

The Cluster Difference Imaging Photometric Survey (CDIPS). I. Light Curves & Planet Candidates in Open Clusters from TESS Sectors 6 & 7

L. G. BOUMA,¹ W. BHATTI,¹ J. D. HARTMAN,¹ G. Á. BAKOS,¹ AND J. N. WINN¹

¹ Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ 08540, USA

(Received June 22, 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. The CDIPS project is processing the TESS full-frame images and making light curves of stars that could be members of stellar associations, moving groups, or open clusters. Our eventual hope is to discover giant transiting planets with known ages, and probe the exoplanet size and separation distributions for planets younger than 1 Gyr. In this study, we made 159,337 light curves of candidate young stars in XXX 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 in agreement with theoretical expectations. We then highlight the 42 planet candidates we identified in an initial search, describe the vetting procedures that they have survived, and emphasize the additional spectroscopic, photometric, and imaging observations that are needed to verify that these candidates are in fact planets. We also comment on the possible use 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.

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 star properties do not appreciably change until the stars evolve 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.*, Kharchenko 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 in open clusters were detected 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 et al. 2007b; Miller et al. 2008a). While these studies led to val-

Corresponding author: L. G. Bouma
luke@astro.princeton.edu

able 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., CITE). 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 also 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 will be larger, as cluster membership catalogs are not yet complete, even at these relatively bright magnitudes (e.g., Cantat-Gaudin et al. 2018, 2019).

One of the major challenges in deriving precise photometry from the TESS images is the relatively poor angular res-

olution ($\approx 21''$ 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 90% of all star clusters¹. This is sensible, given that the large pixel size and the high stellar surface density will make aperture photometry unreliable. However, it also means that a large number of stars in clusters that could yield planet candidates may go unprocessed by the official TESS data reduction pipeline.

If one's concern is to discover and characterize benchmark systems with known ages, hunting for transiting planets in nearby stellar associations is a path that seems to be working. A number of promising targets from TESS are in various stages of follow-up from this approach, notably DS Tuc Ab (TOI 200, Benatti et al. 2019, Newton+2019), as well as TOI-450 and TOI-451. This is good, and these will remain benchmark systems because they are close and bright. However, to do statistics, we need more stars – we might also benefit from investigating the full range of cluster properties (densities, ages, metallicities). This motivates the need to go crowded, and fainter.

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 the process of selecting target stars. § 3 presents the photometric and image processing methods, and § 4 describes results. § 5 discusses, and § 6 concludes.

2. METHOD: STAR SELECTION

A major aim of the CDIPS project is to increase the number of cluster stars for which photometric time-series are available, to facilitate studies of exoplanetary and stellar processes across different times and stellar environments. An essential step is therefore to define a sample of stars that are thought to be young, or members or clusters, or both.

However, a homogeneous membership calculation for every known cluster is outside our scope. So too is a homogeneous search for young stars across the galaxy. Instead, we opt 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 therefore completeness, not accuracy. If there has been a claim in the litera-

¹ TICv8 has updated this to 10° , but for the first year of TESS observations, the number used when selecting target stars was 15° .

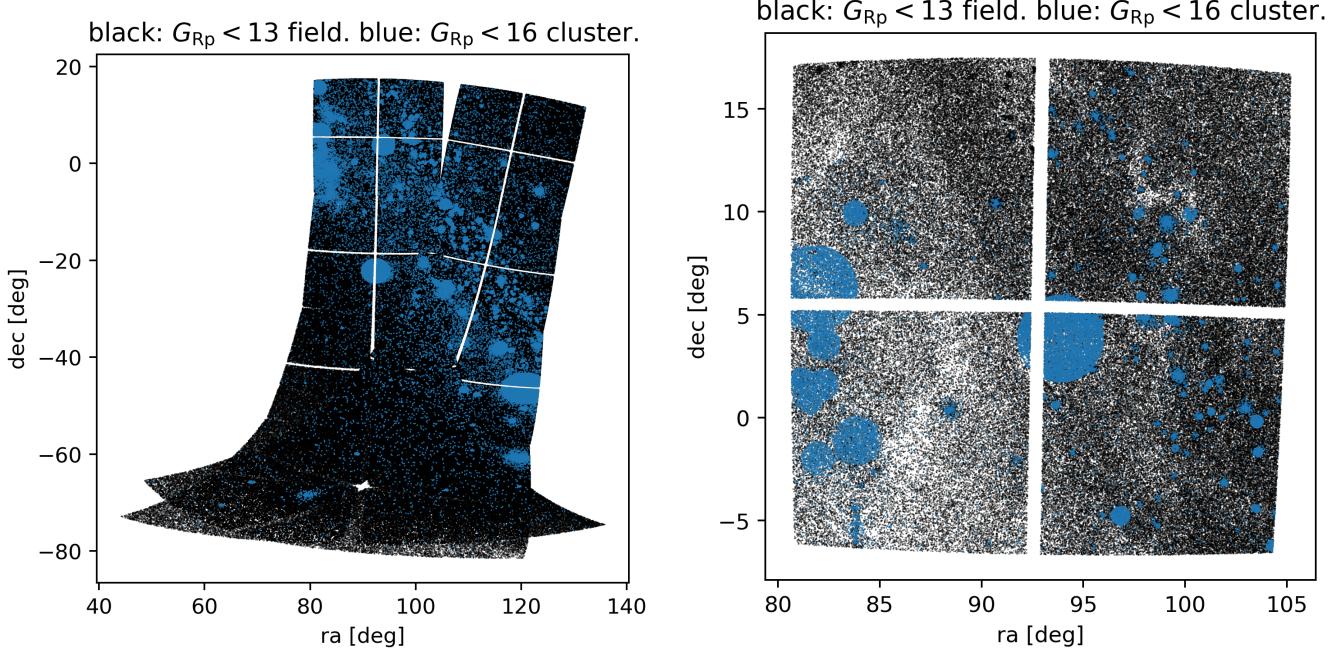


Figure 1. *Left.* Celestial positions of light curves from Sectors 6 and 7. *Right.* Celestial positions of light curves from Sector 6, Camera 1.

ture 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 large catalogs we used to identify candidate members of open clusters. § 2.2 describes the smaller 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 membership 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 preform 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 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 σ " 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 magnitudes², and required that the measured G -magnitude fall within 2 mag-

² See 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)

nitudes 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 ± 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 3. 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 particular 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 Gagne et al. catalogs, we searched the *Gaia*-DR2 archive for sources within 10 arcseconds of the listed positions. If Gagne et al. gave a proper motion, we required that the sign of each the *Gaia* proper motion components match that of the Gagne 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 Gagne 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 Gagne 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 spec-

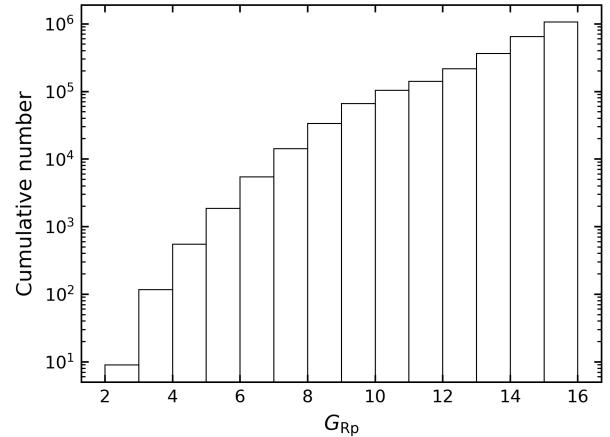


Figure 2. Cumulative number of target stars as a function of Gaia Rp -band magnitude. The selected stars are either candidate members of clusters, or else have other youth indicators. We only consider stars with $G_{Rp} < 16$.

troscopic 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 ≈ 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 `drl_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 selected 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 $G_{Rp} < 16$.

