

## Light Curves & Planet Candidates in Open Clusters from TESS Sectors 6 & 7

L. G. BOUMA (ORDER TBD),<sup>1</sup> J. D. HARTMAN,<sup>1</sup> W. BHATTI,<sup>1</sup> J. N. WINN,<sup>1</sup> AND G. Á. BAKOS<sup>1</sup>

<sup>1</sup> Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ 08540, USA

(Received July 5, 2019; Revised —; Accepted —)

Submitted to AAS journals.

### ABSTRACT

The Transiting Exoplanet Survey Satellite (TESS) has begun to photometrically monitor almost every star cluster in the solar neighborhood. To make light curves of stars that could be members of stellar associations, moving groups, or open clusters, we have begun a Cluster Difference Imaging Photometric Survey (CDIPS). Our aim is to discover giant transiting planets with known ages, and to probe the exoplanet size and separation distributions for planets younger than one gigayear. In this study, we made 159,337 light curves of candidate young stars, across 596 distinct clusters. Each light curve represents between 20 and 25 days of observations of a star brighter than  $G_{Rp} = 16$ , sampled at 30 minute cadence. We describe in detail the image subtraction and time-series analysis techniques we used to create the light curves, which have noise properties that agree with theoretical expectations. We then highlight the 39 planet candidates we identified in an initial search. 29 of these planet candidates are candidate cluster members, and the remaining 10 have indications of youth relative to field stars. We describe the vetting procedures that these objects have survived, and emphasize the additional spectroscopic, photometric, and imaging observations that are needed to determine (i) whether these candidates are planets and (ii) for which cases the host stars are cluster members. We also comment on the possible utility of the light curve sample for studies of stellar rotation evolution, and binary eccentricity damping. The light curves and vetting reports are available at [archive.stsci.edu/prepds/cdips](http://archive.stsci.edu/prepds/cdips).

**Keywords:** planets and satellites: detection – methods: data analysis — techniques: photometric — (Galaxy:) open clusters and associations: general —

### 1. INTRODUCTION

Many evolutionary processes are expected to occur over the lives of stars and exoplanets; it would be interesting to statistically observe these changes as functions of time. However, since most dwarf stars do not appreciably change until evolving off the main sequence, creative approaches must be used to infer stellar ages (Soderblom 2010). One approach beginning to bear fruit in exoplanet studies is to search for planets in coeval groups of stars, hereafter “clusters”. This includes the open clusters of yore, as well as the moving groups and stellar associations that began to be uncovered in the late 1990s (Zuckerman & Song 2004). The ages of these stellar ensembles are inferred through isochrone analyses of color-magnitude diagrams, using an appropriate combination of main-sequence turnoff locations, pre-main-sequence turnon locations, and possibly averaged MS isochronal ages across known cluster members (e.g., Kharichenko et al. 2012, Section 3.4.3). Once calibrated, gyrochronology can also be

used to verify the age of a stellar ensemble (e.g., Barnes et al. 2015; Meibom et al. 2015; Curtis et al. 2019). The ages of planets discovered in these cluster stars are then assumed to equal the age of the larger stellar ensemble.

The first planets to be detected in open clusters were found through radial velocity (RV) measurements of evolved stars in the Hyades and NGC 2423 (Sato et al. 2007; Lovis & Mayor 2007). Subsequent spectroscopic searches in Praesepe led to the detection of the first hot Jupiter in a cluster (Quinn et al. 2012), and also the first multi-planet system in a cluster (Malavolta et al. 2016). The youngest RV detections have come from surveys of T Tauri stars in the Taurus star-forming region, which showed signals of the hot Jupiter (V830 Tau b Donati et al. 2016) and of a giant planet on a 9-day orbit (CI Tau b Johns-Krull et al. 2016; Flagg et al. 2019). A final RV survey of note was conducted by Brucalassi et al. (2017), who followed 88 stars in M 67 over 7 years. They reported detections of three hot Jupiters around main sequence stars, and of one giant planet around an evolved star.

The transit method was comparatively slow to catch up, despite deep searches of M 37 by Hartman et al. (2009b), as well as searches of nine young clusters in the solar neighborhood by the Monitor project (Aigrain et al. 2007; Irwin

Corresponding author: L. G. Bouma  
[luke@astro.princeton.edu](mailto:luke@astro.princeton.edu)

et al. 2007b; Miller et al. 2008a). While these studies led to measurements of stellar rotation periods (Hartman et al. 2009a; Irwin et al. 2007a) as well as the discovery of pre-main-sequence (PMS) eclipsing binaries (EBs) (Irwin et al. 2007c), no transiting planets in open clusters were found. Transit searches in globular clusters were similarly met with null results (Gilliland et al. 2000; Weldrake et al. 2006).

Kepler (Borucki et al. 2010) observed enough stars for long enough and with sufficient precision to detect transiting planets in open clusters: Kepler-66b and 67b, in the gigayear-old NGC 6811 (Meibom et al. 2013). Though a broken reaction wheel ended the prime Kepler mission, the repurposed K2 (Howell et al. 2014) switched between fields along the ecliptic every quarter-year, and was able to observe far more clusters and young stars. The resulting discoveries made by K2 through its surveys of the Pleiades, Hyades, Praesepe, and Upper Sco were a major inspiration for the present work.

K2-25b, discovered in the Hyades, has a radius ( $3.43 R_{\oplus}$ ) that is larger than older Kepler planets at similar insolation levels (Mann et al. 2016a). This inflated radius is perhaps a hint that the planet is contracting, or losing its atmosphere (e.g., Fortney et al. 2007; Owen & Wu 2013; Fulton et al. 2017). The super-Neptune sized K2-33b ( $P \approx 5.4$  d), found in Upper Sco, showed that at least some close-in planets must form within about 10 Myr, excluding long-timescale migration as an explanation for its origin (Mann et al. 2016b; David et al. 2016). Further discoveries included six transiting planets in Praesepe (Obermeier et al. 2016; Mann et al. 2017), a long-period candidate super-Earth in the Hyades (Vanderburg et al. 2018), the three-planet K2-136 system in the Hyades (Ciardi et al. 2018; Livingston et al. 2018; Mann et al. 2018), and later the two-planet system K2-264 in Praesepe (Rizzuto et al. 2018; Livingston et al. 2019). There has also recently been a detection by K2 of a Jupiter-sized planet orbiting every 24 days around V1298 Tau (David et al. 2019).

A significant amount of effort continues to be devoted towards the discovery and characterization of planets around young stars. Sometimes, cluster membership may not even be necessary, if a star's youth can be inferred from combined rotation, activity, and kinematic indicators (e.g., K2-233; David et al. 2018). To investigate demographics in time, the detection sensitivities and reliabilities of these searchers need to be calibrated, likely with a machinery similar to that outlined by Rizzuto et al. (2017). A separate requirement for population studies to be statistically robust, imposed by the geometric rarity of transits, is that many tens of thousands of stars with known ages must be monitored at high precision.

The TESS mission (Ricker et al. 2015), though designed for other purposes, holds the promise to deliver the most homogeneous and comprehensive cluster photometric survey in history. More quantitatively, the Kharchenko et al. (2013) cluster member database indicates that  $\approx 2 \times 10^5$  open cluster members brighter than  $T = 16$  will be observed in the full-frame images over the first two years of TESS observations. The actual number may be larger, as cluster membership catalogs are not yet complete, even at these relatively

bright magnitudes (e.g., Röser et al. 2016; Cantat-Gaudin et al. 2018, 2019).

One of the major barriers to deriving precise photometry from the TESS images is the relatively poor angular resolution ( $\approx 21''$  per pixel). Almost all clusters are within 10 degrees of the Galactic plane. The problems with crowding and complex backgrounds are so severe near the galactic plane that the TESS Candidate Target List has deprioritized all objects at galactic latitudes less than  $\approx 15^\circ$  (Stassun et al. 2018, 2019), which includes almost all star clusters<sup>1</sup>. This is sensible, given that the large pixel size and the high stellar surface density make aperture photometry unreliable. However, it also means that a large number of stars in clusters that could yield transiting planets will go unprocessed by the official *TESS* data reduction pipeline.

Of course, not all young stars are in crowded fields. TESS observations of a few nearby non-crowded stellar associations have already begun bearing fruit, notably DS Tuc Ab (TOI 200; Benatti et al. 2019; Newton et al. 2019), and perhaps the planet candidates TOI-450 and 451. These objects are close and bright, and will be benchmark systems because they are relatively easy to characterize. However, statistical studies of how stellar and planetary properties are affected by ages require tens of thousands of target stars. This motivates us to also search the more crowded, and fainter fields.

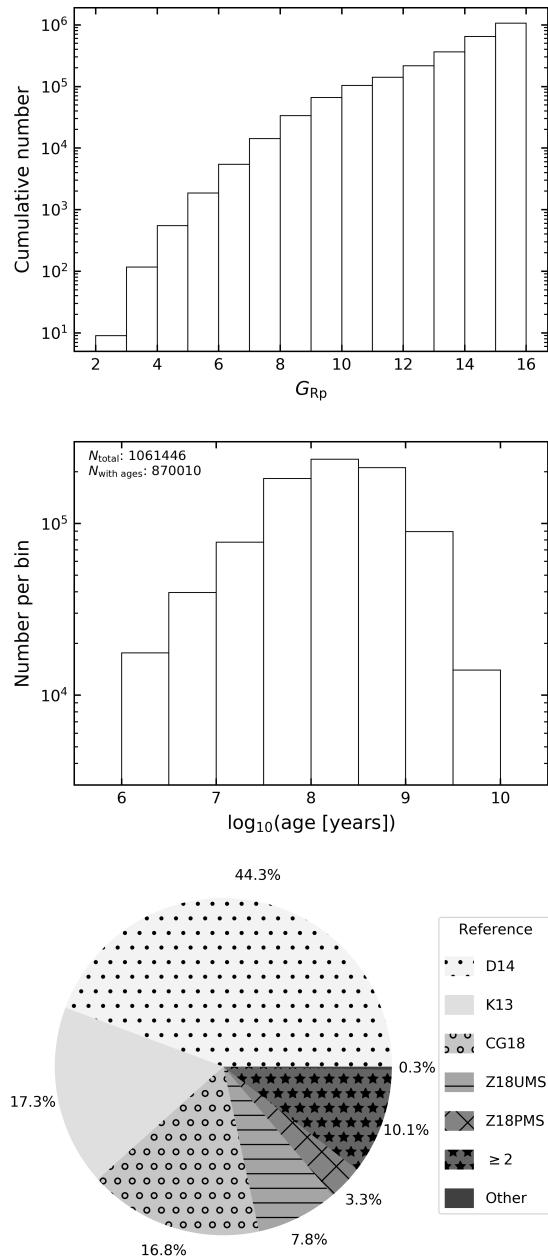
We have therefore opted to produce time-series photometry from the TESS images of any star that could be a member of a coeval group. We are also including some stars that we suspect are young due to combined photometric and astrometric indicators. We are using the difference imaging technique because of its benefits in modelling complex backgrounds in crowded regions. While we hope to discover at least a few giant transiting planets, a separate hope is that the dataset might be useful to those wishing to study the time evolution of stellar processes.

In the following, § 2 describes how we selected target stars. § 3 presents the photometric and image processing methods we used to produce light curves for these stars. The statistical properties of the 159,337 light curves from TESS Sectors 6 and 7 are then summarized (§ 4.1). The procedure we used to identify and vet transiting planets is described in § 4.2, and the candidate planets are presented in § 4.3. § 5 discusses the additional work that is required to confirm or reject the planetary nature of these objects, and § 6 summarizes our findings.

## 2. METHOD: STAR SELECTION

The main aim of the CDIPS project is to increase the number of cluster stars for which photometric time-series are available, and thereby facilitate studies of exoplanetary and stellar processes across different times and stellar environments. An essential step is therefore to define a sample of

<sup>1</sup> TICv8 has updated this to  $10^\circ$ , but for the first year of TESS observations, the number used when selecting target stars was  $15^\circ$ .



**Figure 1.** Target star statistics. *Top*. Cumulative counts as a function of Gaia  $Rp$ -band magnitude. The stars are either candidate members of clusters, or else have other youth indicators (see § 2). *Middle*. Histogram of target star ages, for the subset of stars with ages matched against Kharchenko et al. (2013). *Bottom*. Provenance of cluster membership. Percentages are relative to  $N_{\text{total}} = 1061446$  target stars. Symbols are as follows. D14 is Dias et al. (2014). K13 is Kharchenko et al. (2013). CG18 is Cantat-Gaudin et al. (2018). Z18 is Zari et al. (2018), with upper main sequence and pre-main-sequence samples sub-divided. “ $\geq 2$ ” indicates at least two authors reported a star as a candidate cluster member.

stars that are thought to be young, or members of clusters, or both.

A homogeneous membership calculation for every known cluster is a rather large undertaking, and currently falls outside our scope. So too is a homogeneous search for young stars across the galaxy. Instead, we have opted to collect and concatenate appropriate catalogs from across the literature. We then use the resulting meta-catalog to identify our target stars within the TESS images.

In our initial stellar selection our aim is completeness, not accuracy. If there has been a claim in the literature that a star should be considered a cluster member, or a young star, we would like to report a light curve for the star. For stars that are photometrically interesting, we can perform post-hoc quality checks using Gaia-DR2 astrometry and photometry to assess cluster membership and youth.

§ 2.1 describes the catalogs we used to identify candidate members of open clusters. § 2.2 describes the catalogs we used to identify candidate members of moving groups, stellar associations, as well as young stars identified through combined Gaia photometry and astrometry. § 2.3 then reports summary statistics for the entire sample of about one million target stars.

### 2.1. Big catalogs: open clusters

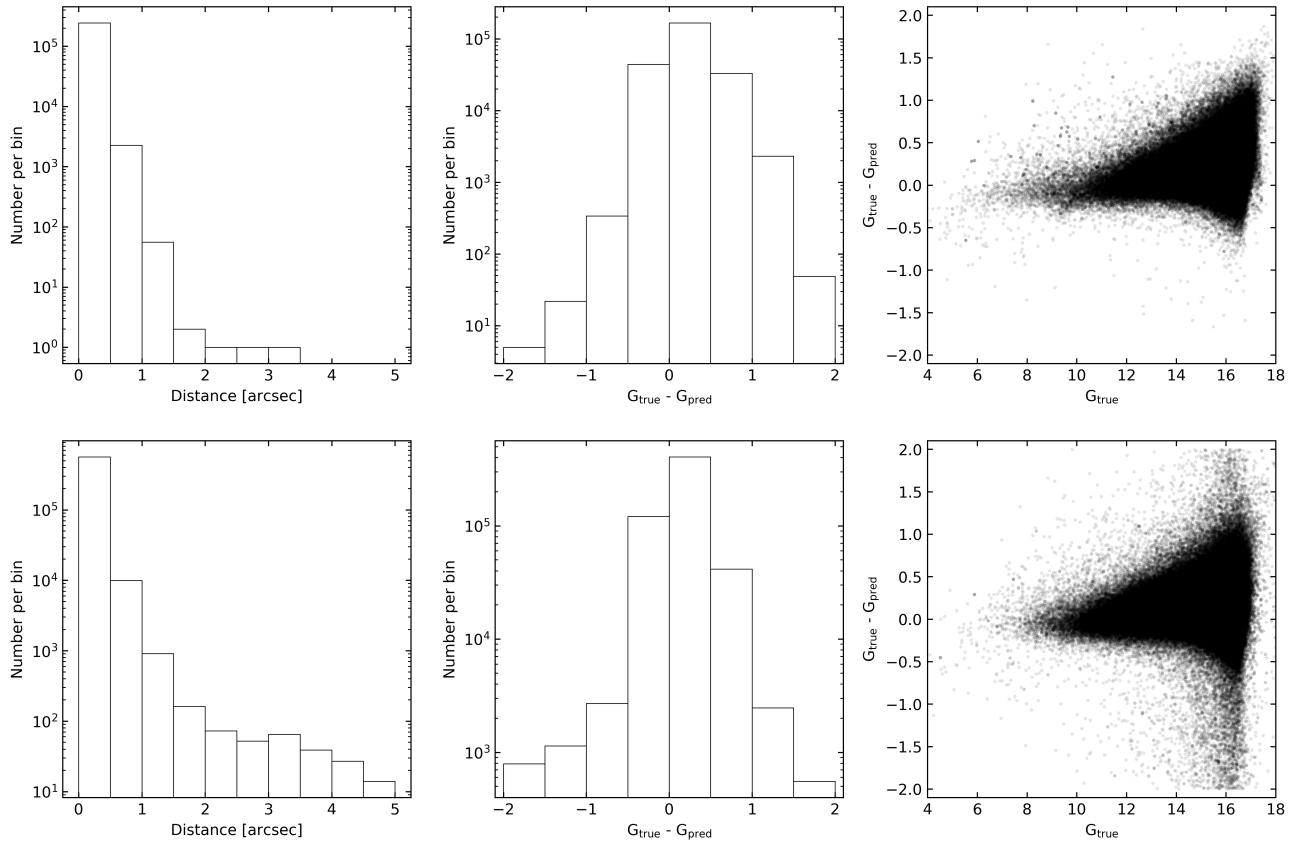
At the time of writing, two relatively large, homogeneous cluster memberships studies had been performed using *Gaia*-DR2: those by Cantat-Gaudin et al. (2018) and Gaia Collaboration et al. (2018a). There were also two large membership studies based on proper motion and photometric catalogs that were of interest: the studies of Kharchenko et al. (2013) and Dias et al. (2014).

*Gaia-derived OC memberships*—Cantat-Gaudin et al. (2018) used an unsupervised membership assignment algorithm to identify clusters in the three-dimensional astrometric space of proper motion and parallax. They used *Gaia* photometry and radial velocities to then verify the claimed membership properties. From their Table 2, we collect an initial 401,448 cluster members, in 1229 clusters, down to their limiting magnitude of  $G = 18$ .

*Gaia Collaboration et al. (2018a)* reported memberships for stars in a smaller, more select group of well-studied open clusters. From their Table A1, we collect 40,903 cluster members, in 41 open clusters, mostly within 500pc. While this work also included memberships for globular clusters, we omitted these from consideration.

