Planet detection with Lomb-Scargle periodogram

John Gillan

1 Introduction

The aim of this project is to analyse the radial-velocity measurements of three stars to see if they have planets orbiting them, and to find the orbital periods of these planets. The data contain measurements taken of the radial-velocities of the stars at a series of different times. If the stars have planets orbiting them, and if those planets are large enough, their gravitational pull on the host star should be apparent in the radial-velocity of the star. Complete measurements of the stars radial-velocities would make this task relatively

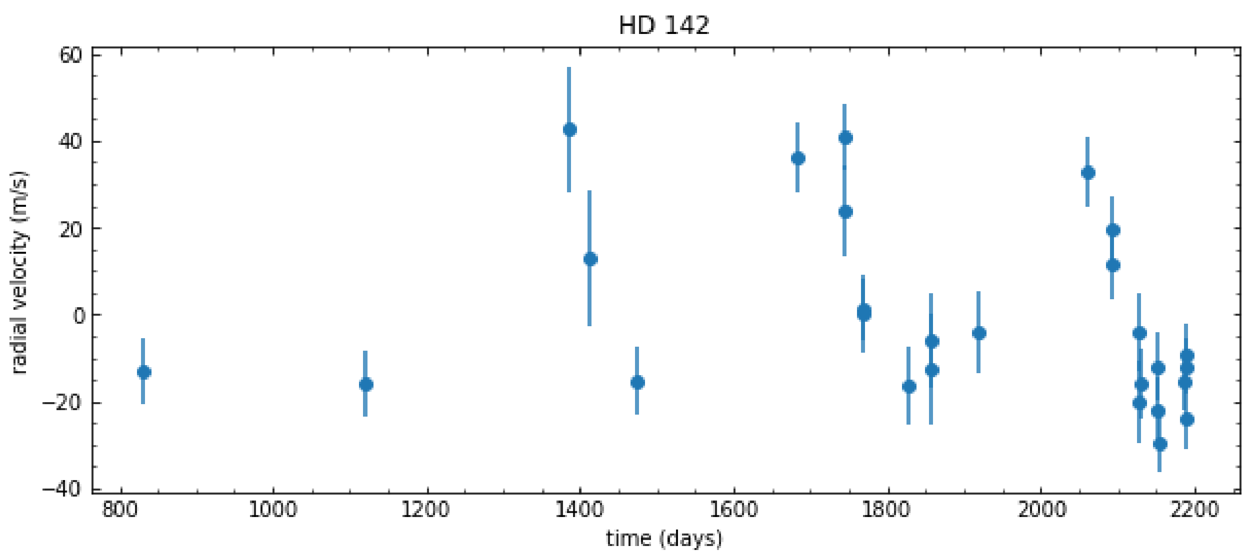

Figure 1: Radial-velocity measurements with uncertainties for HD 142.

easy as a plot of the measurements would make any cyclical motion apparent. The data under investigation here however are not complete. The star data are taken from the Anglo-Australian Planet Search at the Siding Spring Observatory in New South Wales, and because observations can only be taken at certain times, for example during the night and during clear weather, the measurements were not taken at evenly spaced times, and so the data are unevenly sampled. One look at the data for each of these stars show that simply by looking at it, it is not at all obvious that there is a pattern to the data. What is required is a method of assigning probabilities to different orbital periods, by creating a periodogram. A high probability for a period is a sign that the radial-velocities are really cyclical, and that a planet must be causing them. To accomplish this the Lomb-Scargle periodogram will be used. It is designed to find the probabilities of different periods in unevenly sampled data. Important considerations to take into account during the creation of these periodograms will be the range of periods and frequencies investigated, as well as the resolution of the periodogram. A resolution that is not fine enough could potentially miss a peak in the results, and so miss the true likeliest period. A plot of the folded data should then make it clear if the analysis was correct. It is possible that any period found to be likely could be a result of the periodogram finding patterns that do not represent the real motion of the star. Random data will produce some patterns with certain probabilities, and so produce false alarms in the periodogram. It will therefore be necessary to find the significance of any result.