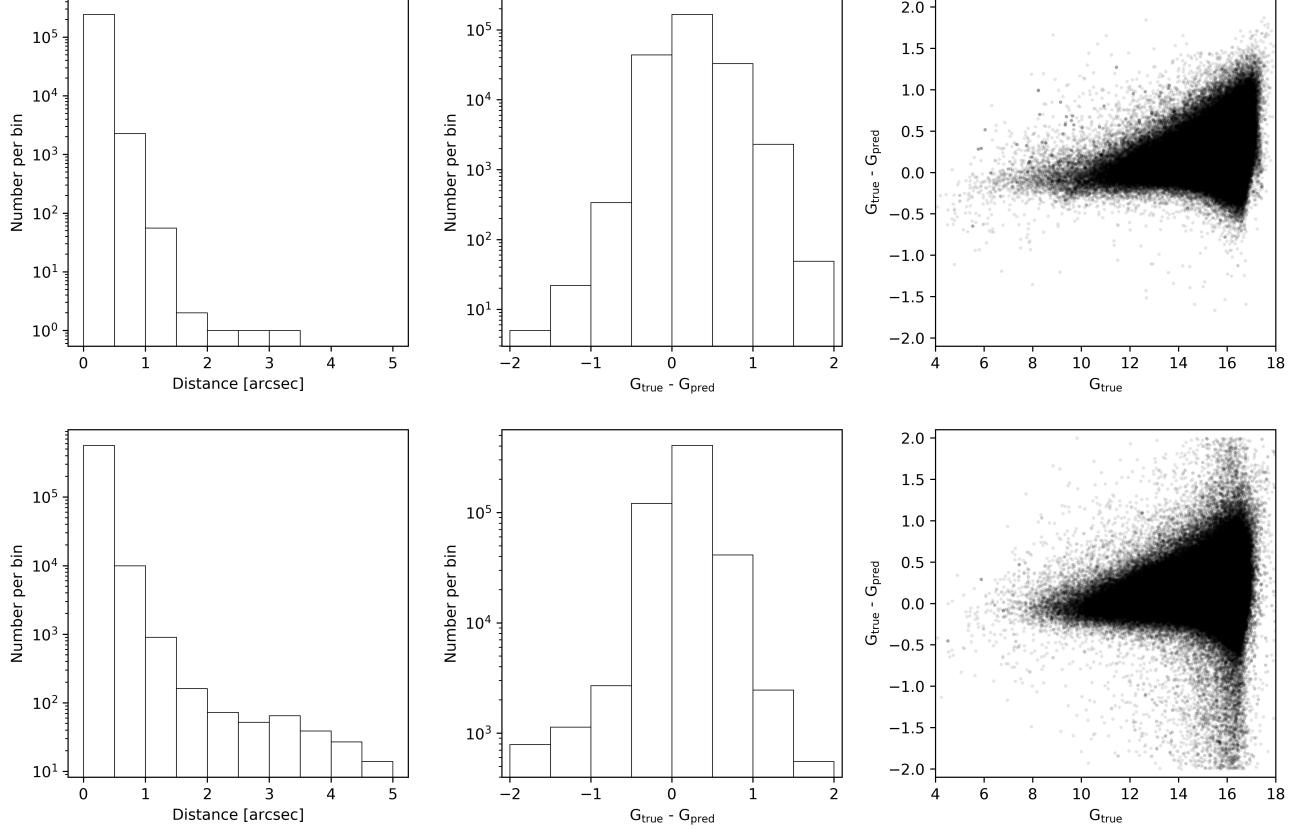


Figure 3. *Top.* Cross-match statistics from Kharchenko et al. (2013) vs. Gaia-DR2. *Bottom.* Ditto, for Dias et al. (2014) vs. Gaia-DR2.

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. The resulting CDIPS target star list is given in Table N. The distribution of target star brightnesses is shown in Figure 2.

3. METHOD: PHOTOMETRY

3.1. Overview

We reduced the TESS images to light curves by performing a sequence of steps using stand-alone programs. Our overall method is in the spirit of the reduction approaches described by Pál (2009), Huang et al. (2015), Soares-Furtado et al. (2017) and Oelkers & Stassun (2018); a conceptual overview is given in Figure 4.

We begin with the calibrated full frame images produced by the Science Processing Operations Center 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 (§ 3.3). Us-

ing the shape values from the initial astrometry, we select an astrometric reference frame to which we transform all of the calibrated images. We construct a photometric reference by stacking a collection of frames, convolve the transformed frames to this photometric reference, and subtract (§ 3.4). We perform aperture photometry on the subtracted images using positions projected onto the frame from Gaia DR2. We detrend the resulting light curves (§ 3.5). 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 ≈ 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 mit-

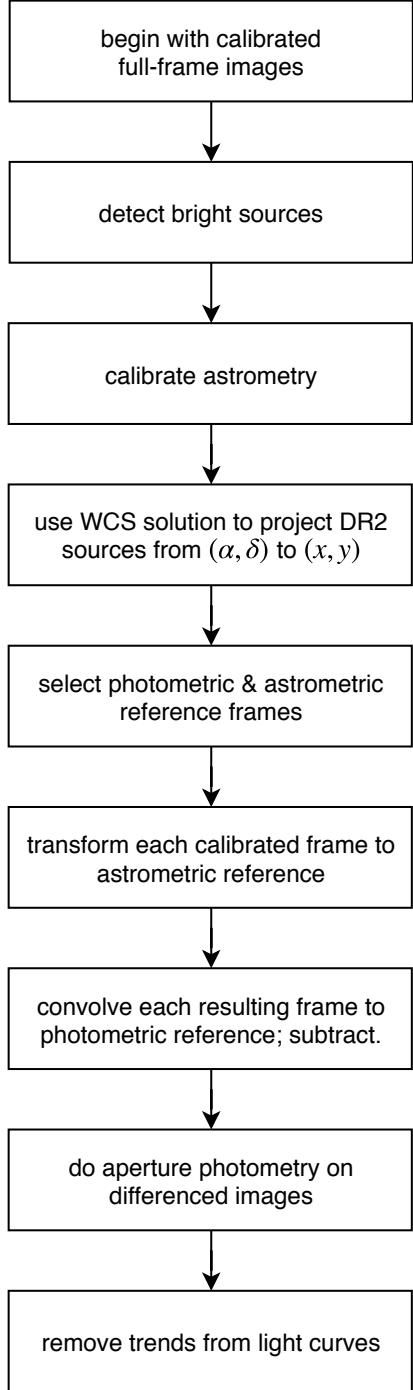


Figure 4. Conceptual overview of photometric reduction pipeline. Details are given in § 3.

