

PREPARED FOR SUBMISSION TO UNIVERSITY OF CAMBRIDGE

Exoplanets - Coursework Assignment

Adnan Siddiquei

University of Cambridge

E-mail: as3438@cam.ac.uk

¹Word Count: XXXXX (including figure captions and appendix).

Contents

1	Q1 - TESS Lightcurve	1
2	Q2 - Planet detection and characterisation via RVs	4
2.1	Initial Exploration	4
2.2	Quantitative Analysis - Evidence of Exoplanets?	5

1 Q1 - TESS Lightcurve

Using photometry data to identify exoplanets works through analysing the light curve of a star and identifying periodic dips in the light intensity. These dips are caused by the planet passing in front of the star, blocking some of the light, resulting in a reduction in measured flux. There are a variety of methods to find and fit for exoplanets, in this report, the TLS algorithm [1] is used to analyse the provided TESS light curve data.

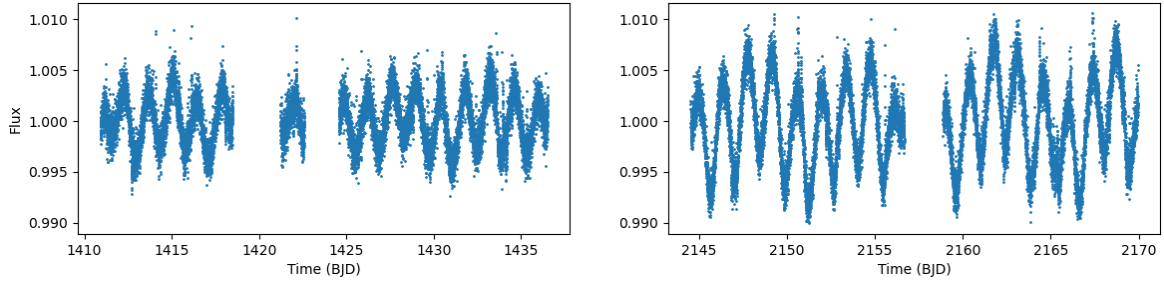
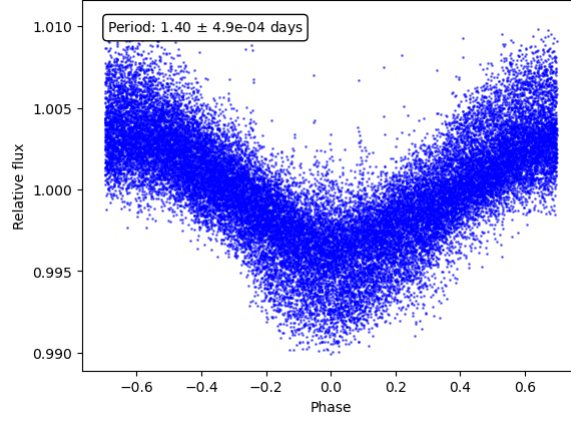


Figure 1: The original TESS light curve data, after outliers have been removed. The data is split into two plots as there is a large time gap between the two sets of data.

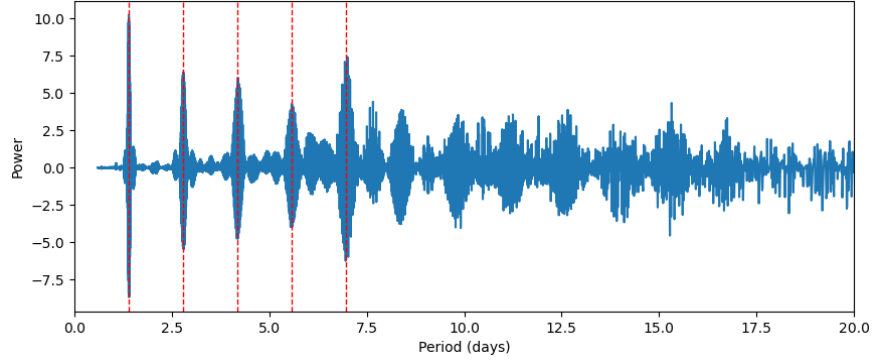
First, the data is pre-processed by removing any outliers, which were classified as points that are more than 3 standard deviations away from the median. This data is shown in Figure (1), where it is apparent that a short term periodicity is present in the data. This periodicity is due to stellar activity, most likely due to the rotation of the star and as such it needs to be normalised out before any exoplanets can be identified. The TLS algorithm is then used to identify the strongest period, which is the aforementioned stellar activity.

Figure (2) shows the results of the TLS algorithm, with the strongest period identified at 1.40 days. The periodogram in Figure (2b) supports this result, showing a clear peak at multiples of this period. A Savitzky-Golay filter was then applied to the data to remove the stellar activity, and the kernel size was chosen to be exactly half this stellar activity period to effectively remove the periodicity. With the stellar activity removed, the TLS algorithm was then run again to identify any exoplanets present in the data.