2 Methods

For each star there were radial-velocity measurements v_i, with an associated time t_i, and with an uncertainty σ_i. The formula for the Lomb-Scargle periodogram to find the probability of a particular frequency in the data is:

$$P(f) = \frac{1}{2\sigma^2}\left[\frac{\left[\sum_i h_i \cos\omega(t_i-\tau)\right]^2}{\sum_i \cos^2\omega(t_i-\tau)} + \frac{\left[\sum_i h_i \sin\omega(t_i-\tau)\right]^2}{\sum_i \sin^2\omega(t_i-\tau)}\right] \tag{1}$$

where τ is:

$$\tau = \frac{1}{2\omega}\arctan\left(\frac{\sum_i \frac{1}{\sigma_i^2}\sin 2\omega t_i}{\sum_i \frac{1}{\sigma_i^2}\cos 2\omega t_i}\right), \tag{2}$$

and h_i is:

$$h_i = v_i - \bar{v}, \tag{3}$$

and σ^2 is the standard deviation, and is used to normalize the periodogram. With this normalization any particular frequency has exponential probability with unit mean. $\bar{v}$, the mean of the radial-velocity measurements, has to be subtracted from each individual measurement in the calculation of h for the Lomb-Scargle periodogram to work. For each star an estimate of the maximum frequency to be analysed up to was made. It was important to make sure this maximum was high enough so that no peaks in the periodogram were cut off. Several different resolutions were used in the calculations for each star.

From the resulting periodograms, the highest peak for each star was taken as the most likely frequency for the orbiting planet, and from this the period of the orbit was found. The period was then used to re-plot the original data, this time folded on the period. These plots allowed the amplitude to be estimated.

To find the significance of the period, or the likelihood that this particular period is not significant, we need to find the probability that this power would appear in random data by chance alone. The probability that a particular power will be greater than some value z is:

$$P(f > z) = 1 - (1 - e^{-z})^M \tag{4}$$

where M is the number of independent frequencies in the data. To find M, Monte Carlo simulations were carried out. 1000 random data sets were created by:

$$v_{i(rand)} = 0 + N(0, \sigma) \tag{5}$$

where σ is the variance. A Lomb-Scargle normalized periodogram was then computed for each of these random sets of data. The maximum peak for each was recorded and the mean of all these peaks was found. This mean was then taken to be the 50% significance level, as half of the random data sets produced peaks that reached this level, and half produced peaks that didn't reach this level. The above equation was then rearranged for M:

$$M = \frac{\ln(1 - P(f > z))}{\ln(1 - e^{-z})} \tag{6}$$

with $z = 50\%$, and the mean of the maximum peaks from the random data $= P(f > z)$. Once M was estimated, it could then be used to calculate any significance level needed.

3 Results

For HD 142 the data showed a start day of measurements on day 830 and a finish on day 2190, so the periods searched were from 1 to 1000 days, as any periods higher than this would not show up in this data, and there were not enough data points to find any period smaller than this. So the maximum frequency searched, (f_{max}), was 1. Several resolutions for the periodogram were tested, without a large affect on the results. The x-axis of all the plots of the frequency periodograms were limited to the regions where the peaks were, in order to make them easier to see. The most likely period for HD 142 is 332 days with frequency 0.003. The plot of the folded data shows that the amplitude is 25 m/s.

