Calculation of Kepler-75b's Radius and Orbital Period Using Kepler Data

Gianfranco Grillo

May 24th, 2017

Abstract

The Kepler spacecraft, launched on March 7, 2009, has been responsible for the detection of thousands of exoplanets by monitoring the brightness of stars, thus constructing light curves that can be analyzed to search for regular dips in brightness due to planet transits. The transit method can be used to determine an exoplanet's orbital period and, if the corresponding star's radius is known, it's radius. Here we aim to use a combination of short and long cadence *Kepler* data for the star Kepler-75 in order to determine the orbital period and planetary radius of Kepler-75b. The approach relies mainly on the implementation of Lomb-Scargle whitening techniques and matched filtering. We find values for the parameters in agreement with those obtained by others using more sophisticated procedures. Using the long cadence data, the radius was calculated as $1.11 \pm 0.05\, R_J$ and the orbital period as 8.885 ± 0.001 days, whereas using the short cadence data, the radius was estimated as$1.08 \pm 0.02\, R_J$ and the orbital period as 8.884 ± 0.001 days.

1 Introduction

The detection of the first exoplanet in 1992 opened up a new field in astronomy that has grown ever since, aided by the tremendous advances made with regards to observing power and computational processing power. As of this writing, the astronomical community has detected 3,611 exoplanets in 2,705 planetary systems, with many more awaiting confirmation. Key to this success has been the deployment of space telescopes, starting with the *Hubble* Space telescope in 1990, and continuing with the *CoRoT* space observatory in 2006, and the *Kepler* spacecraft in 2009. The latter has been responsible for the discovery of more than a thousand exoplanets, via the continued study of more than 156,000 stars with apparent magnitudes ranging from 9 to 16. Its observations have not only been able to determine the existence of these exoplanets, but also have allowed us to study some of their characteristics, especially when the transit observations can be combined with radial velocity observations from other telescopes such as SOPHIE.

This is precisely what Hebrard et al. (2013), and Bonomo et al. (2014) have done with exoplanet Kepler-75b, after its discovery by Batalha et al. in 2013. This provides a good opportunity to develop and implement whitening and matched filtering algorithms and check their performance, as the results of the analysis can be compared with previous results. The parameters found in these earlier studies are shown in Table 1, and the characteristics of the planet's host star Kepler-75 are reported in Table 2. The idea of this work is to find out how well these results can be replicated using more primitive data reduction techniques than those used in more sophisticated studies such as the ones mentioned.

This paper is divided as follows. Section 2 describes the general observational history of the target star, in order to provide a better characterization of the data that was analyzed. Section 3 describes

Table 1: Planetary parameters from previous analyses

	Hebrard et al. (2013)	Bonomo et al. (2014)
Orbital period (days)	8.884924 ± 0.000002	8.8849116 ± 0.0000034
Planetary radius (R_J)	1.03 ± 0.06	1.05 ± 0.03

Table 2: Parameters for host star Kepler-75 (from Hebrard et al. (2013))

Kepler-75	
Kepler ID	757450
RA (J2000)	19:24:33.0
Dec (J2000)	+36:34:39
Spectral type	G8V
Effective temperature (K)	5330.0 ($\pm$120.0)
Distance (pc)	1140.0 ($_{-160.0}^{+250.0}$)
Mass (M_{sun})	0.88 ($\pm$0.06)
Radius (R_{sun})	0.88 ($\pm$0.04)

the methods used for data reduction, starting with the analysis performed using the long cadence data, followed by a description of the approach used to reduce the short cadence data. We present and analyze the results in Section 4, and develop general conclusions in Section 5.

2 Observations

Kepler-75 was observed by the *Kepler* spacecraft starting with the launch of the mission in May 2009, and ending four years later in May 2013. The publicly available data, hosted at the Mikulski Archive for Space Telescopes (MAST) consists of 17 quarters of long cadence data, each spanning approximately 30 days, with a sampling interval of ~ 30 minutes, and 18 quarters of short cadence data, each also spanning about 30 days, but with a sampling interval of ~ 1 minute. Both the long cadence data and the short cadence data are in the form of two sets of time series: one with raw simple aperture photometry (SAP) and another with preprocessed data intended to be used for data validation (DV). The analysis performed in this work used the long cadence DV time series and the short cadence SAP time series. The long cadence DV time series was used to estimate the parameters in a very simple, crude manner, whereas the short cadence SAP time series was subjected to a more sophisticated analysis involving whitening through an iterative Lomb-Scargle approach, and localization of the planetary transits via matched filtering.