In our photometric reduction, our default identifier for all sources is the *Gaia*-DR2 `source_id`. The TIC identifiers are found through a spatial cross-match after the light curves have been made (Stassun et al. 2018, 2019) This is because *Gaia*-DR2 is the base-catalog used to project sources from celestial coordinates to the imaging plane (§ 3). It also has the advantage that for any *Gaia*-derived cluster memberships, we perform cross-matching directly via the source identifiers.

*Pre-Gaia OC memberships*—Kharchenko et al. (2013) used proper motions calculated in PPMXL (Röser et al. 2010, a



**Figure 2.** Top. Quality diagnostics from cross-matching Kharchenko et al. (2013) cluster members against Gaia-DR2. histogram of the distances between matched stars is on the left; a histogram of the difference between the true  $G$ -band magnitude and that predicted from 2MASS photometry is in the middle; a scatter plot of the same magnitude difference as a function of  $G$ -band magnitude is on the right. Bottom. Same, but cross-matching Dias et al. (2014) cluster members to Gaia-DR2.

combination of USNO-B1.0 and 2MASS astrometry) and near-infrared photometry from 2MASS (Skrutskie et al. 2006) to report the existence of 2859 open clusters and stellar associations (globular clusters were omitted by excluding any entry of type 'g'). We selected “ $1\sigma$ ” members according to the combined photometric, kinematic, and spatial criteria described by Kharchenko et al. (2012). Then, to obtain *Gaia*-DR2 source identifiers for the members, we performed a crossmatch for *Gaia*-DR2 sources within 5 arcseconds of the listed positions. As an additional constraint, we used the 2MASS photometry to predict the  $G$ -band magnitudes<sup>2</sup>, and required that the measured  $G$ -magnitude fall within 2 magnitudes of the predicted  $G$ -magnitude. If multiple neighbors matched the position and magnitude constraints, we took the nearest spatial neighbor as the match. From 373,226 stars, this yielded a unique best neighbor for 352,332 stars (94.4% of the sample), and a choice between two neighbors for 17,774 stars.

The second (non-*Gaia* derived) open cluster membership catalog we used was the Dias et al. (2014) catalog, which was based on UCAC4 proper motions. From their 1805 reported open clusters, we selected sources with quoted membership probability above 50%. To obtain *Gaia*-DR2 source identifiers for the members, we performed a similar crossmatch as before, looking for sources within 5 arcseconds of the listed positions, and within  $\pm 2$   $G$ -band magnitudes of the prediction. From 2,034,269 stars, this yielded a unique best neighbor for 1,828,630 stars (89.9% of the sample), and a choice between two neighbors for 8.7% of the remaining sample.

The distributions of various cross-matching statistics are shown in Figure 2. The distances between matches is typically below 1 arcsecond. The Dias catalog shows somewhat stronger crowding effects at the faint end compared to the Kharchenko catalog. The Kharchenko catalog also has a more lop-sided distribution of true vs. predicted  $G$ -band magnitudes.

## 2.2. Smaller catalogs: moving groups and stellar associations

Stars, moving groups and stellar associations are of interest for similar reasons as stars in open clusters. Though fewer stars are known to exist in moving groups, they are of partic-

<sup>2</sup> See [https://gea.esac.esa.int/archive/documentation/GDR2/Data\\_processing/chap\\_cu5pho/sec\\_cu5pho\\_calibr/ssec\\_cu5pho\\_PhotTransf.html](https://gea.esac.esa.int/archive/documentation/GDR2/Data_processing/chap_cu5pho/sec_cu5pho_calibr/ssec_cu5pho_PhotTransf.html), online, 2019-03-29, or Carrasco et al. (2016)

ular interest because moving groups are less crowded than open clusters, and are often closer to the Sun.

We obtained Gaia-DR2 identifiers from the results of the following studies: Gagné et al. (2018b), Gagné et al. (2018a), Gagné & Faherty (2018), Kraus et al. (2014), Röser et al. (2011), Bell et al. (2017), Rizzuto et al. (2011), Oh et al. (2017), and Zari et al. (2018). The methods applied in these studies vary from kinematic analyses, to astrometric analyses included Gaia-DR1 parallaxes, to photometric searches for infrared excesses, to spectroscopic studies including RVs, H $\alpha$  emission, and Li absorption.

For the Gagné et al. catalogs, we searched the Gaia-DR2 archive for sources within 10 arcseconds of the listed positions. If Gagné et al. gave a proper motion, we required that the sign of each the Gaia proper motion components match that of the Gagné values (the stated proper motion uncertainties seemed to have been underestimated). We also imposed a  $G < 18$  cut on any putative matches. Of 3012 moving group members collected from the three combined Gagné et al. catalogs, we found 2702 matches.

The Kraus et al. (2014), Röser et al. (2011), and Bell et al. (2017) studies reported members in Tucana-Horologium, the Hyades, and 32 Ori respectively. Applying the same procedure as for the Gagné catalogs gave 187, 684, and 119 best-neighbors respectively, compared to 205, 724, and 141 initially reported members. Note that Kraus et al. (2014) found that only  $\sim 70\%$  of their listed members have spectroscopic indicators consistent with their membership in Tucana-Horologium.

Rizzuto et al. (2011) also focused on a single moving group: the Sco OB2 association. We used their reported Hipparcos identifiers, and matched against the *Gaia* archive’s `hipparcos2_best_neighbour` table, which gave 319 nearest-neighbor stars from 436 candidate members.

Next, Oh et al. (2017) searched for comoving stars in the  $\approx 2$  million stars that overlapped between Tycho-2 and *Gaia*-DR1. They found many wide binaries, and also identified a large number of comoving groups. We chose the 2,134 stars that they reported were in groups with sizes of at least 3 stars. Using their *Gaia*-DR1 source identifiers, we matched against the *Gaia* archive’s `dr1_neighbourhood` table, which gave 1,881 nearest-neighbor stars in groups of at least three stars (Marrese et al. 2019).

Finally, Zari et al. (2018) constructed a sample of young stars within 500 pc using data from Gaia-DR2. Two subsamples were made: (a) an upper main sequence (MS) sample, with 86,102 stars, and (b) a pre-MS sample, with 43,719 stars. Each was created from a careful combination of distinct astrometric and photometric cuts. These stars are the youngest, closest stars, spread across star-forming complexes in Sco-Cen, Orion, Vela, Taurus, and other regions of the sky. Though many are not strictly identified with moving groups or open clusters, their reported youth and proximity to star-forming regions justifies their inclusion in our search sample.

### 2.3. Summary of target stars

After collecting the aforementioned lists, we merged them into a single table. We then queried the `gaiadr2.gaia_source` table to retrieve their photometric  $G$ ,  $G_{Rp}$ , and  $G_{Bp}$  magnitudes, as well as their astrometric measurements ( $\alpha, \delta, \mu_\alpha, \mu_\delta, \pi$ ). Finally, we required that  $G_{Rp} < 16$ .

All told, this procedure yielded 1,061,447 unique stars, from 13 distinct membership catalogs.

107,647 of these stars, or about 10% of the collection, have cluster memberships reported by multiple authors. The largest number of stars come from Dias et al. 2014 (44.3% of stars), Kharchenko et al. 2013 (17.3%), Cantat-Gaudin et al. 2018 (16.7%), and Zari et al. 2018 (11.1%). Stars reported in multiple catalogs have all available reference information concatenated.

To assign unique cluster names, we adopted the name matched against Kharchenko et al. (2013) whenever possible. Appendix B describes how this was done in detail. For moving groups not identified in Kharchenko et al. (2013), we used the constellation-based naming convention from Gagné et al. (2018b). Otherwise, we used the name reported by the original catalog claiming membership. This process reduced 16,425 name permutations down to 3,216 unique cluster names. Though we have made every effort to avoid duplicates, a small number may remain, so we advise inspection of the `cluster` column as well as the references given in the `reference` column rather than using the `unique_cluster_name` column to analyze individual objects of interest. Nonetheless, 87.7% of the  $\sim$ million unique target stars are matched to clusters within Kharchenko et al. (2013), and 88.8% are assigned a cluster name. The remainder are mostly the young stars from Zari et al. (2018). Ages and their uncertainties are then merged against the parameters reported by Kharchenko et al. (2013).

The resulting CDIPS target star list is given in Table N. The cumulative distribution of target star brightnesses, as well as a histogram of the ages, is shown in Figure 1. Relative to field stars, our target star sample is young, with a typical age of 100 Myr.

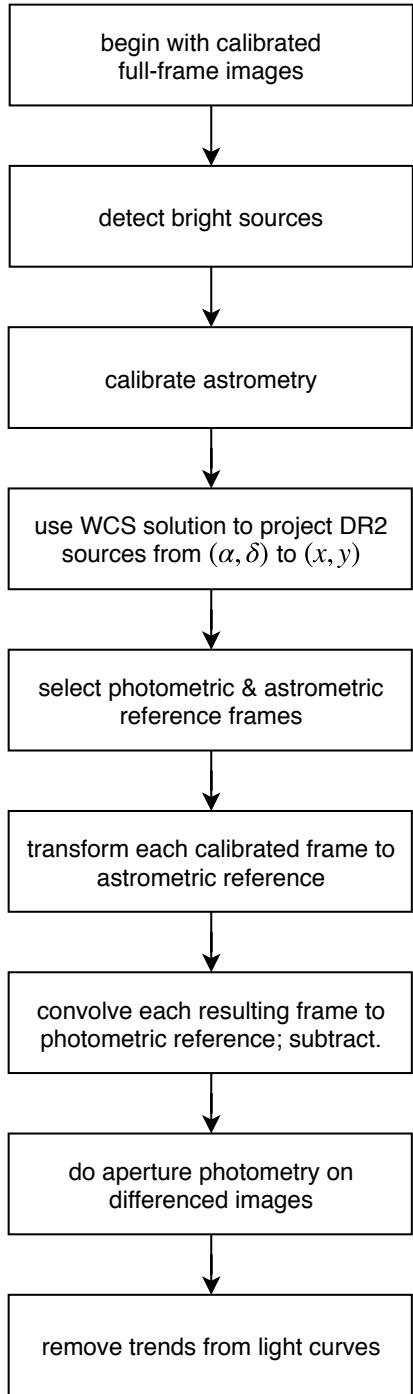
## 3. METHOD: PHOTOMETRY

### 3.1. Overview

To reduce the TESS images to light curves, we adopted an image subtraction approach. The overall method is in the spirit of the reduction pipelines developed by Pál (2009), Huang et al. (2015), Soares-Furtado et al. (2017) and Oelkers & Stassun (2018); a conceptual overview is given in Figure 3. Most sub-modules of our pipeline have been developed over the past decade to process HAT data (e.g., Bakos 2018). The specific high-level framework used for this reduction was adapted from a pipeline called `pipe-trex`, under development primarily for the HATPI project ([hatpi.org](http://hatpi.org)). The code is available online<sup>3</sup>.

We begin our processing with the calibrated full frame images produced by the Science Processing Operations Center

<sup>3</sup> [github.com/waqasbhatti/pipe-trex](https://github.com/waqasbhatti/pipe-trex), commit cd9f9bb.



**Figure 3.** Conceptual overview of photometric reduction pipeline. Details are given in § 3.

at NASA Ames (§ 3.2). We then perform a collection of preparatory steps, including source extraction of bright stars, astrometric verification, and coarse simple aperture photometry of bright stars (§ 3.3). Using the metadata collected from these initial steps, we select an astrometric reference frame to which we transform all of the calibrated images. We construct a photometric reference by stacking a subset of these

transformed frames, after having convolved the frames to have identical stellar profiles. Finally, we subtract each target frame from the photometric reference (§ 3.5). We perform aperture photometry on the subtracted images using positions projected onto the frame from the Gaia-DR2 source catalog. We detrend the resulting light curves (§ 3.7). The resulting white noise and red noise properties of the light curves, and a few interesting cases of variability, are discussed in § 4.

### 3.2. Observations

The TESS spacecraft began science operations on July 25, 2018. To keep its cameras pointed opposite the Sun, the spacecraft advances by  $\approx 28$  degrees east in ecliptic longitude every lunar month. Data acquired throughout each “sector” are downlinked at spacecraft perigee through the Deep Space Network. Verbose descriptions of the spacecraft’s design and operations are given by Ricker et al. (2015) and the instrument handbook (Vanderspek et al. 2018).

For us, the main data product of interest is the calibrated full frame image (FFI). Each TESS camera is read out every 2 seconds. To produce a manageable telemetric load, the resulting pixel values are averaged by the onboard computer into 30 minute exposures. An on-board cosmic ray mitigation algorithm is applied (Vanderspek et al. 2018, §5.1). Once transmitted to the ground, the raw images are calibrated by the Science Processing Operations Center. The calibration process includes an overscan, bias, and dark current correction, and also divides out a flat field. Details are discussed by Clarke et al. (2017), and the resulting science data products are described by Tenenbaum & Jenkins (2018).

We perform our processing using the calibrated images, their uncertainty maps, and their associated headers. The spacecraft has four cameras, and each camera has four CCDs. In the following analysis, all image-level operations are thus performed on the images for each CCD, so that at any instant of time there are 16 independent images that require analysis.

While we performed numerous initial tests on the first sectors of data, by geometric coincidence Sectors 1–5 were pointed away from the galactic plane. Consequently, less than 2% of the CDIPS target star sample was observed in the first five TESS sectors. Though a few interesting clusters are present in these observations (e.g., Blanco 1, NGC 2516, NGC 1901), for the present work we opted to focus on sectors in which there were more stars of interest for our intended science. These begin in Sector 6 (2018-12-12, spacecraft orbit #19). For this run of processing, they conclude at the end of Sector 7 (2019-02-01, spacecraft orbit #22).

### 3.3. Image preparation & background removal

Before we can perform any kind of photometry, a few janitorial tasks are required. First, we convert the calibrated image from MAST into a single-extension FITS image, and trim to remove virtual rows and columns using the SCIRWES, SCIROWE, SCCSA, and SCCED header values.

In order to pre-emptively address the background variations present in some frames due to scattered light from the Earth and Moon (see Vanderspek et al. 2018, §7.3.1–7.3.4),

we then estimate and subtract the large-scale background. We do this by temporarily masking out pixels more than  $2\sigma$  from the image median, and then pass a  $48 \times 48$  median box filter over each pixel in the image, with reflective boundary conditions. We blur the resulting background estimate with a gaussian kernel, which produces a smooth background estimate for each image. These steps also remove a low-level vignetting present in the corners of many images, which remains even after flat-fielding (see Vanderspek et al. 2018, §7.3.5). With the exception of scattered-light caustics, which remain over small areas of less than 10% of the frames, our ad-hoc procedure seems to adequately remove large spatial scale scattered light patches.

After subtracting the background, we mask out saturated stars using a fixed saturation level of  $8 \times 10^4$  ADU. This value was chosen based on the onset of bleeding charge trails in the images, and is slightly greater than the expected saturation level quoted by Vanderspek et al. (2018). As described by Pál (2009), our masks are metadata to the image, and are only applied to the pixel values during the specific image processing steps in which they are necessary (*e.g.*, convolution). We also extend the masks beyond purely saturated pixels to “bloomed” pixels horizontally and vertically adjacent to the saturated pixels (see Figure 6 of Pál 2009).

Finally, for frames with the DQUALITY bit-flag corresponding to the “momentum dumps” and “coarse pointing modes” described by Vanderspek et al. (2018), we omit the entire frame. This removes on average a few frames per sector. Through visual inspection, we see that the stars on these frames are extremely smeared, and are unlikely to produce useful science data. In addition, we use the sector-specific data release notes<sup>4</sup> to identify further times with anomalous spacecraft performance, which we omit from consideration. For Sectors 6 and 7, this included three days at the beginning of Sector 6 dedicated to acquiring pixel response function data, and no additional gaps in Sector 7.

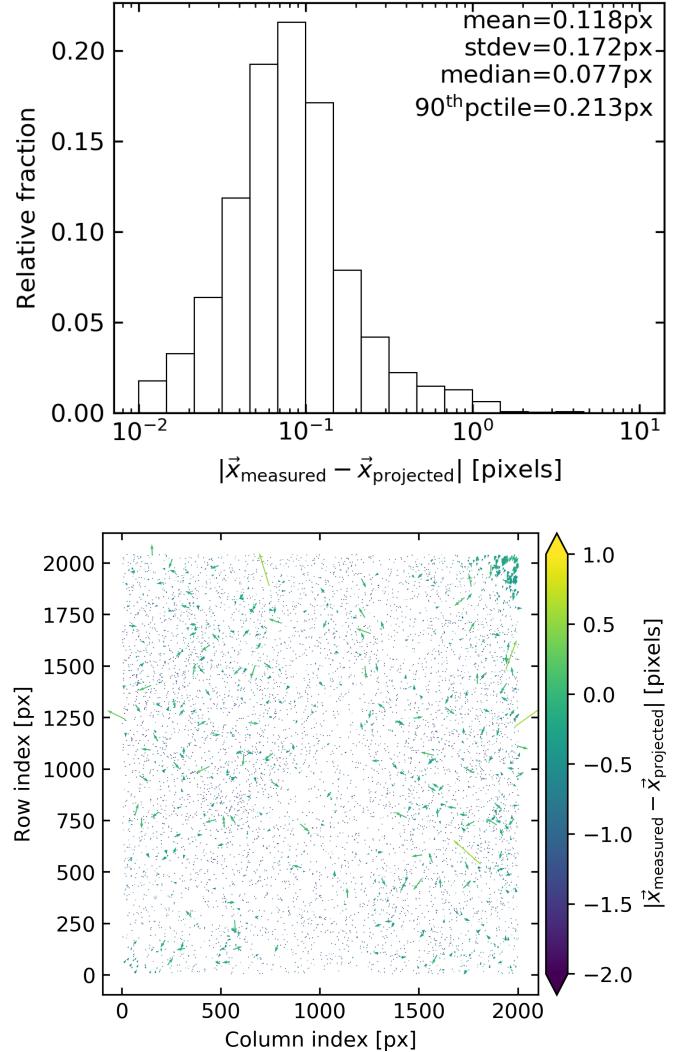
### 3.4. Metadata collection & WCS verification

Having prepared the images, we perform some initial analysis steps to produce metadata needed during image subtraction.

