

Astronomical Image Processing: Exoplanet Detection

M. S. Prasad

Amity Institute of Space Science &
Technology,
Amity University, Noida
India
msprasad@amity.edu

Shivani Verma

Amity Institute of Space Science &
Technology,
Amity University, Noida
India
sverma2@amity.edu

Yulia A. Shichkina

Saint Petersburg Electrotechnical
University "LETI"
Saint Petersburg, Russia
shichkina@co-evolution.ai

Abstract—Time-Series data analysis for space data is flourishing exponentially in nature. This analysis depends upon the brightness of the under-detection or already detected planets, stars and galaxies. The time-series data can be analyzed through light-curves which are interactive in nature. This examination can be done for user selected stars/exoplanets accessible from the NASA Exoplanet Archive. Nasa also provides a Planetary Spectrum generator which is GUI based and can be used for the same. The NASA Exoplanet Archive is full of data for analysis of stars, exoplanets (confirmed, candidate and false positives) and galaxies, which has been collected from several missions like Kepler and TESS. The light-curve generated is then pre-processed to create a normalized flux chart and then processed to see if there is a chance of it being a transiting signal from the transiting planet or not. This is accomplished using the Box Least Squares Methods for a periodogram analysis. The Transit model and its Cadence Mask can be retrieved from this Box least squares (BLS) periodogram which can be again interactive in nature. Also, this pre-processed data can also be fed to a neural network classifier in order to determine if it is likely to be an exoplanet or not.

Keywords—Exoplanet Detection, ML Algorithms, light curves, Kepler mission, TESS Mission, NASA Exoplanet Archive

I. INTRODUCTION

Space data exploration and analysis has been increasing exponentially in today's times. Finding exoplanets by multiple methods or studying other stars and galaxies has become a topic of interest. There are several methods of exoplanet detection but the most convenient way is transit detection. [1]

In the transit method, change in intensity or the brightness of the star is calculated when it travels near the star. A telescope then records its movements over a period of time and records changes in the brightness of the star known as 'dip'. [2] Certain dip in this brightness depicts the presence of the transiting body. Further analysis of this transiting body and calculations of the comparative parameters results in the confirmation of the body being exoplanetary. Else, it is referred to as a false positive. This technique of exoplanet detection has become a handy tool for exoplanet detection and analysis using artificial intelligence and neural networks. Before we jump into conclusions regarding confirmation of any-body being an exoplanet or not a deep analysis regarding its physical and chemical specifications is accomplished and based on that it is justified.

Several space missions have been conducted in order to solve mysteries of the deep space, like the Kepler telescope [3], Transiting Exoplanet Surveying Satellite (TESS) [4], Hubble telescope and very recently James Webb Space

Telescope (JWST). All these telescopes are sent out in the outer space to see what the outer world is and what mysteries it holds. The very first step to exoplanet detection from the data available from these telescopes is the image analysis, which we call as Astronomical Image Processing which helps us in identifying whether a body is transiting or not and based on it a light curve is generated which shows the transiting nature of the body for a certain period of time. Likewise, it is not mandatory if the transition is error free or noise free, therefore, to avoid any error, the light curve thus generated is cleared with any instrumental noise and then the processing moves forward.

The paper has been segregated to five sections, where section II gives a literature survey, section III is based upon design and implementation. Section IV discusses the results and the paper then concludes with the references for the paper.

II. LITERATURE SURVEY

Exoplanet Detection is a major part of space data analysis as it helps in knowing several transiting bodies apart from the planets in our solar system. Further, it helps in determining the likelihood of extra-terrestrial lives. Apart from this, space data analysis has played a great role in helping us understand several other bodies present out in the space. The following figure shows different exoplanet detection techniques.

