

参赛队员姓名： Gavin Wang

中学： 清华大学附属中学国际部

省份： 北京

国家/地区： 中国

指导教师姓名： Shude Mao

指导老师单位： 清华大学

论文题目： Developing an Automated Pipeline  
for Identifying False Positives Among Released  
TESS Objects of Interest

Developing an Automated Pipeline for Identifying False Positives Among Released TESS Objects of Interest

Author:

Gavin Wang, Tsinghua International School

Supervisor:

Prof. Shude Mao, Tsinghua University

S.-T. Yau High School Science Award

September 15, 2021

## ABSTRACT

The Transiting Exoplanet Survey Satellite (TESS) has now revealed over 4,000 planet candidates known as TESS Objects of Interest (TOIs), roughly 10% of which have been identified as nearby eclipsing binaries. Ruling out such false positives, which is typically done through follow-up observations, is a necessary step for discovering extrasolar planets. Another method of identifying nearby eclipsing binaries involves manually reducing TESS full-frame images using AstroImageJ, but this is time-consuming. In this work, Python script and AstroImageJ's macro are used to create a pipeline for automating this method. This automation dramatically reduces the time spent on reduction and has been used to successfully analyze a sample of 12 Quick Look Pipeline (QLP) TOIs, among which one false positive due to a nearby eclipsing binary was identified. Applying this pipeline to a large number of QLP targets will allow many false positives to be identified without the need to conduct ground-based follow-up observations.

*Keywords:* Transiting Exoplanet Survey Satellite, full-frame images, data pipeline, exoplanet candidates, false positives, nearby eclipsing binaries

## TABLE OF CONTENTS

1. Introduction .....	4
2. Pipeline Workflow.....	5
2.1. Preparing Input File.....	5
2.2. Automatic MAST Data Query.....	5
2.3. Download Positions of Nearby Stars.....	6
2.4. Modify Apertures File .....	7
2.5. Multi-Aperture Photometry .....	7
2.6. Smoothing and Phasing of Light Curves.....	8
2.7. Save Light Curves .....	9
3. Results .....	10
3.1. False-positive Events.....	10
3.2. Inconclusive Events.....	12
3.3. Performance.....	13
4. Conclusion and Discussion.....	13

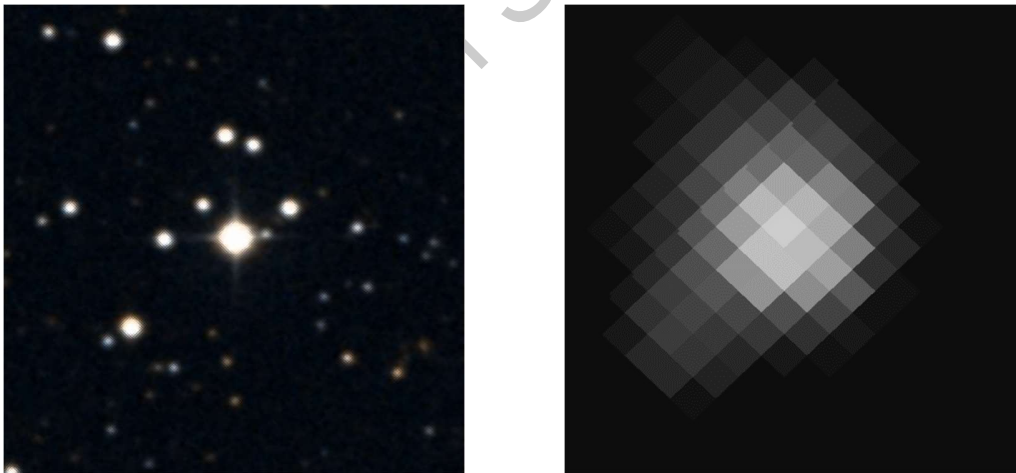
## 1. INTRODUCTION

The Transiting Exoplanet Survey Satellite (TESS, Ricker et al. 2014), which launched in April of 2018, is NASA's latest mission designed to discover transiting exoplanets around bright, nearby stars through an all-sky survey. Prior to discovery, all exoplanets begin as candidates which must then be validated through various methods. In the past three years, TESS has identified a large number of transiting planet candidates, the most promising of which have been publicly released in the TESS Object of Interest (TOI) Catalog.

Identifying and ruling out false positives is a necessary step in the discovery of new planets. For TESS, false positives are typically identified through follow-up observations conducted by the five Working Groups of the TESS Follow-up Observing Program<sup>1</sup>, each of which implements a different technique to determine the nature of planet candidates. Time-series photometry and doppler spectroscopy are the two most major methods by which false positives are revealed. To date, more than 800 false positives have been found among the over 4,000 TOIs.

As was learned from the Kepler mission, nearby eclipsing binaries are common sources of false positives (Morton et al. 2016). In this scenario, a pair of binary stars (either a chance-alignment or a pair physically bound to the target star) creates eclipse signals that masquerade as those produced by transiting planets.