3 Data reduction

3.1 Long cadence data

Figure 1 shows a scatter plot of the complete long cadence light curve, which was previously quarter-stitched and normalized by the *Kepler* science team. The transits are easy to see. The procedure used to analyze this data went as follows. The period was calculated by simply finding the location of the normalized flux minima, finding the time differences between each of these minima, and averaging over

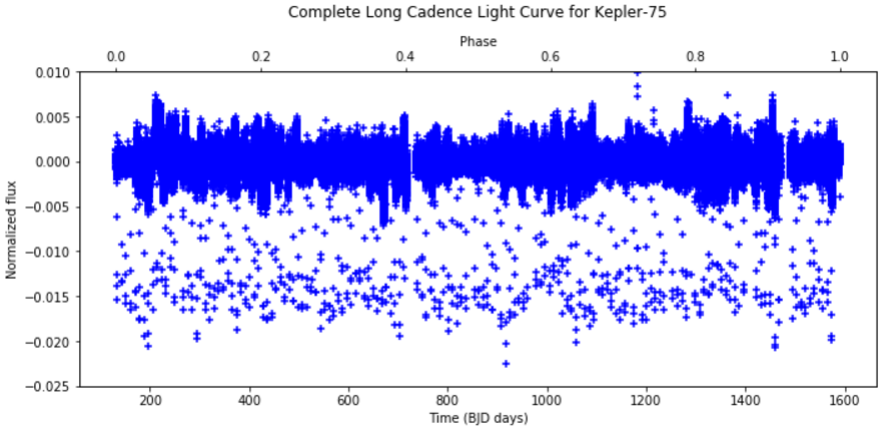

Figure 1: Complete long cadence light curve

all the values. An estimate of the uncertainty Δt was made by taking the standard deviation of the time differences and dividing by the square root of the number of time differences,

$$\Delta t = \sqrt{\frac{\sum_{i=0}^{N-1}(t_i - \bar{t})^2}{N(N-1)}} \tag{1}$$

3.2 Short cadence data

The raw SAP data is of course much harder to analyze properly. As can be seen in Figures 2-5, each of the quarters contains underlying trend lines that essentially rule out the possibility of directly implementing matched filtering to locate the transits, so it is necessary to first remove these trends if such a procedure is to be carried out. There were a few other issues that were dealt with first. One, it became apparent that there are some erroneous data points, as they lied very far away from their neighbors; this was fixed by using an algorithm that calculated the flux difference between one data point and the next, and deleted those data points that were more than 100 units of flux away from the preceding one (this was done before producing the light curves shown, which is why these points cannot be seen). Two, the flux intensity clearly differs from one quarter to the next, so the flux was normalized on a per quarter basis. Finally, there are some plots like the one shown in Figure 5 that have some kind of anomaly that needs more work to be properly fixed. These plots were simply removed from the analysis. Specifically, the ninth observation quarter (Q9) and the seventeenth observation quarter (Q17) were removed from the data pool at this point in the reduction process.

The removal of those outlying data points, together with the fact that some data points were simply missing for certain observation times, implies that performing a straightforward Fourier spectral analysis of the time series is not possible: one must either fill in those missing data points via some kind of procedure, or perform some kind of spectral analysis that applies to data sampled at uneven intervals. In this work, we follow the second approach and use a Lomb-Scargle algorithm to calculate the power spectrum.