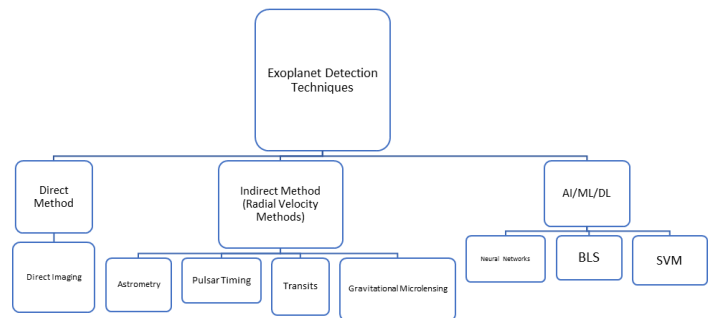


Fig. 1. Exoplanet Detection approaches

Box Least Squares Method:

The Box-fitting least squares periodogram is a statistical tool which helps in identifying eclipsing binaries and/or transiting exoplanets in time-series photometric data. It finds these transit candidates by designing a transit as a periodic signal by defining four parameters specifically: period, depth, duration and a reference time; where the reference time is usually the mid-transit time of the very initial transit in the observational baseline. This has also been shown in the figure below.

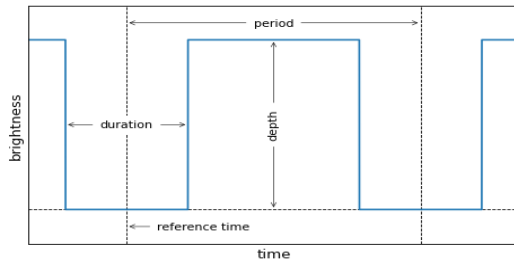


Fig. 2. BLS Parameters

Mathematical Perspective

It is assumed that the measured flux uncertainties are not known, Gaussian and independent. Therefore, the maximum possibility of in-transit flux can be given as:

$$y_{in} = \frac{\sum_{in} \frac{y_n}{\sigma_n^2}}{\sum_{in} \frac{1}{\sigma_n^2}}$$

Here, y_n refers to the measurement of brightness, σ_n depicts its uncertainties. The sums are then calculated for the data points over the in-transit.

Also, the maximum possibility for the out-of-transit flux can be given as

$$y_{out} = \frac{\sum_{out} \frac{y_n}{\sigma_n^2}}{\sum_{out} \frac{1}{\sigma_n^2}}$$

These results can be used for calculating the possibility of a transit model which has been maximized over depth at a particular period P , duration τ , and t_0 be the reference time and can be given as:

$$\log \mathcal{L}(P, \tau, t_0) = -\frac{1}{2} \sum_{in} \frac{(y_n - y_{in})^2}{\sigma_n^2} - \frac{1}{2} \sum_{out} \frac{(y_n - y_{out})^2}{\sigma_n^2} + \text{constant}$$

The above value is proportional to the ‘‘chi squared’’ χ^2 for the model and is also a sequel of the assumption made above i. e., the Gaussian uncertainties. χ^2 is known as the ‘‘signal residue’’ [16], thus, by maximizing the log possibility for duration and reference time will be equivalent to calculating the box least square spectrum form. This can be accomplished by detecting the maximum possibility model for a grid in duration and reference time as set out as ‘durations and ‘oversample’ parameters for the ‘power’ method. [17]

III. KEPLER’S MISSION

The Kepler mission has been notably created for surveying the region of the Milky Way galaxy in order to find multiple hundreds of Earth-sized and/or smaller planets which can be in or near the habitable zone and to identify a part of hundreds of billions of stars in our Milky Way galaxy that may contain these types of planets. The missions’ scientific aim is to survey the layout and the diversity of these planetary systems.

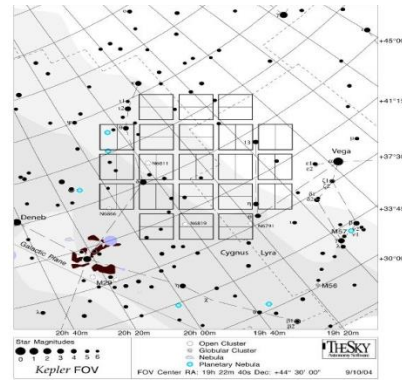


Fig. 3. Kepler’s field of view in the targeted star field

The Kepler spacecraft lost its second of the four reaction wheels in May 2013 ended the Kepler’s four-year mission. [10]

Mathematical Perspective

The Transiting Exoplanet Survey Satellite was sent up into the space by being loaded into a SpaceX Falcon 9 rocket on 18 April, 2018 [10] which had been basically designed to survey and detect exoplanets specifically similar to Earth for in-depth studies. The approach used was to survey for two years by dividing it into 26 separate sectors, each one of them being 24 degrees by 96 degrees across. The equipped cameras will collect the data for each sector for a minimum of 27 days and capture a two-minute cadence for the brightest stars. TESS was designed to cover 400 times larger sky-area than the Kepler mission.