Figure (3) shows the results of the TLS algorithm after removing the stellar activity, with the strongest period identified at 5.360 ± 0.009 days. The periodogram in Figure (3b) supports this result, showing a clear peak at multiples of this period. Further analysis of the data in a similar way yielded little evidence of any other exoplanets in the data. Therefore, Table (1) summarises the parameters of the identified exoplanet, and the derived planetary radius and semi-major axis using Kepler's third law (with errors propagated accordingly).



(a) The folded lightcurve of the strongest period identified by the TLS algorithm. The period is 1.40 days.

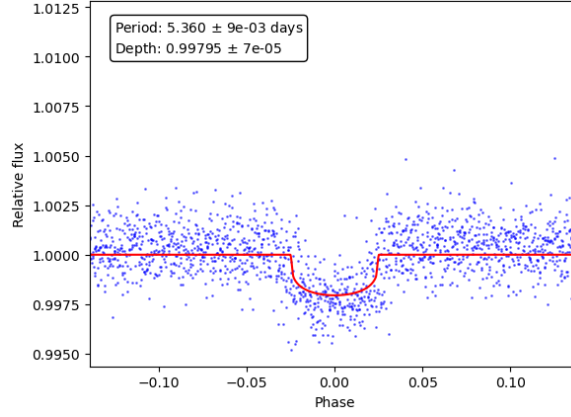


(b) The periodogram of the TLS algorithm, showing the strongest period at 1.40 days, and repeating peaks at multiples of this period.

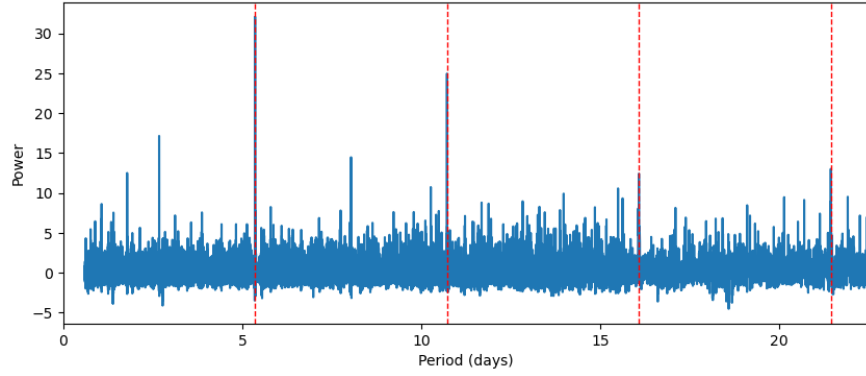
Figure 2: The folded lightcurve and periodogram of the strongest period identified by the TLS algorithm.

Parameter	Units	Value
Orbital Period	<i>days</i>	5.360 ± 0.009
Planetary Radius	R_{\odot}	0.012 ± 0.001
Planet Semi-major Axis a	<i>AU</i>	0.0381 ± 0.0007

Table 1: Summary of the parameters of the identified exoplanet, and the derived planetary radius and semi-major axis.



(a) The folded lightcurve at a period of 5.36 days.



(b) The periodogram of the TLS algorithm, showing peaks at multiples of the 5.36 day period.

Figure 3: The folded lightcurve and periodogram of the strongest period identified by the TLS algorithm after removing the stellar activity.

2 Q2 - Planet detection and characterisation via RVs

2.1 Initial Exploration

Radial velocity measurements are extremely useful for detecting exoplanets and characterising their properties. These measurements are taken by observing the Doppler shift in the spectrum of a star, indicating a shift in the star's velocity tangential to the observer's line of sight. However, a variety of stellar "nuisance" signals can also induce these shifts in the star's spectrum. These include things like oscillations and granulations which occur over the timescales of minutes to hours; rotationally-modulated activity which occur over the timescale of days to weeks, often in line with the star's rotation period; and magnetic cycles which are long-term variations in the star's magnetic field. Likewise, not all periodic signals in the RV data will be due to exoplanets or stellar signals, it is important to note that instrument noise or poor sampling (aliasing) can also induce periodic signals in the RV data.

The radial velocity is extracted from an observed spectrum using a cross-correlation function (CCF) which compares the observed spectrum to a template spectrum of what the star should look like at rest. From this CCF, two stellar activity indicators can also be extracted: the full width at half maximum (FWHM) and the bisector span (BIS). These indicators correlate to the stellar activity signals mentioned above, and can be used to filter out the stellar activity signals from the radial velocity measurements.

Parameter	Units	Value
Rotation Period P_{rot}	days	31 ± 10
Mass M_{\star}	M_{\odot}	0.69 ± 0.01
Effective Stellar Temp. T_{eff}	K	4500

Table 2: Properties of star CB 01223.

