Multi-scale (time and mass) dynamics of space objects
Proceedings IAU Symposium No. 364, 2022
A. Celletti, C. Galeş, C. Beaugé, A. Lemaitre, eds.
doi:10.1017/S1743921322000023

An algorithm for automatic identification of asymmetric transits in the TESS database

M. Vasylenko[1,2], Ya. Pavlenko[1], D. Dobrycheva[1], I. Kulyk[1], O. Shubina[1,3] and P. Korsun[1]

[1]Main Astronomical Observatory of the NAS of Ukraine,
27 Akademika Zabolotny Str., Kyiv, 03143, Ukraine

[2]Institute of Physics of the National Academy of Sciences of Ukraine,
46 avenue Nauka, Kyiv, 03028, Ukraine

[3]Astronomical Observatory of Taras Shevchenko National University of Kyiv,
3 Observatorna Str., Kyiv, 04053, Ukraine

Abstract. Currently, the Transiting Exoplanet Survey Satellite (TESS) searches for Earth-size planets around nearby dwarf stars. To identify specific weak variations in the light curves of stars, sophisticated data processing methods and analysis of the light curve shapes should be developed and applied. We report some preliminary results of our project to find and identify minima in the light curves of stars collected by TESS and stored in the MAST (Mikulski Archive for Space Telescopes) database. We developed Python code to process the short-cadence (2-min) TESS PDCSAP (Pre-search Data Conditioning Simple Aperture Photometry) light curves. Our code allows us to create test samples to apply machine learning methods to classify minima in the light curves taking into account their morphological signatures. Our approach will be used to find and analyze some sporadic events in the observed light curves originating from transits of comet-like bodies.

Keywords. TESS, MAST, exoplanets, exocomets.

1. Introduction

During the last ten years, more than 4,300 exoplanets orbiting their parent stars have been discovered, mostly by the space missions Kepler (Borucki (2010)) and TESS (Ricker (2015)). However, we still have a little information about the populations of extrasolar subplanetary bodies in these systems, such as planetesimals, asteroids, and comets. Meanwhile, the modern theories of the planetary system formation assume a large population of planetesimals, which play an important role in the dynamics and physical evolution of the planetary system (see Lagrange (2020) and references herein). Currently, only a few solid detections of exocomets have been reported by Rappaport (2018), and Zieba (2019) for star KIC3542116 and β Pictoris system based on analysis of the Kepler and TESS data bases, respectively.

Our project is to contribute to studying cometary activity in extrasolar systems based on the TESS high-precision observations. The obtained and pre-proceeded light curves are available in the MAST archive (Ricker (2015)). Our work aims to analyze the light curves of stars to find asymmetric transits caused by the passage of comet-like bodies across the stars' disks. Machine learning (ML) methods can be used to optimize this process.

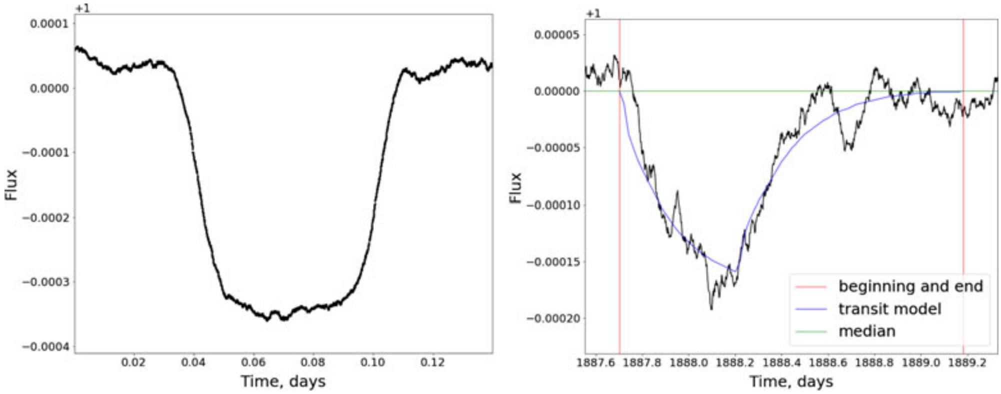

Figure 1. *Left:* The profile of the "55 Cnc e" planet transit, processed by our program "*Lc_cuter*". *Right:* The simulation of comet Hale-Bopp transit encapsulated into the observed light curve is marked by black solid line. The start and the end moments of the comet transit are marked by the red vertical lines; the exocomet transit model is marked by the blue curve; the star flux median calculated apart from the transit interval is marked by the green line.