IV. LIGHT CURVES

The graphs that exhibit the brightness of an object which has been captured for a period of time is known as a light-curve. These graphs are quite helpful when it is required to study about variant brightness of objects over time, such as supernovae, novae, or any other stars. For example, let us consider the following information about a specific source; [22]

The brightness of a specific astronomical object has been shown in the above light-curve which has been captured over a period of time. This change in the intensity of the brightness in a light-curve aids astronomers to study astronomical objects in-depth. This also helps in identifying the nature of the object that is being studied and thus we can identify the nature of the object as well.

The transit light curves help in determining estimates directly related to the body’s mass and radius. A transit light curve is formed when anybody creates a dip in the brightness of its host star due to its eclipsing around that host star. This creates a possibility of presence of an exoplanet. An example of an exoplanet surpassing star spots has been shown in the figure below. A small elevation in flux is visible as the light of the cooler side of the host star gets blocked out. [23]

A high Sound to Noise (S/N0 ratio and high precision, the light curve can also be helpful in inferring the possibility of presence of other exoplanets in the same system that is being studied.

V. DESIGN & IMPLEMENTATION

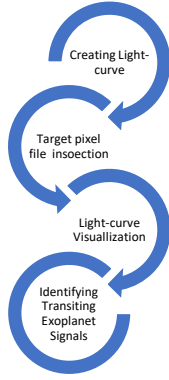


Fig. 4. Flowchart for the algorithm

Dependencies:

- **Numpy:** It is rudimentary package developed for scientific computing in Python that dispenses a multidimensional array object, and multiple derived objects. It also consists of mixture of routines for speedy operations on array.
- **Matplotlib:** It is a library in Python which is used to graph the data on figures for one or more Axes. [18]
- **Astropy:** It is a package comprising key functionality and common tools that are required to perform various functions related to astronomy and astrophysics. It is the vital part of the Astropy Project that are intended towards development of a robust environment of affiliated packages that cover a wide range of functionalities for astronomical research, data processing and analysis.[19]
- **Astroquery:** It is a package within astropy that consists of multiple tools for querying astronomical databases. In addition to this package, there are two more complimentary functionality packages: pyvo and Simple-Cone-Search-Creator.[19]
- **Lightkurve:** The package is designed to aid time-series data on the planets, stars and galaxies which is acquired by telescopes that have been sent into deep space in search for data. The images collected by these telescopes have Infrared and/or visible light wavelength. The light curve object of the lightkurve is an augmentation of Astropy time-series object. [20]

Nasa Exoplanet Archive

The NASA Exoplanet Archive is an online catalogue which consists of the astronomical exoplanet and stellar data. It offers data services that accumulate and cross-correlate the data and aids in performing the analysis on the same. The main aim of the Archive is to gather vital data sets for searching and characterization of exoplanets and their host stars. The data comprises of stellar parameters, exoplanet parameters and characterization data. [21] The data can be directly accessed from this catalogue depending upon the requirement of the project.

VI. RESULTS & DISCUSSION

In order to attain the required results, the light curve has been plotted. The light curves have been plotted for a specific extra solar body from the Kepler mission – 11904151. The initial light curve that has been plotted is shown below:

