IOP Conf. Series: Journal of Physics: Conf. Series **869** (2017) 012074 doi:10.1088/1742-6596/869/1/012074

An optimized search for exoplanets with Kepler data

K Mohamed[1,2], A Ibrahim[2,3] and N Abd-Allah[1]

[1] Sohag University, Physics Department, Sohag, Egypt
[2] Zewail City of Science and Technology, Physics of Earth and Universe, and
[3] Cairo University, Faculty of Science, Physics Department, Giza, Egypt

Email: ai@ZewailCity.edu.eg

Abstract. The search and detection of extra-solar planets (exoplanets) have considerably improved over the past few decades. The launch of space observatories dedicated to exoplanets such as the Kepler mission and the wealth of publicly available data call for the development of efficient search methods. Here we present a search technique with a Graphical User Interface (GUI) to detect true exoplanets signals from Kepler data and filter out false ones using the optimal Box-fitting Least Squares (BLS) algorithm. The algorithm is used to detect transiting exoplanets from their light curves through searching for periodic flux variability of the host star. We present the results of the search to some Kepler catalog objects.

1. Introduction

The prospect of finding extra-solar planets (exoplanets) began to materialize in 1992 when Wolszczan and Frail discovered the first extra-solar planet around the neutron star radio pulsar PSR 1257+12 [1]. A Jupiter-mass planet orbiting around the main-sequence star 51 Pegasi was subsequently discovered by Mayor and Queloz in 1995 [2] and confirmed by Marcy et al in 1997 [3]. In recent years, the development of ground-based and space-borne observatories to search for exoplanets produced a surge in the number of known exoplanets, which is now in the thousands. Exoplanets can be detected from their gravitational influence on the host. This results in observable signatures such as a shift or wobbling in the star's position around the barycentre that is inferred by procession astrometry or by an observed Doppler shift in the star's spectrum due to its radial velocity. Exoplanets can also be detected when they transit the host star, causing a small but observable decline in the star's brightness when observed with sensitive photometers. The transit method is useful because of its ability to detect small size planets in the habitable zone of the host stars, which is the main objective of the Kepler mission.

The Kepler mission was launched to space in 2009 to observe simultaneously the brightness of more than 200,000 stars in a fixed field of view. By the completion of the mission, Kepler aims to assess the likelihood of finding terrestrial and Earth-size planets in the habitable zone of Sun-like stars, allowing to probe the diversity of the exoplanet population. The Kepler mission uses a 0.95-meter diameter telescope and a single instrument in the form of precession photometer consisting of 42 CCDs each is 50 × 25 mm with a resolution of 2200 × 1024 pixels, giving a total resolution of 94.6 megapixels. It is designed to observe a field of view of 115-square-degree of the sky.

The objective of this work is to introduce an efficient tool to search for exoplanets in the Kepler data with enhanced accuracy and reduction in the processing time. The analysis of Kepler data includes several automated stages of processing that ends with a data validation process. A manual/human stage screens thousands of computer-detected signals of candidate planets to identify true planetary signatures.

This cannot be done automatically and hence we built a search method using the Box-fitting Least Squares (BLS) that processes the data and feeds the resultant output to a Graphical User Interface (GUI) to identify the transit signals faster and more accurate. The GUI represents an efficient interactive interface that bridges the output of the search algorithm and the data validation through manual/human screening.