2. Data analysis and results

More than 200,000 stars were selected as primary TESS mission targets and included in the TESS Input Catalogue to produce time sequences of the brightness measurements at every 2 min during the long enough observation sets in 27 days. To analyse so large amount of the data novel approaches should be used, such as ML techniques, which have been already successfully applied for searching for exoplanet transits (Shallue (2018), Malik (2020)). In order to find and identify the asymmetric decrease in the star brightness caused likely by a comet-like body and separate them from exoplanet transits we apply the classical ML Random Forest method, for which we constructed the classifier using the Python module Scikit-learn (Pedregosa (2011)). The two different samples of data were used to train the ML model: a) the light curve profiles caused by the already identified exoplanet transits and b) the simulated brightness profiles due to exocomet transits. The latter is calculated using the Monte-Carlo approach to model the exocomet dusty tail taking into account orbital characteristics of the exocomet and physical properties of particles that populated its dusty atmosphere at the given distance from the star (Korsun (2010)). We developed a program code "*Lc_cuter*", based on the Python package "lightkurve2.0" (Borucki (2010)) to process the short-cadence (2-min) TESS PDCSAP light curves and create the sample of exoplanet transits. The example of the brightness profiles from the two different data samples is presented in Fig. 1.

To generate the sample for the ML we chose about 6000 PDCSAP light curves which have no signs of transits. Then we selected ∼50% of the light curves and artificially put planetary transits with random periods larger than 2 days. If the period is less than the observation sector base line, it means that the planet transit occurs more than once. There may also be a case when no planetary transit can be seen in the light curve for the certain sector. The remaining light curves were populated with the simulated cometary transits. The important step of the analysis is to evaluate the noise level which restrict the ability to detect a transit. As an indicator of the noise we used a simplified proxy algorithm to calculate CDPP (Combined Differential Photometric Precision) metric based on the Lightkurve 2.0 package. The package implements the Savitzky-Golay filter to remove frequency signals. To calculate the CDPP we applied a window of 1515 points for the Savitzky-Golay filter and transit duration of 6.5 hours. For the trained samples we used the initial light curves with the CDPP less than 30 ppm (part per million). We focused

https://doi.org/10.1017/S1743921322000023 Published online by Cambridge University Press

the ML on two options of classifying signals in the light curves: the presence or absence cometary transits. For this, the time series feature extraction ("tsfreash", Christ (2018)) provides features for the classifier. The significance of the features was determined with the Scikit-learn package (Pedregosa (2011)). The most significant features were tested for correlation with each other. To verify the results we divided the total sample into two subsamples, i.e. 80% training and the 20% test one. We found that the Random Forest method allows us to separate the light curves with accuracy of 97% (98% for the light curves with cometary transits and 95% with no cometary transits).

3. Brief conclusions

The ML Random Forest method allows potentially identify asymmetric minima in the light curves due to the transits of comet-like bodies with an accuracy of 97% if the minimum depth is larger than 0.02% of the star's flux and for light curves with a noise level corresponding to CDPP < 30. Since the detection of any signal becomes more difficult with increasing in the noise level, our result is the first step towards the ability to detect shallower comet transits in the light curves of poorer quality, using more advanced data processing methods, a wider set of features for the classifier, and more powerful deep neural networks.

Acknowledgements

This study was performed in the frames of the government funding program for institutions of the National Academy of Sciences of Ukraine (NASU) and supported by the National Research Foundation of Ukraine (№ 2020.02/0228). All data presented in this paper were obtained from the Mikulski Archive for Space Telescopes (MAST). STScI is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555. Support for MAST for non-HST data is provided by the NASA Office of Space Science via grant NNX09AF08G and by other grants and contracts. This paper includes data collected by the TESS mission. We thank the referee for their helpful comments on an earlier version of this contribution.

References

Barentsen, G., Hedges, C. L., *et. al* 2019, *American Astronomical Society Meeting Abstracts*, 233, 109.08

Borucki, W. J., Koch, D., Basri, G., *et. al* 2010, *Science*, 327, 977

Christ, M., Braun, N., Neuffer, J., and Kempa-Liehr, A. W. 2018, *Neurocomputing*, 307, 72

Korsun, P. P., Kulyk, I. V., Ivanova, O. V., *et. al* 2019, *Icarus*, 210, 916

Lagrange, A. M., Rubini, P., Nowak, M., *et. al* 2020, *A&A*, 642, A18

Malik, A., Moster, B. P. and Obermeier, C., 2020, *arXiv e-prints*, arXiv:2011.14135

Pedregosa, F., Varoquaux G., Gramfort A., *et al* 2011, , 12, 2825

Rappaport, S., Vanderburg, A., Jacobs, T.,*et. al* 2018, *MNRAS*, 474, 1453

Ricker, G. R., Winn, J. N., Vanderspek, R., *et. al* 2015, *Journal of Astronomical Telescopes, Instruments, and Systems*, 1, id. 014003

Shallue, C. J. and Vanderburg, A., 2018, *AJ*, 155 (2), 94

Zieba, S., Zwintz, K., Kenworthy, M. A., *et. al* 2019, *A&A*, 625, L13

https://doi.org/10.1017/S1743921322000023 Published online by Cambridge University Press