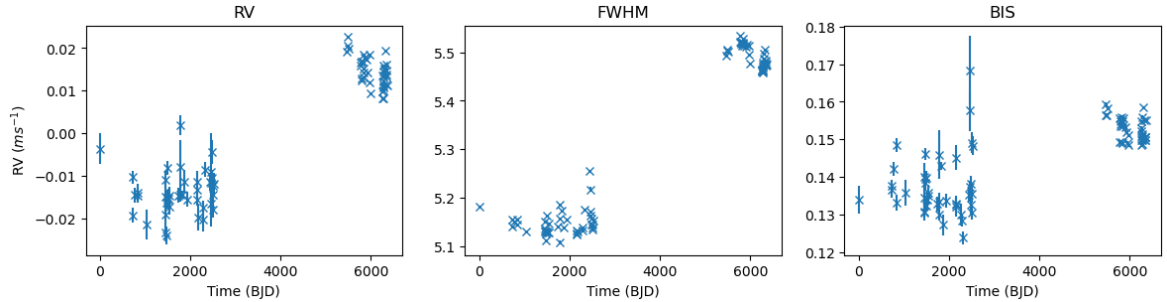


Figure 4: The radial velocity, FWHM and BIS data for star CB 01223.

Table (2) shows the properties of the star CB 01223 that this section will analyse and characterise. Figure (4) shows a plot of the RV, FWHM and BIS data for this star. Initial qualitative inspection of the RV, FWHM and BIS shows a long term correlated positive trend, where all three of the datasets show a jump in the RV between the two groups of measurements (centered around 2000 BJD and 6000 BJD). Given the correlation, this indicates that there is a likely long term stellar activity signal present in the data, and given the time scale, it is likely due to the star's magnetic cycle.

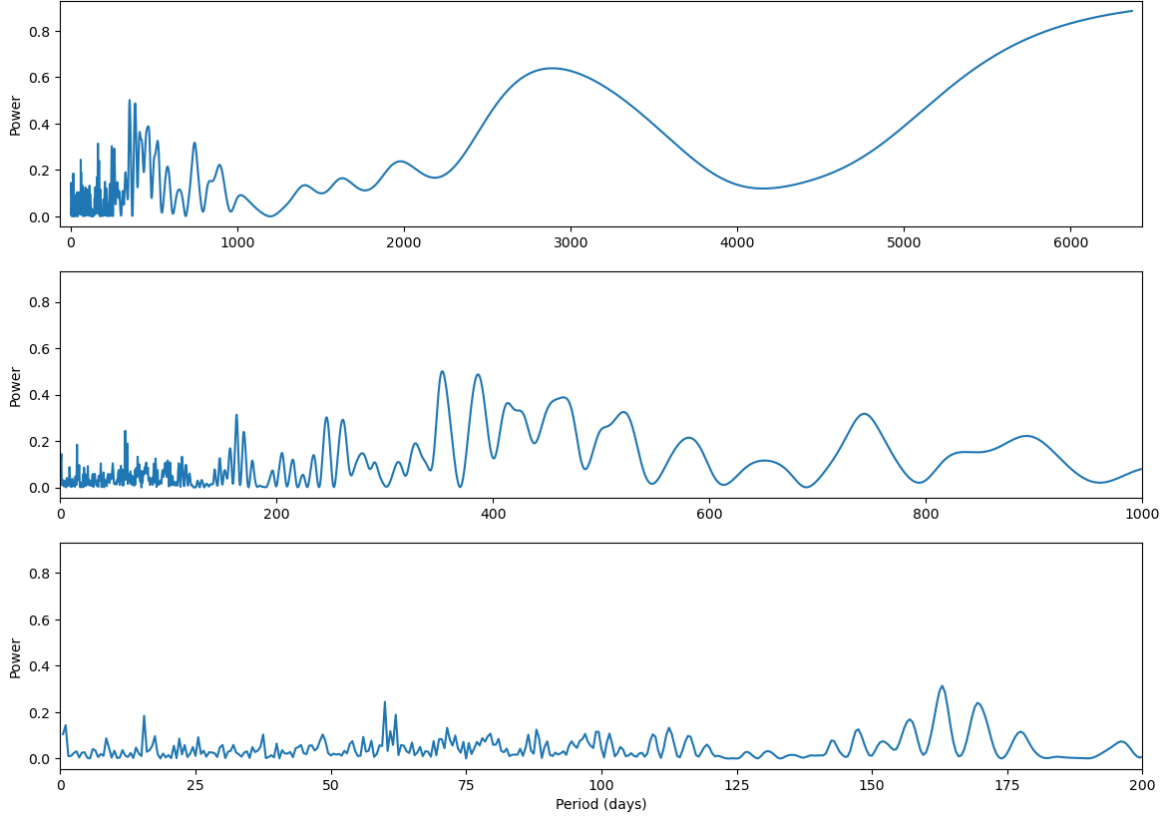


Figure 5: LombScargle periodogram of the RV data for star CB 01223. The periodograms are shown at 3 different period ranges, with the top plot showing the full range of periods, and the bottom two plots showing zoomed in views of the periodogram.