igation algorithm is applied (CITE: BERTA-THOMPSON). 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 begin our analysis using the calibrated images, their uncertainty maps, and their associated headers. The space-craft 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 undergoing 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. Less than 2% of the CDIPS target star sample was therefore observed in the first five TESS sectors. Though a few interesting clusters are present (*e.g.*, Blanco 1, NGC 2516, NGC 1901), for the time being we opted to focus on sectors in which there were 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, & Meta-Data Collection

Before we can perform any kind of photometry, a few janitorial tasks are required.

First, we trim the images. We convert the calibrated image from MAST into a single-extension FITS image, trimmed to remove virtual rows and columns using the SCIROWS, 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 2σ from the image median, and then pass a 48×48 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 in small areas of less than 10% of the frames, this ad-hoc procedure removes large spatial scale scattered light patches.

After subtracting the background, we mask out saturated stars using a fixed saturation level of 8×10^4 ADU. This value was chosen based on the onset of visible trails of bleeding charge, 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 sec-

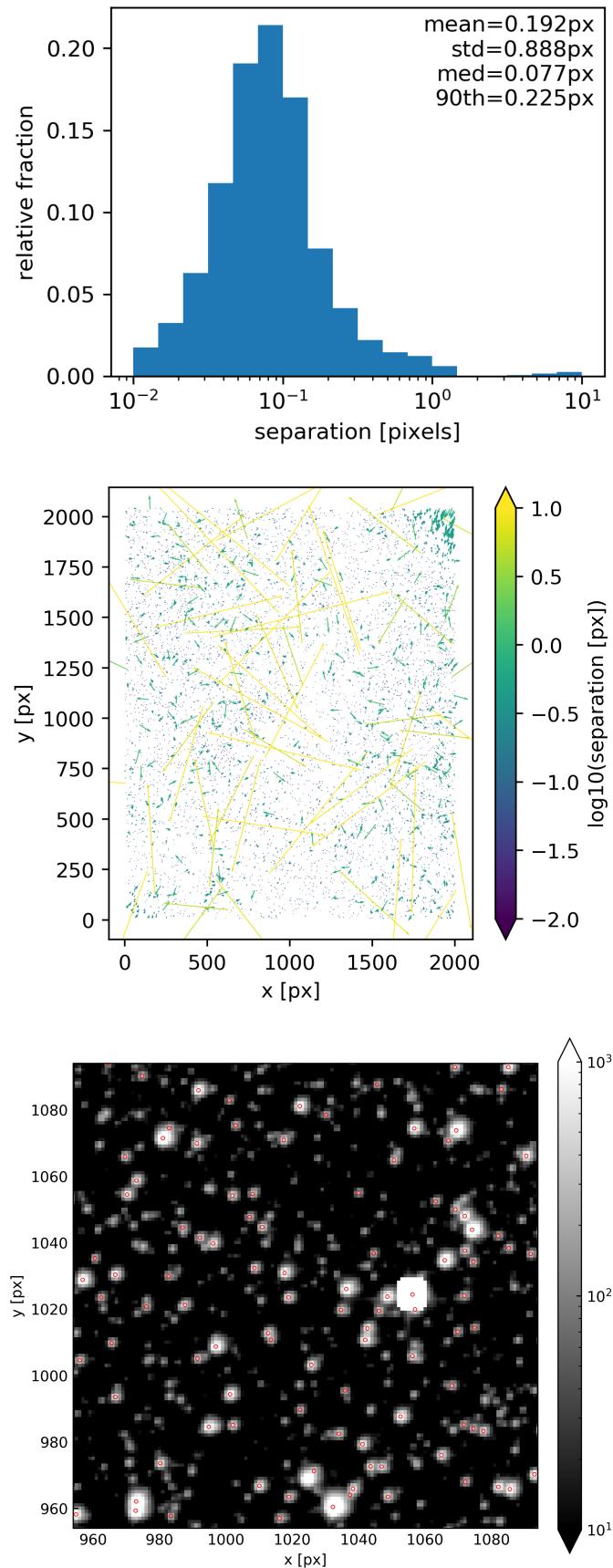


Figure 5. Top. Histogram Middle. Vector Bottom. Apertures

tor. 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 notes³ to identify further times with anomalous spacecraft performance, which we omit from consideration. For Sectors 6 through 8, these include three days at the beginning of Sector 6 dedicated to acquiring pixel response function data, and also about three days during Sector 8 lost to an instrument anomaly.

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 as a function of position is assumed to take the form

$$f_{\text{elong}}(\vec{x}) = B + A \exp\{-0.5 \times (s(\Delta x^2 + \Delta y^2) + d(\Delta x^2 - \Delta y^2) + k(2\Delta x\Delta y))\}, \quad (1)$$

for $\Delta x = x - x_0$, and $\Delta y = y - y_0$. For a nearly circular shape profile, the 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 (§ 3.4). We note that the fast focal ratio of the TESS cameras introduces significant comatic aberrations to the images: stars closer to the center of the field are more round, while stars towards the edges are more triangular.

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 (perhaps attributable to the algorithm used to select stars and measure 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 expected projected of these sources to center the apertures in our photometry, rather than attempting to detect the positions. Such “forced-aperture photometry” is preferable to performing source extraction in the crowded fields that are central in 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`⁴ (Kim 2018).

The astrometric residual is displayed for one photometric reference frame (to be described shortly) in Figure ???. A few salient points can be made. First, the typical median precision of the WCS solution is a bit below 0.1 pixels, and its 90th percentile is typically less than 0.3 pixels⁵. The distribution has a outlier tail of saturated stars which were masked, and thus do not have measurable centroid positions. 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. The final sub-panel of Figure ?? is a sanity check on the location of the projected positions, and also gives some context for our chosen aperture radii of 1, 1.5, and 2.25 pixels.

Finally, 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 from Gaia DR2. The pixel weights $w_{x,y}$ 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 (see Pál 2009 Fig 17). The background levels are measured in annuli around each aperture center. The raw flux of the object after background removal is then (Pál 2009 Eq 65)

$$f = \sum_{x,y} w_{x,y}(I_{x,y} - B) = f_{\text{total}} - Br_0^2. \quad (2)$$

The resulting measurements, for instance of the background level of each aperture, and the number of “good” objects that are detected, are later used to select photometric reference frames.