2. The search algorithm

The Box-fitting Least Squares (BLS) is an algorithm that is used to detect transiting exoplanets from the light curves of their host stars. The algorithm searches for transiting planet signatures (small, periodic drops in the star brightness). The principle of the BLS algorithm is based on direct Least Squares (LS) fits of step functions of the folded signal corresponding to trial periods [4]. The step function is defined as two states function High (H) and Low (L) with a period P_o. Applying that on transiting planets with fractional transit length q, the transit duration is qP_o, where q is assumed to be a small number (~ 0.01 - 0.05). For a given set of data points, the algorithm aims to find the best model while estimating five parameters: P_o, q, L, H and t_o, the epoch of the transit. Let us denote the data set by x_i, $i = 1, 2, ..., n$, and σ_i be the standard deviation. The noise is presented by assigning to each data point a weight w_i, which is defined by $w_i = \sigma_i^{-2} \sum_{j=1}^{n} \sigma_j^{-2}$. For a given trial period, we consider a folded time series, which is a permutation of the original time series. This series is denoted by $\overline{x_i}$ and the corresponding weights by $\overline{w_i}$. A step function is then fitted to the folded time series with the following parameters: L, the flux level in $[i_1, i_2]$; H, the flux level in $[1, i_1]$ and $[i_2, n]$. For a given window (i_1, i_2), the averaged squared deviation of the fit becomes

$$D = \sum_{i=1}^{n} \bar{x}_i^2 \bar{\omega}_i - \frac{s^2}{r(1-r)}, \tag{1}$$

where

$$r = \sum_{i=i_1}^{i=i_2} \bar{\omega}_i \text{ and } s = \sum_{i=i_1}^{i=i_2} \bar{\omega}_i \, \bar{x}_i. \tag{2}$$

Finally, the BLS frequency spectrum can be defined by the value of signal residue (SR) of the time series at any given trial period

$$SR = MAX \left[\frac{s^2(i_1,i_2)}{r(i_1,i_2)[1-r(i_1,i_2)]} \right]^{1/2}, \tag{3}$$

where SR represents the significance of the chosen trial period.

In practical computation, the folded time series is divided into a number of bins, and the significance ratio (SR) is evaluated for each of these binned values. We followed an approach to optimize the efficiency of the BLS algorithm through defining the number of bins in advance [5]. By using this approach, the efficiency of the computation is enhanced and the number of bins will define the time resolution of the evaluated LS solution. The automated BLS algorithm can show false signals. A false alarm is a signal with a planet-like signature that is not a true planet. Most of variable stars and eclipsing binaries are counted as planets because they show periodic change of the brightness in their light curves. Most of the false alarms are eclipsing binaries, thus the output of the BLS is screened manually using the high resolution light curves to filter out such sources.

For binary star systems, the drops in the brightness are not identical, which helps to distinguish binaries from planetary systems. We use the odd/even test outlined in [6] to exclude eclipsing binaries with planet-like signatures. The first and third drops in a light curve are called odds; while, second and fourth drops called evens, and so on. The principle of this test is to compare the odd and even drops in the phase folded light curves. If the light curve of a true exoplanet is folded with the exact period, the odds and the evens will be folded on each other and will be identical.

3. Results

Here we demonstrate the search method on Kepler data. The interface will first identify whether the object represents a true exoplanet candidate or not. The GUI (shown in figure 1) will then present the results for the manual/human screening. The left side of the GUI has the CONTROL PANEL which contains three interactive panels: OBJECTS CONTROL, PERIOD CONTROL, and TYPE CONTROL.

On the right side of the GUI, the INTERACTIVE PLOTS contain eight PLOT PANELS (7 to the right and 1 to the left). They are provided with an option of ZOOM IN/OUT.

The procedure for the search method includes: (a) identifying the object by searching for true planet signals which are characterized by periodic drop in the flux, (b) performing additional tests to distinguish between true signals and false alarms such as (i) using the phased light curve (PLOT #2) and the model (PLOT #3) as a first indication of a true planet signal, (ii) the (odd/even TEST) (PLOT #5), (iii) when the user suspects an overestimation or underestimation of the BLS period, it is preferable to investigate the BLS frequency spectrum (PLOT #1) with trial periods when the BLS period is not exact, (iv) compare the time series (in Julian dates) (PLOT #4) with the Centroids (PLOT #7 and #8), which shows if the drop in the flux is related to the position of the object on the CCD, (v) by examining (PLOT #4) (IN transit) one can assesses if "transits" are systematic effects of bad filtering of the data, and finally (vi) by comparing the (IN transit) points in (PLOT #4) with the model (PLOT #3) and phase folded light curve (PLOT #2).

