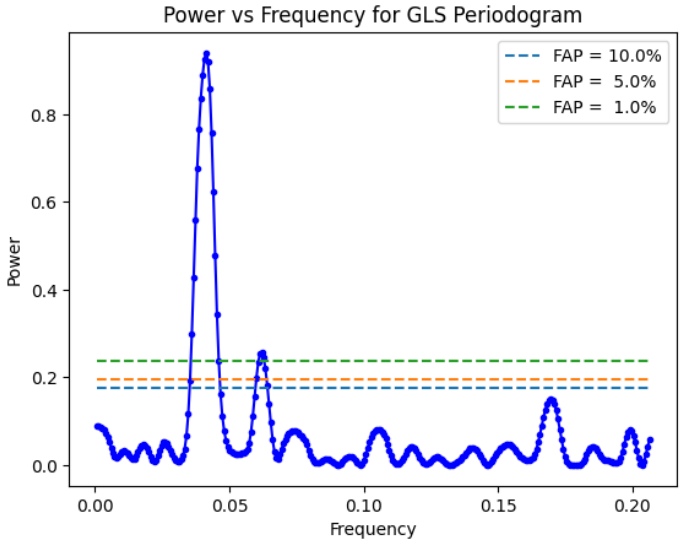

Figure 3: Power as a function of frequency, along with False-Alarm-Probability (FAP) levels at 1%, 5% and 10%.

5 EXERCISE 2

In Exercise 2, the objective is to draw the shape of the Keplerian function for 4 different sets of parameters, shown in table 3. All of the eccentricities are not null, so the orbit of the hypothetical planet is eccentric.

Table 3: Four different sets of orbital parameters P, e, K, γ, T and ω.

Orbit	**P**	**e**	**K**	γ	**T**	ω
1	100	0.3	0.1	0	0	0
2	100	0.3	0.1	0	0	π
3	100	0.6	0.1	0	0	0
4	100	0.9	0.1	0	0	0

To compute the eccentric anomaly E by solving Kepler's equation (equation 7), I used the function *fsolve* from the *scipy* package, as this equation is transcendental. The planet signals in one period (100 days) resultant from the different Keplerian orbits is depicted in figure 5.

I also plotted the true anomaly ν as a function of the time and the eccentric anomaly E as a function of the mean anomaly M for all the orbits, seen in figures 6 and 7, respectively.

Looking at figure 5, one can see that the higher the eccentricity, the more abrupt the profile changes, but the most constant it stays between periods. The closer the eccentricity is to zero, the more the signal resembles a sinusoidal wave. One can also see that the parameter ω performs a reflection in the time axis, which is expected for the argument of periastron, and as $\cos(\pi + x) = -\cos(x)$. In the same line of thinking, one can see in figure 7 that the relation between the eccentric anomaly and the mean anomaly is almost linear, or a very smooth sinusoidal, for lower eccentricities, being more wobbly as e goes up.

6 EXERCISE 3

In this exercise, the orbit of the planet is not circular, but eccentric, which means that we can't use most of the simplifications considered in Exercise 1. So the equation for the radial-velocity used is equation 4, along with the inherent expressions for the mean, true and eccentric anomalies.

6.1 Orbital parameters

To find the orbital parameters K, P, T, e, ω and γ, the procedure was the same as in Exercise 1.

The obtained parameters from the *curve_fit* were also used as the initial parameters for the Markov-Chain Monte Carlo (MCMC) algorithm. The number of walkers and steps used was 500 each, as in Exercise 1. Figure 8 shows a corner plot for