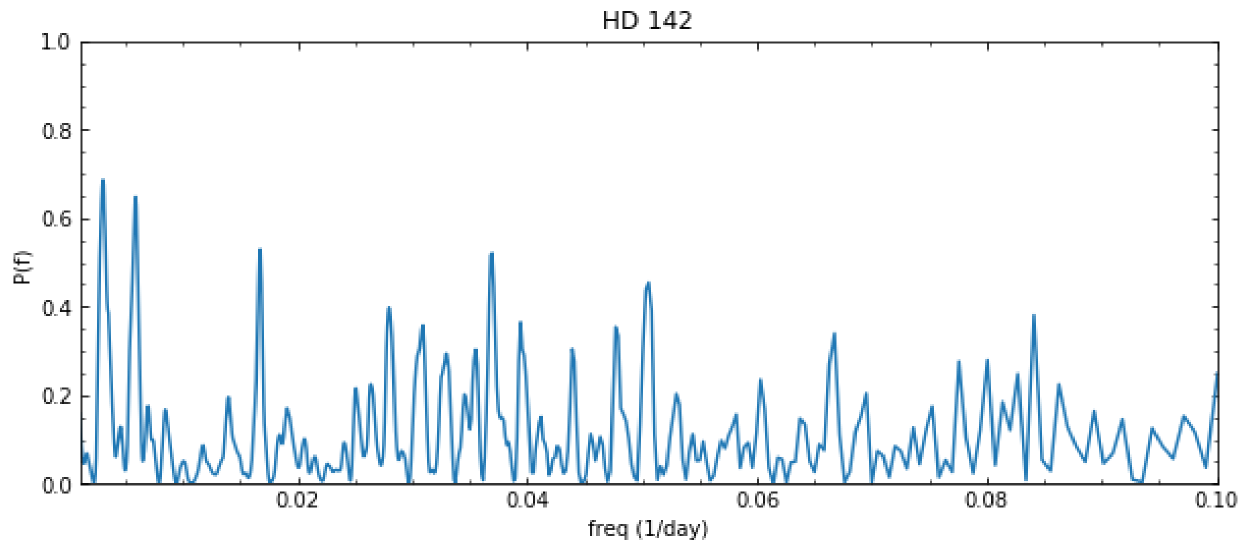

Figure 2: Normalized Lomb-Scargle periodogram showing probabilities of different frequencies for HD 142. There are several peaks, with the largest at 0.003.

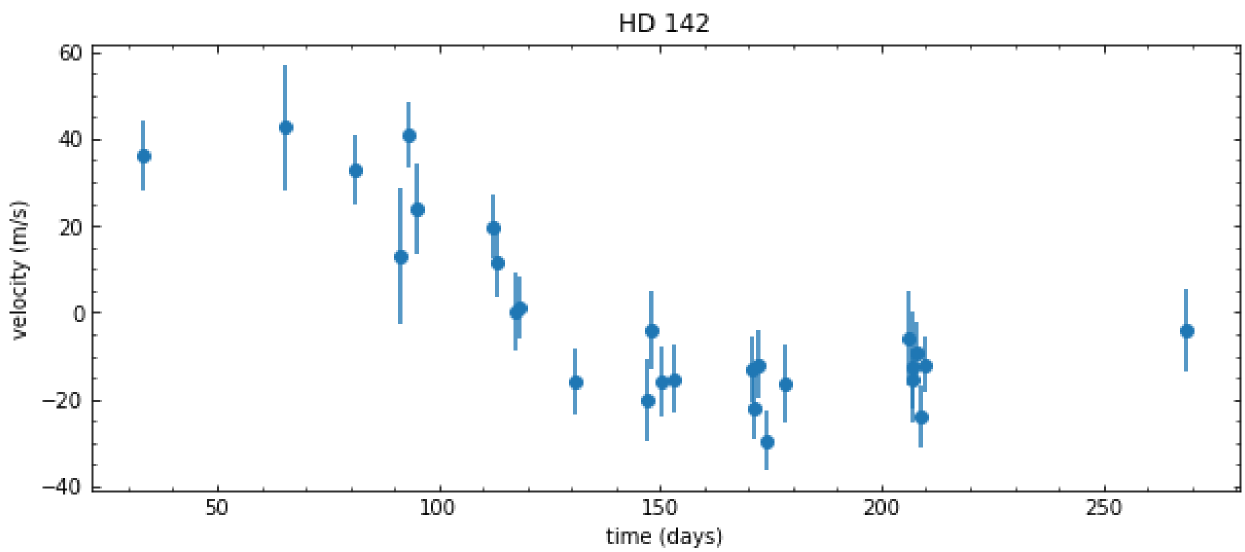

Figure 3: Folded radial-velocity measurements for HD 142 generated from period found with the Lomb-Scargle periodogram. The amplitude is about 25 m/s.

