

The Cluster Difference Imaging Photometric Survey (CDIPS). I. Method, Light Curves, and Planet Candidates from Sectors 6–8

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 May 22, 2019; Revised —; Accepted —)

Submitted to AAS journals.

ABSTRACT

Lorem ipsum.

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

1. INTRODUCTION

The exoplanet field is evolving. At the beginning, Doppler spectroscopy by the Geneva and California groups led to the first exoplanet detections (CITE: Mayor 95, Butler, Marcy). Some of these planets transited (CITE: Henry, Charbonneau). Despite the geometric rarity of transits, photometric monitoring can be performed in parallel; with precise spectroscopy this has not yet been achieved. Parallel photometric monitoring enabled thousands of planet detections with the Kepler spacecraft (CITE: Borucki, Morton). The abundance of planets from Kepler taught us about statistics (CITE: Howard, Fressin, Petigura). By nature of its design, Kepler left open two major avenues for improvement: (i) brighter stars, (ii) rare weirdos. TESS (CITE: Ricker), by nature of its small, wide-field cameras, and nearly all-sky coverage, can capitalize on both.

One class of rare weirdo to which TESS is sensitive is stars in clusters. For brevity, we use the term “cluster” to refer to open clusters, moving groups, associations, and star-forming regions. Each of the ~ 1000 star clusters of the Milky Way is a gift to astrophysics, providing a sample of stars that vary widely in mass but all have approximately the same age and composition.

Cluster stars have been photometered from the ground for a while (CITE: Joel thesis, Cambridge group). A few clusters, NAME1 and NAME2, were observed by Kepler in its prime mission (CITE: Meibom). And a few more, most notably X, Y, and Z, were observed in K2’s ecliptic mission (CITE).

TESS holds the promise to deliver the most homogeneous and comprehensive cluster photometric survey in history. Based on the cluster membership database of Kharchenko et al. (2013) and the TESS apparent magnitude calculator of

Jaffe & Barclay (2017), we estimate that 10^5 cluster members brighter than $T = 16$ will be observed in the northern ecliptic hemisphere’s full-frame images.

However, there are formidable obstacles to deriving precise photometry from the TESS images because of the relatively poor angular resolution. Almost all of the clusters are within 10 degrees of the Galactic plane (see Figure 1), where the problems with crowding and complex backgrounds are so severe that the TESS Science Team has neglected that portion of the sky in their planet simulations (Sullivan et al. 2015). Similarly, the TESS Candidate Target List deprioritizes all objects within 15° of the galactic plane (?), which includes 90% of all star clusters. The large pixel size and the high stellar surface density will make aperture photometry unreliable. Yet, aperture photometry is the basis of the official TESS data reduction pipeline and almost all reduction methods that have been applied to Kepler and K2 data.

We have begun to produce photometry from the TESS images. For the time being, we are adopting the difference imaging (or optionally, “image subtraction”) technique, because of its benefits in modelling complex backgrounds in crowded regions.

In what follows, § 3 presents BLAH, and § 4 describes BLAH BLAH. § 5 discusses, and § 6 concludes.

2. METHOD: STAR SELECTION

A major motivator for the CDIPS project is to increase the number of cluster stars for which continuous photometric timeseries are available, and thus facilitate studies of exoplanetary and stellar processes across different times and stellar environments. An essential aspect of the project is therefore to define a *cluster star sample*, made of stars that are thought to be members of clusters.

Performing a homogeneous membership determination for every known cluster is beyond the scope of the current work. Instead, we collect various cluster membership catalogs from the literature. We then use them to identify cluster members