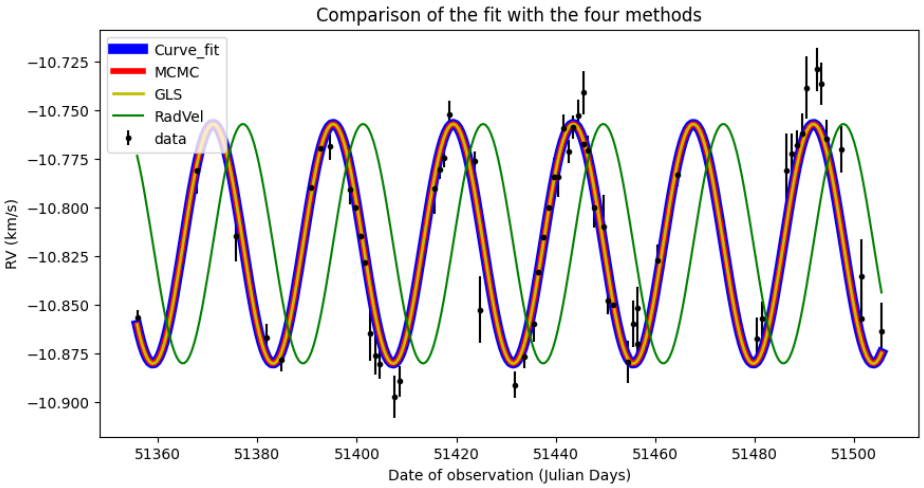

Figure 4: RV model fitted using each method. The thickness of the lines was changed for better contrast, as the fits for $Curve\_fit$, MCMC and GLS are overlapped.

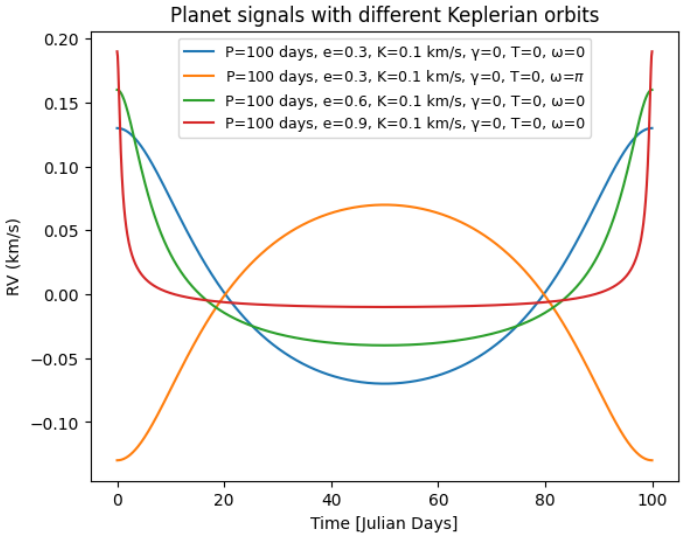

Figure 5: Planet signals for different Keplerian orbits, shown for one period (100 days).

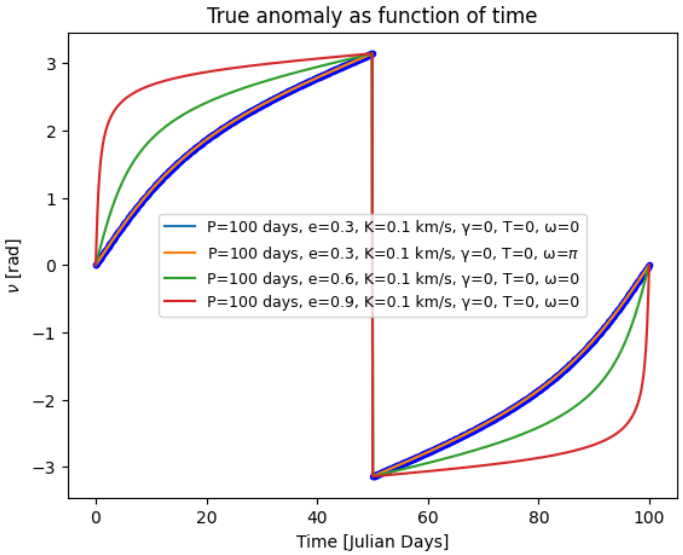

Figure 6: True anomaly of each orbit as a function of time. Blue and orange lines are overlapped, so the marker for the blue one was set to circles, for better contrast.

the parameters obtained and figure 9 shows the three σ posterior spread.

Then, the GLS was applied in the same manner as Exercise 1. The way it was applied is inherently wrong, as the GLS assumes a simple sinusoidal signal, while for an eccentric orbit the RV signal is much more complex (see equation 4). Theoretically, the GLS could be correctly applied, but I opted not to. For reference, the results of the applied GLS (incorrect) are still shown, not displaying any estimate for the eccentricity and for the argument of periastron. Figure 10 shows the power as a function of the frequency of the signal, along with False-Alarm-Probability (FAP) levels at 1%, 5% and 10% to help distinguish between significant and spurious peaks.