3.4. Image Subtraction

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 PSF variations. The kernel, once applied to the high signal-to noise reference, produces a model image, M_{xy} ,

$$M_{xy} = (R \otimes K)_{xy} + B_{xy}, \quad (3)$$

where B_{xy} is a component of the model image that allows for background variations. We model the background separately,

⁴ github.com/samuelyewl/gaia2read

⁵ In fact, in our reduction we automatically impose that the median residual and 90th percentile remain below 0.2 and 0.4 pixels, respectively.

so 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

$$\chi^2 = \sum_{xy} \left(\frac{I_{xy} - M_{xy}}{\sigma_{xy}} \right)^2, \quad (4)$$

where σ_{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 this work, the uncertainty in each target image pixel is 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 clarified 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 Section 2.8).

For posterity, we note a few implementation details. We must select two “reference frames” for image subtraction: (1) the astrometric reference; (2) the photometric reference.

To choose the astrometric reference, we search for frames with the largest and roundest stars (big s , small d and k values). We also require that the frame must minimal background noise, as measured in annuli around the bright stars selected in § 3.3. Finally, the astrometric reference frame must have, relative to the other frames being considered, a large number of detected sources. We sort the frames using these metrics, and then select the astrometric reference from successive intersections of each sorted list. Using the algorithm presented by Pál & Bakos (2006), we then use the `grtrans` tool to calculate an affine transformation that maps each calibrated frame to the astrometric reference. We use the flux-conserving interpolation scheme described by Pál (2009) to minimize photometric errors incurred during this step – standard bilinear or bicubic interpolation do not achieve flux conservation. The largest component of the transformation is typically a translation, of order 2 arcseconds, or about 0.1 TESS pixels.

The second 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 minimize Poisson errors, our photometric reference is a median average of $N = 50$ selected frames. The sub-frames are chosen to have low scatter in their photometry, and low background measurements, using the metadata collected in § 3.3. We convolve the 50 best candidate photometric references to the best photometric reference, and then perform a median combination of the frames to make the photometric reference.

To measure the reference flux for each star, we go a bit further. 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. If we were to stop here, *it would be a bad mistake*. The reference flux 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 fit for an for aperture-size specific zero point that relates each star's catalog T -band magnitude to the measured flux. The T -band magnitude is calculated according to Equation 1 of Stassun et al (2019, CITE). Crowding is minimized down to Gaia's resolution limit. We then use the known catalog magnitude to predict the expected photometric reference flux.

The final instrumental flux values we report (for example, the `IFL1` column) are similar to Pál 2009 Equation 83:

$$f = f_{\text{subtracted}} + f_{\text{reference}} \quad (5)$$

$$= \frac{1}{||K||_1^2} \sum_{x,y} S_{x,y} (w \otimes K)_{x,y} + g(T_{\text{cat}}), \quad (6)$$

for the subtracted image S , which is equal to $I - R \otimes K$, where I is the original 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 matched to the image convolution. The function g takes as input the target star's catalog magnitude T_{cat} , and returns the reference flux. Its coefficients have been fit per the procedure discussed in the preceding paragraph.

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 up 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 4.

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 main metrics for success included the measured RMS as a function of magnitude, and also the recovered SNR of transits from known TOIs (CITE: Guerrero et al in prep 2019).

Varying the kernel box-size, we found that kernels with box-length smaller than the typical TESS FWHM at field center (≈ 3 pixels) produced light curves with the highest average scatter. Increasing the kernel box-size from a 3×3 box to a 7×7 box led to about a 50% reduction in RMS for bright stars, and no difference for faint stars. The largest kernels, of (11×11) 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 (7×7) pixels, in part because the FWHM at the camera corners can grow by factors of ≈ 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 (7×7) kernel with second-order spatial polynomial weights in the basis functions. 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 would have lead to non-detections of objects 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 subtraction 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 on each frame per Equation 6. We do 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 in the pre-flight Sullivan et al. (2015) work. To finally convert from a list of sources on each frame to a list of flux values at any given time, we use the `grcollect` transposition tool.

3.5. 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 typically proceeds in a few sequential stages (see discussions from e.g. Bakos et al. 2010; Huang et al. 2015; Zhang et al. 2016). Given the heritage of our pipeline, we briefly discuss these, and then explain why we omit most of the steps described.

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, Sec. 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, Appendix). For ground-based data this may include zenith angle, or changing PSF parameters. For TESS data, this might include CCD temperature, or perhaps some measure of the scatter in the spacecraft pointing (CITE: Vanderburg 2019).

The final step is to then decorrelate the measured brightnesses of stars against each other (TFA, Kovács et al. 2005a). This accounts for variations due to unknown systematic instrument changes that affect many stars.

In § 3.5.1, we show that the “external parameter” dependence visible in the TESS data is rather complex: ordinary

linear model-fitting, as well as an initial attempts at non-linear model fitting, are poor descriptions of the data. In a similar vein “magnitude-fitting” is minimally justified, given how the TESS magnitudes correlate against these external parameters. We go on to show (§ 3.5.2) that a plausible detrending approach for the purpose of transit discovery is to simply decorrelate against other nearby stars with standard TFA.

3.5.1. Flux versus external parameters

The traditional approach to EPD is to fit and subtract a model for the magnitudes m of the form

$$m = \text{const.} + \sum_i c_i \theta_i, \quad (7)$$

where $\vec{\theta}$ is a vector of parameters such as the shape parameters (s, d, k) , their products $(s^2, s \cdot d, d^2, \dots)$, the temperature T of the instrument or environment⁶, the centroid positions (x, y) , the fractional part of the centroid positions $(\{x\}, \{y\})$, and any other parameters that are expected⁷ to correlate strongly with the observed flux. The coefficients c_i are calculated through linear least-squares, and subtracted to produce “EPD” light curves.

The premise of this model is that the correlations between the magnitudes and the external parameters are linear. For ground-based CCD data (e.g., HATNet, HATS, and Nikon DSLRs), Bakos et al. (2010) and Zhang et al. (2016) have verified that this model is a good description to the data. To discern whether such a model extends to the TESS data, we examined scatter plots of each parameter, as a function of all the other parameters. We also examined plots of each parameter as a function of time. A few prominent trends were present.

1. *Flux vs. time.* Most of the light curves we examined showed a secular drift with amplitude 0.01 mag over the timescale of each orbit. Sharper trends (“hooks”) at the beginning of each orbit seemed to be less prominent for stars at the corners of the fields than stars at the center. The periodicity incurred by the 2.5 day momentum dumps was also noticeable in more of the light curves at the center of the field than at the corners.
2. *Flux vs. centroid positions.* The flux as a function of centroid position often showed non-linear “hooks” (see Figure 7). Most of the data points from a given orbit reside at a given level, but about 10% are in a tail. This was seen in light curves all across the TESS field of view.