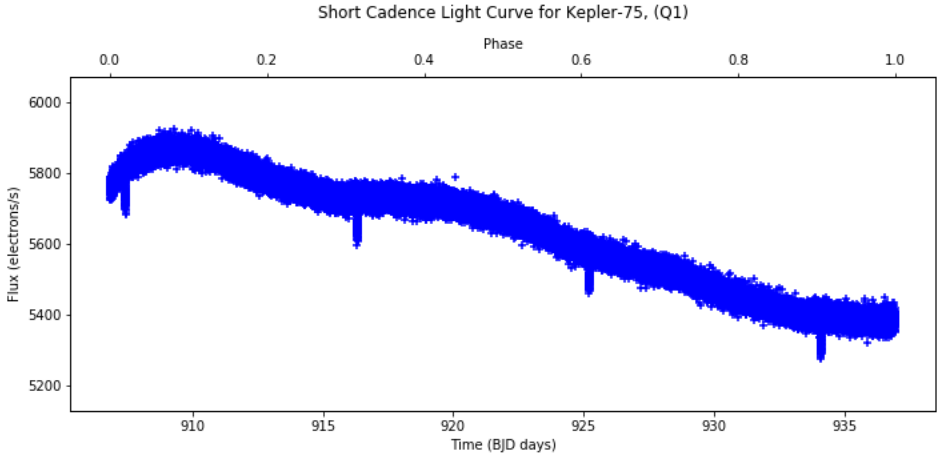

Figure 2: SC light curve for Q1

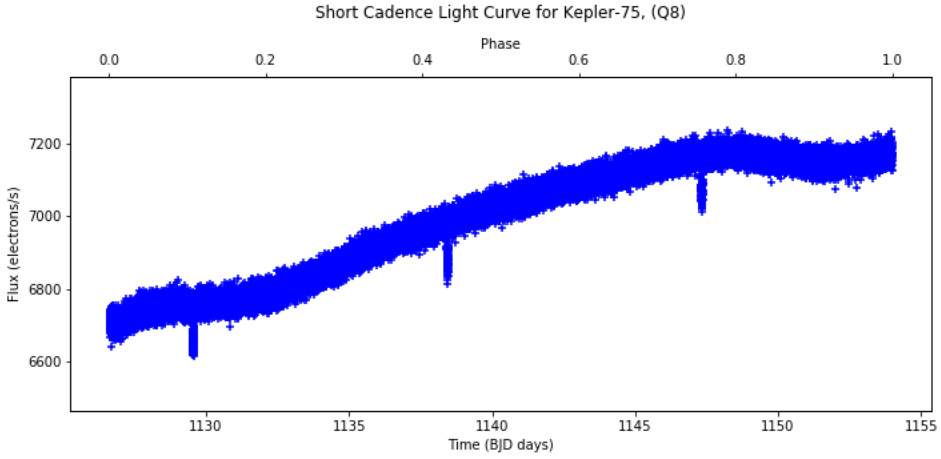

Figure 3: SC light curve for Q8

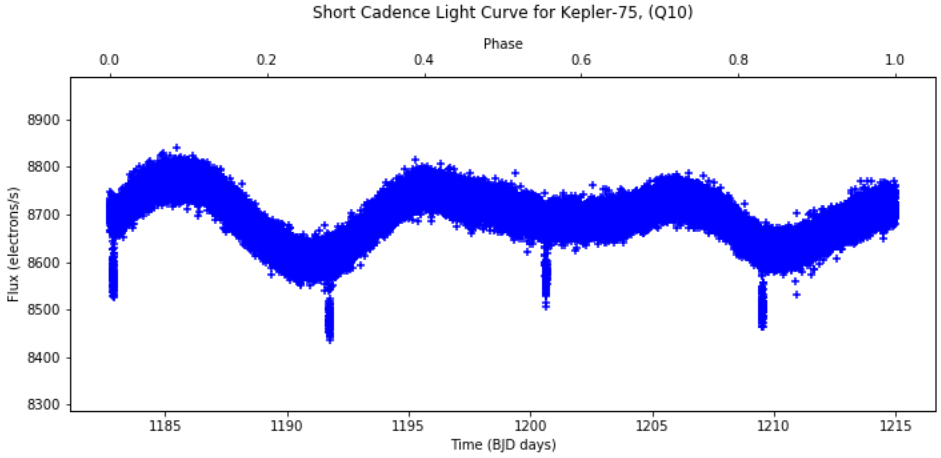

Figure 4: SC light curve for Q10

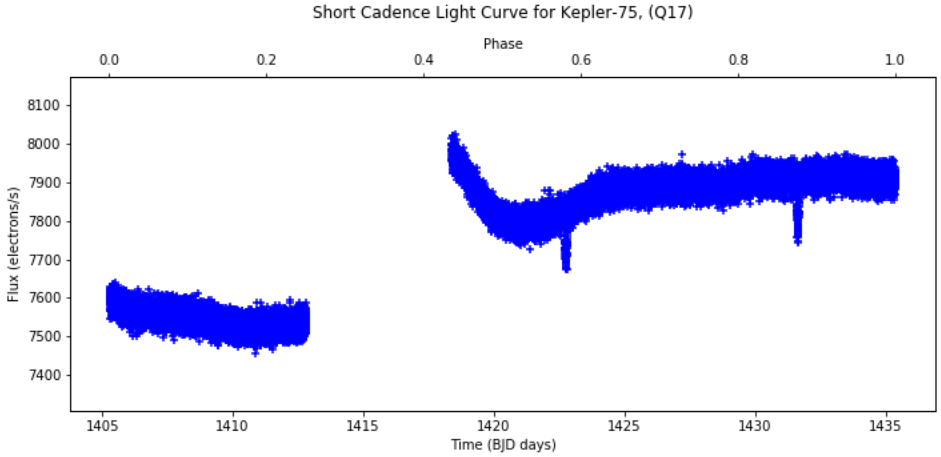

Figure 5: SC light curve for Q17

3.2.1 Lomb-Scargle Analysis

What follows is a brief overview of the mathematics underlying the Lomb-Scargle analysis of a time series. As mentioned above, Lomb-Scargle spectral analysis differs from the more common Fourier spectral analysis because it can be applied to unevenly spaced data points. The method evaluates data to calculate the spectrum only at the times t_i at which the points are actually measured. More technically, given N data points $h_i \equiv h(t_i)$, $i = 0, \ldots, N-1$, the Lomb-Scargle algorithm first finds the mean and the variance of the data via the usual definitions,

$$\bar{h} \equiv \frac{1}{N} \sum_{i=0}^{N-1} h_i \tag{2}$$

$$\sigma^2 \equiv \frac{1}{N-1} \sum_{i=0}^{N-1} (h_i - \bar{h})^2 \tag{3}$$

And proceeds to calculate a normalized periodogram (spectral power as a function of angular frequency ω), which is defined by

$$P_N(\omega) = \frac{1}{2\sigma^2} \left\{ \frac{\left[\sum_j (h_j - \bar{h}) \cos(\omega(t_j - \tau))\right]^2}{\sum_j \cos^2(\omega(t_j - \tau))} + \frac{\left[\sum_j (h_j - \bar{h}) \sin(\omega(t_j - \tau))\right]^2}{\sum_j \sin^2(\omega(t_j - \tau))} \right\} \tag{4}$$