The data for HD 27442 covered a range of about 1200 days so 1000 was again used as the maximum period. The most likely period is 423 days with frequency 0.002. The plot of the folded data shows that the amplitude is 35 m/s.

The data for HD 102117 covered a range of about 3000 days, however when a large maximum period was used no high periods showed any significant likelihood so the plotted results

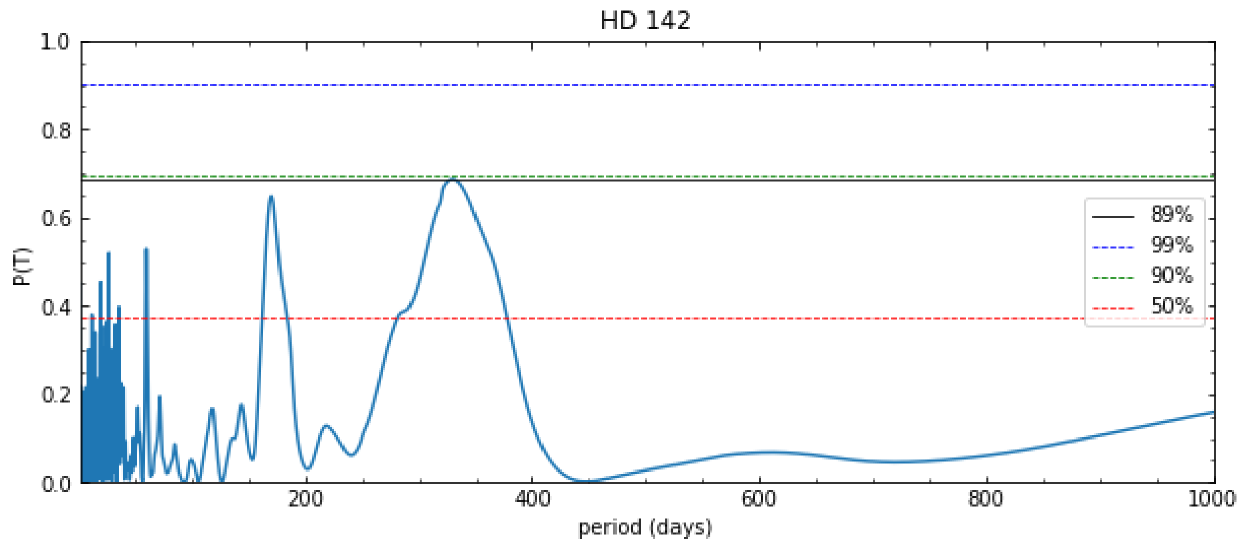

Figure 4: Lomb-Scargle periodogram for HD 142 with the probability significance levels shown.