Figure 1 shows an example with the Kepler data of a confirmed planet with Kepler Object of Interest and ID identification: KOI = 17.01 and KIC = 10874614.01. The transit planet signal can easily be identified from (PLOT #2) and (PLOT #3), and the (odd/even TEST) (PLOT #5) shows that the BLS period is exactly estimated. For further investigations, we can compare the centroids (PLOT #7 and #8) with the (IN transit) points (PLOT #4). Also, PLOT #1 and #6 show a very short period planet orbiting a star close to an F-type.

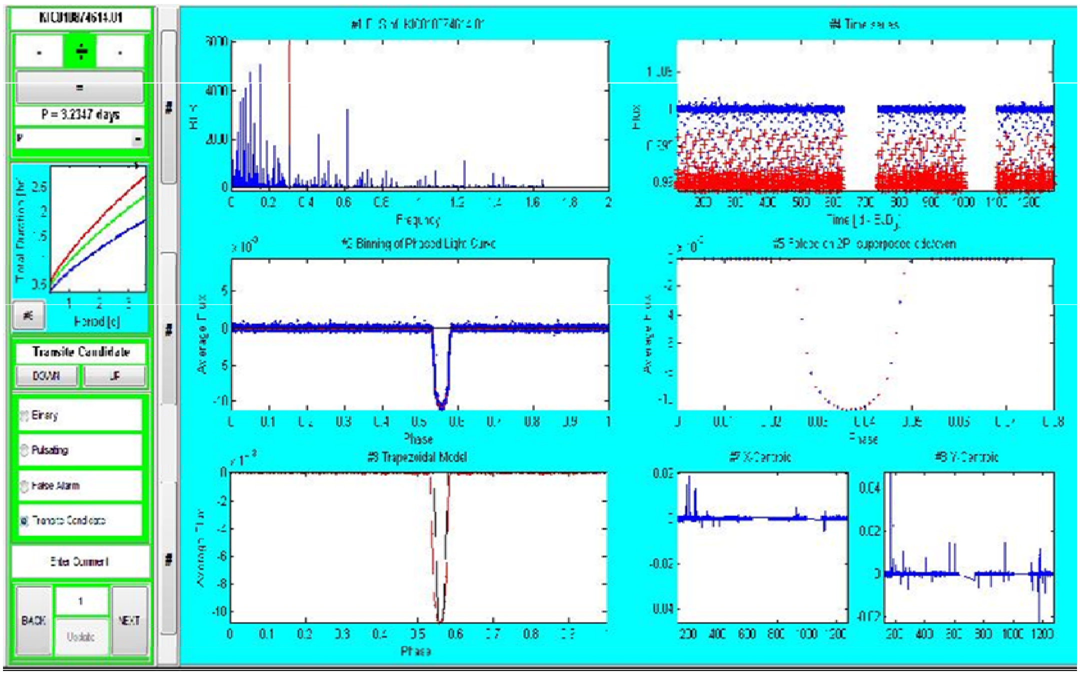

Figure 1. The output of the search method in the GUI for Kepler Catalog Object KIC= 10874614.01.

In figure 2, we show another example from the Kepler catalog of Kepler Object of Interest KOI = 509.01 and KIC = 6381846.01. The planet signal is again clearly identified from PLOT #2 and PLOT #3. The odd/even TEST (PLOT #5) shows that the BLS period is exactly estimated and the odds/evens are the same in these plots. For additional investigations, we can compare the centroids (PLOT #7 and #8) with the (IN transit) points (PLOT #4). PLOT #1 and PLOT #6 show that this is a short period planet orbiting a star close to a K-type.