⁶ We used the temperature from the on-chip aluminum-copper sensor measurements included in the engineering data: archive.stsci.edu/missions/tess/engineering/.

⁷ The fractional centroid positions might matter because intra-pixel quantum efficiency variations could affect the measured stellar brightness. The varying temperature T of the CCD electronics might matter.

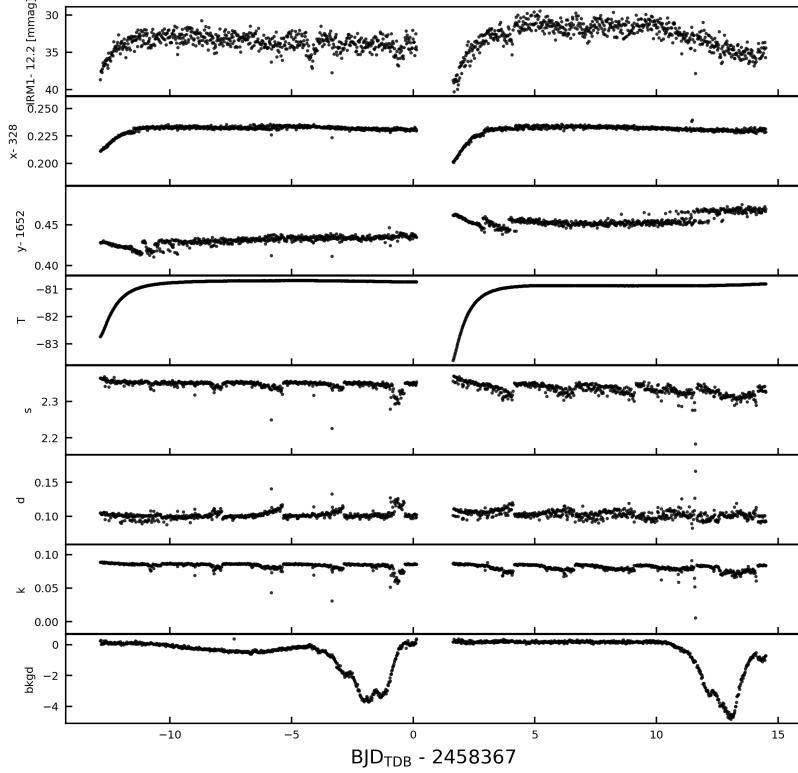


Figure 6. Timeseries of “external parameters” for a representative star. *Top:* Instrumental raw magnitude (with a particular aperture size), as a function of time. *Second and third from top:* x and y centroid positions as a function of time. Continuing in order are the CCD temperature, the (s, d, k) shape parameters, and the measured background value. Most of the apparent variability is instrumental: see § 3.5.1.

3. *Flux vs. temperature* exhibited similar hooks, with most of the flux values residing at a particular level, and perhaps 10% following a non-linear path (often resembling the Nike “swoosh”) away from the bulk of points.
4. *Flux vs. shape parameters.* For light curves in the corner of the field of view, similar hooks are present in flux vs. (s, d, k) , though the hooks are less sharp. In the center of the field of view, gaussian ellipses are a better description of the flux vs. the shape parameters.

Considering the timeseries of parameters other than flux (Figure 6):

1. *Centroid positions vs time.* The main variability in the centroid positions as a function of time is a secular drift, that is reset every orbit. The 2.5 day momentum wheel dump is superimposed on this secular drift, and has smaller amplitude than the drift.
2. *Temperature vs. time.* The main variability in temperature vs. time is a secular drift of the same timescale as that for the centroid positions timeseries.
3. *Shape parameters vs. time.* The main variability in the shape parameters as a function of time is the 2.5 day momentum wheel dump periodicity, with hooks before each momentum dump.

4. *Background value vs. time.* The background is typically stable, except when scattered light from the Earth or Moon enters the frame (visible towards the end of each orbit in Figure 6).

Given the characteristics of the variability, a linear model of the form given in Equation 7 is not applicable. To fit out the correlations between flux and parameters which most commonly exhibited “hooks”, we explored fitting a parametric open curve (an N -dimensional B-spline, Dierckx 1996) to the flux, centroid positions, and temperatures simultaneously. We selected the number of knots through brute-force, by calculating χ^2 for the model fit over a grid of possible knots, and minimizing the Bayesian Information Criterion. Though this approach showed some initial promise, even with “optimal” knot-selection (in the BIC sense) it introduced undesirable residuals in the light curves, and also distorted transits. One thought that we did not try but may explore in future work is to use the decorrelate against the scatter of the quaternion time-series (CITE: Vanderburg, 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.