Nearby eclipsing binaries are especially prevalent among TOIs, a consequence of TESS's large pixels that typically blend multiple stars. Indeed, nearby eclipsing binaries account for 50% of all TESS false positives, with well over 400 identified at the time of writing. Most of these have been identified by ground-based follow-up telescopes with high angular resolution ( $< 1''$ ), with the data analyzed using AstroImageJ (AIJ, Collins et al. 2017).



**Figure 1.** Left: image of star field from DSS2. Right: same field from TESS full-frame images exhibiting pixelation due to large pixels.

TESS's full-frame images (FFIs) contain the entire field of view collected by each of the TESS cameras, stacked to 30-minute cadence for the primary mission. Despite TESS's cameras having large pixels measuring  $\sim 21''$  by  $21''$  (see Figure 1), it was shown in a separate study (Wang, G. and Collins, K. 2021) that analyzing full-frame

<sup>1</sup> <https://tess.mit.edu/followup/>

images using AstroImageJ is a feasible method of identifying nearby eclipsing binaries among bright ( $T_{mag} < 10.5$ ) TOIs identified by the TESS Quick Look Pipeline (QLP). This method saves expensive follow-up telescope time by filtering out false positives before they are followed-up; however, it is time-consuming and requires large amounts of repetitive labor. This paper presents a pipeline that automates this method and greatly reduces the time spent on analysis.

## 2. PIPELINE WORKFLOW

There are seven major steps involved in the automation of the extraction of light curves from TESS full-frame images: prepare a TOI input file, download the full-frame images, download an apertures file containing the positions of all nearby stars, modify the apertures file, perform multi-aperture photometry, post-process the resulting light curves, and save the light curves to a local disk. After the input file is created, the subsequent three steps are completed via a Python script, and the final three are realized via an AstroImageJ macro.

### 2.1. Preparing Input File

The pipeline requires a single input file, containing a list of eight values for each TOI (see Table 1).

Value	Notes
TOI Number	Unique ID of target; integer greater than 100
TESS Magnitude	Calculated from Gaia magnitude as in Stassun et al. 2019
T0	Barycentric Julian Date
Period	Days
Duration	Hours
Transit Depth	Parts per million (ppm)
Right Ascension	Decimal degrees
Declination	Decimal degrees

**Table 1.** Parameters for each TOI in the input file.

### 2.2. Automatic MAST Data Query

After calibration by the TESS Science Processing and Operations Center (Jenkins et al. 2016), full-frame images are released to the public through the Mikulski Archive for Space Telescopes (MAST<sup>2</sup>). The pipeline queries MAST by sending the RA and Dec of the target star to the TESScut API,<sup>3</sup> which makes use of Astrocut. All data associated with the target star is downloaded in the form of Flexible Image Transport System (FITS) files, with cutout FFIs for each sector compressed into individual files. These are then extracted (method courtesy of Dr. Karen Collins) into individual FITS files which can be handled by AstroImageJ. To ensure the accuracy of the results, all images with a non-zero quality flag are rejected. After extraction, the pipeline permanently removes all residual files, preventing too much disk space from being consumed.

<sup>2</sup> <https://mast.stsci.edu/>

<sup>3</sup> <https://mast.stsci.edu/tesscut/>

### 2.3. Download Positions of Nearby Stars

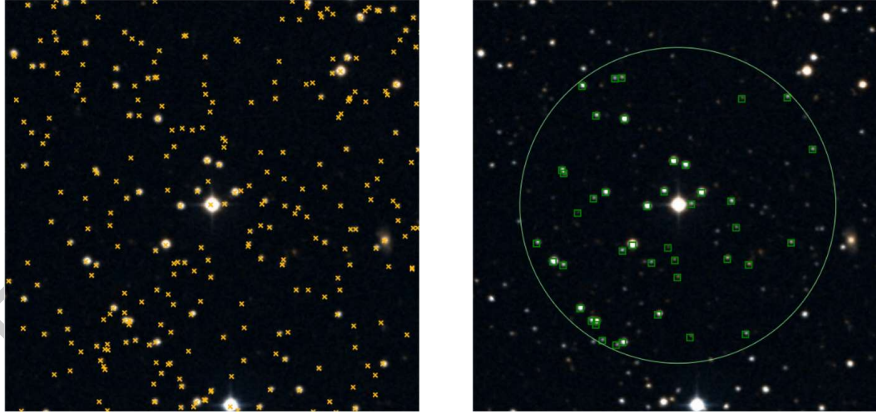
The pipeline then downloads an apertures file containing the right ascension and declination of all stars capable of producing the TESS-detected signal, in ascending order of their angular separation from the target star. This is done to inspect the light curves of all nearby stars and to determine whether there are signals off-target, a clear sign that the TOI is a nearby eclipsing binary. Due to the large size of TESS pixels, all stars located within 2.5' of the target star should be investigated. However, stars that are too faint to have caused a signal the magnitude of that detected by TESS need not be included. Whether a star is bright enough is determined using the following formula, which is an application of the fact that an object of any brightness can, at most, experience a flux deficit  $\delta$  of 1000 ppt (i.e., completely disappear):