Where τ is given by

$$\tan(2\omega\tau) = \frac{\sum_j \sin(2\omega t_j)}{\sum_j \cos(2\omega t_j)} \tag{5}$$

This quantity, τ, acts as an offset that effectively renders $P_N(\omega)$ independent to a constant shift in all the t_j's, and implies that (4) is identical to the equation governing the least-squares fitting to a sinusoidal model ie. one described by the equation

$$h(t) = A\cos(\omega t) + B\sin(\omega t) \tag{6}$$

The Lomb-Scargle approach is convenient when one does not want to artificially close the gaps in the available data, but implementations of the algorithm are significantly slower than the Fast Fourier Transform.

3.2.2 Data whitening

The data whitening procedure was performed by iteratively implementing the Lomb-Scargle algorithm, substracting the calculated fit over each iteration, as shown in Figures 6-7 for the case of quarters 1 and 8. The initial Lomb-Scargle periodograms for the corresponding quarters shown in Figures 6-7 are displayed in Figures 8-9. The fit was performed by calculating the Lomb-Scargle periodogram, finding the peak of the periodogram, and using the location of the peak to establish the frequency of the fitting sinusoid, which can be inserted into the algorithm in order to find the best fitting values for parameters A and B from equation (6). The number of iterations was adjusted manually; after a certain point, to continue iterating would result in dewhitening.