First, we use `fistar` to perform source extraction on bright stars in each image. During the initial source extraction, we also fit elongated gaussians to the bright stars, yielding the shape parameters ( $s, d, k$ ), where the flux  $f$  as a function of position  $(x, y)$  in the CCD image plane is assumed to take the form

$$f(x, y) = B + A \exp\{-0.5 \times [s(\Delta x^2 + \Delta y^2) + d(\Delta x^2 - \Delta y^2) + k(2\Delta x\Delta y)]\}, \quad (1)$$

for  $(x_0, y_0)$  arbitrary reference positions,  $\Delta x = x - x_0$ ,  $\Delta y = y - y_0$ ,  $B$  is the background level, and  $A$  is an arbitrary flux-scaling constant. For a nearly circular shape profile, the



**Figure 4.** *Top.* Histogram of astrometric residual. The  $x$ -axis shows the distance between the measured centroid positions of stars, compared to the predicted positions from the WCS solution. *Bottom.* Vector plot of astrometric residual. Each arrow is the vector from the measured to the projected star position. Directions are correct, but lengths are 50 times their true size for visualization purposes. The systematic error in the top-right corner is a typical problem generic to wide-field astrometry. The frame chosen for this plot is the photometric reference frame used for Sector 6, Camera 1, CCD 1; we automatically impose cutoffs on the median and 90<sup>th</sup> percentile of the astrometric residual in order to ensure similar levels of astrometric precision are maintained throughout the reduction.

sharpness  $s$  is related to the FWHM as  $\text{FWHM} \approx 2.35\sqrt{s}$  (*e.g.*, Pál 2009). These shape parameters are later used when selecting an astrometric reference frame (§ 3.5). In agreement with what is obvious upon visual inspection, this fitting process shows that stars closer to the center of each camera’s field are round, while stars near the field edges are more triangular.

<sup>4</sup> [archive.stsci.edu/tess/tess\\_drn.html](http://archive.stsci.edu/tess/tess_drn.html)

For the astrometric solution, we use the WCS and SIP coefficients derived by SPOC and included in the FFI headers (Pence et al. 2010, Sec. 8). We explored the possibility of using `astrometry.net` (Lang et al. 2010) to derive our own astrometric solutions for each frame, but found that the astrometric residual (the mean separation between projected and measured positions) was consistently a factor of 1.5-2 times higher in our WCS solutions than in those given by SPOC. This was perhaps because we did not invest major amounts of time developing a robust algorithm to select non-blended stars of intermediate brightness before measuring their positions.

With the resulting WCS information, we then project a source catalog onto each frame. Initially, we planned to photometer all Gaia-DR2 sources in each field down to a cutoff of  $G_{Rp} < 16$ . However, we found that for the galactic plane fields this produces an excessively large number of sources (millions of stars per  $12^\circ \times 12^\circ$  CCD). We therefore limited our source catalog for each frame to be a combination of the CDIPS target stars ( $G_{Rp} < 16$ ), and all Gaia-DR2 sources down to  $G_{Rp} < 13$ . We use the projected positions of these sources to center the apertures in our photometry, rather than attempting to detect the positions. Such “forced-aperture photometry” is preferable to source extraction in the crowded fields that are central to this work. The Gaia-DR2 epoch is J2015.5, so even the fastest-moving stars with proper motions of  $\sim 1\text{ arcsecond yr}^{-1}$  are still well within one pixel of their predicted positions in the TESS images. The projection from catalog sky-coordinate positions to pixel coordinates is performed using an analog of the `wcs-rd2xy` program that performs the standard matrix algebra (Lang et al. 2010). The source catalog look-up is performed using `gaia2read`<sup>5</sup> (Kim 2018).

The astrometric residual is displayed for one photometric reference frame in Figure 4. The construction of this frame will be described shortly, however a few salient points can first be made. First, the typical median precision of the WCS solution is a bit below 0.1 pixels, and its 90<sup>th</sup> percentile is typically less than 0.3 pixels<sup>6</sup>. The errors are typically largest in the corners of the field of view, where the non-linearity of the focal plane is most significant, and where the corrections required by the SIP coefficients are largest.

As a final step of metadata collection, we use `fiphot` to perform aperture photometry on the bright stars from the source list, by summing the counts inside appropriately-weighted circular apertures centered on the projected positions. The pixel weights are equal to the fraction of the pixel that falls within the aperture. They are thus unity for pixels entirely within the aperture, and fractional along the aperture boundary. The background levels are measured in annuli around each aperture center. The resulting measurements, for instance of the background level of each aperture, and the

number of stars with “good” photometry, are later used to select photometric reference frames.

### 3.5. Image subtraction

#### 3.5.1. History of image subtraction method

The core operation of “classical” image subtraction attempts to match a photometric reference  $R$  and a target image  $I$  by computing and applying a convolution kernel. For ground-based data, this “match” typically corrects for differences in seeing or transparency between the reference and target; for space-based data, the match might correct for spacecraft jitter, or thermal and corresponding point-spread function (PSF) variations. The kernel, once applied to the high signal-to noise reference, produces a model image,  $M_{xy}$ ,

$$M_{xy} = (R \otimes K)_{xy} + B_{xy}, \quad (2)$$

where  $B_{xy}$  is a component of the model image that allows for background variations, and  $\otimes$  denotes convolution. Since we modeled the background separately (§ 3.3), we set  $B_{xy} = 0$ . The convolution kernel  $K$  is typically decomposed onto a basis,  $K = \sum_i c_i K_i$ , and the coefficients are found by minimizing

$$\chi^2 = \sum_{xy} \left( \frac{I_{xy} - M_{xy}}{\sigma_{xy}} \right)^2, \quad (3)$$

where  $\sigma_{xy}$  is the uncertainty in the target image pixel values. Photometry is then performed on the difference image  $D_{xy}$ , where  $D_{xy} = I_{xy} - M_{xy}$ . For the present reduction, the uncertainty in each target image pixel was taken to be a constant.

The procedure described above was first proposed by Alard & Lupton (1998). It was reviewed and clarified by Miller et al. (2008b). The choice of how to decompose the kernel was further explored by Bramich (2008), who showed that using a delta function had a few select advantages compared to a basis of gaussians. We perform the convolution using `ficonv`, and opt for the implementation of Bramich’s method (see Pál 2009 § 2.8).

#### 3.5.2. Astrometric registration

We now describe a few implementation details we used to make the above high-level picture work. First, we needed to select two “reference frames” for image subtraction: (1) the astrometric reference; and (2) the photometric reference.

To choose the astrometric reference, we searched for frames with large, round stars (big  $s$ , small  $d$  and  $k$  values). We also required that the frame have minimal background noise, as measured in annuli around the bright stars selected in § 3.3. Finally, the astrometric reference frame needed to have, relative to the other frames being considered, a large number of detected sources. We sorted the frames using these metrics, and then selected the astrometric reference from successive intersections of each sorted list.

We then used the `grmatch` and `grtrans` tools to compute and apply a spatial transformation of each calibrated frame to match the astrometric reference frame. The transformation is calculated using the measured source positions

<sup>5</sup> [github.com/samuelyeewl/gaia2read](https://github.com/samuelyeewl/gaia2read)

<sup>6</sup> In fact, in our reduction we automatically impose that the median residual and 90th percentile remain below 0.2 and 0.4 pixels, respectively.

found in § 3.4, and the symmetric triangle point-matching scheme described by Pál (2009, § 2.5.2). The transformation – a combination of rotation, dilation, and translation – typically moves stars by less than a pixel, since the TESS spacecraft is quite stable. The precision of the transformation is guaranteed by forcing its unitarity  $\Lambda$  to always be below 0.01, where  $\Lambda$  is defined by Pál (2009, Eq. 54). We reduced the photometric errors incurred during this step by using the flux-conserving interpolation scheme described by Pál (2009), which was necessary because polynomial interpolation schemes do not conserve stellar flux.

### 3.5.3. Photometric reference frame construction & reference flux measurement

The second required reference frame is the photometric reference, which is used both to calculate the convolution kernel, and to obtain a reference flux for each star. To make it, we first chose 50 sub-frames with low scatter in their photometry and low background measurements, according to the metadata collected in § 3.3. We then convolved these candidate photometric reference frame to the best frame, and performed a median average over the entire stack to make the photometric reference.

Measuring the reference flux for each star is a slightly non-trivial operation. First, we perform forced simple aperture photometry to measure the flux for each source. The local background is estimated in annuli, with neighboring stars masked out during the background measurement. However if we were to stop here, *it would be a bad mistake*. The reference fluxes for faint stars would be overestimated, due to crowding. The relative amplitude of photometric signals would correspondingly be biased small, hindering exoplanet detection.

Therefore, after performing simple aperture photometry on the reference frame, we fitted an aperture-size specific zero point between the catalog  $T$ -band magnitude of each star, and its measured flux. The  $T$ -band magnitude was calculated according to Equation 1 of Stassun et al. (2019). We then use the known catalog magnitude to predict the expected reference flux for each star. In effect, this minimizes crowding down to Gaia’s resolution limit of  $\approx 1''$ , rather than the TESS limit of  $\approx 20''$ .

Following Pál (2009) Equation 83, the final instrumental flux values  $f$  we report<sup>7</sup> are then equal to

$$f = f_{\text{reference}} + f_{\text{subtracted}} \quad (4)$$

$$= g(T_{\text{cat}}) + \frac{1}{||K||_1^2} \sum_{x,y} D_{x,y} (w \otimes K)_{x,y}. \quad (5)$$

The function  $g$  takes as input the target star’s catalog magnitude  $T_{\text{cat}}$ , and returns the reference flux. Its coefficients have been fit per the procedure just discussed. The difference image  $D$  is equal to  $I - R \otimes K$ , where as in Equation 2,  $I$  is the

target image transformed to the astrometric reference,  $R$  is the photometric reference, and  $K$  is the convolution kernel. The weights  $w$  from the circular aperture mask are included in the convolution. The norm  $||K||_1$  is defined by Pál (2009) Equation 81.

### 3.6. Choice of convolution kernel

A few brief notes on the algorithm to actually solve for the coefficients to the convolution kernel. The procedure implemented in `ficonv` is to grid the image, and within each grid element find the brightest non-saturated star. These isolated “stamp” stars are then used to solve for the coefficients of the kernel, by minimizing Equation 3 over the sum of all stamps.

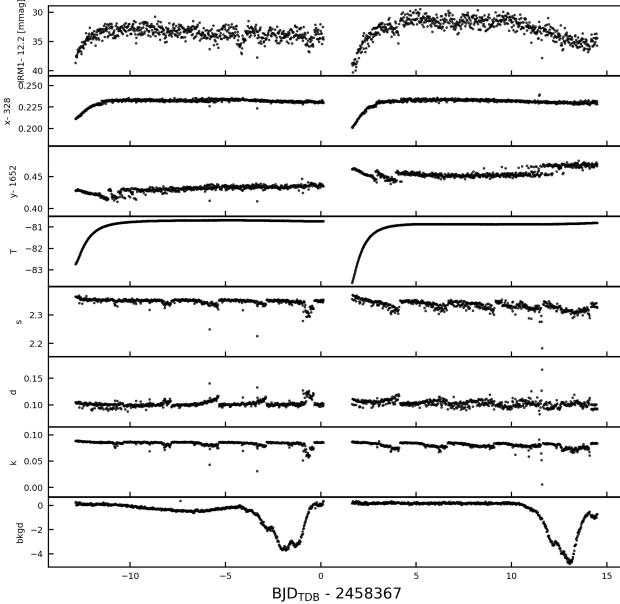
For the choice of basis, we opt for a delta-function kernel with an additive flux scaling term (Soares-Furtado et al. 2017 Section 3.3.1 gives the equations). The spatial variations of the PSF are captured by weighting the delta function and the flux scaling terms with varying polynomial orders across the image. Choosing this kernel introduces three additional free parameters: (1) the kernel box-size; (2) the maximum order of the polynomial weighting the delta function terms; (3) the maximum order of the polynomial weighting the flux scaling. We performed a grid-search to tune these parameters, in which our metrics for success included the measured RMS as a function of magnitude, and also the recovered SNR of transits from the catalog of known TOIs (TESS Objects of Interest; N. Guerrero, in prep).

Varying the kernel box-size, we found that kernels with box-length smaller than the typical TESS FWHM at field center ( $\approx 3$  pixels) produced light curves with the highest average scatter. Increasing the kernel box-size from a  $3 \times 3$  box to a  $7 \times 7$  box led to about a 50% reduction in RMS for bright stars, and no difference for faint stars. The largest kernels, of  $(11 \times 11)$  pixels, had on average slightly lower signal-to-noise for recovered transits than for kernels of intermediate size. We settled on a kernel box-size of  $(7 \times 7)$  pixels, in part because the FWHM at the camera corners can grow by factors of  $\approx 2$  relative to the field center.

Varying the polynomial orders, we found that the highest order polynomials retrieved transits with  $\approx 10\%$  worse SNR compared to lower order polynomials. Varying the polynomial orders between 1 and 4 did not produce large differences.

Averaging over all TOIs present in the camera we used for these experiments, we found that different choices of kernel parameters produced variations of  $\lesssim 12\%$  in the retrieved transit SNR. For computational expediency, we therefore chose a single  $(7 \times 7)$  kernel with second-order spatial polynomial weights in the basis functions for the remainder of our reduction. However, we caution that within our parameter-tuning experiments, the recovered SNR of perhaps a third of the TOIs varied by up to a factor of two, which in a few cases led to non-detections of transit signals near the noise floor. In the longer term, developing an image-subtraction method that marginalizes over uncertainties of how to chose “optimal” kernels would be desirable. Pixel-level image sub-

<sup>7</sup> For example, the `TFL1` column.



**Figure 5.** Timeseries of “external parameters” for a representative star over one sector. *Top:* Instrumental raw magnitude (with a particular aperture size), as a function of time. *Second and third from top:*  $x$  and  $y$  centroid positions as a function of time. Continuing in order are the CCD temperature, the  $(s, d, k)$  shape parameters, and the measured background value. Most of the apparent variability is instrumental.

traction methods that omit these parameters entirely are also worth exploring (Wang et al. 2017).

With a kernel selected, and the convolution and subtraction performed, we calculate the instrumental fluxes for each frame per Equation 5. We did this with three different aperture sizes: for this work, circles of radii 1 pixel, 1.5 pixels, and 2.25 pixels. These sizes were chosen to roughly span the range of optimal aperture sizes for stars in our sample, as calculated from the pre-flight Sullivan et al. (2015) noise model. Finally, to convert from a list of flux measurements for each source on a frame to a list of flux values at any given time, we used the `grcollect` transposition tool.

### 3.7. Light curve detrending

The preceding steps produce light curves that include both instrumental systematics as well as astrophysical variability. The detrending approach adopted by the HAT group to remove systematic variability typically proceeds in a few sequential stages, as discussed by *e.g.* Bakos et al. (2010), Huang et al. (2015), and Zhang et al. (2016).

The first detrending step commonly performed on ground-based data is “magnitude-fitting”: the raw magnitudes measured from the difference images are fit by a polynomial that depends on a combination of CCD position, sub-pixel position, and optionally catalog magnitude and color (Zhang et al. 2016, § 5.5).

The second step is to decorrelate against external parameters that are known to affect the stellar flux measurements

(EPD, Bakos et al. 2010). For ground-based data this may include zenith angle, or changing PSF parameters. For TESS data, this might include CCD temperature, or the angles of the Moon and Earth relative to each camera’s boresight. Some example “external” parameters that we include with our light curves are shown as functions of time in Figure 5.

The final step is to then decorrelate the measured brightnesses of stars against each other (TFA, Kovács et al. 2005). This accounts for variations due to unknown systematic instrument changes that affect many stars. This latter step is effectively a principle component analysis of all the light curves, in which the components are chosen as a subset of the light curves. In this sense TFA is similar to the PDC-MAP and ARC2 methods (Smith et al. 2012; Aigrain et al. 2017) except that these latter two methods are much more careful in their choice of basis vectors.

We explored the possibility of fitting linear models of flux as functions of *e.g.*, temperature, shape parameters, centroid positions, and stellar colors. We also briefly explored non-linear model fitting using  $N$ -dimensional B-splines to fit the flux, centroid positions, and temperatures simultaneously (Dierckx 1996). The linear models typically underfit the light curves, particularly during the large shifts that happen as the spacecraft nears perigee. The non-linear models showed some promise, but often seemed to overfit stellar variability signals. One thought that we did not try but may explore in future work is to decorrelate against the scatter of the spacecraft quaternion time-series (Vanderburg et al. 2019).

Given these complications, for the time being we omit the step of “detrending” as a function of external parameters. To enable further exploration of the issue, we include all the necessary vectors of *e.g.*, centroid positions, temperatures, and shape parameters in our reported light curves.

Since most of the external parameter dependence is shared between stars, we opt instead to decorrelate the flux time-series of each star against other stars in the frame. We use the TFA algorithm proposed by Kovács et al. (2005), which for self-consistency we describe here.

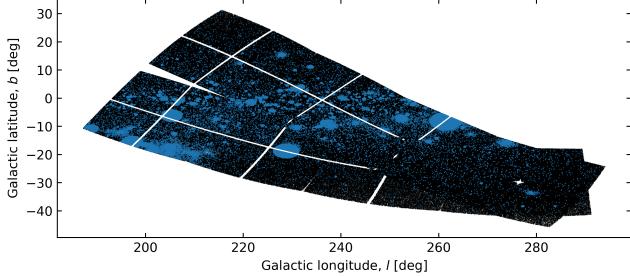
The idea of the method is as follows. Suppose we have  $M$  “template stars”, which are a subsample of stars that represent all types of systematics across the dataset. Each template star has a light curve with  $N$  data points. Denote the template time-series  $X_j(i)$ , where  $j = 1, \dots, M$  and  $i = 1, \dots, N$  is the time index. We then want to find periodic signals in a target time-series  $Y(i)$ . This is done by defining a filter function

$$F(i) = \sum_{j=1}^M c_j X_j(i), \quad (6)$$

for which the coefficients  $c_j$  are found by minimizing

$$\mathcal{D} = \sum_{i=1}^N [Y(i) - A(i) - F(i)]^2. \quad (7)$$

When trying to find periodic signals,  $A(i)$  represents our prior knowledge of the light curve’s shape. This prior is simply



**Figure 6.** Positions of light curves from TESS Sectors 6 and 7 in galactic coordinates. Black:  $G_{Rp} < 13$  field stars. Blue:  $G_{Rp} < 16$  target stars. Target stars are mostly near the galactic plane. The data for Sectors 6 and 7 cover about one-sixth of the galactic plane.

that stars on average maintain a constant brightness:

$$A(i) = \langle Y \rangle = \frac{1}{N} \sum_{i=1}^N Y(i) = \text{const.} \quad (8)$$

If a signal is eventually found, for instance using the box-least squares method (Kovács et al. 2002), this detrending process must then be repeated while accounting for our updated knowledge about the light curve’s shape.