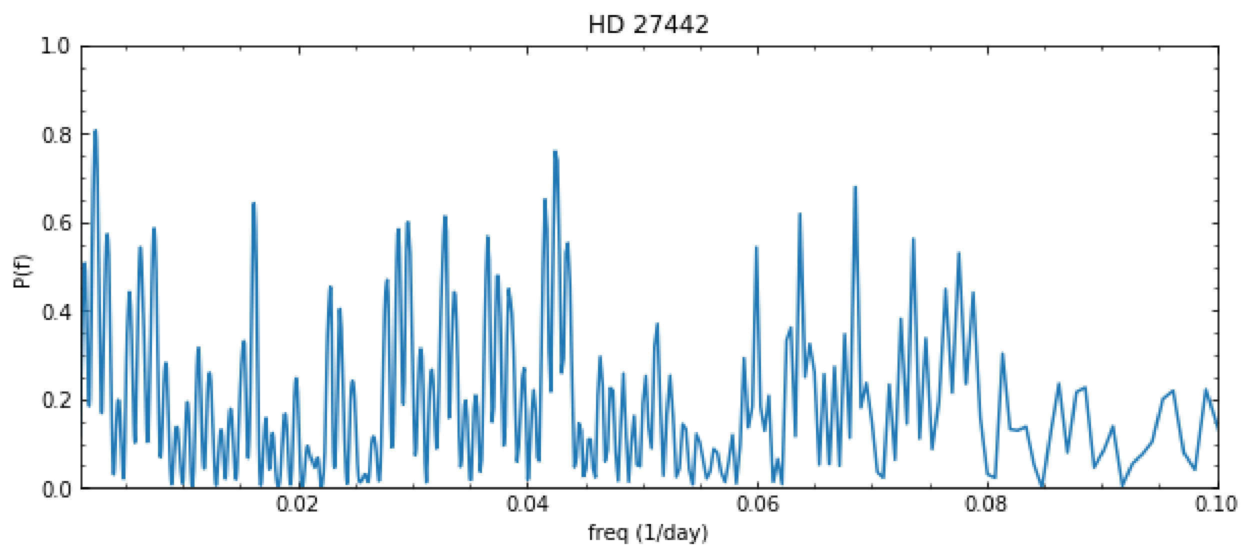

Figure 5: Normalized Lomb-Scargle periodogram showing probabilities of different frequencies for HD 22742. There are many peaks in the data, with several almost the same power. However the largest is found at 0.002.

took 200 as a maximum period. A higher resolution was required for this star. When the same resolution was used as in the previous two cases, the folded periodogram did not show any clear pattern. When the resolution was increased the new results improved the folded periodogram. The results show that the most likely period is 21 days with frequency 0.05. The plot of the folded data shows that the amplitude is 15 m/s.

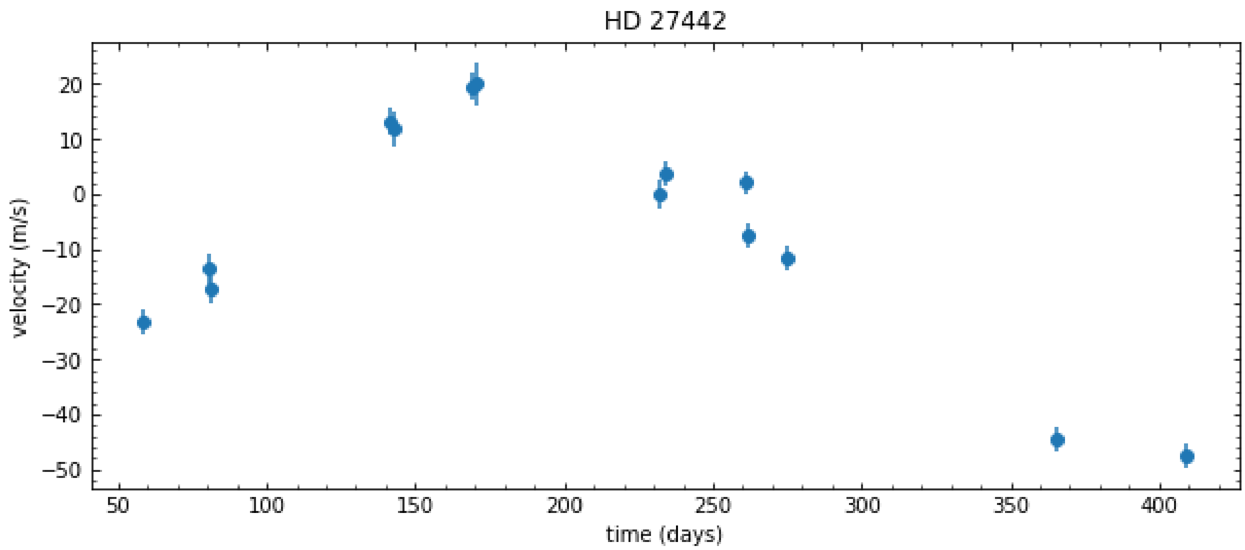

Figure 6: Folded radial-velocity measurements for HD 22742 generated from period found with the Lomb-Scargle periodogram. The amplitude is about 35 m/s.