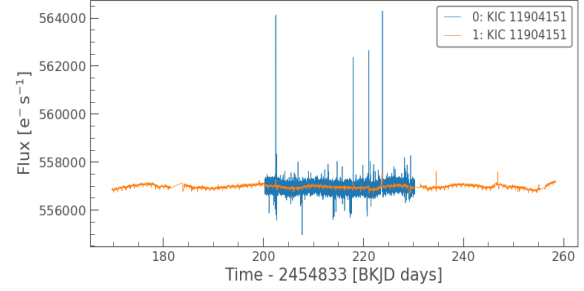


Fig. 5. Light curve for KIC 11904151

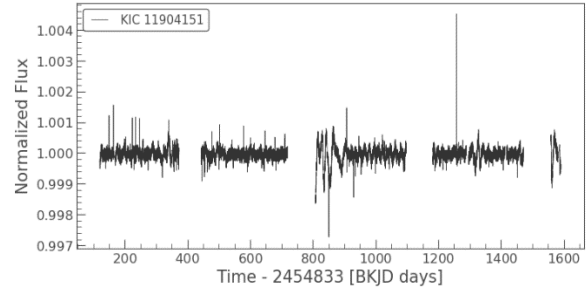


Fig. 6. Transiting Time Variation Plot

This object reveals an effective transiting timing variation (TTVs), i.e., the transit time of the target candidate varies with respect to time. It occurs when the planet's orbit is not quite periodic due to, for example, the presence of several moving stars or planets in the system. On plotting the data, a long-term trend can be seen, likely from star-spots. This can be removed with a Savitzky–Golay filter using Lightkurve's flatten method. A Savitzky–Golay filter is a digital filter which is implemented over a set of digital data points for smoothing the data, that is, to enlarge the precision of the data without influencing the signal tendency.

The standard deviation based light curve has been shown below for our target star:

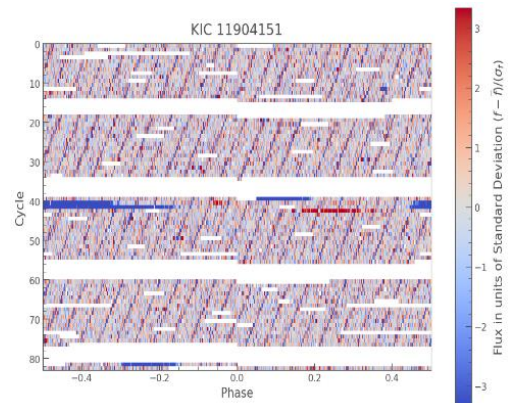


Fig. 7. Standard deviation curve for KIC 11904151

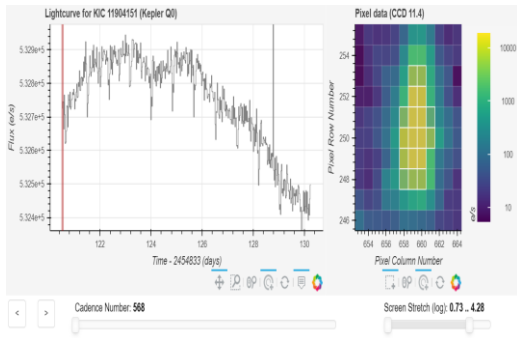


Fig. 8. The interactive light curve plot

The above plot is an interactive plot, a feature provided by the lightcurve package. This makes easy to inspect the data points in the analysis just by scrolling through the light curve.

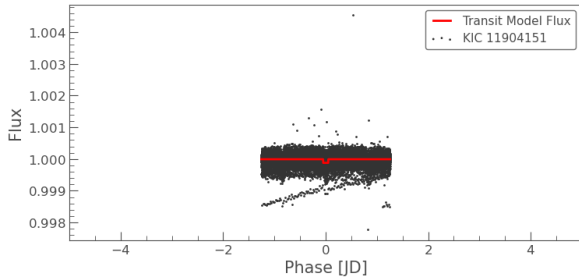


Fig. 9. BLS and Scatter plot for KIC 11904151

The above figure shows the combination of the BLS method being applied to the scatter plot to understand the presence of any transit through the particular phase.

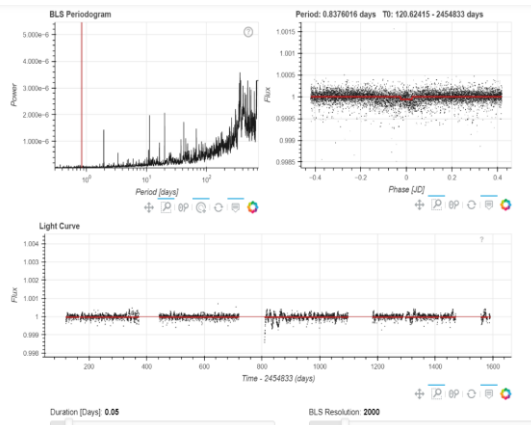


Fig. 10. Interactive BLS plot for analysis.