$$\Delta T_{mag} \leq -2.5 * \log(\delta_{TOI}) \quad (1)$$

However, the calculated TESS magnitudes for all stars in the TESS Input Catalog (TIC) v8 (Stassun et al. 2019) have uncertainty. To account for this, stars that are an extra 0.5 magnitudes fainter are also included in the set of stars to be investigated:

$$\Delta T_{mag} \leq -2.5 * \log(\delta_{TOI}) + 0.5 \quad (2)$$

It is also necessary to take into account the proper motions of stars, as many targets in the TOI catalog have high proper motion and the ongoing TESS mission already spans a few years. In this project, the proper motions are taken from the TIC and the positions of the stars are adjusted to the date of observation.



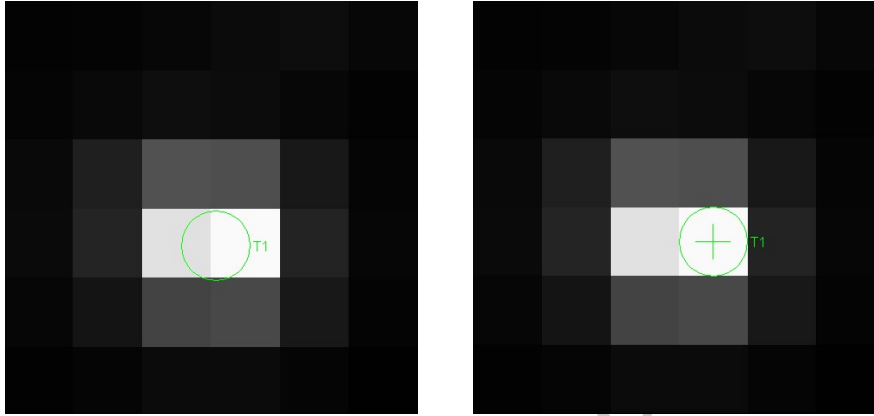
**Figure 2.** The left panel shows all stars (marked with orange crosses) in the TIC surrounding TOI 2150. The right panel shows all stars (green boxes) that are possible sources of a nearby eclipsing binary. The green circle has a radius of 2.5'.

To perform the above steps and calculations, the pipeline makes use of the online Gaia-to-AIJ tool<sup>4</sup>. The pipeline passes the target star's TESS magnitude, its transit depth, and its RA/Dec to this tool, and the Python module `urllib.request` is used to download the apertures file. Figure 2 shows the stars around TOI 2150.

<sup>4</sup> [https://www.astro.louisville.edu/gaia\\_to\\_aij/](https://www.astro.louisville.edu/gaia_to_aij/)

### 2.4. Modify Apertures File

As the Gaia-to-AIJ tool was developed primarily with follow-up observations in mind, the AstroImageJ centroid setting is toggled on for the target star by default. However, since the apertures are all 1 pixel in diameter for full-frame image analysis, centroiding will force the aperture onto the center of the brightest pixel (see Figure 3), displacing the aperture away from the true position of the target star. Thus, the pipeline modifies the apertures file and toggles centroiding off, ensuring that the aperture is centered at the exact location of the target star.



**Figure 3.** Centroiding off (left) vs centroiding on (right). The aperture is correctly centered on the star in the left panel but not in the right.

To increase computational efficiency, the pipeline trims the apertures file to the 60 apertures closest to the target star (corresponding to the first 60 rows). Although AstroImageJ can quickly handle full-frame image cutouts, which are only 20,160 bytes each, there are thousands and sometimes tens of thousands of files for each target. It is not uncommon for there to be a few hundred nearby stars that meet the criteria for inclusion in the apertures file for a single TOI, and investigating all stars will require up to millions of photometric measurements for each target.

For TOIs identified by the Quick Look Pipeline, off-target signals are typically fairly close to the target star. Manual results have shown that nearby eclipsing binaries are almost always among the closest 50 stars to the target. This informed the decision of choosing 60 as the limit.