Table 4 shows the orbital parameters K, P, T, e, ω and γ for the four methods, and the respective uncertainties. Figure 11 shows the fitted models for each algorithm and one can see that the models obtained from *Curve_fit*, MCMC and RadVel practically overlap, while the one obtained from GLS is completely inaccurate.

6.2 Mass of planet and semi-major axis of the orbit

I again assume that $M_{\rm pl} << M_{\rm st}$ and that the stellar mass is one solar mass, resulting in a reduced expression for the minimum mass function:

$$M_{pl} \sin i = \left(1.036 \times 10^{-7} K^3 (1-e^2)^{3/2} P\right)^{1/3} \quad [M_\odot]. \quad (15)$$

To compute the semi-major axis of the orbit a in astronomical units, the equation 14 is applied again.

Table 5 shows the resultant minimum mass estimate for each method applied (*curve_fit*, MCMC, GLS and RadVel) in solar,

Table 4: Orbital parameters K, P, T and γ and respective uncertainties, obtained with *Curve_fit*, MCMC, GLS and RadVel.

Parameter	Curve_fit	MCMC	GLS	RadVel	Unit
K	$0.041692^{0.001592}_{0.001592}$	$0.041703^{+0.000968}_{-0.000895}$	$0.034708^{+0.001581}_{-0.001581}$	0.040925	km/s
P	$777.259283^{+3.681798}_{-3.681798}$	$777.242063^{+1.894780}_{-1.914279}$	$768.602749^{+10.011560}_{-10.011560}$	776.239000	days
T	$53792.740494^{+10.954298}_{-10.954298}$	$53792.728025^{+5.240416}_{-5.063577}$	$52912.489531^{+5.570760}_{-5.570760}$	54574.900000	Julian days
e	$0.405610^{0.020624}_{0.020624}$	$0.406753^{+0.013304}_{-0.013124}$	—	0.393834	adimensional
ω	$-0.625043^{0.061984}_{0.061984}$	$-0.625120^{+0.030566}_{-0.030237}$	—	-0.588982	radians
γ	$39.307962^{+0.000848}_{-0.000848}$	$39.307973^{+0.000476}_{-0.000464}$	$39.303476^{+0.001118}_{-0.001118}$	39.307620	km/s

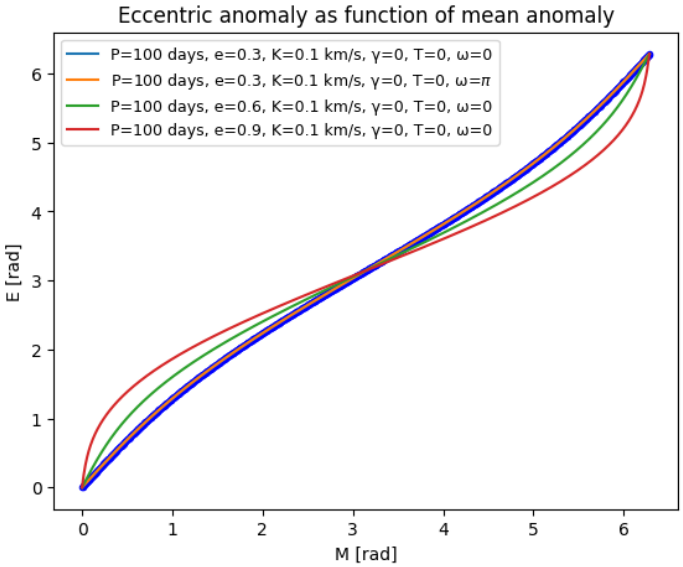

Figure 7: Eccentric anomaly of each orbit as a function of its mean anomaly. Blue and orange lines are overlapped, so the marker for the blue one was set to circles, for better contrast.

Jupiter and Earth masses, as well as the semi-major axis of the orbit, in AU. As one can see, the spread around the results is small: all the methods return a value for the minimum mass and orbital semi-major axis within the uncertainty interval of each other, on the exception of GLS, which is disregarded.