The above figure shows an interactive plot that can be used for further analysis of the light curve obtained just by scrolling over it. It is an interactive feature provided by lightkurve suite. Following are the parameter calculations for the target star:

RA: 19h02m43.03s
 DEC: +50d14m29.34s
 DISTANCE: 185.506+0.811-0.805
 PARALLAX: 5.3618500±0.0233009 mas
 ECLIPTIC LATITUDE: 71.85007 deg
 ECLIPTIC LONGITUDE: 303.70098 deg
 GALACTIC LATITUDE: 18.81626 deg
 GALACTIC LONGITUDE: 80.48988 deg
 TOTAL PROPER MOTION:
 45.3461486±0.0457027 mas/yr

PROPER MOTION (RA): -18.3937000±0.0448046 mas/yr
 PROPER MOTION (DEC): 41.4481000±0.0458775 mas/yr

This is just an analysis over one particular body. Likewise, multiple light curves can be created in order to analyse the data and see the presence of other features as well. It is not only to analyse the data of the target exoplanets, but also the older data of the confirmed exoplanets, novae and supernovae as well.

VII. CONCLUSION

From the above results it can be interpreted that the target star KIC 11904151 has a body eclipsing around it. The standard deviation plot also confirms the change in flux values for particular transits depicting the same possibility. Moreover, the interactive plots can be adjusted and analysed to compare the transit time for the dip occurring. Applying the BLS method, the presence of the exoplanet body gets confirmed. Thus, there is a presence of an extrasolar body around the target star. This is just an analysis over one particular body. Likewise, multiple light curves can be created in order to analyse the data and see the presence of other features as well. It is not only to analyse the data of the target exoplanets, but also the older data of the confirmed exoplanets, novae and supernovae as well.

ACKNOWLEDGEMENT

The authors are grateful for facilitating research in this area to University VC. It's a part of the work done as dissertation by Post grad student.

REFERENCES

- [1] Johnson, John (2015). How Do You Find an Exoplanet? 41 William Street, Princeton, NJ 08540: Princeton University Press. p. 65. ISBN 978-0691156811.
- [2] Chintarungruangchai, Pattana & Jiang, Ing-Guey. (2019). Detecting Exoplanet Transits through Machine-learning Techniques with Convolutional Neural Networks. Publications of the Astronomical Society of the Pacific. 131. 064502. 10.1088/1538-3873/ab13d3.
- [3] Borucki, W. J., Koch, D., Basri, G., Batalha, N., Brown, T., Caldwell, D., Caldwell, J., Christensen-Dalsgaard, J., Cochran, W. D., DeVore, E., Dunham, E. W., Dupree, A. K., Gautier, T. N., Geary, J. C., Gilliland, R., Gould, A., Howell, S. B., Jenkins, J. M., Kondo, Y., Latham, D. W., Marcy, G. W., Meibom, S., Kjeldsen, H., Lissauer, J. J., Monet, D. G., Morrison, D., Sasselov, D., Tarter, J., Boss, A., Brownlee, D., Owen, T., Buzasi, D., Charbonneau, D., Doyle, L., Fortney, J., Ford, E. B., Holman, M. J., Seager, S., Steffen, J. H., Welsh, W. F., Rowe, J., Anderson, H., Buchhave, L., Ciardi, D., Walkowicz, L., Sherry, W., Horch, E., Isaacson, H., Everett, M. E., Fischer, D., Torres, G., Johnson, J. A., Endl, M., MacQueen, P., Bryson, S. T., Dotson, J., Haas, M., Kolodziejczak, J., Van Cleve, J., Chandrasekaran, H., Twicken, J. D., Quintana, E. V., Clarke, B. D., Allen, C., Li, J., Wu, H., Tenenbaum, P., Verner, E., Bruhweiler, F., Barnes, J., and Prsa, A. (2010). Kepler Planet-Detection Mission: Introduction and First Results. Science, 327:977.
- [4] Ricker, G. R., Winn, J. N., Vanderspek, R., Latham, D. W., Bakos, G. A., Bean, J. L., Berta-Thompson, Z. K., Brown, T. M., Buchhave, L., and Butler, N. R. (2015). Transiting Exoplanet Survey Satellite (TESS). Journal of Astronomical Telescopes, Instruments, and Systems, 1:014003.
- [5] Kovács, G., Zucker, S., and Mazeh, T. (2002). A box-fitting algorithm in the search for periodic transits. , 391:369–377.
- [6] Obermeier, C. (2016). Searching for hot jupiter transits around cool stars.
- [7] Wright, Jason & Gaudi, B. (2012). Exoplanet Detection Methods. Planets, Stars and Stellar Systems. Volume 3: Solar and Stellar Planetary Systems. 10.1007/978-90-481-8818-5_59.