### 2.5. Multi-Aperture Photometry

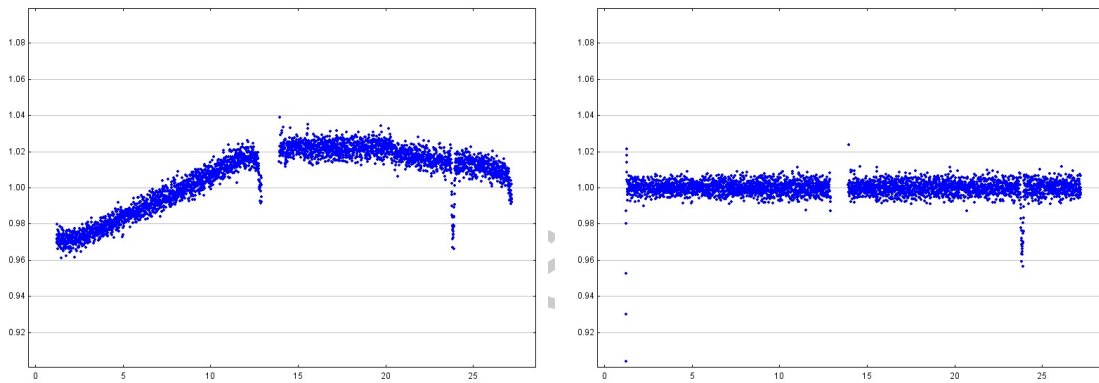
The reduction of full-frame images is importantly different from that of ground-based time-series photometry. While the latter calls for the manual determination of the optimal aperture size, detrending, and reference stars on a case-by-case basis, the process of analyzing full-frame images does not vary between different targets and thus can be automated without compromising the quality and accuracy of results.

After importing the previously mentioned images and apertures into AstroImageJ, the pipeline makes adjustments to a few settings before performing multi-aperture photometry. For reduction of full-frame images, all apertures used are 1 pixel ( $\sim 21''$ ) in diameter. The background annuli for all apertures are also set to have an inner

radius of 5 pixels and an outer radius of 10 pixels. Furthermore, no detrending parameters and no comparison stars are used.

### 2.6. Smoothing and Phasing of Light Curves

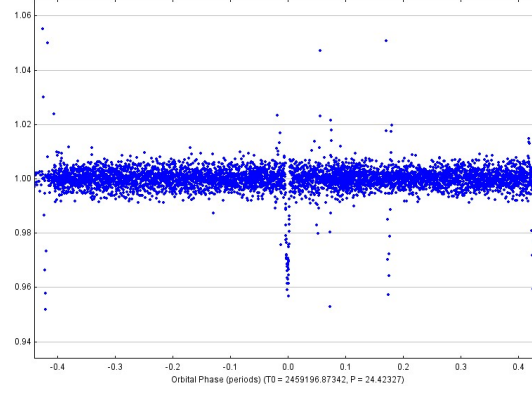
All generated light curves are post-processed by the pipeline through two steps: smoothing and phasing. Although TESS’s cameras are photometrically very precise, an additional challenge is that TESS data exhibits long-term variations due to scattered light from objects such as the Earth and Moon<sup>5</sup>. These have to be carefully removed while preserving the short-term transit signals to yield useful light curves. To combat this, the pipeline implements AstroImageJ’s “smoothing” function with a smoothing length of 31 for all full-frame image light curves. Figure 4 shows the light curve before and after smoothing.



**Figure 4.** Sample light curve before (left) and after (right) smoothing. A transit is clearly visible in the raw light curve; however, after smoothing its depth can be measured with greater accuracy.

Since TESS observes the sky in individual “sectors,” spending ~27 days observing each sector, and since at least two transits must be detected in order for a candidate to be released as an active TOI in the Catalog (and not as a Single Transit Planet Candidate), light curves must be phase-folded to the correct ephemeris to see periodic transits. This is done using the  $T_0$  (time of first-detected transit) and period values, which are calculated by the TESS vetting pipelines, released in the TOI Catalog, and refined through follow-up observations. Figure 5 shows the light curve before and after phasing.

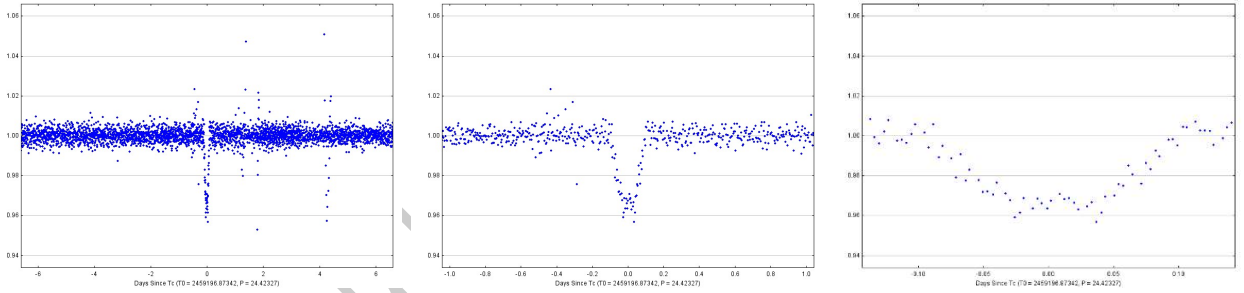
<sup>5</sup> <https://heasarc.gsfc.nasa.gov/docs/tess/observing-technical.html>



**Figure 5.** Light curve in right panel of Figure 4 after phasing. The x-axis is scaled in orbital phase, centering the transit at phase 0.