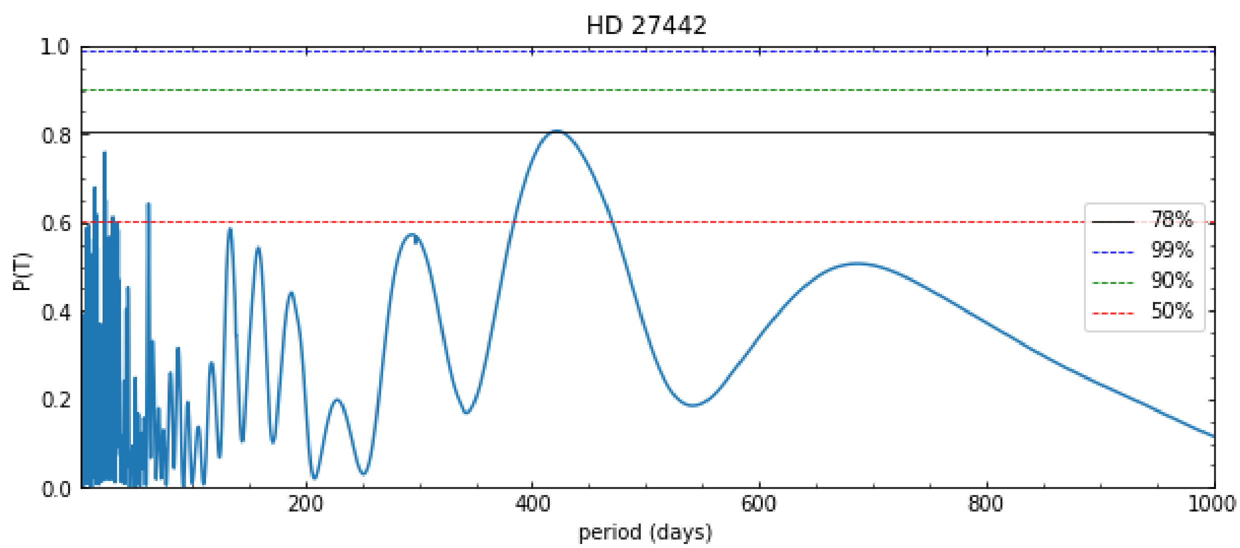

Figure 7: Lomb-Scargle periodogram for HD 27442 with the probability significance levels shown.

4 Discussion

	HD 142	HD 27442	HD 102117
period	332 days	423 days	21 days
amplitude	25 m/s	35 m/s	15 m/s
significance	89%	78%	91%