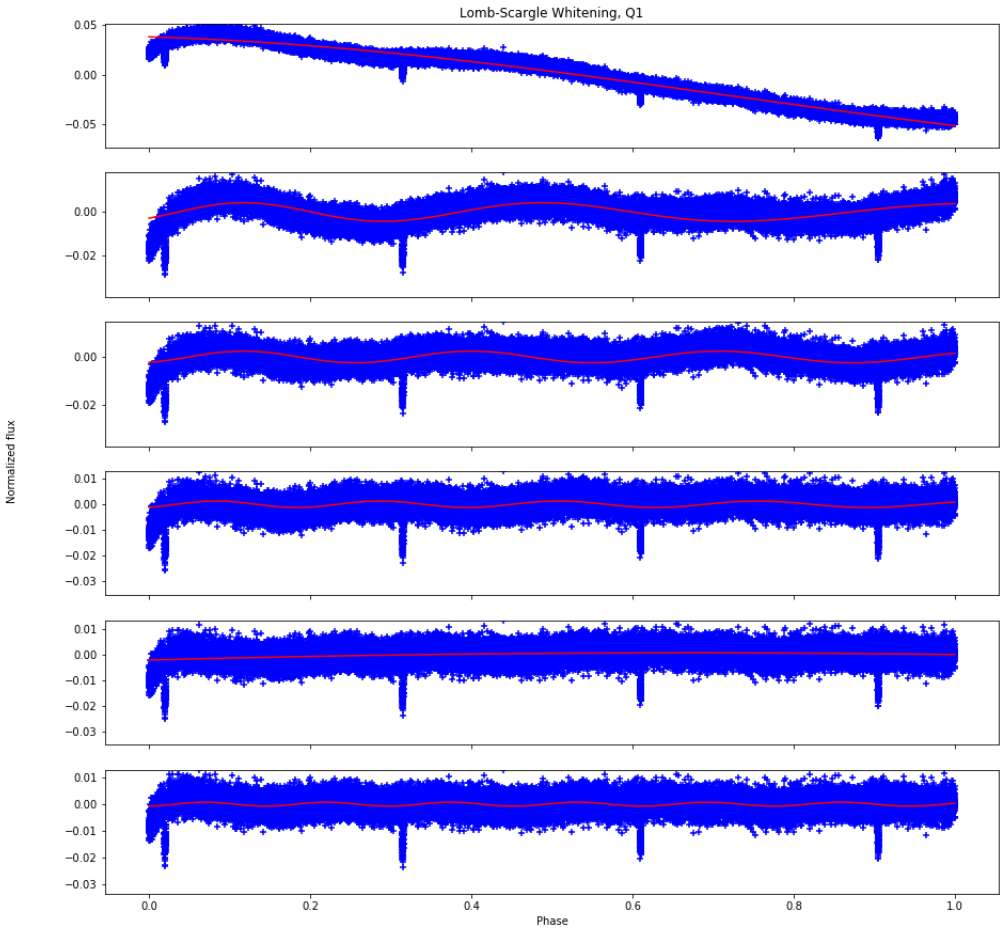

Figure 6: Whitening procedure for Q1

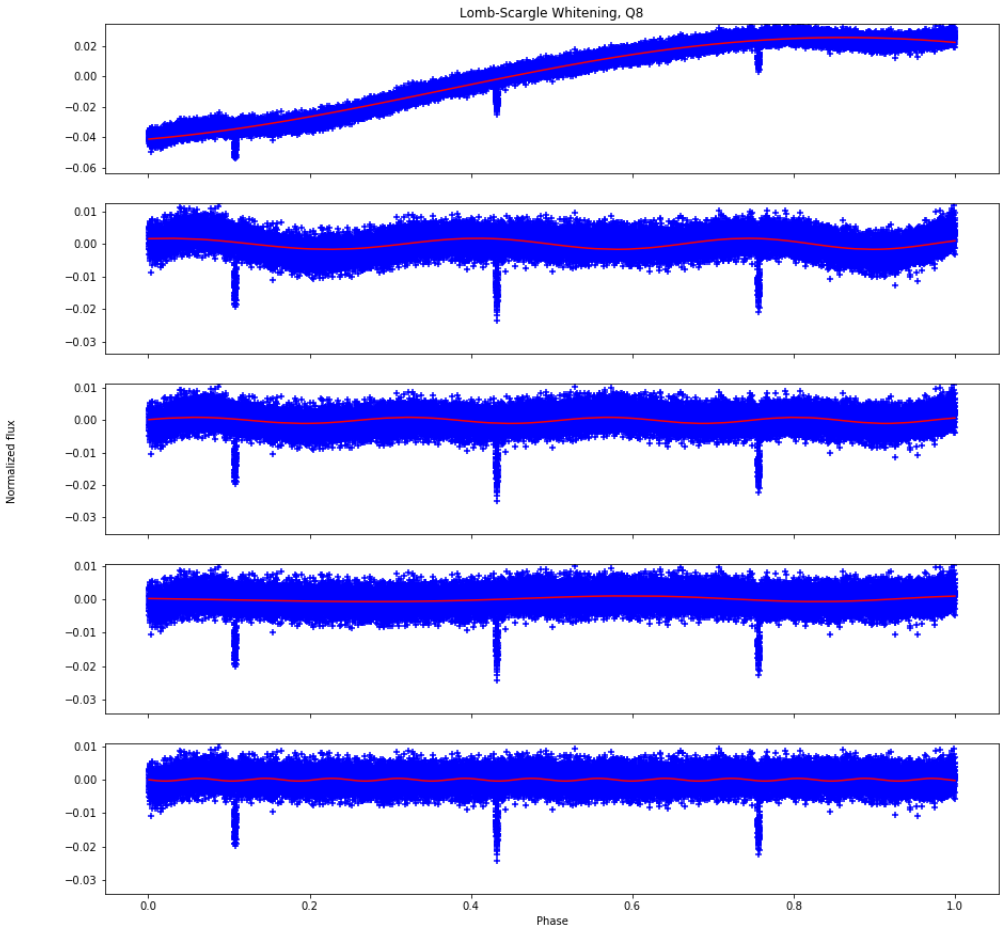

Figure 7: Whitening procedure for Q8

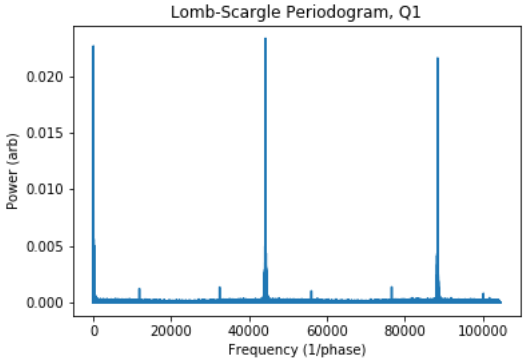

Figure 8: Periodogram for Q1

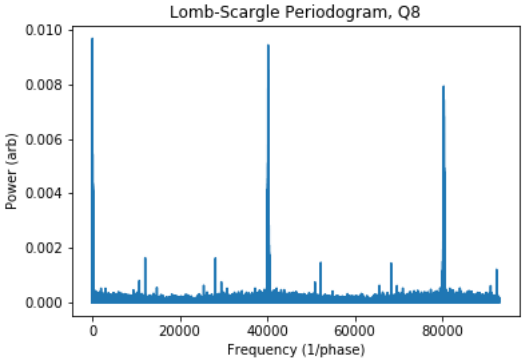

Figure 9: Periodogram for Q8

3.2.3 Matched filtering

Matched filtering involves the construction of a template, followed by cross-correlating that template with the data in order to localize the points in which the data most resembles the template. In this work, we use an inverted Gaussian as our model. This is an imperfect way of characterizing a planetary transit, since transit curves are not exactly Gaussian, but it is good enough as an approximation. The template was constructed as follows. An estimate of the location of the transit points was first found by finding the quarter's minima, similar to what was done for the case of the long cadence data. Then the width of the transit was estimated by searching for the neighboring points whose magnitude was greater than half of the magnitude at the corresponding minimum. The widths found across the quarter were then averaged, with that average used as the FWHM for the Gaussian. The rough amplitudes of the dips, given by the value of the minima, were averaged and used as the template amplitude.