3.5.2. Trend filtering algorithm

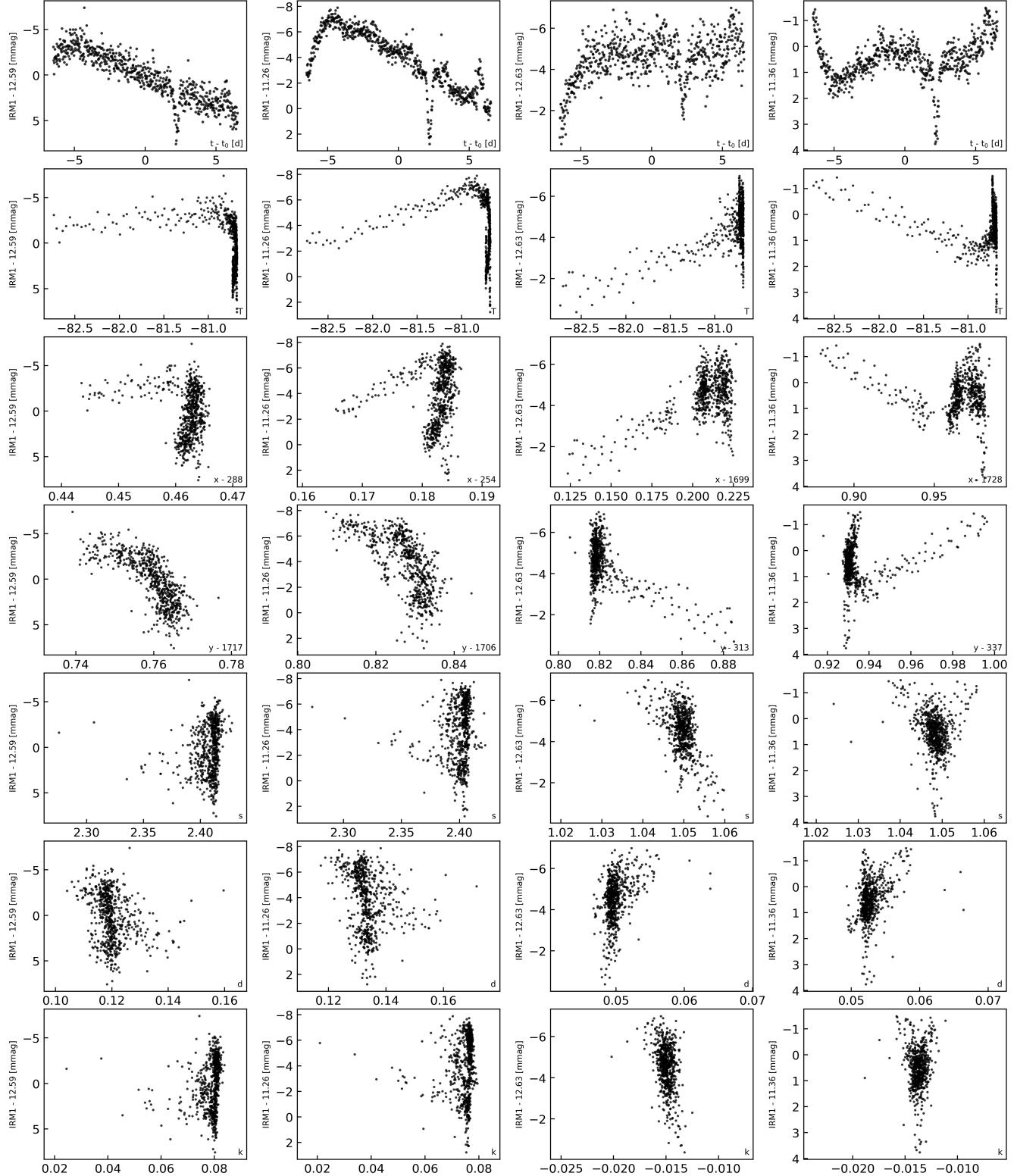


Figure 7. Flux as a function of “external parameters” for four representative stars. The left two columns are stars at the corner of a camera’s field; the right two columns are from the centers. Each row shows a different parameter along the x -axis, given in text at the bottom right of each subplot. “Hooks” are common features in flux as a function of temperature and centroid position. § 3.5.1 gives a verbose description.

Since most of the external parameter dependence is shared between stars, we opt to decorrelate the flux timeseries of each star against other stars in the frame. We use the TFA algorithm proposed by Kovács et al. (2005a), which for self-consistency we reproduce here. The idea of the method is somewhat simpler than the PDCMAP algorithm used in the SPOC pipeline (CITE). 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

$$F(i) = \sum_{j=1}^M c_j X_j(i), \quad (8)$$

for which the coefficients c_j are found by minimizing

$$\mathcal{D} = \sum_{i=1}^N [Y(i) - A(i) - F(i)]^2. \quad (9)$$

When trying to find periodic signals, $A(i)$ represents our prior knowledge of the light curve’s shape. This prior is simply that stars on average maintain a constant brightness:

$$A(i) = \langle Y \rangle = \frac{1}{N} \sum_{i=1}^N Y(i) = \text{const.} \quad (10)$$

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 notes on our implementation follow. We select template stars in two stages. In the first stage, we fit a parabola in the RMS-magnitude plane, and discard stars more than 2σ away from the prediction of the fit. We also require 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 perform an initial iteration of TFA, on only the candidate template stars. We inspect the resulting detrended light curves for residual structure by computing a Lomb-Scargle periodogram. If the maximum-power peak has a false alarm probability below 0.1%, we exclude the star from the list of candidate template stars, on the basis of its presumed periodic variability. We then randomly select at most 200 template stars from the remaining non-variable candidates. The choice of number of template stars was discussed by Kovács et al. (2005b), and is another free parameter in the broad problem of light curve production. While it can be optimized by constructing and minimizing a BIC-like quantity, a little overfitting is acceptable for our purposes.

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

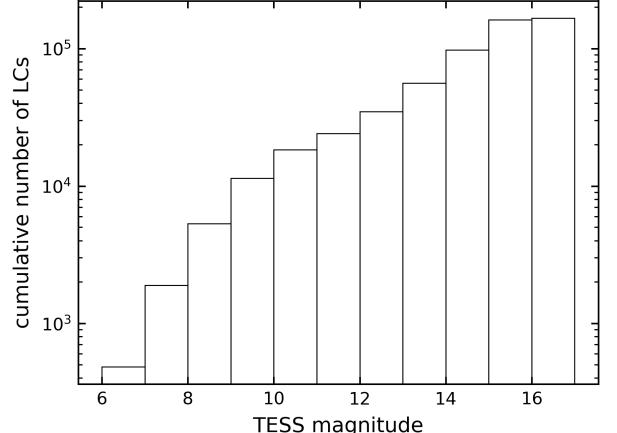


Figure 8. Cumulative distribution of TESS-band magnitudes for CDIPS light curves.

4. RESULTS

4.1. Light Curve Statistics

4.1.1. Stellar properties

Describe Figure 8 Figure 9 Figure 10

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 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 (1σ) 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 this sample was completeness, rather than fidelity. The Dias et al. (2014) stars in particular were included if they were listed with membership probability exceeding 50%. To create cleaner sub-samples, we advise use of the CDEXTCAT header keyword, which can be cross-matched 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 ...

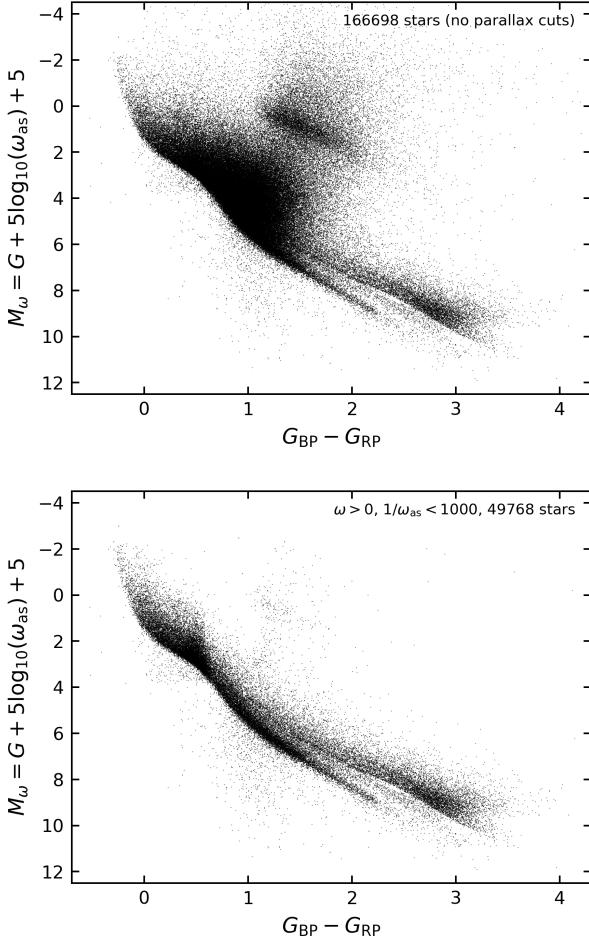


Figure 9. *Top.* HR diagram of CDIPS stars with light curves in this data release. *Bottom.* HR diagram of close CDIPS stars with light curves. The wedge separating the pre-MS sample from the MS stars was discussed by Zari et al. (2018), who introduced it in order to avoid contamination by photometric binaries.