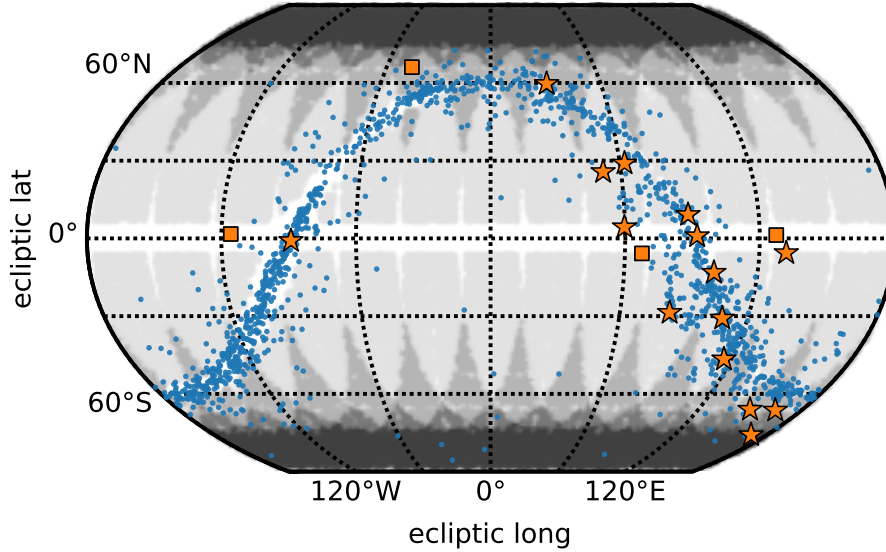


Figure 1. There are many clusters; TESS looks at them. Placeholder sky map.

within the TESS images. In initially selecting stars, we aim for completeness, not accuracy. If there has been a claim in the literature that a star should be considered a cluster member, we would like to report a light curve for the star. For stars that are photometrically interesting, we can perform post-hoc cluster membership quality checks using Gaia DR2 astrometry and photometry.

Table N describes the cluster membership catalogs we have used to identify candidate cluster members. Since our photometric reduction is based on positions reported in *Gaia* DR2, each of our light curve sources by default is associated with a Gaia identifier. Correspondingly, the cluster membership determinations performed by Cantat-Gaudin et al. (2018) and Gaia Collaboration et al. (2018a) are the easiest case for us to merge against our light curve database. For the TESS Sectors 1–5 light curves produced in this study, this yielded X and Y light curves respectively.

2.1. Big catalogs: open clusters

Gaia-derived OC memberships—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).

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 500 pc. While

this work also included memberships for globular clusters, we omitted these from consideration.

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. We selected their “ 1σ ” 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 magnitudes¹, 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 ± 2 G -band magnitudes of the prediction. From 2,034,269 stars, this yielded a unique best neighbor

¹ 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)

bor 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 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 α 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 32Ori 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 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 ≈ 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 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 $G_{\text{Rp}} < 16$.

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. The resulting CDIPS target star list is given in Table N.

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), Soares-Furtado et al. (2017) and Oelkers & Stassun (2018); a conceptual overview is given in Figure 3.

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 calibration, and coarse simple aperture photometry (§ 3.3). Using 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