### 2.7. Save Light Curves

All light curves are then saved in a local folder by the pipeline. To ensure readability and usefulness of these light curves, the pipeline calculates the range of the x-axis according to the transit duration. Figure 6 shows three plots of the same light curve with different x-axis ranges: the ranges of the left and right panels are respectively too large and too small, while the central plot is scaled at the optimal range.



**Figure 6.** The same phase-folded light curve with different x-axis scales. The center plot is scaled according to the formula implemented in the pipeline and allows the transit to be seen most clearly.

For the y-axis, the amount of scatter present in different light curves varies greatly, requiring different ranges for different plots. To combat this issue, the pipeline first calculates a base value for the y-axis range based on the transit depth and saves five light curves for each star, one of each scaled at 1x, 3x, 5x, 10x, and 25x this base value, ensuring that there is at least one plot for each star that allows the features in the light curve to be clearly seen.

### 3. RESULTS

To gauge the performance of the pipeline, it is used to vet 12 QLP bright-star targets whose TOI IDs are listed in Table 2. For each of 191 apertures, five light curves are generated, for a total of 955. These light curves are then examined to determine whether or not each TOI is a false positive.

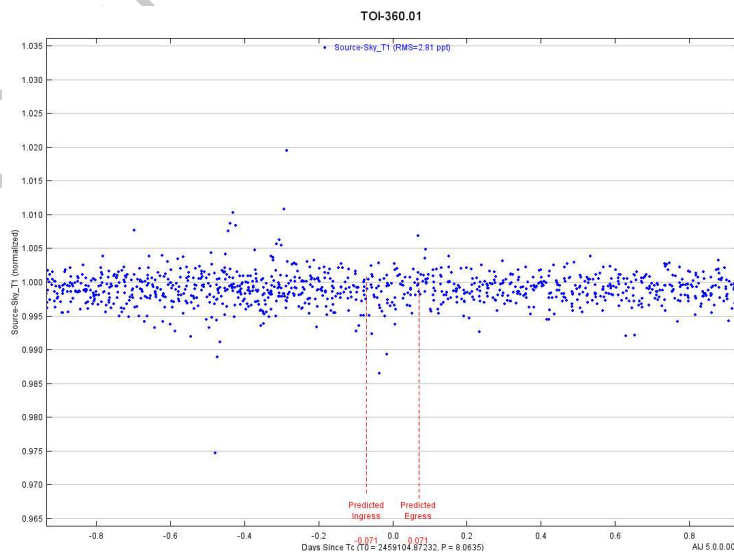
Target Number	TOI ID	Expected Transit Depth (ppm)	Number of apertures
1	297	5840	9
2	301	1750	16
3	314	1190	19
4	329	1070	11
5	331	4450	14
6	356	3830	6
7	357	620	14
8	359	3320	18
9	360	5180	6
10	438	960	9
11	518	977	43
12	579	810	26

**Table 2.** List of all 12 TOIs analyzed using the pipeline and their parameters. The expected transit depth is obtained from the TOI Catalog.

#### 3.1. False-positive Events

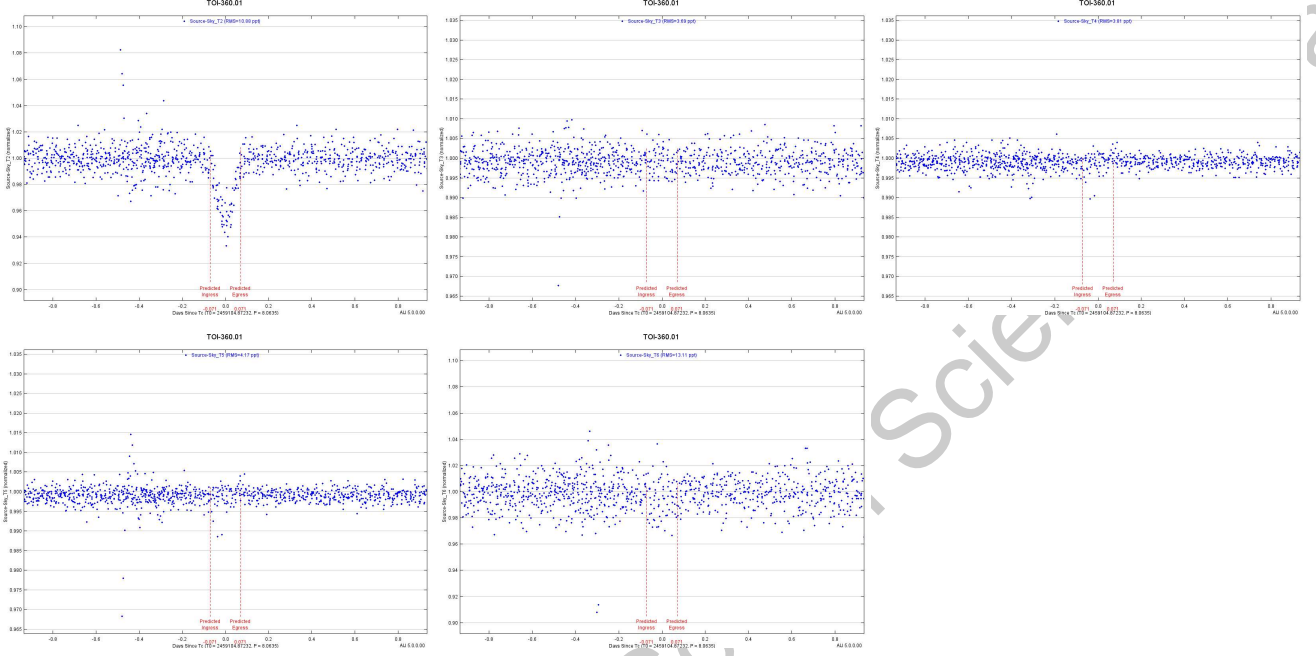
For false-positive events, transit signals with depth comparable to the transit depth in the TOI catalog are not observed in the target star light curve, and clear signals are instead seen in nearby stars. Since the star showing the strongest signals is the most likely source of the TESS detection, this situation suggests that the source is not the target star and further analysis of the transiting object is required.