An initial LombScargle periodogram [2, 3] as shown in Figure (5) yields little useful information. A peak near the 3000-day mark is likely due to the 2 groups of measurements, and the zoomed in views show no significant peaks worth exploring.

2.2 Quantitative Analysis - Evidence of Exoplanets?

The exoplanet search conducted in this report follows closely with the methodology used by Faria, et al. (2022) [4] to search for exoplanets in the radial velocity data of Proxima Centauri. They initially utilised a joint RV and FWHM Gaussian process (GP) model to model the stellar activity signals in the data, and subtracted this from the RV data before fitting a Keplerian model to the residuals. This methodology is extended slightly and a joint GP model is fit to all three of the RV, FWHM and BIS datasets to model the stellar activity.

$$\mathcal{K}_{\text{QP}}(\tau) = \theta_1^2 \exp \left[-\frac{\tau^2}{2\theta_2^2} - \frac{\sin^2 \left(\frac{\pi|t_i - t_j|}{\theta_3} \right)}{\theta_4^2} \right] \quad (2.1)$$

Equation (2.1) shows the Quasi-Periodic (QP) kernel used in the GP model, where all θ_2 , θ_3 and θ_4 were all shared between the RV, FWHM and BIS datasets (hence the joint GP model), but θ_1 was allowed to float for each dataset.

Parameter	Units	Prior
Amplitude θ_1 RV	ms^{-1}	$\mathcal{U}(1e^{-4}, 3\Delta RV)$
Amplitude θ_1 FWHM	ms^{-1}	$\mathcal{U}(1e^{-4}, 3\Delta FWHM)$
Amplitude θ_1 BIS	ms^{-1}	$\mathcal{U}(1e^{-4}, 10\Delta BIS)$
Exp Length Scale θ_2	s	$\mathcal{U}(1e^{-4}, 3\Delta t)$
Stellar Period θ_3	s	$\mathcal{G}(31, 10)$
Periodic Length Scale θ_4	s	$\mathcal{U}(1e^{-4}, 3\Delta t)$
Jitter j	ms^{-1}	$\mathcal{LU}(1e^{-10}, 1e^{-6})$

Table 3: Parameters and priors for the joint QP GP model on the RV, FWHM and BIS datasets. $\mathcal{U}(a, b)$ denotes a uniform prior between a and b ; $\mathcal{G}(\mu, \sigma)$ denotes a Gaussian prior with mean μ and standard deviation σ ; and $\mathcal{LU}(a, b)$ denotes a log-uniform prior between a and b .

Table (3) shows a list of all the parameters and priors used for the joint stellar activity GP model. Note that the data was standardised before fitting the GP model to help with convergence, the table lists the priors in the non-standardised scale, but these were converted to the standardised scale before fitting. Bringing the data to a common scale also allowed for a single jitter term to be used across all three datasets, reducing the number of parameters that needed to be fit. The jitter term was given a log-uniform prior to allow for a wide range of values, but the upper limit was set to $1e^{-6}$. In general, relatively uninformative priors were used for all the remaining parameters, all priors covered a multiple of the entire extent of their respective dataset ranges. The only exception was the stellar period θ_3 which was given a Gaussian prior centered around the rotation period derived from the TESS photometry measurements.

A Markov Chain Monte Carlo (MCMC) simulation was then run, drawing initial samples from these priors with 250 walkers and 2000 iterations per walker. Samples were extracted with a burn in of 100 iterations and a thinning factor of 20 to reduce autocorrelation. Figure (6) shows the corner plot of this MCMC simulation, and Figure (7) shows the GP model fit to the RV, FWHM and BIS datasets from the maximum likelihood sample of the MCMC simulation. The MCMC simulation found 2 distinct distributions for the stellar period θ_3 in the posterior space, as shown in Figure (8) which plots a 2D histogram of the stellar period θ_3 and the log-likelihood of the MCMC samples, illustrating the two distinct distributions. The maximum likelihood sample of the MCMC simulation had a period of 35.3 days.

As per the methodology used by Faria, et al. (2022) [4], the GP model was subtracted from the RV data as shown in the residual plot in Figure (7). A simplified Keplerian model, in the form of a sine wave was then fit to the residuals to identify any potential exoplanets. The single planet model consisted of 3 parameters: the semi-amplitude K , the orbital period P and the phase offset ϕ . Likewise, the 2 planet model consisted of 6 parameters: the semi-amplitudes K_1, K_2 , the orbital periods P_1, P_2 and the phase offsets ϕ_1, ϕ_2 . Both of these models used the priors shown in Table (4), the priors were chosen to be relatively uninformative, covering at least the entire range of values that the parameters could take.