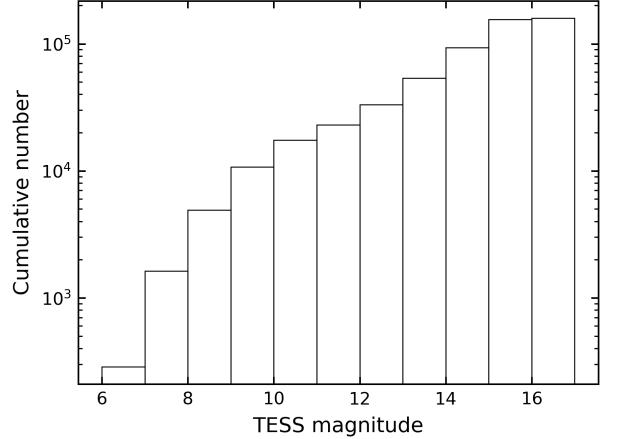
Some implementation notes follow. We selected template stars in two stages. In the first stage, we fitted a parabola in the RMS-magnitude plane, and discarded stars more than  $2\sigma$  away from the prediction of the fit. We also required that these initial candidate stars have intermediate brightness ( $8.5 > T > 13$ ), and have a relatively large number of time-series data points. We then performed an initial iteration of TFA, on only the candidate template stars. We inspected the resulting detrended light curves for residual structure by computing a Lomb-Scargle periodogram. If the maximum-power peak had a false alarm probability below 0.1%, we excluded the star from the list of candidate template stars, on the basis of its presumed periodic variability. We then randomly selected at most 200 template stars from the remaining non-variable candidates. The choice of number of template stars was discussed by Kovács et al. (2005), and is another free parameter in the broad problem of light curve production. This choice is analogous to the issue of how many cotrending basis vectors to choose in PDC-MAP or ARC2 (Aigrain et al. 2017). While the number of template stars can be optimized by constructing and minimizing a BIC-like quantity, a little overfitting is acceptable for our purpose of finding planetary transits.

Once the template stars were selected, we use the VARTOOLS program to perform the detrending (Hartman & Bakos 2016). This process was performed for each photometric aperture separately.

## 4. RESULTS

### 4.1. Light Curve Statistics

#### 4.1.1. Stellar properties



**Figure 7.** Cumulative number of CDIPS light curves a function of TESS  $T$ -band magnitude. Light curves were made for the target stars (Figure 1) that were observed in Sectors 6 or 7.

The on-sky locations of the light curves from Sectors 6 and 7 are shown in Figure 6. Black stars are the field stars for which we performed photometry, and of which a subset were used as template stars during TFA. The blue stars are the 159,337 stars with light curves available from archive.stsci.edu/prepds/cdips. Most of the target stars are quite close to the galactic plane. Gaps between CCD chips are also visible. The highly clumped nature of the target stars is also apparent.

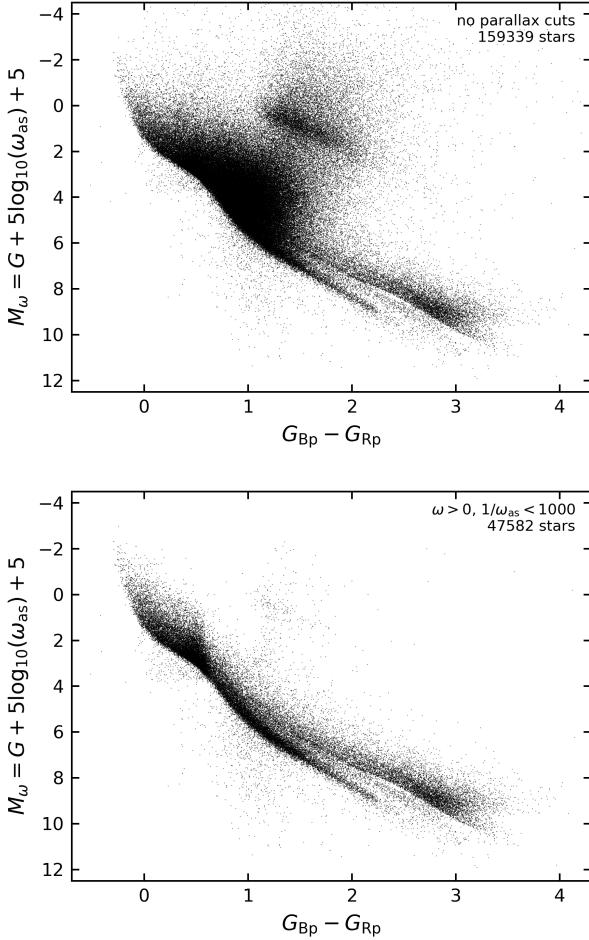
Figure 7 shows the cumulative distribution of TESS  $T$ -band magnitudes for the target stars. Though most of the targets are faint,  $\approx 3 \times 10^4$  are brighter than  $T = 13$ . These targets are more promising for any detailed follow-up observations.

An HR diagram for the entire sample of stars on silicon is shown in Figure 8 (top). The sub-sample of stars with measured positive parallaxes and naive distances less than 1 kpc is also shown (bottom). About one-third of the stars are in this latter sample. The close stars are predominantly on the main-sequence, or the pre-main-sequence. A relatively large fraction of these come from Zari et al. (2018), and are either on the PMS or upper main-sequence. The latter set of OBA dwarf stars, while “younger” than the typical field dwarf, are likely the least interesting subset of our target sample from the perspective of age analyses. In the entire sample (Figure 8 top), a much larger fraction of stars are sub-giants, red giants, and helium-burning red-clump stars. We failed to obtain light curves for roughly 5% of the stars on silicon, primarily due to our masking of saturated stars and their nearby pixels.

Finally, Figure 9 shows the proper motions of the entire and close samples of stars on silicon. Each clump signifies a different star cluster. The central blob in the top figure is in fact two separate blobs, due to the different directions in the Galaxy being observed in Sectors 6 and 7.

#### 4.1.2. Cluster membership provenance

**Sector 6**—In Sector 6, 67,601 light curves of candidate cluster stars were made. The provenance of the claimed cluster

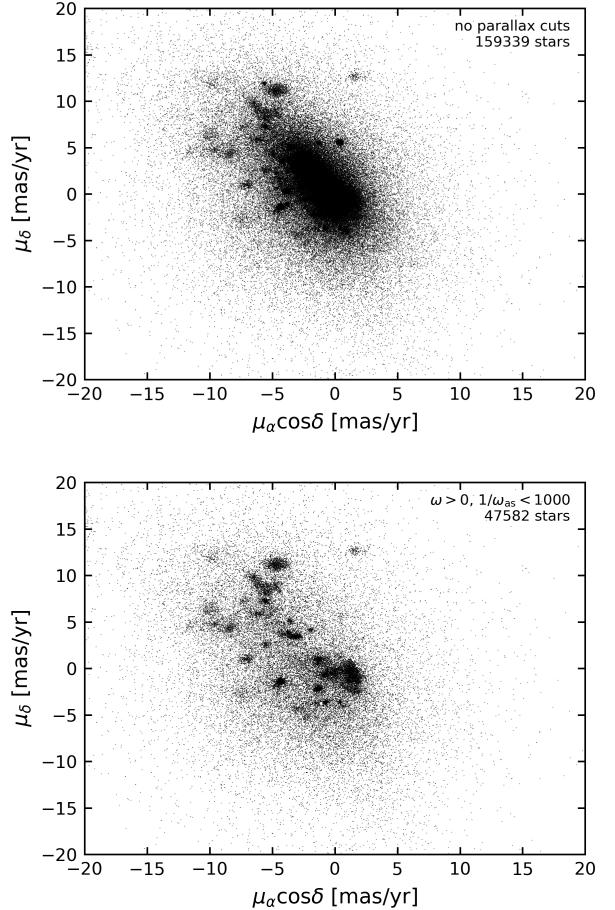


**Figure 8.** *Top.* HR diagram of CDIPS stars on silicon in this data release. *Bottom.* HR diagram of close CDIPS stars on silicon. The wedge separating the pre-MS sample from the MS stars was discussed by Zari et al. (2018), who introduced it in order to avoid contamination by photometric binaries.

origin of these sources is Dias et al. (2014) for 59% of the sources; Zari et al. (2018) for 16% of sources from their upper main-sequence table and 1% of sources from their PMS table; Kharchenko et al. (2013) for 8% of sources, Cantat-Gaudin et al. (2018) for 6% of sources, and more than two catalogs for the remaining 10% of sources.

The clusters with the largest claimed numbers of sources are the Platais 6, Platais 5, and Mamajek 3 moving groups, all from Dias et al. (2014), composing about 7000, 7000, and 3500 sources respectively. These membership claims should be regarded with some skepticism on a source-by-source basis. For Platais 6, Kharchenko et al. (2013) claimed only about 400 probable members ( $1\sigma$ ) to exist within the angular radius of the cluster. Mamajek 3 (= 32 Ori) has only about 50 confirmed members (Bell et al. 2017).

We remind the reader that our goal in creating our target catalog was completeness, rather than fidelity. The Dias et al. (2014) stars in particular were included if they were



**Figure 9.** *Top.* Proper motions of CDIPS stars on silicon in this data release. Many of the stars in the central “blob” are possible field-contaminants. *Bottom.* Proper motions of close CDIPS stars on silicon.

listed with membership probability exceeding 50%. To create cleaner sub-samples, we advise use of the CDEXTCAT header keyword, which can be used to merge against the original source catalog to obtain the membership probabilities claimed by the original catalog.

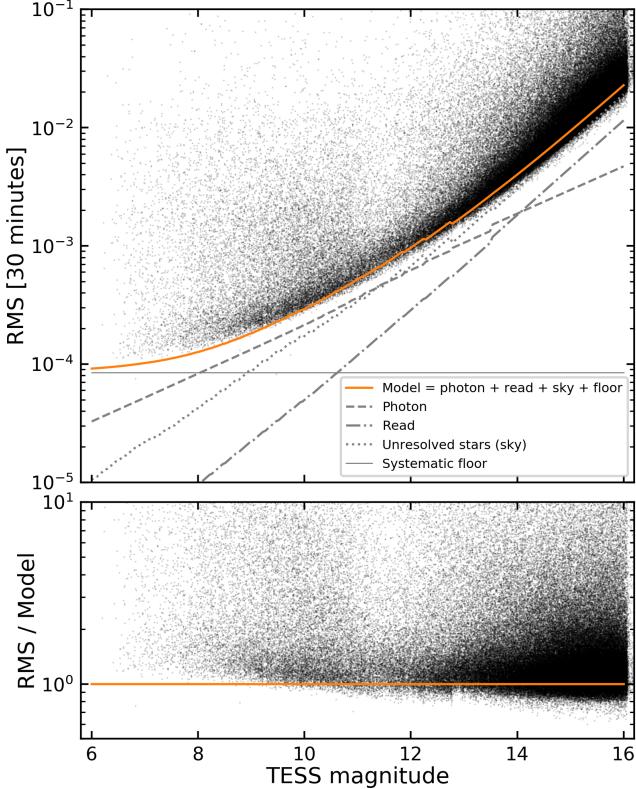
*Sector 7*—In Sector 7, 91,736 light curves of candidate cluster stars were made. The provenance of the claimed cluster origin of these sources is ... LGBTODO

The clusters with the largest claimed numbers of sources are ... LGBTODO

#### 4.1.3. Light curve noise properties

*Observed RMS vs magnitude.*—The standard deviation of the TFA-detrended light curves is plotted as a function of the catalog  $T$ -band magnitude for CDIPS light curves in Figure 10. For the y-axis of this plot, we have taken

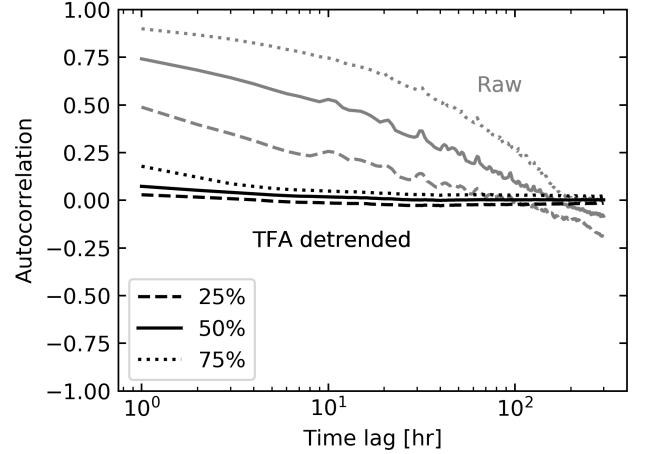
$$\text{RMS} = \left[ \frac{1}{N-M-1} \sum_{i=1}^N (f_i - \bar{f})^2 \right]^{1/2}, \quad (9)$$



**Figure 10.** Standard deviation of trend-filtered CDIPS light curves as a function of catalog TESS-band magnitude. Black points correspond to the minimum RMS across the available three aperture sizes. The model (orange and gray lines) assumes aperture sizes reported by Sullivan et al. (2015), and the effective area from Vanderspek et al. (2018). The noise from unresolved background stars (dotted gray line) is a function of galactic latitude, and dominates over zodiacal light for faint stars near the galactic plane; the line shown assumes a sight-line towards the center of Sector 6, Camera 1 (further details are in § 4.1.3).

where  $f_i$  is the value of the flux at the  $i^{\text{th}}$  point in the time-series,  $\bar{f}$  is the mean flux,  $N$  is the number of points in the time-series, and  $M$  is the number of template light curves used during TFA detrending. The correction in the denominator penalizes the natural degree of overfitting inherent to the TFA algorithm.

The observed RMS (black points) follows the usual shape, with photon noise dominating from  $T = 9$  to  $T = 12$ , beyond which the onset of the “sky” background changes the overall slope of the curve to be more steep. For the brightest stars ( $T \lesssim 9$ ), a “systematic floor” was an expected part of the mission’s error budget (Ricker et al. 2015), but has not been observed in early reports of the photometric performance of various aperture photometry pipelines (e.g., the SPOC pipeline Jenkins et al. 2010, the MIT-QLP Huang et al. 2018, and eleanor Feinstein et al. 2019). The fact that our light curves for the brightest stars are above this purported “floor”, rather than below it, suggests that our image subtraction techniques could be introducing some small degree of



**Figure 11.** Average autocorrelations of  $10^4$  randomly selected light curves from each sector. The 25<sup>th</sup>, 50<sup>th</sup>, and 75<sup>th</sup> percentiles at time lags spaced by one hour are shown by dashed, solid, and dotted lines respectively. The raw light curves (column name `IRM`) are shown in gray, and the TFA-detrended light curves are in black.

noise to the light curves of the brightest stars. It is also true however that our largest aperture contains only about 16 pixels, which is sub-optimal for stars brighter than  $T \approx 9$  (see Sullivan et al. 2015, Figure 14). Since the brightest stars are not the focus of the present work, we leave this issue unaddressed for the time being.

Importantly, the faint stars do not noticeably exhibit the typical effects of crowding at the faint end (e.g., Feinstein et al. 2019, Figure 5). In aperture photometry pipelines, stars in very crowded regions typically have their flux overestimated relative to the actual number of photons that fall on silicon. This leads to systematic underestimates of the uncertainty in the relative fluxes, as well as “flux contamination” (the reduction in amplitude of say, transit signals, due to diluting flux from neighbor stars). Our method to work around this problem – using the catalog magnitudes to predict the reference flux values, and measuring deviations from these reference fluxes on the subtracted images – seems to be performing well.

*Expected RMS vs magnitude.* —The noise model shown in Figure 10 is quite similar to that of Sullivan et al. (2015), save for two changes. The first change is that the effective area of the telescope is updated to be  $86.6\text{ cm}^2$ , per the measurements by Vanderspek et al. (2018).

The second change is that we have explicitly included the estimated noise contribution from unresolved faint stars. The brightness of the diffuse sky is dominated by different sources at different wavelengths. For instance, the CMB is most important in the microwave, and thermal radiation from dust grains in the solar system (zodiacal light) is dominant in the far infra-red (Leinert et al. 1998). In the TESS-band, both zodiacal light and faint stars can play a role, depending on the line of sight under consideration. The zodiacal light is brightest near the ecliptic plane, and the faint star

background is brightest near the galactic plane (and towards the galactic center). When performing pre-launch noise estimates, Winn (2013) estimated the photon-counts from each component. His zodiacal light model was presented by Sullivan et al. (2015), but the faint star component was not emphasized since the Sullivan simulations were performed away from the galactic plane.

The diffuse sky model we have used for Figure 10 is adopted explicitly because most of our target stars are near the galactic plane. Stars are judged to be “unresolved” and part of the background if their surface density exceeds the angular resolution of the telescope. TESS has an angular resolution of  $\Delta\theta \sim 1'$ , set by a combination of the pixel size as well as the typical stellar FWHM. Sources with sky surface density exceeding  $\Delta\theta^{-2}$  therefore contribute to the background.