Among the 12 QLP TOIs, TOI 360 is the only candidate identified as a false positive. The expected transit of depth 5180 ppm is not observed on the target star's light curve (Figure 7).



**Figure 7.** Light curve of TOI 360, the target star, showing no transits at the predicted ephemeris.

The light curves of all 5 nearby stars were then inspected (Figure 8). A strong V-shaped signal is observed in the light curve of the star TIC 13023737 (upper-left panel), which has a separation of 41'' (2 pixels) from the target star. No signals are observed in any other stars; therefore, TIC 13023737 is the most likely source of the signal, and the next step is to determine the size of its transiting companion.



**Figure 8.** Light curves of all 5 nearby stars for TOI 360, with the light curve in the first panel showing an eclipse.

The following formula, which assumes that the target star and the off-target source are fully blended in the TESS aperture, is a good approximation of the true transit depth in the source:

$$\delta_{source} = \delta_{TOI} * 10^{\Delta T_{mag}/2.5} \quad (3)$$

From the TIC, TOI 360 has  $T_{mag} = 12.184$ ,  $\delta_{TOI}$  is 5180 ppm, and TIC 13023737 has  $T_{mag} = 17.131$ . Using Eq. 3 gives:

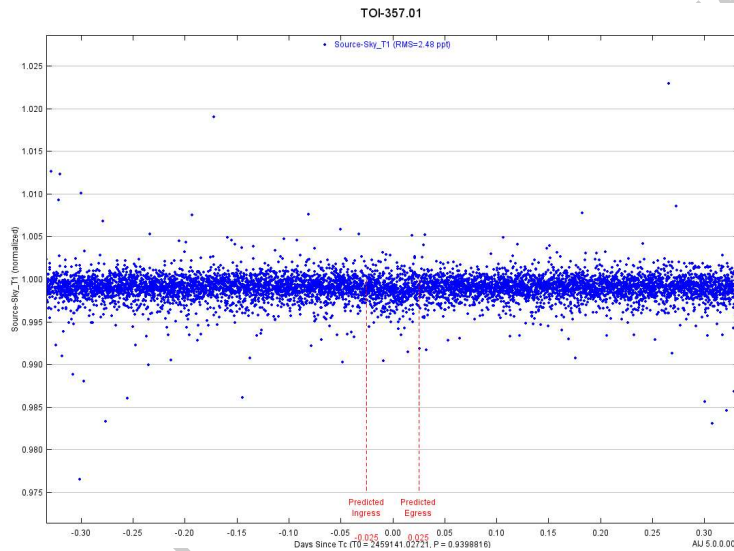
$$\delta_{source} = 493,000 \text{ ppm}$$

The flux deficit as estimated from the light curve without taking into account the contamination ratio,  $\sim 50,000$  ppm, is significantly smaller than this value; this agrees with expectations, as the much-brighter target star contaminates the aperture and dilutes the eclipse.

Given  $\delta_{source}$  and the stellar radius of TIC 13023737, which is  $0.929 R_{Sun}$ , the minimum radius of the companion is calculated to be  $71.2 R_E$  using  $R_{companion} \geq \sqrt{\delta_{source}} * R_{hos}$ . Some of the largest exoplanets discovered to date (Mandushev et al. 2007; Hebb et al. 2009; Anderson et al. 2009) have radii of  $\sim 20 R_E$ . Since the object transiting TIC 13023737 is much larger than this threshold, it is a star and the source of the signal is an eclipsing binary. Therefore, TOI 360 is deemed as a false positive and has been retired from the TOI Catalog as a result of this work.