Seeing that the minimum mass of the planet is around 1.7 Jupiter masses and the semi-major axis of the orbit is approximately 1.6 astronomical units, one can conclude that the planet consists in a giant gas planet.

7 CONCLUSIONS

In this work, I was able to retrieve the orbital parameters, as well as the mass and the semi-major axis of the orbit of two different planets, as well as studying the RV signal of four different configurations of orbital parameters. To perform this task, different types of methods were used: Markov Chain Monte Carlo, Generalised Lomb-Scargle periodogram, Python's *curve_fit* function and a tool called RadVel.

The real spotlight in this work belongs to the MCMC method, as it was the most developed upon and the one that returned the most precise results (lowest uncertainties). One could say that the main limitations of this method are the calibration of the number of walkers and steps to be used and the care that needs to be applied in defining the priors and the initial parameters. The GLS, as it is, is not a method that can be directly applied to a non-circular ($e > 0$) orbit, being the only parameter that it can accurately estimate is the period, P. This was expected due to the nature of the GLS as a periodogram. On the other side, it performed really well for the circular orbit, though the reliability of its direct application is not confirmed. The RadVel tool here was used as a kind of test to cross check if the values obtained by the other methods make sense. In the circular orbit, RadVel returned a displaced T, but in the eccentric orbit it performed really well. The main limitation was not being able to obtain the uncertainties, which could be corrected with some prior knowledge of the inner machinery of the package. The *curve_fit* function from *scipy* would be a very reasonable method to bet on, since it delivered good parameters in both orbits and its use is very simple and straight-forward.

For the first planet, a mass of around 0.88 Jupiter masses was obtained, as well as a semi-major axis of the orbit of around 0.1636 astronomical units. The mass being close to Jupiter indicates that it consists in a gas giant, and the fact of it being so close to the star it is orbiting narrows it down to a Hot Jupiter. For the second planet, the mass was approximately 1.72 Jupiter masses and the semi-major axis approximately 1.66 astronomical units, once again indicating that it is a gas giant.

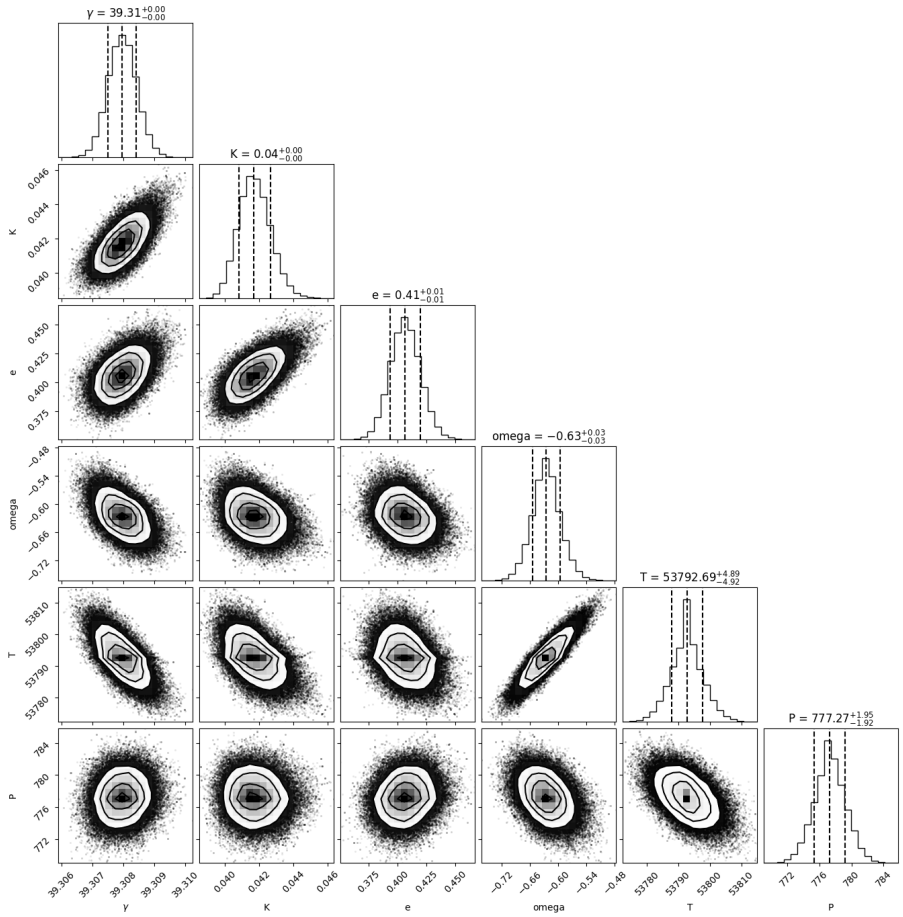

Figure 8: Corner plot of the parameters obtained with MCMC. It shows 2D Gaussian functions between the parameters and their uncertainties.