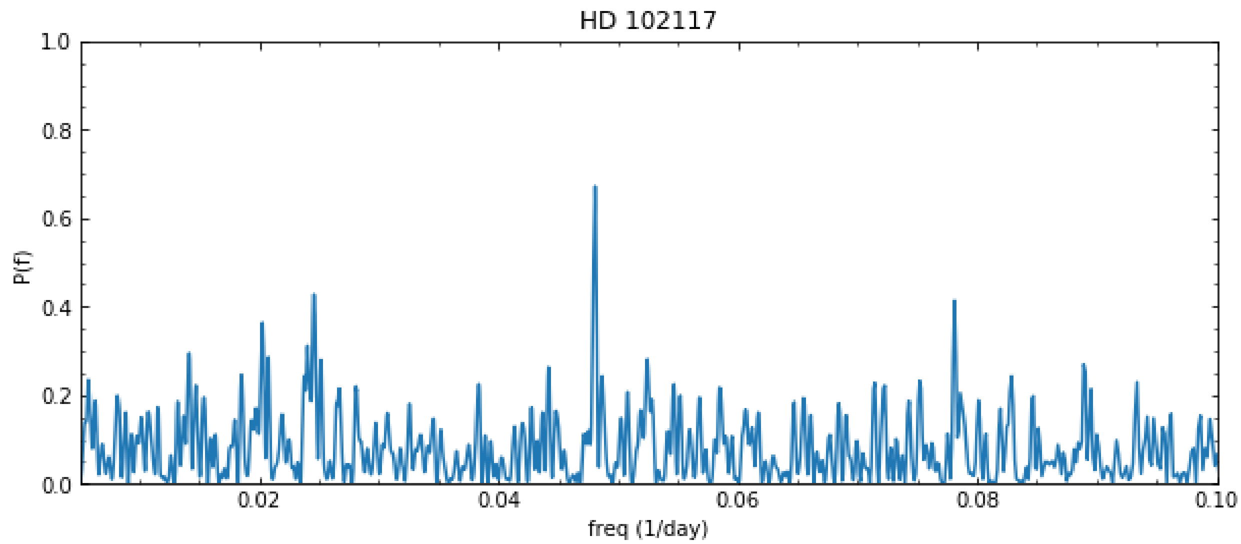

Figure 8: Normalized Lomb-Scargle periodogram showing probabilities of different frequencies for HD 102117. There is a large peak at 0.048.

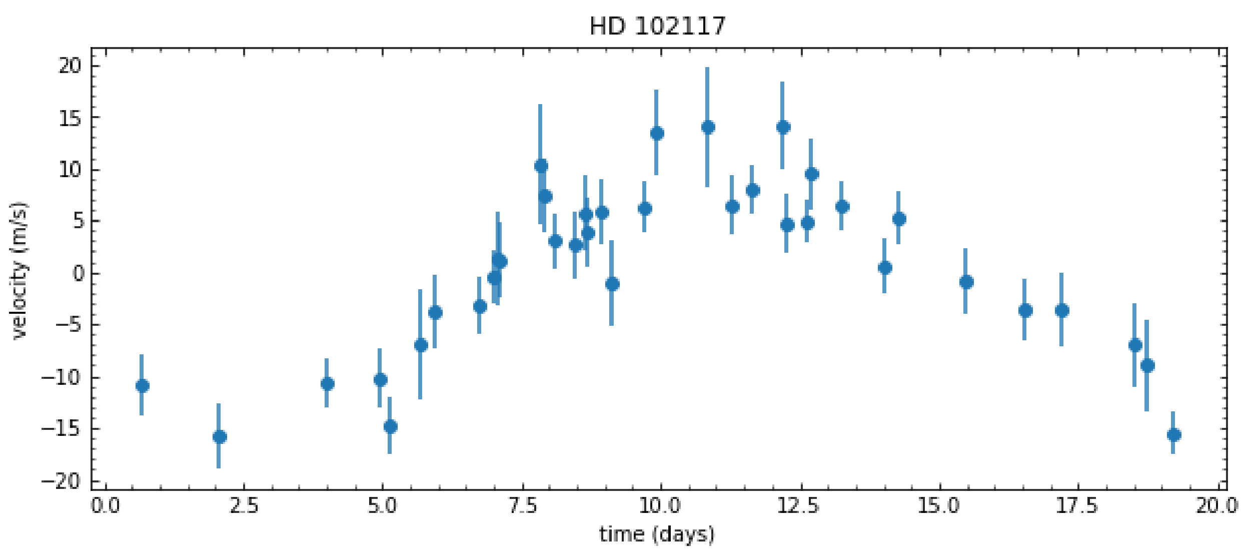

Figure 9: Folded radial-velocity measurements for HD 102117 generated from period found with the Lomb-Scargle periodogram. The amplitude is about 15 m/s.

Table 1. Summary of the most likely period for each star, and the significance of the period power.