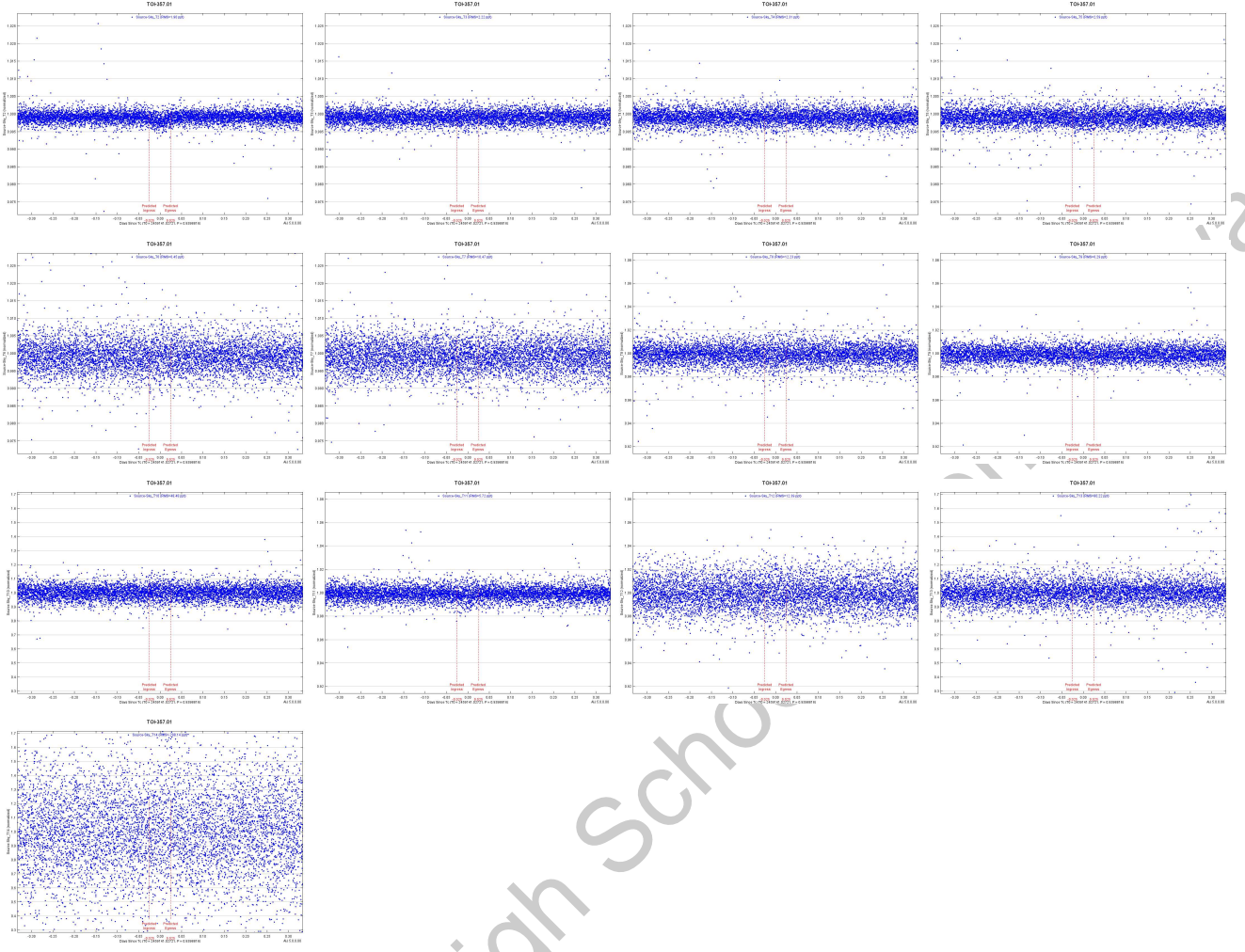
### 3.2. Inconclusive Events

Oftentimes, transit signals comparable to the transit depth listed in the TOI catalog are observed in the target star's light curve, and/or no deep transits are observed in any of the nearby stars. This is consistent with the signal being on-target and is considered to be an inconclusive event. The light curves for TOI 357 (Figure 9) and its neighboring stars (Figure 10) were generated by the pipeline and show that TOI 357 is such an event. According to the TOI Catalog, the expected transit depth of TOI 357 is 620 ppm. The phase-folded light curve of the target star shows a slight dip at the center between the two dashed vertical markers. Due to the noise in the un-binned light curve, it is not possible to accurately measure the depth. Furthermore, due to significant blending among nearby stars in the full-frame images, the TIC contamination ratio must be factored in to accurately measure the depth of the transit from the light curve. Nevertheless, the depth of this transit is visually estimated to be on the order of 1000 ppm, consistent with the expected 620 ppm.



**Figure 9.** Light curve of the target star, TOI 357, showing a slight dip of  $\sim 1000$  ppm at the predicted transit time.