The MCMC simulation was run in a nested manner by running a simulation, and then

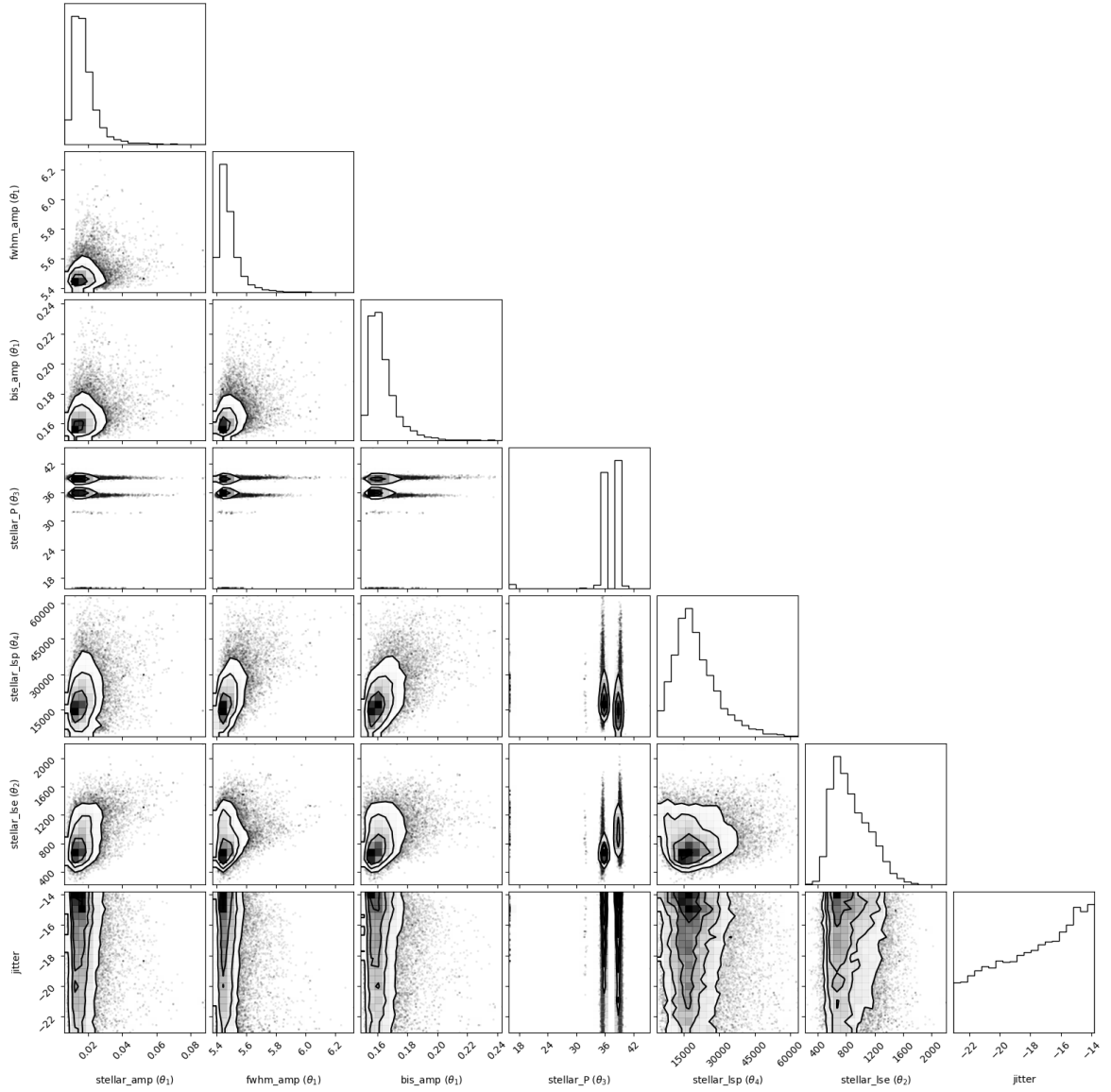


Figure 6: A corner plot of the MCMC simulation for the joint QP GP model on the RV, FWHM and BIS datasets. The bottom 10th percentile of the samples (based on the log-likelihood) were dropped to remove the tails of the distributions, for better visualisation.

Parameter	Units	Prior
Semi-amplitude K	ms^{-1}	$\mathcal{U}(0, 1.5\Delta\text{Residuals})$
Orbital Period P	s	$\mathcal{U}(0, 3\Delta t)$
Phase Offset ϕ	s	$\mathcal{U}(0, 2\pi)$

Table 4: Parameters and priors for the sinusoidal orbital model on the residuals of the RV data. Both of the single and 2 planet models used these same priors.