Figures 10 and 11 show the results of performing the matched filtering procedure on Q1 and Q5. Notice that the procedure works well for Q5, with the peaks of the CCF corresponding to the transit points, but for Q1 the procedure does not work so well, because the whitening algorithm used was unable to remove all trends, and as a result the CCF shows an extra peak at the beginning that does not correspond to a transit. This could potentially be fixed by dividing the original quarter data into smaller pieces and performing the Lomb-Scargle whitening separately on them, but this was not attempted here, and instead we discarded problematic quarters like Q1. In the end, we perform the analysis using only 9 of the original 18 quarters.

Using the updated values for the transit points, a better estimate of the transit amplitude was found by averaging over flux values neighboring the transit center. This was followed by recalculating the transit width using the updated value of the amplitude. Figures 12 and 13 show the resulting transit fittings for quarters 5 and 10.

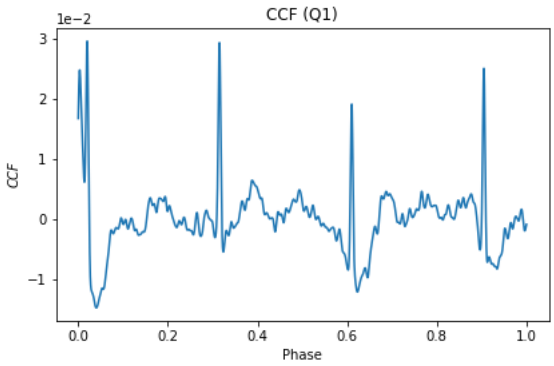

Figure 10: CCF for Q1

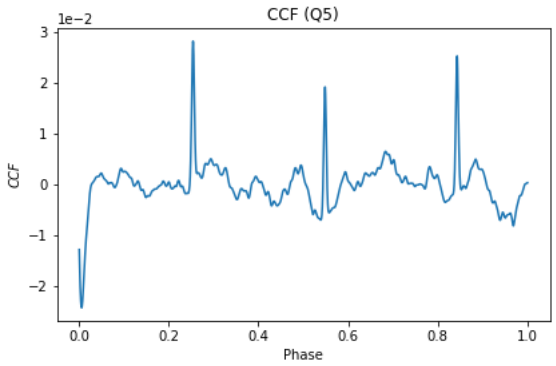

Figure 11: CCF for Q5

The orbital period was then calculated by looking at the difference between transit times within each individual quarter, with the uncertainty calculated in the same way as for the long cadence data, using equation (1). The planetary radius was calculated using the average value of the dip amplitudes A_{av} across each quarter, using the relationship

$$R = R_s\sqrt{A_{av}} \tag{7}$$

Where

$$R_s = (0.88 \pm 0.04)R_{sun} = (0.88 \pm 0.04) \times 9.951 R_J = (8.8 \pm 0.4)R_J \tag{8}$$

The uncertainty in R was calculated by propagating the initial error in the flux as given by the original data.