Again, the light curves of the nearby stars are inspected. Of the 13 light curves, the light curve of the star closest to the target star (upper-left panel of Figure 10) shows a transit-like signal with depth  $\sim 2000$  ppm, comparable to the 1000 ppm event seen in the target star. However, due to the fact that the two stars are separated by only  $14''$  ( $< 1$  pixel), these two signals are caused by blending between the two stars. The light curves of other stars are mostly flat so there are no clear nearby eclipsing binaries and TOI 357 cannot be concluded as a false positive event.



**Figure 10.** Light curves of all 13 nearby stars surrounding TOI 357. The light curve in the upper-left panel shows a transit of  $\sim 2000$  ppm, while the other stars show no apparent transits.

Of the 12 QLP TOIs, 11 yield inconclusive results and cannot be ruled out as nearby eclipsing binaries. This is within expectations, as nearby eclipsing binaries only account for a small percentage of TOIs.

### 3.3. Performance

Using this pipeline, the generation of all light curves for 12 TOIs took roughly an hour of manual work. Compared to the  $\sim 6$  hours that would be needed to complete the same task entirely manually, this pipeline greatly reduces the time needed for manual work to 1/6 the original amount and significantly increases the analysis efficiency.

## 4. CONCLUSION AND DISCUSSION

This work presents a pipeline developed with Python script and AstroImageJ macro that automates the reduction of full-frame images to identify nearby eclipsing binaries using AstroImageJ software. This pipeline is

applied to a sample of 12 bright-star QLP TOIs, among which one false positive was found. This demonstrates that the pipeline is capable of efficiently automating the process from downloading images to outputting batches of light curves.

This pipeline is most suited for analyzing bright QLP targets, as the high signal-to-noise ratio accentuates features in the light curves. A limitation, however, arises from the full-frame images' large pixels: significant blending is present among nearby eclipsing binaries with separations of under  $\sim 20''$ . As such, the pipeline cannot conclusively identify such false positives.

As TESS's extended mission continues, more TOIs are continually being released in the TOI Catalog every month, with  $\sim 20$ -30 new QLP bright-star candidates released each sector. Applying this pipeline to these targets shortly after they are released acts as a second filter that further vets TOIs, saving ground-based follow-up observing time for targets more likely to be bona fide planets.

## REFERENCES

- Anderson, D., et al. 2010. "WASP-17b: An Ultra-Low Density Planet in a Probable Retrograde Orbit." *The Astrophysical Journal* **709** 159.
- Collins, K. et al., 2017. "AstroImageJ: Image Processing and Photometric Extraction for Ultra-precise Astronomical Light Curves." *The Astronomical Journal* **153** 77.
- Hebb, L., et al. 2009. "WASP-12b: The Hottest Transiting Extrasolar Planet Yet Discovered." *The Astrophysical Journal* **693** 1920.
- Jenkins, J., et al., 2016. "The TESS Science Processing Operations Center." *Software and Cyberinfrastructure for Astronomy IV, Proceedings Volume* **9913**.
- Mandushev, G., et al. 2007. "TrES-4: A Transiting Hot Jupiter of Very Low Density." *The Astrophysical Journal* **667** L195.
- Morton, T. et al., 2016. "False Positive Probabilities for all *Kepler* Objects of Interest: 1284 Newly Validated Planets and 428 Likely False Positives." *The Astrophysical Journal* **822** 86.
- Ricker, G. et al., 2014. "Transiting Exoplanet Survey Satellite (TESS)." *Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave, Proceedings Volume* **9143**.
- Stassun, K. et al., 2019. "The Revised TESS Input Catalog and Candidate Target List." *The Astronomical Journal* **158** 138.
- Wang, G., Collins, K., 2021. "Analyzing FFIs to Identify False Positives within TESS Candidates." *TESS Science Conference II*.

## ACKNOWLEDGEMENTS

I analyze time-series photometric datasets collected by the Las Cumbres Observatory as a Junior Member of the TESS Follow-up Observing Program (Sub-Group 1). My work has led to the discovery of nine exoplanets and has contributed to 11 peer-reviewed papers (three published; eight in prep or submitted). While working in this collaboration, I looked into methods of saving expensive follow-up telescope time and thought of the topic of this project. I would like to thank my advisor Prof. Shude Mao for supporting this topic when I brought it up to him. I would also like to thank him for reviewing my paper. These were done without compensation.

Additionally, I would like to thank Dr. Karen Collins for sharing the instructions for manual FFI analysis and verifying the false-positive result I obtained. I would like to thank Kevin Eastridge for making changes to the AstroImageJ source code, without which my macros would not have been functional.

Funding for the TESS mission is provided by NASA's Science Mission directorate. This research has made use of the Exoplanet Follow-up Observation Program website, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program. This paper makes use of data collected by the TESS mission, which are publicly available from the Mikulski Archive for Space Telescopes (MAST).

All parts of the pipeline were developed solely by the author, with the exception of an existing Python script used to unpack files from MAST into individual FITS files that can be handled by AstroImageJ. All analyses were performed solely by the author. All parts of this paper were written solely by the author.