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

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 11. For the y-axis of this plot, we have taken

$$\text{RMS} = \left[\frac{1}{N-M-1} \sum_{i=1}^N (f_i - \bar{f})^2 \right]^{1/2}, \quad (11)$$

where f_i is the value of the flux at the i^{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

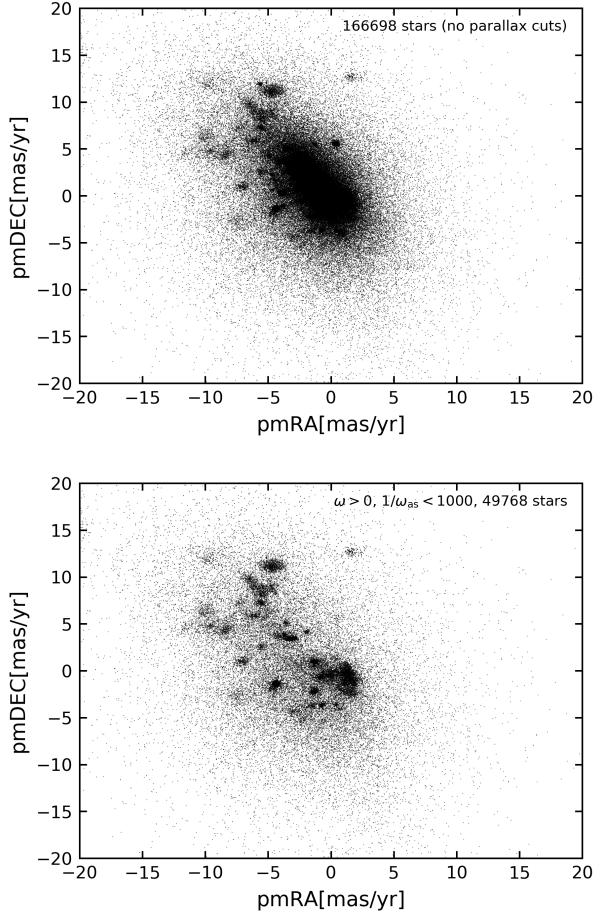


Figure 10. *Top.* Proper motions of CDIPS stars with light curves in this data release. *Bottom.* Proper motions of close CDIPS stars with light curves. A few additional clusters are visible.

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 expected 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 a slightly steeper scaling. 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 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

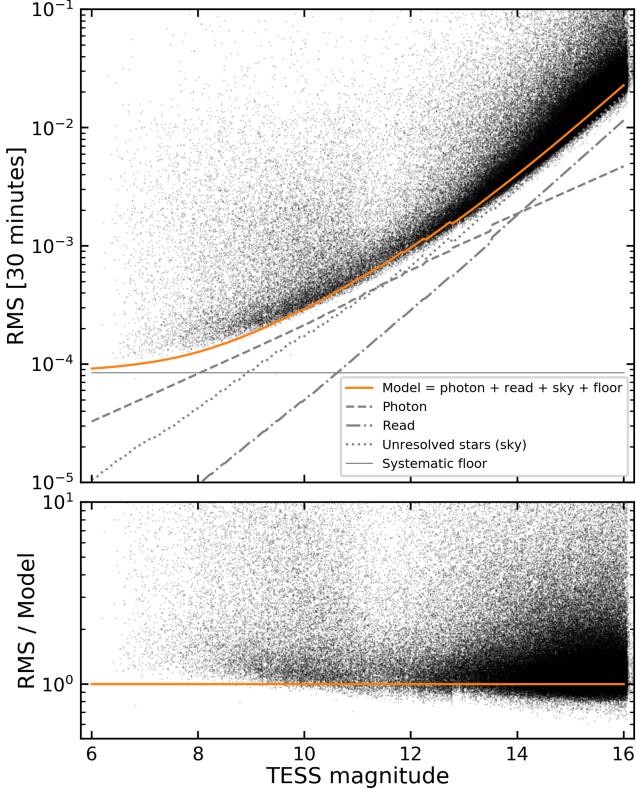


Figure 11. Standard deviation vs catalog TESS-band magnitude for trend-filtered CDIPS light curves. Black points correspond to the minimum RMS across the available three aperture sizes. The model (orange and gray lines) uses the optimal aperture sizes reported by Sullivan et al. (2015), and the effective 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 and read noise 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).

$T \approx 9$ (see Sullivan et al. 2015, Figure 14). Since the brightest stars are not the focus of the present work, we leave this as a point to improve in future work.

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 what would be predicted from their catalog magnitudes. 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). 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 as we had hoped.

Expected RMS vs magnitude. —The noise model shown in Figure 11 is quite similar to that of Sullivan et al. (2015), save

for two changes. The first change is minor: the effective area of the telescope is updated to be 86.6 cm^2 , per the measurements of Vanderspek et al. (2018).

The second point is more important: we explicitly include 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). Winn (2013) estimated the photon-counts from each component that would be observed by TESS. His zodiacal light model was presented by Sullivan et al. (2015), but the faint star model was not emphasized since the Sullivan simulations were performed away from the galactic plane.

A few words are therefore pertinent about the diffuse sky model we have used for Figure 11, which 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) . Winn (2013) queried the Besançon model (Robin et al. 2003) along a grid of galactic sight-lines, and converted the resulting $N(< m, l, b)$ surface densities into I -band surface brightnesses. Fitting a smooth function to the results, Winn (2013) found

$$I \text{ mag arcsec}^{-2} = a_0 + a_1 \left(\frac{|b|}{40^\circ} \right) + a_2 \left(\frac{|l|}{180^\circ} \right)^{a_3}, \quad (12)$$

where the galactic longitude l is measured from -180° to 180° , 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 systematically underestimate the background brightness by an even larger degree.

ACF statistics —ACF statistics before and after detrending(?)
SNR of retrieved HJs.

Maybe movies of subtracted images?

Some stellar variability plots (perhaps of known stellar variables).

Some focus on particular clusters.

4.2. Objects of Interest

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 curve described in the preceding sections. This processing was entirely independent of the data release, since it was quite specific to our own scientific interests.