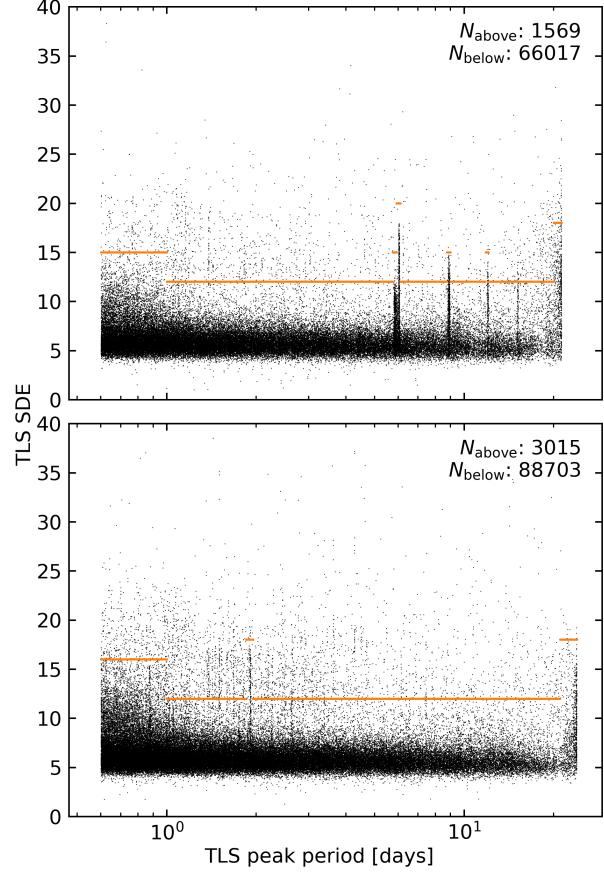
The relevant quantity needed to calculate the integrated photon counts from faint sources is  $N(< m, l, b)$  — the number of stars per square arcsecond brighter than magnitude  $m$ , along a line of sight with galactic longitude and latitude  $(l, b)$ . To calculate this surface density, Winn (2013) queried the Besançon model (Robin et al. 2003) along a grid of galactic sight-lines, and then converted to  $I$ -band surface brightnesses. Fitting a smooth function to the results, Winn (2013) found

$$I \text{ mag arcsec}^{-2} = a_0 + a_1 \left( \frac{|b|}{40^\circ} \right) + a_2 \left( \frac{|l|}{180^\circ} \right)^{a_3}, \quad (10)$$

where the galactic longitude  $l$  is measured from  $-180^\circ$  to  $180^\circ$ , and the empirical coefficients were found to be  $a_0 = 18.9733$ ,  $a_1 = 8.833$ ,  $a_2 = 4.007$ , and  $a_3 = 0.805$ . This fit was cautioned to be *very approximate*. It is sensitive to the threshold used to select “unresolved” stars, and likely no more accurate than 0.5 mag on average. In regions with exceptionally high extinction (*e.g.*, star forming regions) it is expected to systematic underestimate the background brightness by an even larger degree. Nonetheless, this model for the diffuse sky background seems to agree reasonably well with the observed trend of standard deviation versus stellar magnitude.

**ACF statistics**—Beyond the white noise properties of the light curves, the red noise properties are also important. In Figure 11 we show the average autocorrelation of the raw and TFA-filtered light curves. The raw light curves have substantial systematic red noise; their average autocorrelation, evaluated at various time lags, is positive. The TFA-filtered light curves are much closer to gaussian — the average autocorrelation between any two points in these light curves is close to zero. However, the average TFA-filtered light curve is not *exactly* gaussian. At time lags of a few hours or less, there is some excess power. This suggests that additional detrending may be necessary to maximize the effectiveness of planet searches (or the discovery of any other signals via matched-filter techniques).

#### 4.2. Identifying Planet Candidates



**Figure 12.** Transiting planet search periodogram summary. Each point represents the TLS periodogram peak signal detection efficiency (SDE) and corresponding period from one light curve. Significance thresholds (orange lines) are empirically defined for the transiting planet search described in § 4.2. *Top.* Sector 6. *Bottom.* Sector 7.

To identify an initial set of transiting planets, strong stellar rotators, and eclipsing binaries, we performed a few steps of post-processing on the light curves. This processing was done independently of the data release, since it was quite specific to our own scientific interests. The code used to do this additional processing is also available online<sup>8</sup>.

For simplicity, we first chose a single aperture size – “aperture 2” – with a radius of 1.5 pixels. Then, to identify periodic transit-like signals, we used Hippke & Heller (2019)’s transit least-squares (TLS) tool. The algorithm is the same as the canonical box least-squares (Kovács et al. 2002), except in place of a box template, a transit template is used for a marginal improvement of the detection efficiency. In addition, the search grids in `tls`<sup>9</sup> are slightly more efficient than in most BLS implementations, since `tls` uses the cubic-

<sup>8</sup> [github.com/lgbouma/cdips](https://github.com/lgbouma/cdips), commit 348eea7.

<sup>9</sup> <https://github.com/hipanke/tls>

in-frequency sampling advocated by [Ofir \(2014\)](#), rather than standard linear-in-frequency sampling. Our grids typically consisted of about 30 different durations, and about 5000 periods between 0.5 and 21 days.

Before performing the period search, we rejected 6 hours at the beginning and end of each spacecraft orbit, to mitigate the presence of correlated red noise in the results. This shrank the data volume by about 5%, but also lowered the number of systematic false positives in subsequent vetting. We then performed an asymmetric sigma-clipping of  $(50\sigma, 5\sigma)$  to preserve transits while omitting flares and other positive flux excursions.

A number of the light curves show residual variability, often of stellar origin. This is expected for a sample of young stars; the problem of finding transits in the face of large stellar rotation signals has been explored by both [Rizzuto et al. \(2017\)](#) and also [Hipke et al. \(2019\)](#). The former adopted an approach that we have yet to explore, which is to pass a sliding window over the light curve that, within each step, performs a Bayesian model comparison between a spline and a spline-plus-notch model. If the spline-plus-notch model is favored, the transits are preserved for subsequent discovery. [Hipke et al. \(2019\)](#), conversely, described a number of different detrending methods, and found that most performed more or less the same at recovering planetary transits for a sample of young stars with strong rotation signals.

Our approach in this work is essentially identical to one of the methods described by [Hipke et al. \(2019\)](#). First, before sigma-clipping or omitting points near the edges of orbits we run a generalized Lomb-Scargle periodogram on each TFA light curve ([Lomb 1976](#); [Scargle 1982](#); [VanderPlas & Ivezić 2015](#)). If a peak is found with false alarm probability below  $10^{-5}$ , we consider the star “variable”, and opt to detrend the flux timeseries with robust penalized B-splines, which are splines with knot-length automatically determined via cross-validation ([Eilers & Marx 1996](#)). The idea behind the cross-validation is that more knots leads to smaller residuals on training data, but larger errors when tested on the entire dataset. We used the `wotan` implementation, which is a wrapper to the `pyGAM` spline fitter, with  $2\sigma$  clipping of outliers from the fit residuals at each iteration ([Servén et al. 2018](#); [Hipke et al. 2019](#)). The maximum number of spline knots is set to 50, which for each TESS sector ( $\approx 25$  days) is commensurate with a  $\approx 0.5$  day window.

We verified by injecting and recovering a 2.5 mmag central-transit planet with variable period that this detrending approach marginally improved the recovery probability, compared to not doing anything. The main cases for which it helps are those with large residual stellar variability in the TFA light curve.

After (optionally) detrending, we run the TLS search. To select “significant” signals for visual inspection, we performed a cut on the TLS signal detection efficiency defined piecewise over the signal periods. The overall lower limit is  $SDE > 12$ , with higher limits imposed in regions with heavy contamination from systematic signals. The signals for sec-

tors 6 and 7 are shown in the space of SDE vs period in Figure 12.

This yielded a few thousand light curves. About two-thirds already had been detrended with splines, and were not processed further. For the remaining third we performed TFA signal reconstruction using `VARTOOLS`. In this process, the model lightcurve  $A(i)$  in Equation 7 is set to the phase-binned signal from the most powerful peak in the TLS spectrum, rather than being a constant function. This helps lower the scatter of the light curve at frequencies away from the TLS peak.

We then make a multi-page .pdf-format document report with the information necessary to make classifications for vetting. These vetting reports are released along with the light curves, and are a useful summary for anyone interested in the subset of objects that we have analyzed. A full description of each page of the report is given in Appendix C. When classifying these reports, we assessed:

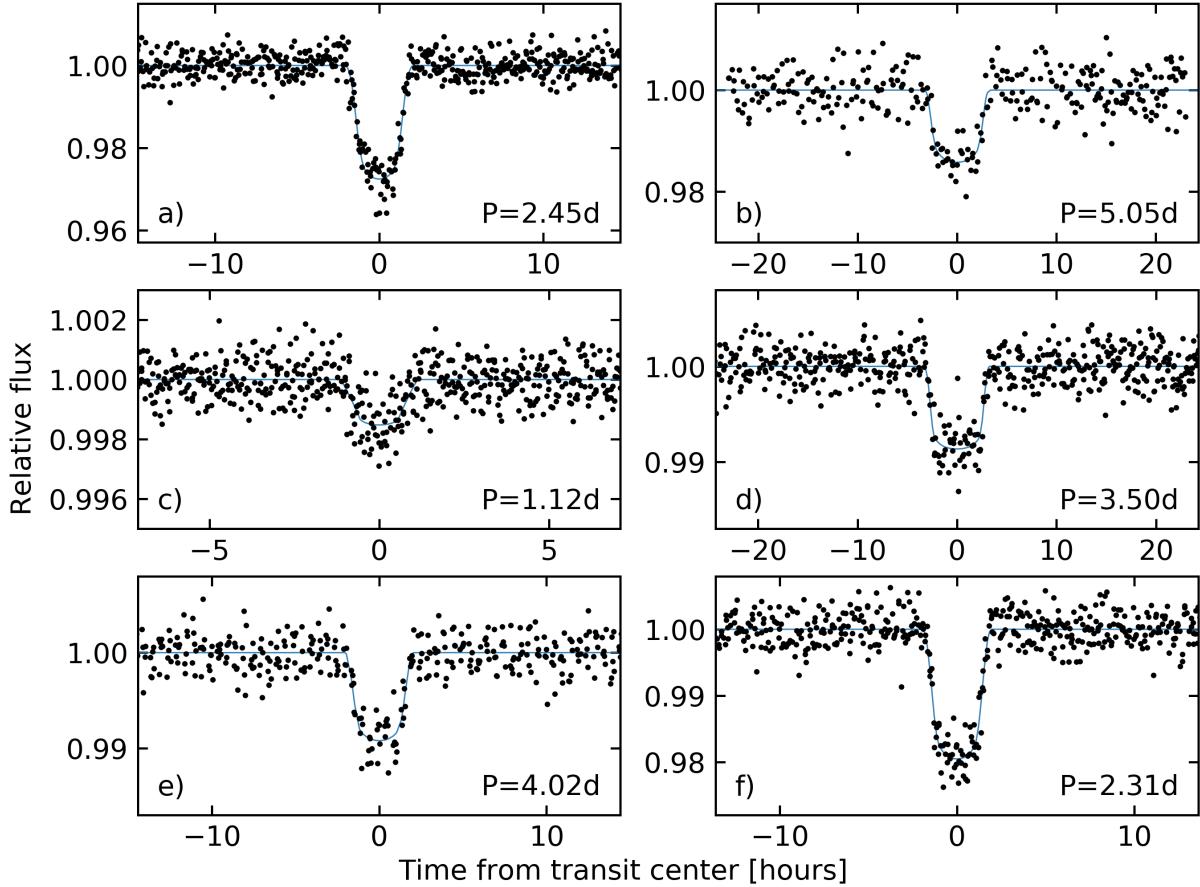
1. *Morphology of photometric variability.* Designations include tags for planet candidates, eclipsing binaries, instrumental variations, stellar variations, and “weirdos”.
2. *Cluster membership status.* By default, all light curves were made for stars with at least one literature claim of cluster membership. However, some of these claims pre-dated the availability of Gaia-DR2 parallaxes or proper motions. We therefore re-assess the membership status on a case-by-case basis through inspection of the available astrometry and photometry. Known asterisms are also omitted at this point (e.g., [Sulentic et al. 1973](#); [Baumgardt 1998](#); [Kos et al. 2018](#)).
3. *Photometric blends.* Tags are created to highlight whether the depth of the photometric signal shows a strong dependence on aperture size, and also whether the in-transit minus the out-of-transit images reveal that the source of variability is in fact far from the target star.

The initial classifications were performed by LGB, and possible planet candidates were then vetted by JH and JNW. The TagSpaces software was used to easily assign labels to the documents.

### 4.3. Candidate TESS Objects of Interest

If all vetters identified an object as a planet candidate, we then proceeded to fit a transit model to the data. The procedure closely mirrors that used by [Bouma et al. \(2019\)](#) to create a phase-folded light curve for WASP-4.

First, we selected points within 5 transit durations of each transit midpoint, using the orbital period, transit duration, and reference epoch found from applying TLS to the detrended light curve. We then fitted a line to the out-of-transit flux measurements around each transit window, and divided it out. We then fitted a standard transit model using the formulae given by [Mandel & Agol \(2002\)](#) and implemented by



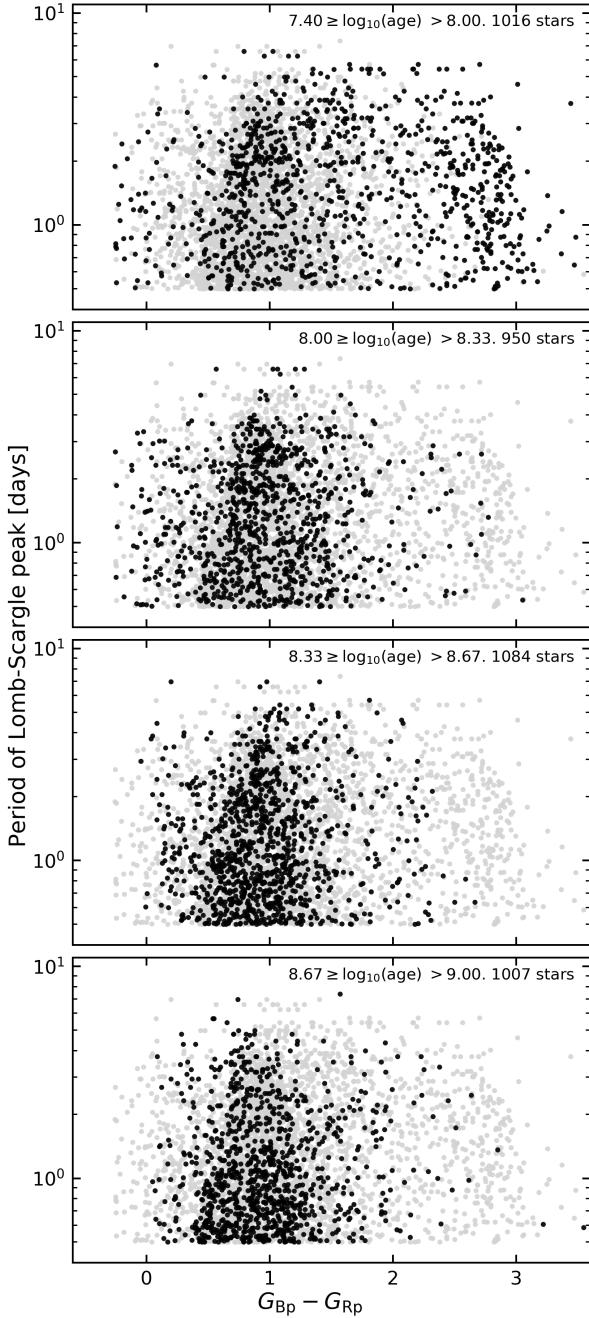
**Figure 13.** Phased light curves for six of the 39 candidate transiting planets identified in this work. Black points are detrended flux measurements, phased to the best-fit orbital period indicated in the lower right of each panel. Transit models are fitted as described in § 4.3, and plotted with blue lines. For display purposes, a window of  $\pm 4$  transit durations centered on the transit mid-point is displayed. The vetting reports for these objects are available at XXX. The Gaia source identifiers are a): 5605128927705695232, b): 3125263468681400320, c): 5510676828723793920, d): 3007171311355035136, e): 5617126180115568256, f): 3220266049321724416.

Kreidberg (2015, BATMAN). The free parameters were the orbital period, the reference epoch, the planet to star radius ratio  $R_p/R_*$ , the orbital distance to stellar radius ratio  $a/R_*$ , and the inclination  $i$ . We assumed circular orbits. To fix the quadratic limb-darkening parameters, we collected  $T_{\text{eff}}$  and  $\log g$  from the TIC, and then interpolated against the Claret (2017) tables. For stars hotter than the maximal effective temperature of these tables (12,000K), we simply used the coefficients of the hottest star.

We then sampled the posterior probability distribution using the Goodman & Weare (2010) algorithm, as implemented by Foreman-Mackey et al. (2013, emcee). Our main goal in this procedure was to provide an accurate ephemeris. To ensure that the reported uncertainties are valid, after performing the initial fit using the photometric uncertainties reported in our light curves, we rescaled the the uncertainty in each photometric data point to equal the root-mean-square deviation of each transit-model subtracted light curve. We then refitted each light curve.

The resulting parameters are given in Table YYY, and are also available as cTOIs NNN.01 - MMM.01 on exofop-TESS (LINK). Figure 13 shows the phase-folded lightcurves and their best-fitting models.

The physical parameters quoted in Table YYY, and on exofop-TESS, should be considered preliminary, and could be substantially improved with an object-by-object analysis. A few additional assumptions were required to report some of the values. For the stellar radii, we assumed the  $R_*$  values and uncertainties reported in the TIC. There were a few cases with null uncertainties in the stellar radii – for these, we simply invented a relative uncertainty of 30%, and note in the comments that this was done. There were also a few cases for which the TIC stellar temperatures and/or radii were not available, but the Gaia values were. In these cases, we used the Gaia  $T_{\text{eff}}$  and  $R_*$  values (Andrae et al. 2018), and a stellar mass interpolated from the Pecaut & Mamajek (2013) table (assuming the star is a dwarf) to estimate  $\log g$ . The uncertainties in e.g., the planetary radii and impact parameters were found using linear error propagation theory, under the



**Figure 14.** Time evolution of Lomb-Scargle periodogram peak period versus Gaia color. Red stars are on the right, blue stars are on the left. Each point on this diagram is a star for which the false alarm probability of the most powerful Lomb-Scargle peak is below  $10^{-15}$ . Additionally, we required (1) that the membership claim originate from [Cantat-Gaudin et al. \(2018\)](#), [Kharchenko et al. \(2013\)](#), or [Gaia Collaboration et al. \(2018a\)](#); and (2) that the cluster with claimed membership had an age reported by [Kharchenko et al. \(2013\)](#). Our approach for avoiding photometric blends is described in § 4.4.

idealized assumption that the covariance between all stellar and planetary parameters was zero.