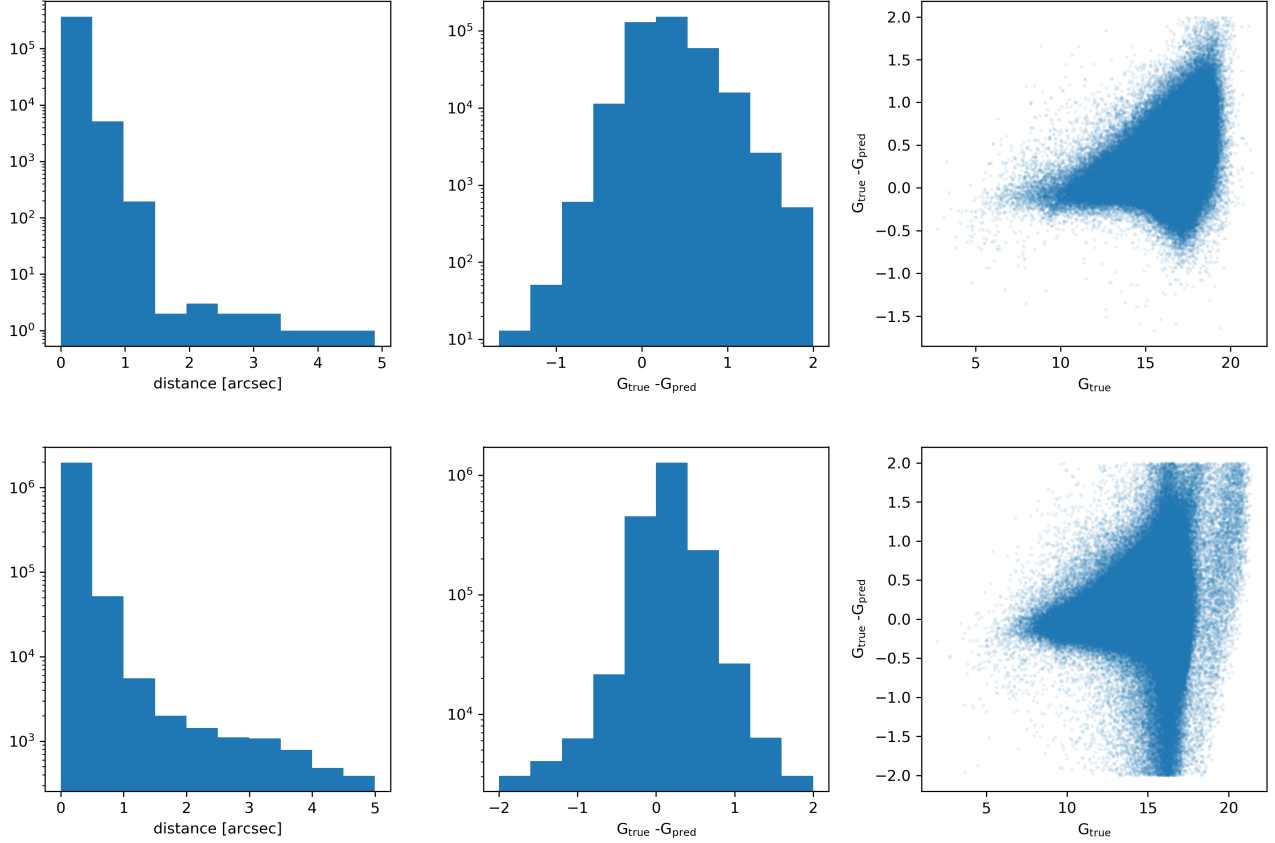


Figure 2. *Top.* Cross-match statistics from Kharchenko et al. (2013) vs. Gaia-DR2. *Bottom.* Ditto, for Dias et al. (2014) vs. Gaia-DR2.

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 telemetry load, the resulting pixel values are averaged by the onboard computer into 30 minute exposures. An on-board cosmic ray mitigation 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 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 undergoing analysis.

While we performed numerous initial tests on the first sectors of data, by geometric coincidence Sectors 1–5 fell away from the galactic plane. Less than 2% of the CDIPS target star sample is 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).

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.

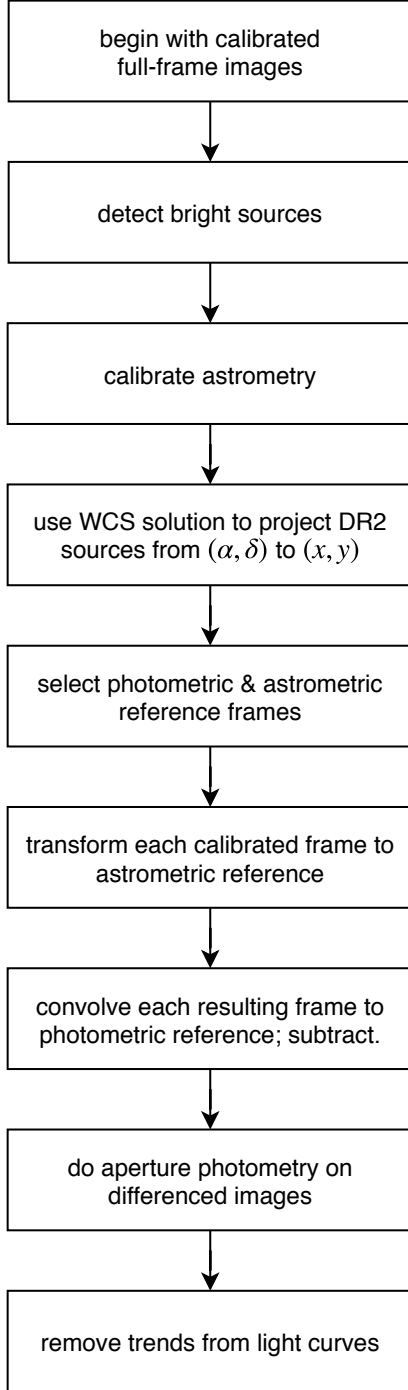


Figure 3. Conceptual overview of photometric reduction pipeline. Details are given in § 3.