The folded plots for each star show that the periodograms appear to be successful. In each plot it is possible to see the pattern of a sine wave. The values found for the periods also match the actual values in the literature. The significance of the power of the peak for

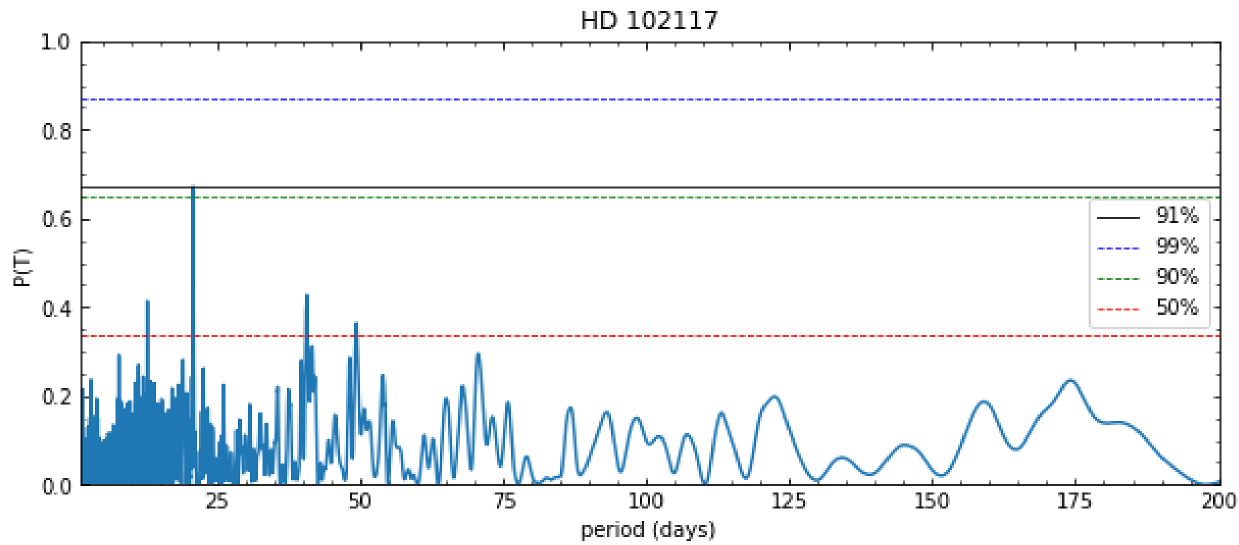

Figure 10: Lomb-Scargle periodogram for HD 102117 with the probability significance levels shown.

HD 142 is 89%. This is high, but it can be seen in the plot of the significance that another peak at around 170 days reaches almost as high. The significance plot for HD 27442 shows a lower significance for the power of the highest peak there. It reaches only 78%, with another peak at around 10 days almost reaching almost as high, and several other peaks reaching a significance over 50%. For HD 102117 the results are slightly clearer, with the highest peak reaching a significance of 91%, and no other peak being very close to this one.

The significance of the power of the true period in each of these cases may not be very surprising as the data were not complete, as can be seen in the plots of the original data. Each of the periodograms show many peaks, indicating that the function found many potential false signals. This is particularly the case for HD 27442, with the second highest peak reaching almost as high as the highest. A different resolution could possibly have found this second highest power as the highest, and so return an incorrect value for the period of the orbit. The results show the power of the Lomb-Scargle periodogram. It would be very difficult to find a period in the data just from the original plots without this sort of analysis, but with the Lomb-Scargle periodogram the periods can be found from quite incomplete data.

However there are problems with this method of finding the periods of unevenly sampled data. It seems as though it would be easy to find a period that was not actually in the data as several of the peaks in the results for the planets are almost as high as the true period. Also, the planets that are being measured here have periods similar to earths, but many planets have periods that are much larger that this and that would make it more difficult to measure their periods with this method as not enough data could be recorded in a convenient time. Any planet discovered using the radial velocity method by necessity has a relatively small period and is close to its host star.