Finally, to assign the names of the planet candidates – and determine whether they had already been identified as TOIs or cTOIs – we used lists downloaded from the [tev.mit.edu](#) alert site for TOIs, and from [exofop.ipac.caltech.edu/tess](#) for cTOIs. We cross-referenced based on a combination of the TIC identifier, spatial position, and the ephemeris of each object.

#### 4.4. Eclipsing binaries, stellar rotation

To assess the prospects for studying forms of stellar variability other than that caused by transiting planets, we performed a brief additional exploration. Using the Lomb-Scargle periodogram results from § 4.2, we selected light curves for which the periodogram peak power had a corresponding false alarm probability below  $10^{-15}$ . We chose this cutoff to be small enough to give us a reasonable number of objects, but large enough so that there would be no question that the signal being detected was astrophysical.

Further, we then restricted our focus to stars with a cluster membership claim in [Cantat-Gaudin et al. \(2018\)](#), [Kharchenko et al. \(2013\)](#), or [Gaia Collaboration et al. \(2018a\)](#), since from our experience these works have a relatively low false positive probability. We then further restricted our sample of stars to be those for which the claimed cluster also had an age reported by [Kharchenko et al. \(2013\)](#).

To remove obvious photometrically-blended duplicates, for any signal, if any nearby stars had matching periods, we kept only the signal with the lowest false alarm probability. While this does not ensure that the detected signals originate from the target star, it removed a large number of the duplicate signals, and suits our exploratory purposes.

All told, this yielded 4057 stars between the ages of about 25 Myr and 1 Gyr. In Figure 14, we plot the Lomb-Scargle peak period versus  $G_{\text{BP}} - G_{\text{RP}}$  color over four different bins of time. The bins were chosen to match the sub-sample sizes within reasonable proximity.

The youngest sample of stars is the reddest, as might be expected given that many of these stars have not yet reached the main sequence, and may exist along lines of sight with heavy extinction. Stars younger than about 200 Myr (the first two age bins) also tend to be bluer. The oldest stars, between about 450 Myr and 1 Gyr, show a hint in the detected periods of being “bottom-heavy”, *i.e.*, there are relatively more short-period objects in this sample. The reason why is unclear – the effect could be due to systematic differences in the stellar samples, rather than being astrophysical. More investigation will be required.

## 5. DISCUSSION

Maybe JH or JNW could work on these subsections?

### 5.1. The planet candidates require more observations

Additional observations are needed for two major goals: a) cluster membership confirmation (age indicators, across the

board); b) verify the planetary nature of the signals (seeing-limited photometry, RV mass measurement).

Different types of observations and analyses are needed to confirm (i) the planetary nature of these signals, and (ii) whether the host stars are bona fide cluster members. For (i), follow-up photometry, high-resolution imaging, and in some cases spectroscopic mass measurement are needed. For (ii), one or two radial velocity measurements might kinematically exclude cluster membership in a number of cases. A full assessment of the target star's age indicators, including photometric rotation signatures, spectroscopic Li and H $\alpha$  measurements, and even an isochronal analysis, would be helpful.

### 5.2. Eclipsing binary and stellar variability analyses may also be feasible

JH?

## 6. CONCLUSION

In this study, we first collected an all-sky sample of about one million candidate young stars brighter than  $G_{\text{Rp}}$  of 16. 82% of the target star sample resides in a “cluster”, a generic term we adopted for open clusters, stellar associations, and moving groups. The remaining stars have photometric or astrometric indications of their youth, and either reside on the pre-main-sequence or upper main sequence.

We then reduced TESS full frame images taken over the course of about two months (Sectors 6 and 7; the first fields very close to the galactic plane). We performed difference imaging in order to minimize the complications of the crowded background. Using forced aperture photometry, we made light curves for all stars brighter than  $G_{\text{Rp}}$  of 13, and went three magnitudes deeper for our target star sample. This yielded 159,337 light curves of candidate young stars across 596 distinct clusters.

The light curves seem to be limited in precision by photon noise at the bright end, and by unresolved background stars at the faint end. Though the raw light curves show significant red noise, decorrelating against a set of template stars led to an ensemble of light curves with very nearly gaussian noise properties.

An initial search for transiting planets, along with subsequent vetting to remove obvious non-cluster members and photometric blends, has revealed 39 planet candidates. 29 of these candidates are candidate cluster members, and the remaining 10 have indications of youth, relative to field stars. The latter sub-sample is the less interesting one for this work, but may be of interest for other studies. Substantial follow-up efforts on these objects are needed to determine (i) whether the candidates are planets and (ii) whether the host stars are truly cluster members. We suspect that a sizeable fraction of our planet candidates may be eclipsing binaries, however it likely also includes real planets. The vetting reports are available at [archive.stsci.edu/prepdscdips/vetting](https://archive.stsci.edu/prepdscdips/vetting), and also documented in Appendix C. The planet candidate parameters are on exoFOP-TESS.

We also briefly explored other forms of photometric variability in the dataset. The age-evolution of extremely high-

SNR variability seems to show some evolution in the space of LS peak-period versus color (Figure 14), but interpreting this result will require a more careful exploration. In searching for transiting planets, we found many eclipsing binaries, as well as stellar rotation and pulsation signals. Those who wish to explore the time-evolution in these and any other systems are invited to explore the data at [archive.stsci.edu/prepdscdips](https://archive.stsci.edu/prepdscdips).

L.G.B. gladly acknowledges helpful discussions with C Huang, M Soares-Furtado, ... ..., and is grateful to the people who have turned TESS from an idea into reality. J.N.W. thanks ... This paper includes data collected by the TESS mission, which are publicly available from the Mikulski Archive for Space Telescopes (MAST). Funding for the TESS mission is provided by NASA's Science Mission directorate. 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. This work made use of NASA's Astrophysics Data System Bibliographic Services. This research has made use of the VizieR catalogue access tool, CDS, Strasbourg, France. The original description of the VizieR service was published in A&AS 143, 23. This work has made use of data from the European Space Agency (ESA) mission *Gaia* (<https://www.cosmos.esa.int/gaia>), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, <https://www.cosmos.esa.int/web/gaia/dpac/consortium>). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. The Digitized Sky Surveys were produced at the Space Telescope Science Institute under U.S. Government grant NAG W-2166. The images of these surveys are based on photographic data obtained using the Oschin Schmidt Telescope on Palomar Mountain and the UK Schmidt Telescope. The plates were processed into the present compressed digital form with the permission of these institutions.

**Facility:** 2MASS (Skrutskie et al. 2006), Gaia (Gaia Collaboration et al. 2016, 2018b), TESS (Ricker et al. 2015), UCAC4 (Zacharias et al. 2013)

**Software:** astrobase (Bhatti et al. 2018), astropy (Collaboration et al. 2018), astroquery (Ginsburg et al. 2018), BATMAN (Kreidberg 2015), corner (Foreman-Mackey 2016), emcee (Foreman-Mackey et al. 2013), fitsh (Pál 2012), IPython (Pérez & Granger 2007), matplotlib (Hunter 2007), numpy (Walt et al. 2011), pandas (McKinney 2010), pyGAM (Servén et al. 2018) psycopg2 ([initd.org/psycopg](http://initd.org/psycopg)), scipy (Jones et al. 2001), TagSpaces ([tagspaces.org](http://tagspaces.org)), tesscut (Brasseur et al. 2019), VARTOOLS (Hartman & Bakos 2016) wotan (Hipke et al. 2019),

## REFERENCES

- Aigrain, S., Hodgkin, S., Irwin, J., et al. 2007, *Monthly Notices of the Royal Astronomical Society*, **375**, 29
- Aigrain, S., Parviainen, H., Roberts, S., Reece, S., & Evans, T. 2017, *Monthly Notices of the Royal Astronomical Society*, **471**, 759
- Alard, C., & Lupton, R. H. 1998, *ApJ*, **503**, 325
- Andrae, R., Fouesneau, M., Creevey, O., et al. 2018, *Astronomy and Astrophysics*, **616**, A8
- Bakos, G. A. 2018, in *Handbook of Exoplanets*, ed. H. J. Deeg & J. A. Belmonte (Springer International Publishing), 1
- Bakos, G. A., Torres, G., Pál, A., et al. 2010, *The Astrophysical Journal*, **710**, 1724
- Barnes, S. A., Weingrill, J., Granzer, T., Spada, F., & Strassmeier, K. G. 2015, *Astronomy & Astrophysics*, **583**, A73, arXiv: 1511.00554
- Baumgardt, H. 1998, *Astronomy and Astrophysics*, **340**, 402
- Bell, C. P. M., Murphy, S. J., & Mamajek, E. E. 2017, *Monthly Notices of the Royal Astronomical Society*, **468**, 1198
- Benatti, S., Nardiello, D., Malavolta, L., et al. 2019, arXiv: 1904.01591 [astro-ph], arXiv: 1904.01591
- Bhatti, W., Bouma, L. G., & Wallace, J. 2018, astrobase
- Borucki, W. J., Koch, D., Basri, G., et al. 2010, *Science*, **327**, 977
- Bouma, L. G., Winn, J. N., Baxter, C., et al. 2019, *The Astronomical Journal*, **157**, 217
- Bramich, D. M. 2008, *Monthly Notices of the Royal Astronomical Society*, **386**, L77
- Brasseur, C. E., Phillip, C., Fleming, S. W., Mullally, S. E., & White, R. L. 2019, *Astrophysics Source Code Library*, ascl:1905.007
- Brucalassi, A., Koppenhoefer, J., Saglia, R., et al. 2017, *Astronomy and Astrophysics*, **603**, A85
- Bryson, S. T., Jenkins, J. M., Gilliland, R. L., et al. 2013, *Publications of the Astronomical Society of the Pacific*, **125**, 889
- Cantat-Gaudin, T., Jordi, C., Vallenari, A., et al. 2018, *Astronomy & Astrophysics*, **618**, A93
- Cantat-Gaudin, T., Krone-Martins, A., Sedaghat, N., et al. 2019, *Astronomy and Astrophysics*, **624**, A126
- Carrasco, J. M., Evans, D. W., Montegriffo, P., et al. 2016, *Astronomy and Astrophysics*, **595**, A7
- Ciardì, D. R., Crossfield, I. J. M., Feinstein, A. D., et al. 2018, *The Astronomical Journal*, **155**, 10
- Claret, A. 2017, *Astronomy & Astrophysics*, **600**, A30, arXiv: 1804.10295
- Clarke, B. D., Caldwell, D. A., Quintana, E. V., et al. 2017, *Kepler Science Document*, 5
- Collaboration, T. A., Price-Whelan, A. M., Sipocz, B. M., et al. 2018, arXiv:1801.02634 [astro-ph], arXiv: 1801.02634
- Curtis, J. L., Agüeros, M. A., Mamajek, E. E., Wright, J. T., & Cummings, J. D. 2019, arXiv:1905.10588 [astro-ph], arXiv: 1905.10588
- David, T., Hillenbrand, L., & Petigura, E. 2016, *Nature*, **534**, 658
- David, T. J., Crossfield, I. J. M., Benneke, B., et al. 2018, *The Astronomical Journal*, **155**, 222
- David, T. J., Cody, A. M., Hedges, C. L., et al. 2019, arXiv:1902.09670 [astro-ph], arXiv: 1902.09670
- Dias, W. S., Monteiro, H., Caetano, T. C., et al. 2014, *Astronomy and Astrophysics*, **564**, A79
- Dierckx, P. 1996, Curve and surface fitting with splines, repr edn., Monographs on Numerical Analysis (Oxford: Clarendon Press), oCLC: 245719230
- Donati, J. F., Moutou, C., Malo, L., et al. 2016, *Nature*, advance online publication
- Eastman, J., Siverd, R., & Gaudi, B. S. 2010, *Publications of the Astronomical Society of the Pacific*, **122**, 935, arXiv: 1005.4415
- Eilers, P. H. C., & Marx, B. D. 1996, *Statistical Science*, **11**, 89
- Feinstein, A. D., Montet, B. T., Foreman-Mackey, D., et al. 2019, arXiv:1903.09152 [astro-ph], arXiv: 1903.09152
- Flagg, L., Johns-Krull, C. M., Nofi, L., et al. 2019, arXiv:1906.02860 [astro-ph], arXiv: 1906.02860
- Foreman-Mackey, D. 2016, *The Journal of Open Source Software*, **24**
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *Publications of the Astronomical Society of the Pacific*, **125**, 306
- Fortney, J. J., Marley, M. S., & Barnes, J. W. 2007, *ApJ*, **659**, 1661
- Fulton, B. J., Petigura, E. A., Howard, A. W., et al. 2017, *AJ*, **154**, 109
- Gagné, J., & Faherty, J. K. 2018, *The Astrophysical Journal*, **862**, 138
- Gagné, J., Roy-Loubier, O., Faherty, J. K., Doyon, R., & Malo, L. 2018a, *The Astrophysical Journal*, **860**, 43
- Gagné, J., Mamajek, E. E., Malo, L., et al. 2018b, *The Astrophysical Journal*, **856**, 23
- Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, *Astronomy and Astrophysics*, **595**, A1
- Gaia Collaboration, Babusiaux, C., van Leeuwen, F., et al. 2018a, *Astronomy and Astrophysics*, **616**, A10
- Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018b, *Astronomy and Astrophysics*, **616**, A1
- Gilliland, R. L., Brown, T. M., Guhathakurta, P., et al. 2000, *The Astrophysical Journal Letters*, **545**, L47
- Ginsburg, A., Sipocz, B., Madhura Parikh, et al. 2018, *Astropy/Astroquery: V0.3.7 Release*
- Goodman, J., & Weare, J. 2010, *Communications in Applied Mathematics and Computational Science*, **5**, 65

- Hartman, J. D., & Bakos, G. A. 2016, *Astronomy and Computing*, 17, 1
- Hartman, J. D., Gaudi, B. S., Pinsonneault, M. H., et al. 2009a, *The Astrophysical Journal*, 691, 342
- Hartman, J. D., Gaudi, B. S., Holman, M. J., et al. 2009b, *The Astrophysical Journal*, 695, 336
- Hippke, M., David, T. J., Mulders, G. D., & Heller, R. 2019, arXiv:1906.00966 [astro-ph], arXiv: 1906.00966
- Hippke, M., & Heller, R. 2019, arXiv:1901.02015 [astro-ph], arXiv: 1901.02015
- Howell, S. B., Sobeck, C., Haas, M., et al. 2014, *Publications of the Astronomical Society of the Pacific*, 126, 398
- Huang, C. X., Penev, K., Hartman, J. D., et al. 2015, *Monthly Notices of the Royal Astronomical Society*, 454, 4159
- Huang, C. X., Burt, J., Vanderburg, A., et al. 2018, *The Astrophysical Journal*, 868, L39
- Hunter, J. D. 2007, *Computing in Science & Engineering*, 9, 90
- Irwin, J., Hodgkin, S., Aigrain, S., et al. 2007a, *MNRAS*, 377, 741
- Irwin, J., Irwin, M., Aigrain, S., et al. 2007b, *Monthly Notices of the Royal Astronomical Society*, 375, 1449
- Irwin, J., Aigrain, S., Hodgkin, S., et al. 2007c, *Monthly Notices of the Royal Astronomical Society*, 380, 541
- Jenkins, J. M., Caldwell, D. A., Chandrasekaran, H., et al. 2010, *The Astrophysical Journal Letters*, 713, L87
- Johns-Krull, C. M., McLane, J. N., Prato, L., et al. 2016, *The Astrophysical Journal*, 826, 206
- Jones, E., Oliphant, T., Peterson, P., et al. 2001, Open source scientific tools for Python
- Kharchenko, N. V., Piskunov, A. E., Schilbach, E., Röser, S., & Scholz, R.-D. 2012, *Astronomy and Astrophysics*, 543, A156
- . 2013, *Astronomy and Astrophysics*, 558, A53
- Kim, J. 2018, Querying Gaia for Wide Binary Companions to Exoplanet Hosts, Princeton Junior Thesis (Unpublished)
- Kos, J., de Silva, G., Buder, S., et al. 2018, *Monthly Notices of the Royal Astronomical Society*, 480, 5242
- Kostov, V. B., Schlieder, J. E., Barclay, T., et al. 2019, arXiv:1903.08017 [astro-ph], arXiv: 1903.08017
- Kovács, G., Bakos, G., & Noyes, R. W. 2005, *Monthly Notices of the Royal Astronomical Society*, 356, 557
- Kovács, G., Zucker, S., & Mazeh, T. 2002, *Astronomy and Astrophysics*, 391, 369
- Kraus, A. L., Shkolnik, E. L., Allers, K. N., & Liu, M. C. 2014, *The Astronomical Journal*, 147, 146
- Kreidberg, L. 2015, *Publications of the Astronomical Society of the Pacific*, 127, 1161
- Lang, D., Hogg, D. W., Mierle, K., Blanton, M., & Roweis, S. 2010, *The Astronomical Journal*, 139, 1782
- Leinert, C., Bowyer, S., Haikala, L. K., et al. 1998, *Astronomy and Astrophysics Supplement Series*, 127, 1
- Livingston, J. H., Dai, F., Hirano, T., et al. 2018, *The Astronomical Journal*, 155, 115
- . 2019, *Monthly Notices of the Royal Astronomical Society*, 484, 8
- Lomb, N. R. 1976, *Astrophysics and Space Science*, 39, 447
- Lovis, C., & Mayor, M. 2007, *Astronomy and Astrophysics*, 472, 657
- Majaess, D. 2013, *Astrophysics and Space Science*, 344, 175
- Malavolta, L., Nascimbeni, V., Piotto, G., et al. 2016, *Astronomy and Astrophysics*, 588, A118
- Mandel, K., & Agol, E. 2002, *The Astrophysical Journal*, 580, L171, arXiv: astro-ph/0210099
- Mann, A. W., Gaidos, E., Mace, G. N., et al. 2016a, *ApJ*, 818
- Mann, A. W., Newton, E. R., Rizzuto, A. C., et al. 2016b, *AJ*, 152, 61
- Mann, A. W., Gaidos, E., Vanderburg, A., et al. 2017, *AJ*, 153, 64
- Mann, A. W., Vanderburg, A., Rizzuto, A. C., et al. 2018, *The Astronomical Journal*, 155, 4
- Marrese, P. M., Marinoni, S., Fabrizio, M., & Altavilla, G. 2019, *Astronomy & Astrophysics*, 621, A144
- McKinney, W. 2010, in *Proceedings of the 9th Python in Science Conference*, ed. S. van der Walt & J. Millman, 51
- Meibom, S., Barnes, S. A., Platais, I., et al. 2015, *Nature*, 517, 589
- Meibom, S., Torres, G., Fressin, F., et al. 2013, *Nature*, 499, 55
- Miller, A. A., Irwin, J., Aigrain, S., Hodgkin, S., & Hebb, L. 2008a, *Monthly Notices of the Royal Astronomical Society*, 387, 349
- Miller, J. P., Pennypacker, C. R., & White, G. L. 2008b, *Publications of the Astronomical Society of the Pacific*, 120, 449
- Newton, E. R., Mann, A. W., Tofflemire, B. M., et al. 2019, arXiv:1906.10703 [astro-ph], arXiv: 1906.10703
- Obermeier, C., Henning, T., Schlieder, J. E., et al. 2016, *The Astronomical Journal*, 152, 223
- Oelkers, R. J., & Stassun, K. G. 2018, *The Astronomical Journal*, 156, 132
- Ofir, A. 2014, *Astronomy and Astrophysics*, 561, A138
- Oh, S., Price-Whelan, A. M., Hogg, D. W., Morton, T. D., & Spergel, D. N. 2017, *The Astronomical Journal*, 153, 257
- Owen, J. E., & Wu, Y. 2013, *ApJ*, 775, 105
- Pál, A. 2009, PhD thesis, arXiv: 0906.3486
- . 2012, *MNRAS*, 421, 1825
- Pecaut, M. J., & Mamajek, E. E. 2013, *The Astrophysical Journal Supplement Series*, 208, 9
- Pence, W. D., Chiappetti, L., Page, C. G., Shaw, R. A., & Stobie, E. 2010, *Astronomy and Astrophysics*, 524, A42
- Pérez, F., & Granger, B. E. 2007, *Computing in Science and Engineering*, 9, 21
- Quinn, S. N., White, R. J., Latham, D. W., et al. 2012, *ApJ Letters*, 756, L33
- Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2015, *Journal of Astronomical Telescopes, Instruments, and Systems*, 1, 014003