Table 5: Minimum mass estimate of the planet $M_{pl} \sin i$ and semi-major axis of the orbit a estimate according to each method applied. The minimum mass is shown in solar, Jupiter and Earth masses.

Algorithm	**Mass** $[M_\odot]$	**Mass** $[M_\mathrm{J}]$	**Mass** $[M_\oplus]$	**a** [AU]
Curve_fit	$0.00164562^{+0.00004787}_{-0.00005000}$	$1.724^{+0.0502}_{-0.0524}$	$547.9288^{+15.9402}_{-16.6473}$	$1.6553^{+0.0052}_{-0.0053}$
MCMC	$0.00164513^{+0.00002841}_{-0.00002653}$	$1.7235^{+0.0298}_{-0.0278}$	$547.7649^{+9.4599}_{-8.8348}$	$1.6553^{+0.0027}_{-0.0027}$
GLS	$0.0014932^{+0.00007477}_{-0.00007423}$	$1.5643^{+0.0783}_{-0.0778}$	$278.8223^{+9.5270}_{-9.4959}$	$1.6429^{+0.0143}_{-0.0143}$
RadVel	0.00162372	1.701	540.6348	1.6539

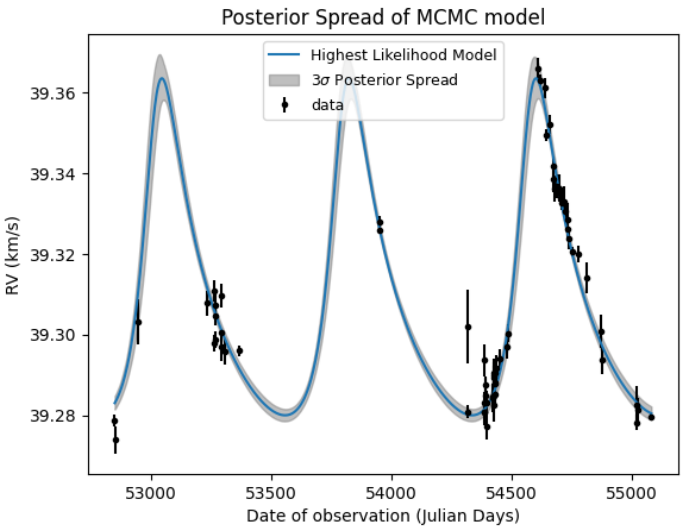

Figure 9: Three σ posterior spread of the fitted RV model with the highest likelihood.

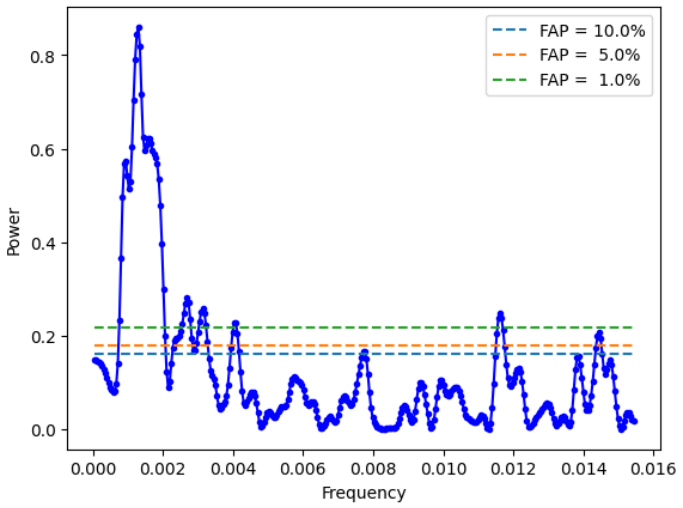

Figure 10: Power as a function of frequency, along with False-Alarm-Probability (FAP) levels at 1%, 5% and 10%.