- [8] [8] Dai, Ziqi & Ni, Dong & Pan, Lizhuang & Zhu, Yiheng. (2021). Five Methods of Exoplanet Detection. *Journal of Physics: Conference Series*. 2012. 012135. 10.1088/1742-6596/2012/1/012135.
- [9] Wright, J.T. & Gaudi, B.S.. (2013). *Exoplanet Detection Methods*. 10.1007/978-94-007-5606-9_10.
- [10] NASA : <https://exoplanets.nasa.gov/>
- [11] Izuagbe, Kennedy Jnr. (2015). The Transit Method for Exoplanet Detection. 10.13140/RG.2.1.1297.0323.
- [12] M.A. Khan and M. Dixit, "Discovering Exoplanets in Deep Space using Deep Learning Algorithms," 2020 12th International Conference on Computational Intelligence and Communication Networks (CICN), 2020, pp. 441-447, doi: 10.1109/CICN49253.2020.9242636.
- [13] Kyle A. Pearson, Leon Palafox, Caitlin A. Griffith, searching for exoplanets using artificial intelligence, *Monthly Notices of the Royal Astronomical Society*, Volume 474, Issue 1, February 2018, Pages 478–491, <https://doi.org/10.1093/mnras/stx2761>
- [14] <https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning>
- [15] Kovacs, Zucker, & Mazeh (2002), *A&A*, 391, 369 (arXiv:astro-ph/0206099)
- [16] Hartman & Bakos (2016), *Astronomy & Computing*, 17, 1 (arXiv:1605.06811)
- [17] John Hunter, Darren Dale, Eric Firing, Michael Droettboom and the Matplotlib development team; 2012 - 2022 The Matplotlib development team.
- [18] <https://exoplanet.nasa.gov>
- [19] Lightkurve Collaboration, "Lightkurve: Kepler and TESS time series analysis in Python", *Astrophysics Source Code Library*, 2018. ascl:1812.013.
- [20] NASA Exoplanet Archive "This research has made use of the NASA Exoplanet Archive, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program."
- [21] <https://imagine.gsfc.nasa.gov/science/toolbox/timing1.html>
- [22] Charbonneau, D., Brown, T. M., Burrows, A., and Laughlin, G., "When Extrasolar Planets Transit Their Parent Stars", in *Protostars and Planets V*, 2007, p. 701.
- [23] Mandel, K. and Agol, E., "Analytic Light Curves for Planetary Transit Searches", *The Astrophysical Journal*, vol. 580, no. 2, pp. L171–L175, 2002. doi:10.1086/345520.
- [24] *Transiting exoplanets*, by C. Haswell. Cambridge, UK: Cambridge University Press, 2010, 335 p. \$60, softcover (ISBN #978-0-521-13938-0); \$130, hardcover (ISBN #978-0-521-19183-8).
- [25] Shallue C. J. and Vanderburg A., "Identifying Exoplanets with Deep Learning: A Five-planet Resonant Chain around Kepler-80 and an Eighth Planet around Kepler-90", *The Astronomical Journal*, vol. 155, no. 2, 2018. doi:10.3847/1538-3881/aa9e09.