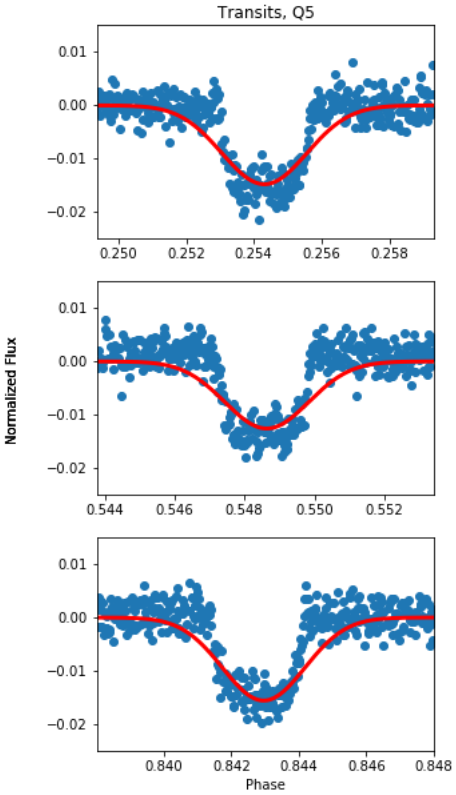

Figure 12: Fitted transit points for Q5

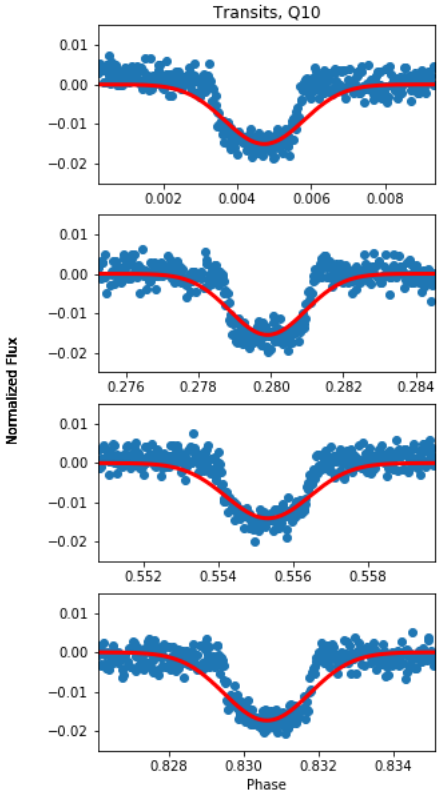

Figure 13: Fitted transit points for Q10

4 Results

Table 3: Parameters for Kepler-75b

Work	Orbital period (days)	Planetary radius (R_J)
Hebrard et al. (2013)	8.884924 ± 0.000002	1.03 ± 0.06
Bonomo et al. (2015)	8.8849116 ± 0.0000034	1.05 ± 0.03
This work (LC data)	8.885 ± 0.001	1.11 ± 0.05
This work (SC data)	8.884 ± 0.001	1.08 ± 0.02

Table 3 shows the results of the analysis. There is excellent agreement between the calculated orbital period using both the long cadence and short cadence data with the results of previous analyses,

although the uncertainty obtained is larger. This makes sense given the crude method for calculating the orbital period using the LC data and the fact that only half of the SC data was ultimately used. The values for the planetary radius obtained are also in agreement with previous works, although the nominal values found are slightly larger. Still, the agreement is quite good considering that a simple Gaussian template was used to model the transits, and the calculation of the amplitude was performed in a relatively simple manner. Improved agreement would probably we reached by using a more sophisticated data reduction procedure with whitening techniques that can be applied to all the data quarters, a more accurate transit template, and a more sophisticated way of finding the dip amplitude.

5 Conclusion

The purpose of this work was to apply whitening and matched filtering techniques to *Kepler* data in order to calculate the orbital period and planetary radius of Kepler-75b. Even though the analysis performed was not very sophisticated in comparison to other studies of the same object, the results obtained were similar, and show that reasonable estimates for exoplanet parameters from transit data can be obtained relatively easily. Furthermore, it is reasonable to think that even better results can be obtained with relatively small changes in the data reduction methods.

References

[1] Batalha, N. M., Rowe, J. F., Bryson, S. T., et al. *Planetary Candidates Observed by Kepler. III. Analysis of the First 16 Months of Data*, 2013, ApJS, 204, 24.

[2] Hebrard, G., Almenara, J.-M., Santerne, A., et al. *KOI-200 b and KOI-889 b: Two transiting exoplanets detected and characterized with Kepler, SOPHIE, and HARPS-N*, 2013, A&A, 554, A114.

[3] Bonomo, A. S., Sozzetti, A., Santerne, A., et al. *Improved parameters of seven Kepler giant companions characterized with SOPHIE and HARPS-N*, 2015, A&A, 575, A85.

[4] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery *Numerical Recipes, The Art of Scientific Computing, 3rd edit.*, Camb. Univ. Press 2007.