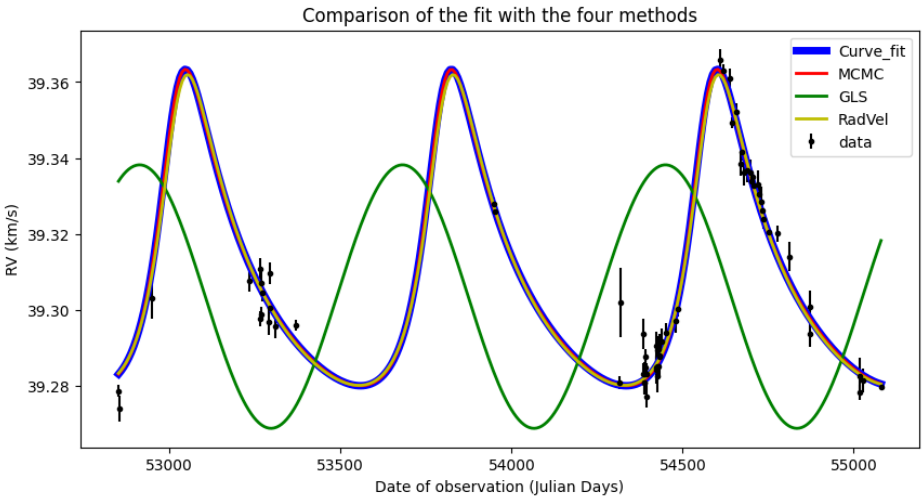

Figure 11: RV model fitted using each method. The thickness of the lines was changed for better contrast, as the fits for $Curve\_fit$, MCMC and RadVel are overlapped.

REFERENCES

Benjamin J. Fulton, Erik A. Petigura, Sarah Blunt, and Evan Sinukoff. RadVel: The Radial Velocity Modeling Toolkit. , 130(986):044504, April 2018. doi:10.1088/1538-3873/aaaaa8.

Paul Gagniuc. *Markov Chains: From Theory to Implementation and Experimentation*. 05 2017. ISBN 978-1-119-38755-8. doi:10.1002/9781119387596.

Nathan C. Hara and Eric B. Ford. Statistical methods for exoplanet detection with radial velocities. *Annual Review of Statistics and Its Application*, 10(1): 623–649, 2023. doi:10.1146/annurev-statistics-033021-012225. URL https://doi.org/10.1146/annurev-statistics-033021-012225.

David W. Hogg and Daniel Foreman-Mackey. Data Analysis Recipes: Using Markov Chain Monte Carlo. , 236(1):11, May 2018. doi:10.3847/1538-4365/aab76e.

Maltby, Henry and others. Markov chains, 2023. URL https://brilliant.org/wiki/markov-chains/. [Online; accessed 13-November-2023].

C. D. Murray and A. C. M. Correia. Keplerian Orbits and Dynamics of Exoplanets. In S. Seager, editor, *Exoplanets*, pages 15–23. 2010. doi:10.48550/arXiv.1009.1738.

Bernt Oksendal. *Stochastic differential equations*. Springer, 5 edition, 2000.

Imad Pasha. MCMC: A (very) Beginnner's Guide. github, 2020. https://prappleizer.github.io/Tutorials/MCMC/MCMC_Tutorial.html.

J. D. Scargle. Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data. , 263:835–853, December 1982. doi:10.1086/160554.

David A. Vallado. *Fundamentals of Astrodynamics and Applications*. Microcosm Press, 4 edition, 2013. ISBN 9781881883180,9781881883197,9781881883203.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi:10.1038/s41592-019-0686-2.

Wikipedia contributors. Markov chain monte carlo — Wikipedia, the free encyclopedia, 2023. URL https://en.wikipedia.org/w/index.php?title=Markov_chain_Monte_Carlo&oldid=1172552258. [Online; accessed 13-November-2023].

M. Zechmeister and M. Kürster. The generalised Lomb-Scargle periodogram. A new formalism for the floating-mean and Keplerian periodograms. , 496(2):577–584, March 2009. doi:10.1051/0004-6361:200811296.