- Rizzuto, A. C., Ireland, M. J., & Robertson, J. G. 2011, *Monthly Notices of the Royal Astronomical Society*, 416, 3108
- Rizzuto, A. C., Mann, A. W., Vanderburg, A., Kraus, A. L., & Covey, K. R. 2017, *AJ*, 154, 224
- Rizzuto, A. C., Vanderburg, A., Mann, A. W., et al. 2018, arXiv:1808.07068 [astro-ph], arXiv: 1808.07068
- Robin, A. C., ReylÁl, C., DerriÁre, S., & Picaud, S. 2003, *Astronomy and Astrophysics*, 409, 523
- Röser, S., Demleitner, M., & Schilbach, E. 2010, *The Astronomical Journal*, 139, 2440
- Röser, S., Schilbach, E., & Goldman, B. 2016, *Astronomy & Astrophysics*, 595, A22
- Röser, S., Schilbach, E., Piskunov, A. E., Kharchenko, N. V., & Scholz, R.-D. 2011, *Astronomy & Astrophysics*, 531, A92
- Sato, B., Izumiura, H., Toyota, E., et al. 2007, *ApJ*, 661, 527
- Saurin, T. A., Bica, E., & Bonatto, C. 2015, *Monthly Notices of the Royal Astronomical Society*, 448, 1687
- Scargle, J. D. 1982, *The Astrophysical Journal*, 263, 835
- Servén, D., Brummitt, C., & Abedi, H. 2018, dswah/pyGAM: v0.8.0
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *The Astronomical Journal*, 131, 1163
- Smith, J. C., Stumpe, M. C., Cleve, J. E. V., et al. 2012, *Publications of the Astronomical Society of the Pacific*, 124, 1000
- Soares-Furtado, M., Hartman, J. D., Bakos, G. Á., et al. 2017, *Publications of the Astronomical Society of the Pacific*, 129, 044501
- Soderblom, D. R. 2010, *Annual Review of Astronomy and Astrophysics*, 48, 581
- Stassun, K. G., Oelkers, R. J., Pepper, J., et al. 2018, *The Astronomical Journal*, 156, 102
- Stassun, K. G., Oelkers, R. J., Paegert, M., et al. 2019, arXiv:1905.10694 [astro-ph], arXiv: 1905.10694
- Stellingwerf, R. F. 1978, *The Astrophysical Journal*, 224, 953
- Sulentic, J. W., Tifft, W. G., & Dreyer, J. L. E. 1973, The revised new catalogue of nonstellar astronomical objects
- Sullivan, P. W., Winn, J. N., Berta-Thompson, Z. K., et al. 2015, *ApJ*, 809, 77
- Tenenbaum, P., & Jenkins, J. 2018, TESS Science Data Products Description Document, EXP-TESS-ARC-ICD-0014 Rev D, <https://archive.stsci.edu/missions/tess/doc/EXP-TESS-ARC-ICD-TM-0014.pdf>
- Vanderburg, A., Mann, A. W., Rizzuto, A., et al. 2018, arXiv:1805.11117 [astro-ph], arXiv: 1805.11117
- Vanderburg, A., Huang, C. X., Rodriguez, J. E., et al. 2019, arXiv:1905.05193 [astro-ph], arXiv: 1905.05193
- VanderPlas, J. T., & Ivezić, Á. 2015, *The Astrophysical Journal*, 812, 18
- Vanderspek, R., Doty, J., Fausnaugh, M., et al. 2018, *TESS Science Document*, 73
- Walt, S. v. d., Colbert, S. C., & Varoquaux, G. 2011, Computing in Science & Engineering, 13, 22
- Wang, D., Hogg, D. W., Foreman-Mackey, D., & SchÃºlkopf, B. 2017, arXiv:1710.02428 [astro-ph], arXiv: 1710.02428
- Weldrake, D. T. F., Sackett, P. D., & Bridges, T. J. 2006, arXiv:astro-ph/0612215, arXiv: astro-ph/0612215
- Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, *Astronomy and Astrophysics Supplement Series*, 143, 9
- Winn, J. N. 2013, TESS Science Memo No. 2, Version 1. Available upon request.
- Zacharias, N., Finch, C. T., Girard, T. M., et al. 2013, *The Astronomical Journal*, 145, 44
- Zari, E., Hashemi, H., Brown, A. G. A., Jardine, K., & de Zeeuw, P. T. 2018, *Astronomy and Astrophysics*, 620, A172
- Zhang, M., Bakos, G. A., Penev, K., et al. 2016, *Publications of the Astronomical Society of the Pacific*, 128, 035001
- Zuckerman, B., & Song, I. 2004, *Annual Review of Astronomy and Astrophysics*, 42, 685

## APPENDIX

### A. TIME SYSTEM & BARYCENTRIC CORRECTION

The time-stamps included with the calibrated TESS Full Frame Images produced by SPOC include a barycentric correction at a single reference pixel given at the middle of every frame. The barycentric correction is at maximum 16 minutes, corresponding to points on the sky separated by 180 degrees. The angular distance from a TESS camera's center of field to the corners is  $\approx 17$  degrees, so naively one might incur at worst an error of  $\approx 90$  seconds on the time-stamps due to using a barycentric correction in a direction that is slightly wrong. Perhaps due to the lead author's obsession with getting time-stamps correct (Bouma et al. 2019), we perform our own barycentric correction using the appropriate sky coordinates for each light curve. We advise use of our TMID\_BJD column, which gives the mid-time of each exposure in the BJD<sub>TDB</sub> time system, which is the defacto standard in exoplanet and stellar astronomy (Eastman et al. 2010).

### B. ASSIGNING UNIQUE NAMES TO EACH CLUSTER

In assigning a single unique cluster name to each star, we matched against the Kharchenko et al. (2013) name whenever possible, since this was the largest available catalog, and it also included homogeneous age determinations for many of the clusters. To find the matching name, in order of precedence we

1. Checked for direct string matches from Kharchenko et al. (2013) clusters with determined parameters;
2. Checked whether the SIMBAD online name resolving service (Wenger et al. 2000) had any direct string matches against Kharchenko et al. (2013) clusters with determined parameters;
3. Checked for string matches in the full Kharchenko et al. (2013) index (including clusters without determined parameters);
4. Searched for spatial matches between each star and cluster centers from Kharchenko et al. (2013) within 10 arcminutes. In cases with multiple cluster matches, we ignored candidate matches to avoid assigning incorrect names;
5. Checked the WEBDA double name list<sup>10</sup>, and repeated Steps 1-4 with any matches.

A few edge-cases, including for instance sub-clusters larger star-forming complexes like in Sco-Cen or Collinder 33, were manually resolved to the extent feasible (Rizzuto et al. 2011 and Saurin et al. 2015 give detailed pictures of the complex morphologies that frequently arise in young star-forming regions).

The procedure described above failed to yield matches for a few of the infrared clusters identified by Majaess (2013) and included in the Dias et al. (2014) catalog. For these cases, we used the name given by Dias et al. (2014).

The Hyades and a number of other nearby moving groups were also missed, since they were not in the Kharchenko et al. (2013) catalog. For moving groups not identified in Kharchenko et al. (2013), we adopted the constellation-based naming convention from Gagné et al. (2018b).

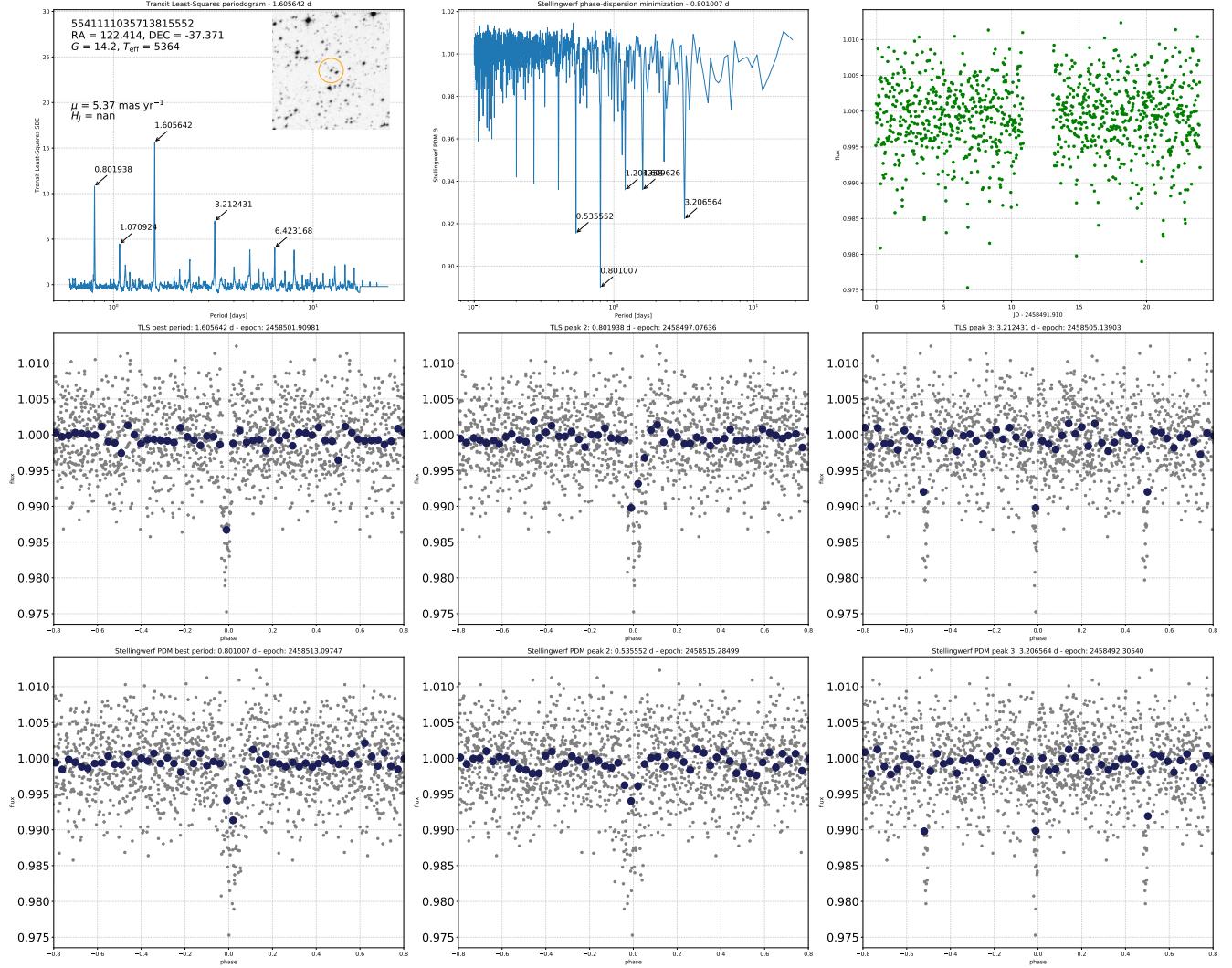
Finally, the procedure enumerated above did not yield matches for recently discovered clusters, such as the “RSG” clusters found by Röser et al. (2016) and the “Gulliver” clusters from Cantat-Gaudin et al. (2018). In these cases, we used the names given by the original authors.

### C. VETTING DOCUMENT DESCRIPTION

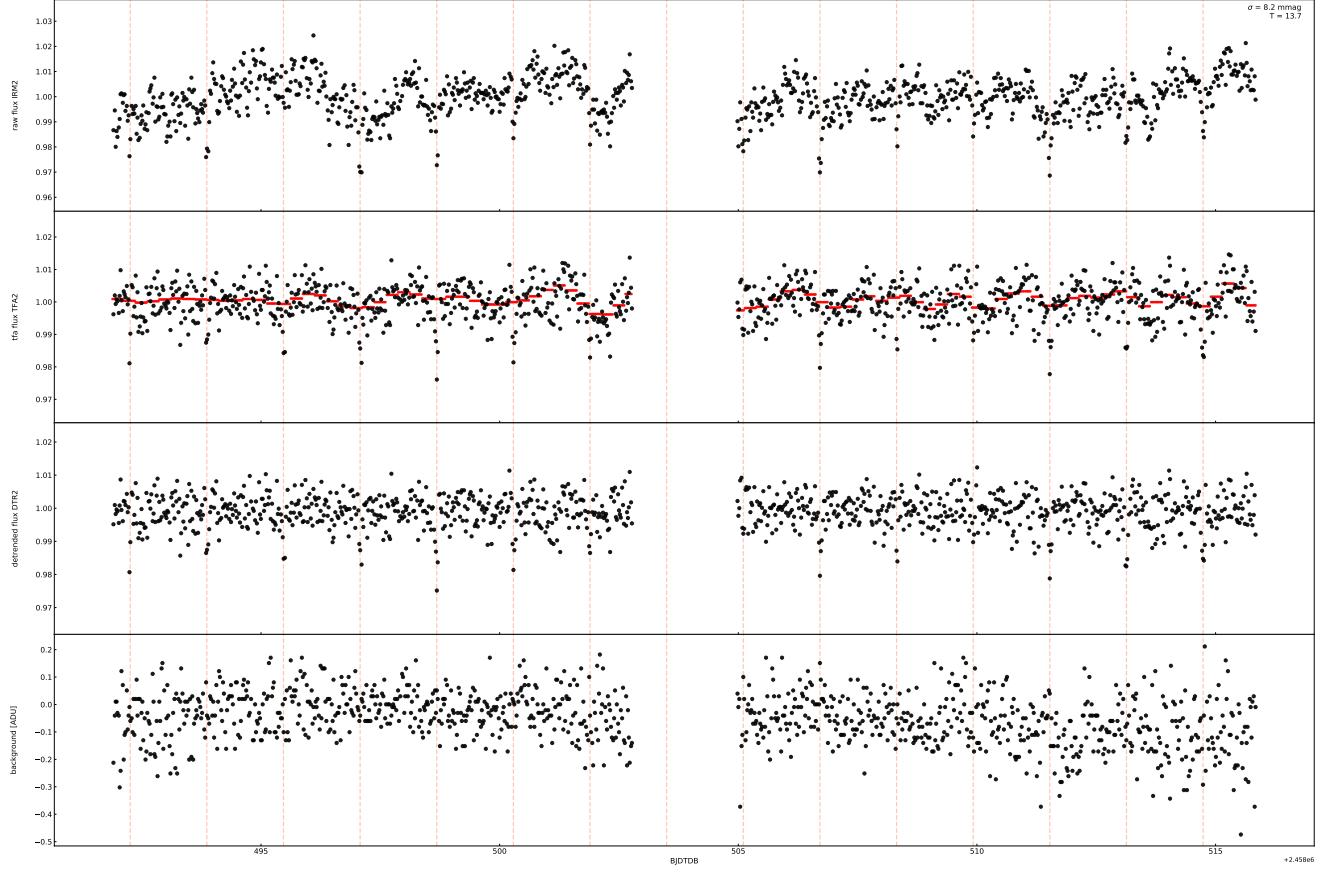
In § 4.3, we described the process by which we made vetting reports suitable for assessing which objects were interesting enough to merit further study.

Figures 15 to 20 of this section summarize this document. Updated versions and their README files will live at this web-address: [mast.stsci.edu/CDIPS](https://mast.stsci.edu/CDIPS). The planet candidate chosen for these figures (Gaia-DR2 554111035713815552 = TIC 134528212) was chosen for the demonstration in part because it passes all the tests. It is one of a number of the giant planet candidates that we note may be in open clusters.