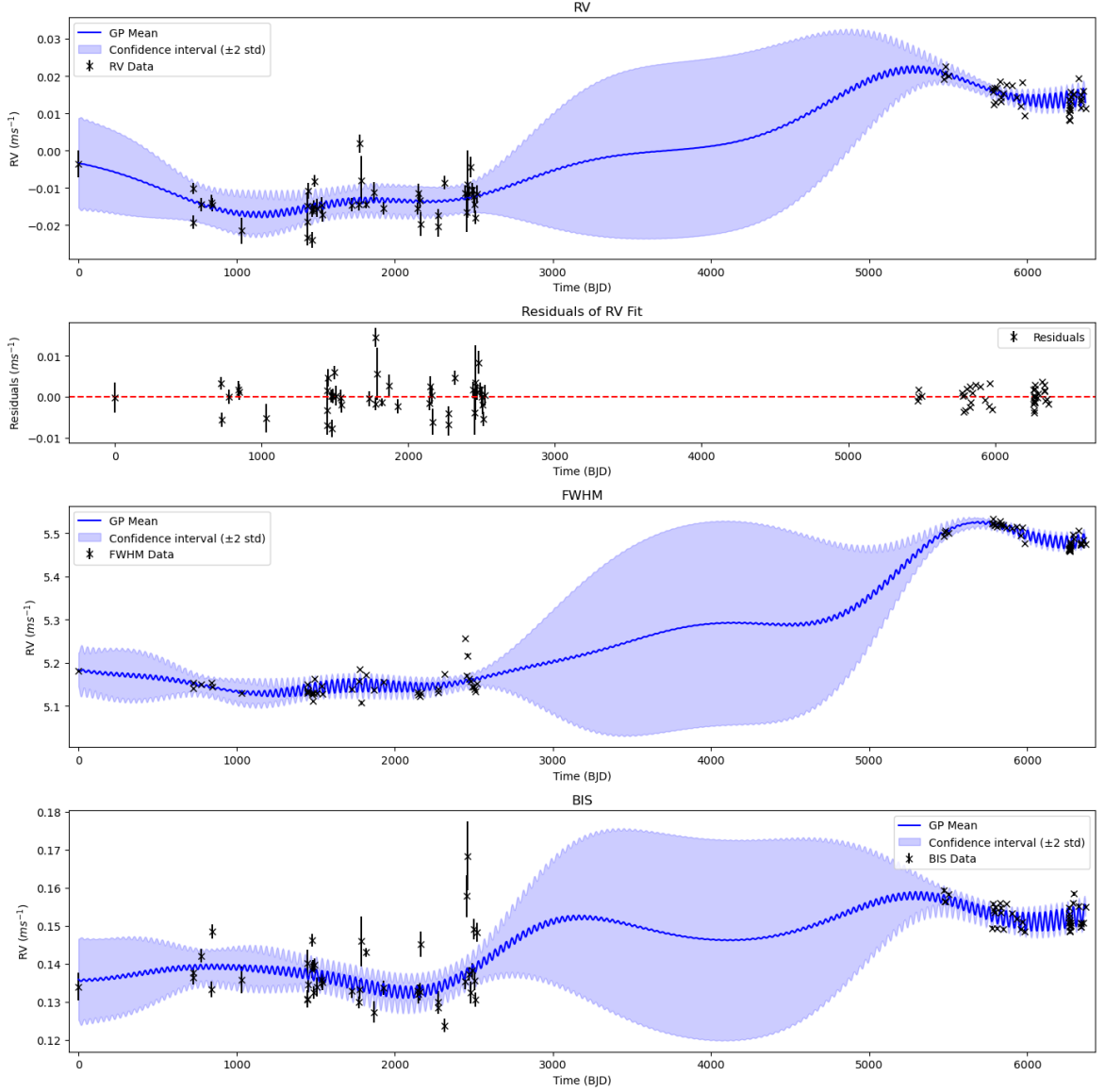


Figure 7: The GP model fit to the RV, FWHM and BIS datasets for star CB 01223. The fitted GP uses the parameters from the maximum likelihood sample of the MCMC simulation. The second plot shows the residuals of the GP model fit to the RV data.

shrinking the uniform prior on the orbital period P , with the new prior being informed by the posterior distribution of the previous simulation. This was done to ensure that no bias was added into the priors initially, but to also allow the MCMC simulation to sequentially explore the higher likelihood regions of the posterior space to yield a better final estimate of the parameters.