IOP Conf. Series: Journal of Physics: Conf. Series **869** (2017) 012074 doi:10.1088/1742-6596/869/1/012074

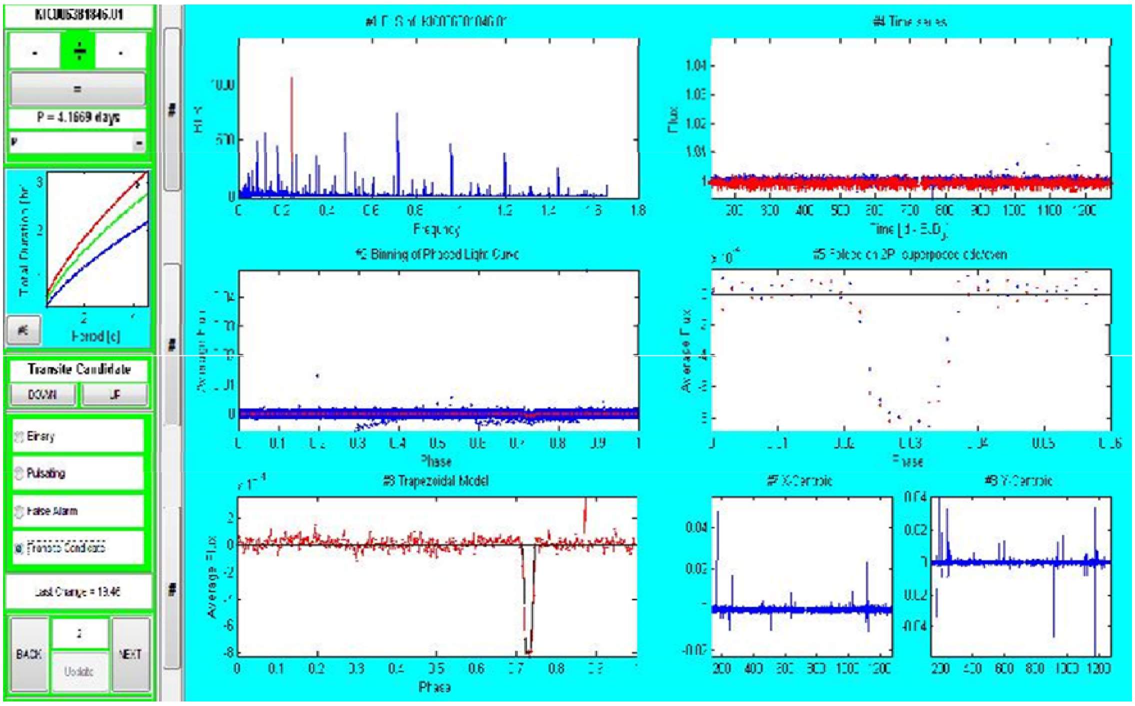

Figure 2. The output of the search method in the GUI for Kepler Catalog Object KIC= 6381846.01.

4. Conclusion

We have successfully built an exoplanet search tool for Kepler data and tested its viability in identifying true exoplanet signals through a versatile Graphical User Interface that allows for additional and customized investigations. The tool showed consistent results with ten known objects from the Kepler catalog and we showed here the results of the first two here for space limitation. Upgrading this tool to utilized large data is currently underway.

References

[1] Wolszczan A and Frail D 1992 *Nature* **355** 145
[2] Mayor M and Queloz D 1995 *Nature* **378** 355
[3] Marcy G, Butler R, Williams E, Bildsten L, Graham J, Ghez A and Jernigan J 1997 *ApJ* **481** 926
[4] Kovács G, Zucker S and Mazeh T 2002 *A&A* **391** 369
[5] Ofir A 2014 *A&A*, **561** A138
[6] Tenenbaum P, Bryson S, Chandrasekaran H, Li J, Quintana E, Twicken J and Jenkins J 2010 *Proc. SPIE* **7740** 77400J