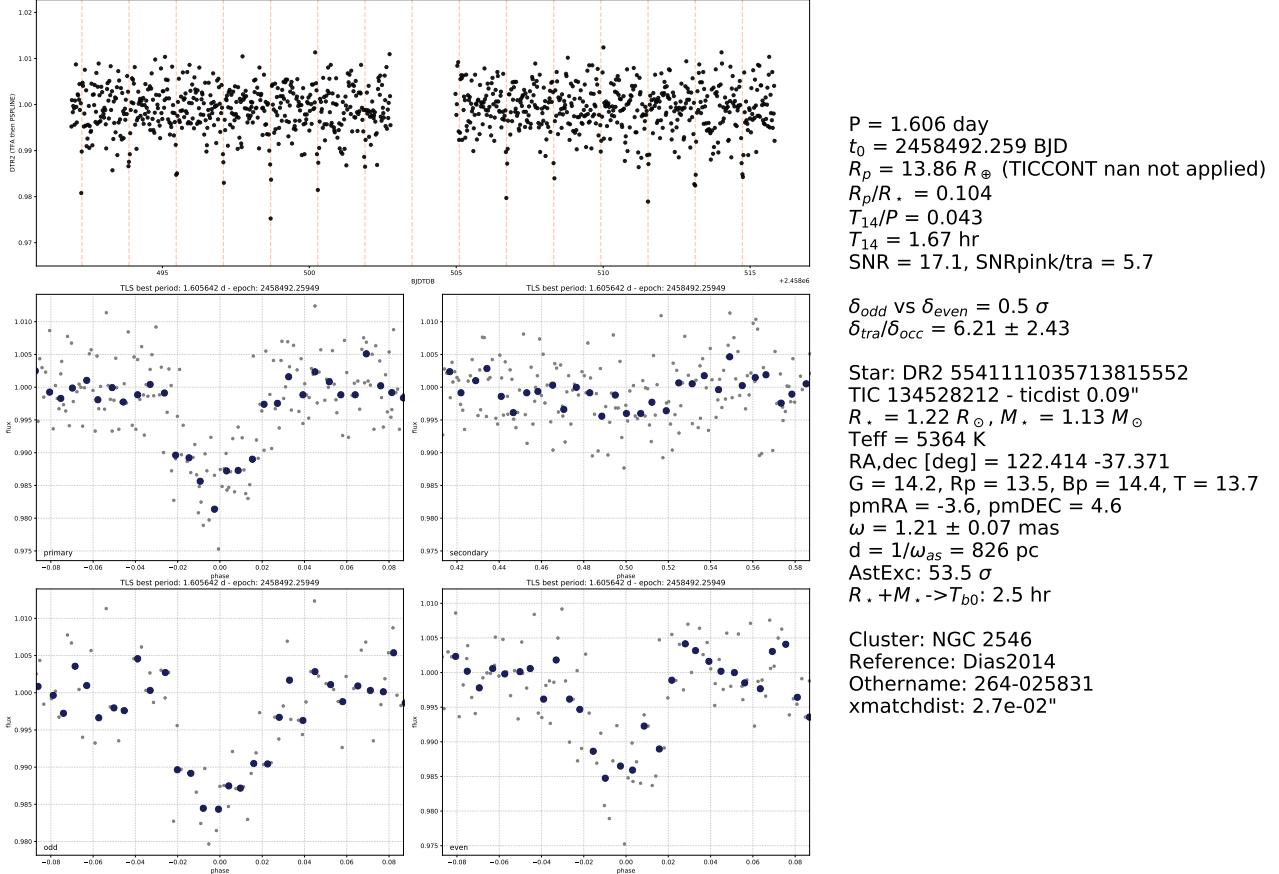
<sup>10</sup> [https://webda.physics.muni.cz/double\\_names.html](https://webda.physics.muni.cz/double_names.html)



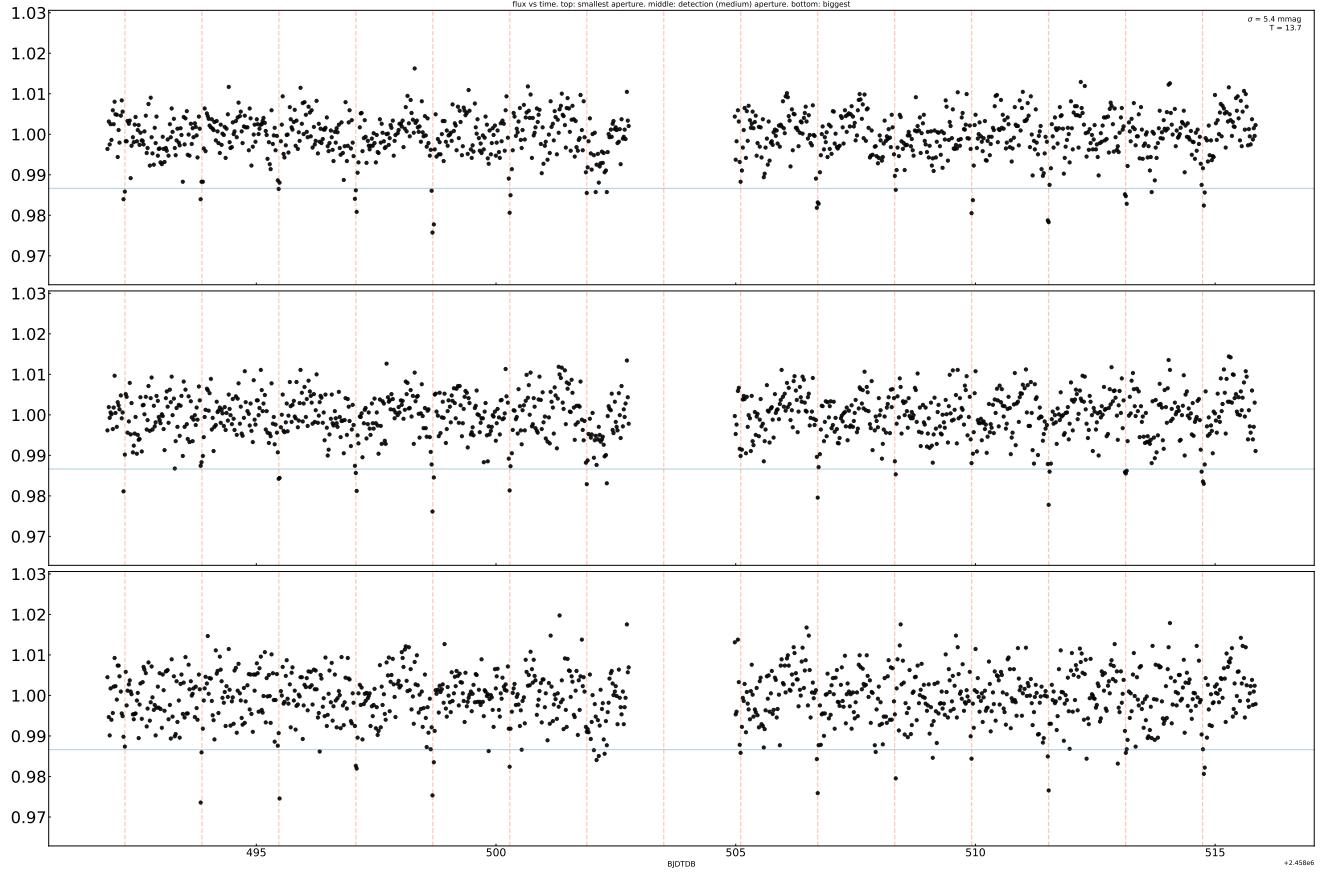
**Figure 15. Page 1: period-search summary.** Periodograms from TLS and phase-dispersion minimization, as calculated with `astrobase.periodbase`, are shown in the top left and top center (Bhatti et al. 2018; Hippke & Heller 2019; Stellingwerf 1978). The top three peaks from each method are shown in the second and third rows; the raw light curve is in the top-right. A small finder chart from DSS is inset to the top left, with the 1.5-pixel radius aperture used to extract the light curve in orange (CITE).



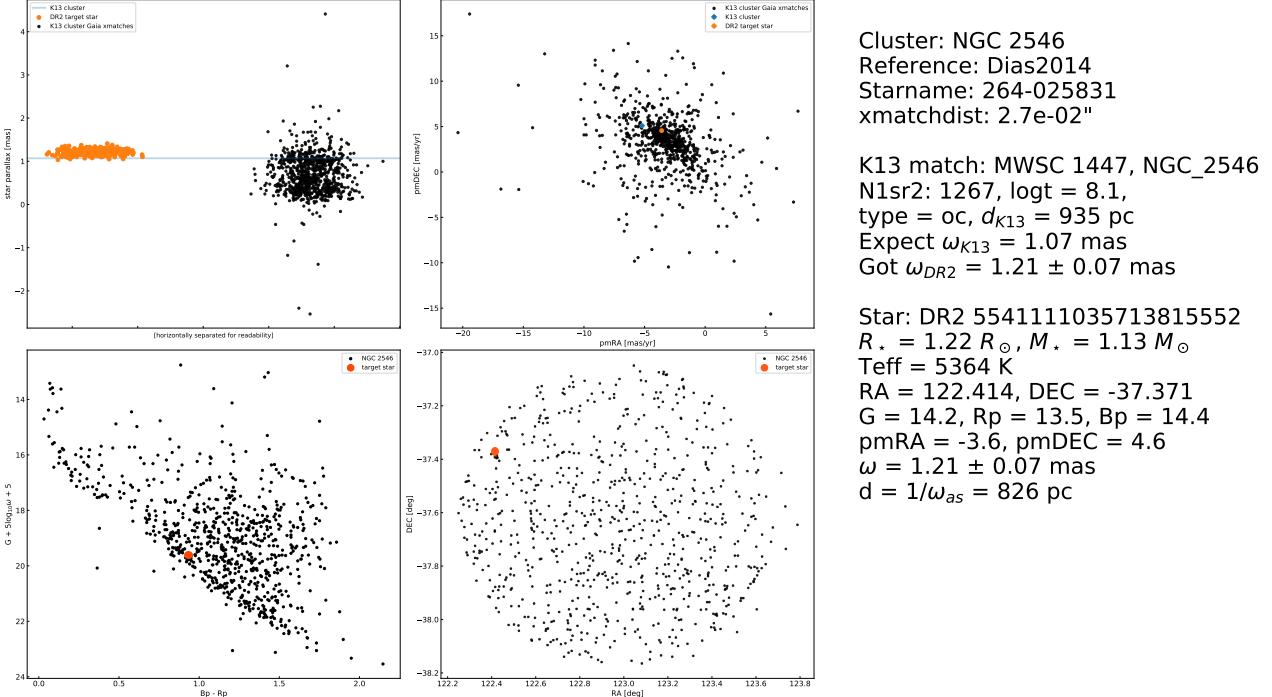
**Figure 16. Page 2: light curve diagnostics.** Time-series of raw flux (IRM2), TFA-detrended flux (TF2), stellar-variability detrended flux, and the background are shown as a function of barycentric Julian date. The overplotted dashed vertical lines are the ephemeris of the highest-power TLS peak from Figure 15. An important check is whether the flux dips are correlated with changes in the background level – in this case, they are not. The standard deviation and TESS magnitude are quoted in the upper right. The red line in the second from the top plot is the windowed spline described in Section 4.3 – in this case it is near its maximal frequency.



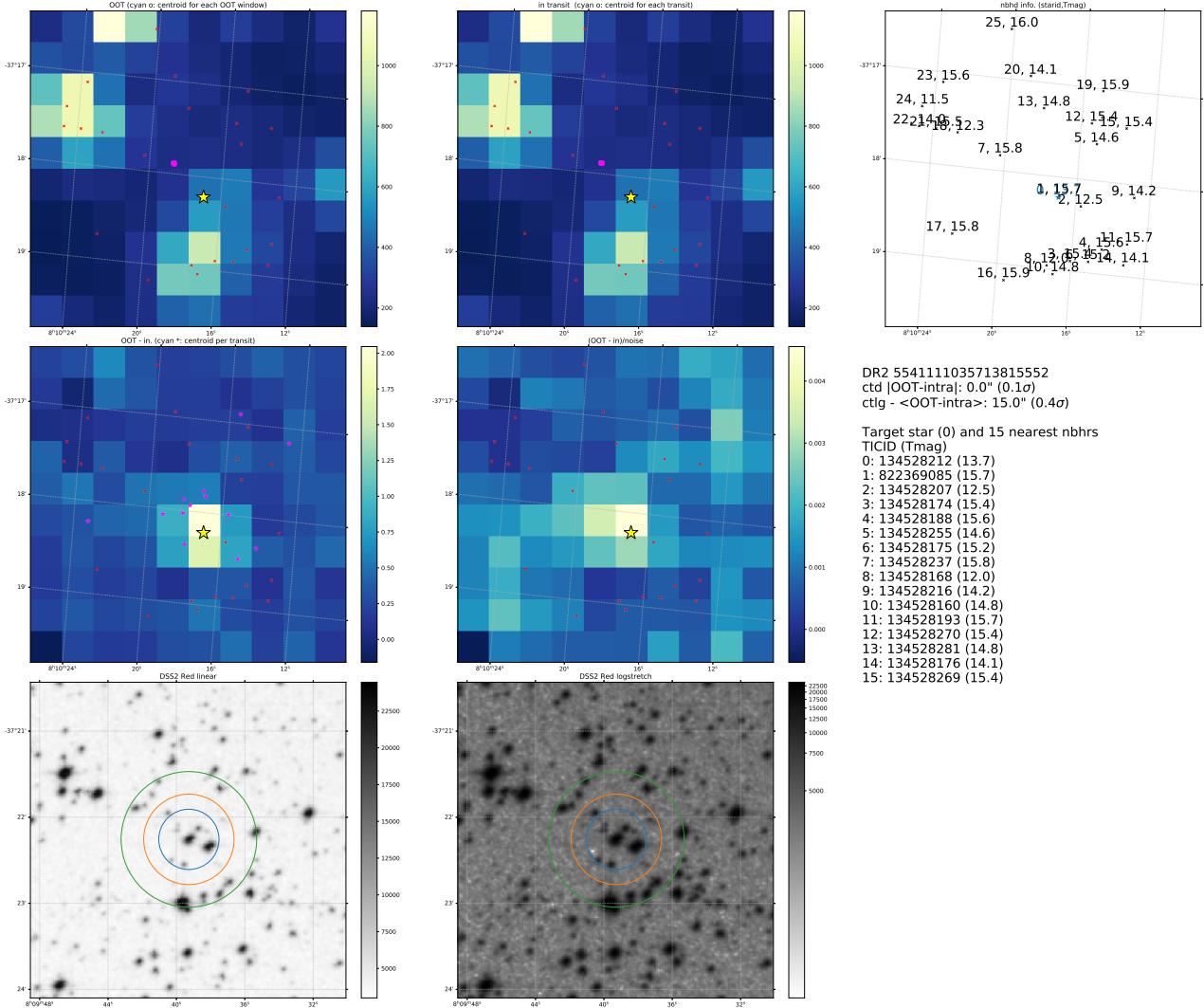
**Figure 17. Page 3: Transit diagnostics.** The plots show the maximally-detrended light curve (top); the phase-folded light curve centered over  $\pm 3$  transit durations of the primary transit (middle left); the secondary eclipse (middle right); the odd-numbered transits (lower left); and the even-numbered transits (lower right). The stellar parameters ( $T_{\text{eff}}, R_*$ ) are taken from Gaia-DR2 Apsis pipeline results when available (Andrae et al. 2018). The first eight lines of text are parameters determined from the best-fitting TLS model. The one exception is the planet radius, which uses the stellar radius as noted above. The “flux contamination” (TICCONT) from neighboring stars is *never* taken into account, because transit depth dilution does not affect image subtraction analyses in the same manner as aperture-photometry reductions. To estimate the transit to occultation depth ratio, the phase-folded light curve is also coarsely fit via maximum likelihood by a sum of two gaussians. “AstExc” refers to the Gaia-DR2 astrometric excess, which can indicate hints of astrometric binarity in the system. “ $R_* + M_* \rightarrow T_{b0}$ ” gives the duration of a zero-eccentricity central transit based on the Gaia-DR2 stellar radius, and a stellar mass interpolated from the Pecaut & Mamajek (2013) table, under the assumption that the star is a dwarf.



**Figure 18. Page 4: Light curves for increasing aperture sizes.** Apertures of radius 1, 1.5, and 2.25 pixels are shown from top to bottom. The blue line is the reference transit depth from the best-fitting TLS model. Any changes in depth with aperture size might indicate the need for a detailed pixel-level analysis.



**Figure 19. Page 5: Cluster membership assessment diagnostics.** The star was considered a candidate cluster member by the source(s) listed under “Reference”, in this case [Dias et al. \(2014\)](#). The name used in the Dias catalog in this case was 264-025831, a UCAC-4 identifier, which can be back-referenced to find that [Dias et al. \(2014\)](#) assigned this star a membership probability in NGC 2546 of 98%. The base catalog for the plots is chiefly that of [Kharchenko et al. \(2013\)](#), due to its homogeneous parameter determination procedure (particularly for age). If a match to the [Kharchenko et al. \(2013\)](#) catalog is found, then the remaining plots are populated. Top-left shows the parallax, with orange points sampled from the Gaia-DR2 posterior, black points the other cluster members in the Kharchenko catalog, and the blue line the claimed Kharchenko parallax for the cluster. A large number of field contaminants are visible in this case. Top-right are the Gaia proper motions, where against black points are cluster members from Kharchenko, and the orange is the target star. Bottom-left is the HR diagram, and bottom-right are the on-sky positions. In the text,  $N1sr2$  is the number of  $1\sigma$  cluster members claimed by [Kharchenko et al. \(2013\)](#) within the cluster angular radius;  $\log t$  is the base-10 logarithm of the age in years;  $\text{type}$  matches the type codes provided by [Kharchenko et al. \(2013\)](#).



**Figure 20. Page 6: Imaging variability diagnostics.** This page is intended to help diagnose which stars are producing the observed variability. Top-left and top-center are the mean out-of-transit (OOT) and mean in-transit calibrated images (separate from any of our image-subtraction analysis). The OOT images are based on the same number of exposures as the in-transit images and split evenly before and after each transit (following Bryson et al. 2013; Kostov et al. 2019). The yellow star is the target; cyan dots are the flux-weighted centroid of the entire image for each transit event; small red crosses are WCS-projected locations of neighbor stars. Middle-left is the most important sub-panel: the difference between the OOT and in-transit mean images. If the variability shown in background map (units: ADU) is off-target, the transit is typically not from the target star. Middle-center is the same, normalized by the uncertainty map. Lower left and lower center show the DSS field in linear and log scales at roughly the same pixel scale as the TESS image, with the 1, 1.5, and 2.25 pixel-radius apertures in blue, orange, and green respectively. The brightness of neighborhood stars is given on the far right. Note the slight rotation difference between DSS and TESS images; DSS images are aligned north-up, east-left; TESS images are oriented as closely as possible to this system without actually performing the rotation.