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`⁸ are slightly more efficient than in most BLS implementations, since `tls` uses the cubic-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 expected for a sample of young stars; the issue of finding transits in the face of large stellar rotation signals has been explored by both Rizzuto et al. (2017) and also Hippke 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. Hippke et al. (2019), conversely, described a large number of possible detrending methods, and found that many 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 Hippke et al. (2019). First, we run a generalized Lomb-Scargle periodogram (CITE) on each TFA light curve. If a peak is found with false alarm probability below 10^{-5} , we consider the star “variable”, and opt to detrend the light curve 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 a larger number of 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 2σ clipping of outliers from the fit residuals at each iteration (Servén et al. 2018; Hippke et al. 2019). The maximum number of spline knots is set to 50, which for each TESS sector (≈ 25 days) is commensurate with a ≈ 0.5 day window.

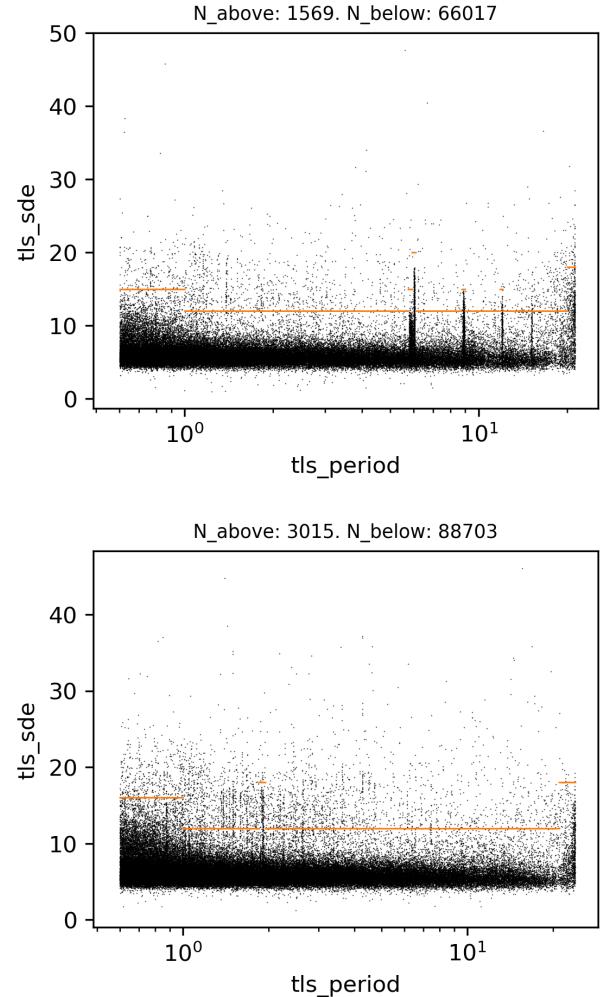


Figure 12. Significance thresholds empirically defined for the transiting planet search described in § 4.2. Each point represents the TLS periodogram peak signal detection efficiency from one light curve. *Top.* Sector 6. *Bottom.* Sector 7.

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

⁸ <https://github.com/hippke/tls>

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 9 is set to the phase-binned signal from the most powerful peak in the TLS spectrum, rather than being a constant function. This helps improve lower the scatter of the light curve at frequencies near the TLS peak.

We then make a multi-page .pdf document report with the information necessary to make classifications for vetting. These documents 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 summary .pdf is given in the Appendix.

The classifications we used assess:

1. *The source 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. We therefore include tags only for non-cluster members, and possible non-cluster members. Typically the primary source of this information is the Gaia-DR2 parallax.
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 actual classification process was performed by LGB, JH, and JNW. The TagSpaces software was used – an extremely helpful tool for the purposes of easily assigning labels to documents.

5. DISCUSSION

Lorem ipsum.

6. CONCLUSION

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.

Facility: TESS (Ricker et al. 2015), Gaia (Gaia Collaboration et al. 2016, 2018b), 2MASS (Skrutskie et al. 2006), DSS (CITE)

Software: astrobase (Bhatti et al. 2018), astropy (Collaboration et al. 2018), astroquery (Ginsburg et al. 2018), astroquery.gaia CITE, astroquery.simbad CITE, astroquery.mast CITE, astroquery.nasaexoplanetar CITE, 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) scipy (Jones et al. 2001), TagSpaces (CITE), tesscut (Brasseur et al. 2019), wotan (Hippke et al. 2019)

REFERENCES

- Aigrain, S., Hodgkin, S., Irwin, J., et al. 2007, *Monthly Notices of the Royal Astronomical Society*, 375, 29
- Alard, C., & Lupton, R. H. 1998, *ApJ*, 503, 325
- 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
- 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
- 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
- Ciardi, D. R., Crossfield, I. J. M., Feinstein, A. D., et al. 2018, *The Astronomical Journal*, **155**, 10
- Clarke, B. D., Caldwell, D. A., Quintana, E. V., et al. 2017, *Kepler Science Document*, 5
- Collaboration, T. A., Price-Whelan, A. M., Sipöcz, 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
- 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*
- 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
- Jaffe, T. J., & Barclay, T. 2017, *Tessgi/Ticgen: V1.0.0*, DOI: 10.5281/zenodo.888217
- 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)
- Kovács, G., Bakos, G., & Noyes, R. W. 2005a, *Monthly Notices of the Royal Astronomical Society*, **356**, 557
- . 2005b, *MNRAS*, **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
- Lovis, C., & Mayor, M. 2007, *Astronomy and Astrophysics*, 472, 657
- Malavolta, L., Nascimbeni, V., Piotto, G., et al. 2016, *Astronomy and Astrophysics*, 588, A118
- 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
- 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
- Pál, A. 2009, PhD thesis, arXiv: 0906.3486
- . 2012, *MNRAS*, 421, 1825
- Pál, A., & Bakos, G. A. 2006, *Publications of the Astronomical Society of the Pacific*, 118, 1474
- 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., DerriÁlre, 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., 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
- 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
- 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
- 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
- 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., & SchÁulkopf, 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
- Winn, J. N. 2013, *TESS Science Memo No. 2, Version 1*. Available upon request.

- Zari, E., Hashemi, H., Brown, A. G. A., Jardine, K., & de Zeeuw, P. T. 2018, *Astronomy and Astrophysics*, 620, A172
- Zhang, M., Bakos, G. Á., 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 ≈ 17 degrees, so naively one might incur at worst an error of ≈ 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_{TDB} time system, which is the defacto standard in exoplanet and stellar astronomy (Eastman et al. 2010).

B. VETTING DOCUMENT DESCRIPTION

In § 4.2, we described the process by which we made .pdf documents suitable for assessing which objects were interesting enough to merit further study.

This section briefly summarizes this document. Updated versions and their README files will live at this web-address: mast.stsci.edu/CDIPS.

- *Page 1: period-search summary.* Periodograms from TLS and phase-dispersion minimization (Hippke & Heller 2019; Stellingwerf 1978), as calculated with astrobase.periodbase are shown. 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 shown (CITE).
- *Page 2: light curve diagnostics.*