For the single planet model, the initial MCMC simulation found that the top 5th percentile of the samples (based on the log-likelihood) had a period in the range $[1.72, 53.10]$, and so a second simulation was conducted with a new uniform prior on the orbital period P being $\mathcal{U}(0, 58)$. This yielded a best fit model as shown in Figure (9), with the corner plot of

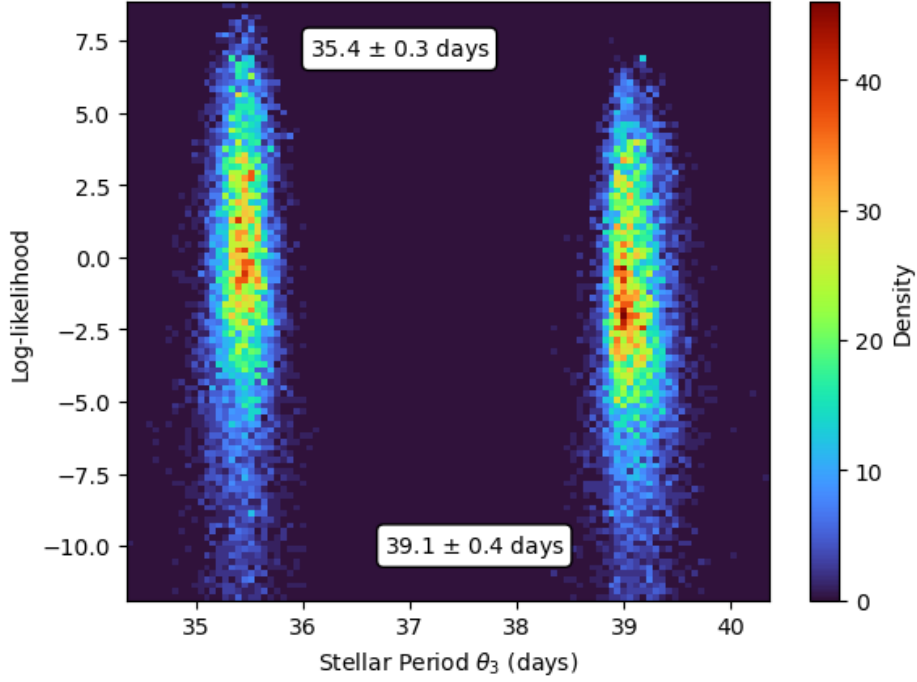


Figure 8: The joint distribution of the stellar period θ_3 and log-likelihood of the MCMC samples. The text boxes show the mean and standard deviation of the stellar period θ_3 for each of the two distributions.

the simulation samples given in Figure (10). Figure (9) also shows the posterior distribution of the orbital period P and semi-amplitude K . The best fit model had a period of 5.09 ± 0.87 days and a semi-amplitude of $2.24e^{-3} \pm 0.61e^{-3} ms^{-1}$. The best fit model was taken as the maximum likelihood sample of the MCMC simulation, and the errors were given as 2 sigma on the posterior distributions, to indicate a 95% confidence region. As evident on the corner plot in Figure (10), there were several peaks in the posterior distribution of the orbital period P in multiples of roughly 5 days (which is to be expected), and as such, the errors on the orbital period P was calculated using only the samples centered around 5.09 days.

Similar exploration on the 2 planet model did not yield any significant results indicating the possibility of a second exoplanet and as such, the single planet model was taken as the best fit model to the RV data for star CB 01223.

Table (5) shows a summary of the final posterior distributions for all key parameters and derived properties for star CB 01223 in the single planet model. The planet was computed to have a semi-major axis of 0.051 ± 0.020 AU, indicating that it lives very close to the host star, making it likely to be a hot Jupiter.

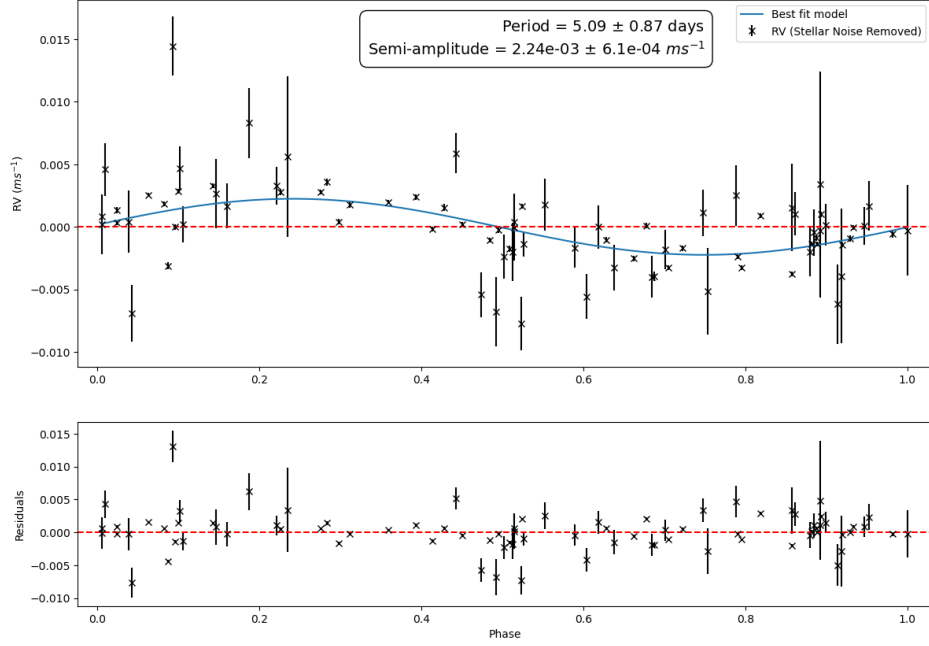


Figure 9: The best fit single planet model to the RV data for star CB 01223. The posterior distribution of the orbital period P and semi-amplitude K are shown in the top right corner of the plot, the errors given are 2 sigma errors on the parameters.

Parameter	Units	Prior
Stellar Period P_{\star}	<i>days</i>	35.4 ± 0.3
Planet Orbital Period P_{planet}	<i>days</i>	5.09 ± 0.87
Planet Semi-amplitude K	ms^{-1}	$2.24e^{-3} \pm 0.61e^{-3}$
Planet Semi-major Axis a	<i>AU</i>	0.051 ± 0.017

Table 5: Final summary of posterior distributions for all key parameters and derived properties for star CB 01223 in the single planet model. All errors are quoted as 2 sigma errors, to indicate a 95% confidence region.

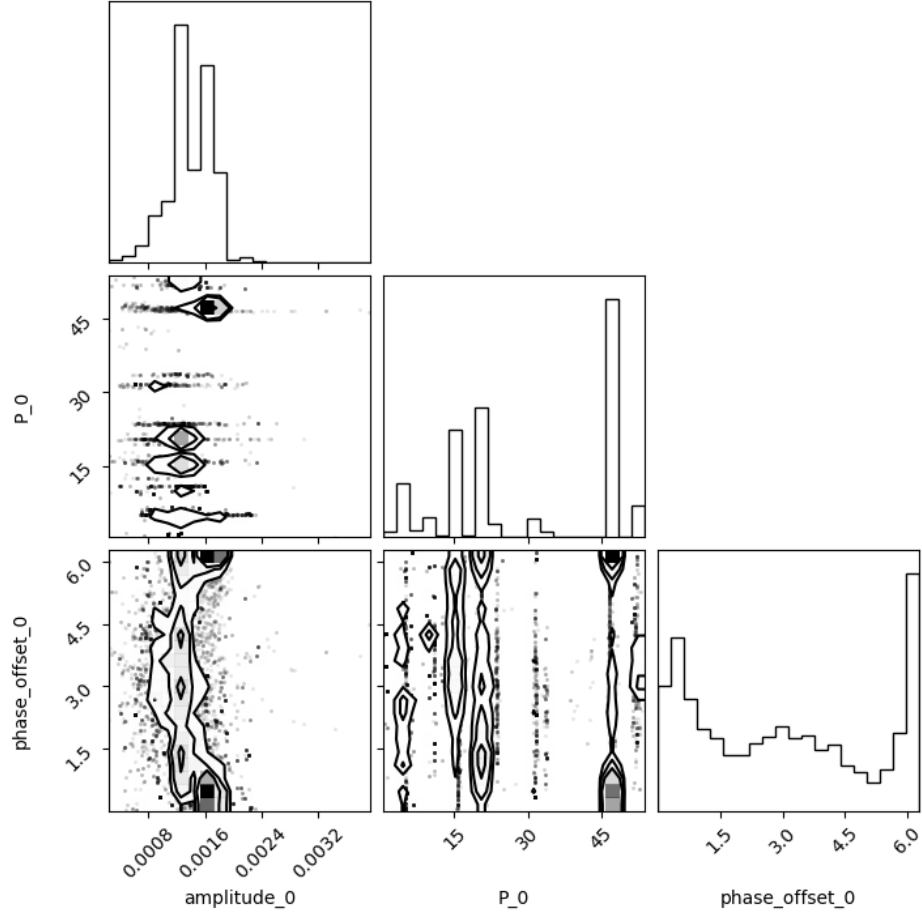


Figure 10: A corner plot of the MCMC simulation for the single planet model on the RV data.

References

- [1] Hippke, M., and Heller, R., *Transit Least Squares: Optimized transit detection algorithm to search for periodic transits of small planets*. arXiv preprint arXiv:1901.02015, 2019. Available at <https://arxiv.org/abs/1901.02015> [Accessed: 20-Jun-2024].
- [2] Lomb, N.R., *Least-squares frequency analysis of unequally spaced data*. Astrophysics and Space Science, 39(2), pp.447-462, 1976. Available at <https://doi.org/10.1007/BF00648343> [Accessed: 20-Jun-2024].
- [3] Scargle, J.D., *Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data*. The Astrophysical Journal, 263, pp.835-853, 1982. Available at <https://doi.org/10.1086/160554> [Accessed: 20-Jun-2024].
- [4] Faria, J.P., Suárez Mascareño, A., Figueira, P., et al., *A candidate short-period sub-Earth orbiting Proxima Centauri*. Astronomy & Astrophysics, 658, A115, 2022. Available at <https://doi.org/10.1051/0004-6361/202142337> [Accessed: 20-Jun-2024].