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 re-

mains 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 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 notes² 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. To obtain an astrometric solution (independent from the solution included with the frames), we use `fistar` to perform source extraction on bright stars in each image. We pass the resulting source lists through `astrometry.net` (Lang et al. 2010), which returns an astrometric solution for each frame packaged in the WCS format (Pence et al. 2010, Sec. 8). 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.

² archive.stsci.edu/tess/tess_drn.html

With the resulting WCS information, we then project a source catalog onto each frame. Initially, we intended 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. Our source catalog for each frame is therefore a combination of the CDIPS target stars ($G_{Rp} < 16$), and all Gaia-DR2 sources down to $G_{Rp} < 13$. We use the expected 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 performing source extraction simply because the TESS pixels are large, and we are operating in crowded fields. The Gaia-DR2 epoch is J2015.5, so even the fastest-moving stars with proper motions of ~ 1 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 standard WCS transformations, and the source catalog look-up is performed using `gaia2read`³ (Kim 2018).

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)

$$f = \sum_{x,y} w_{x,y}(I_{x,y} - B) = f_{\text{total}} - B r_0^2. \quad (2)$$

The resulting measurements, for instance of the background level of each aperture, and the number of “good” objects that are detected, are later used as input for selecting 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 a convolution kernel. 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, 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. (2008). 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 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

³ github.com/samuelyeewl/gaia2read

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 re-

trieved 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 led 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 choose “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. To finally convert from a list of sources on each frame to a list of flux values at any given time, we use the `gcollect` 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

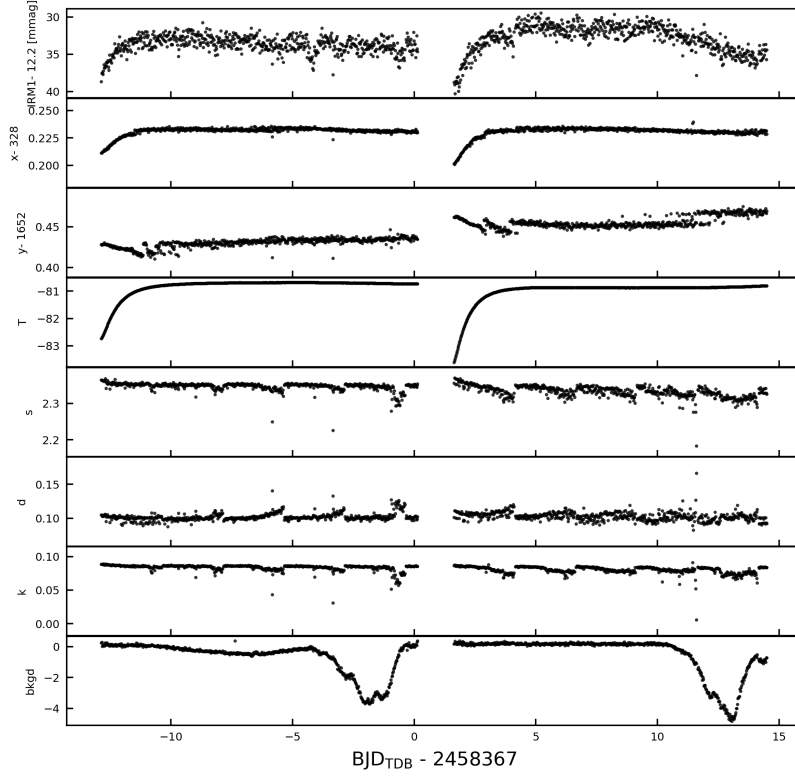


Figure 4. Timeseries of “external parameters” for a representative star. *Top:* Instrumental raw magnitude (with a particular aperture size), as a function of time. *Second and third from top:* x and y centroid positions as a function of time. Continuing in order are the CCD temperature, the (s, d, k) shape parameters, and the measured background value. Most of the apparent variability is instrumental: see § 3.5.1.

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 environment⁴, the centroid positions $(\{x\}, \{y\})$, the fractional part of the centroid positions $(\{x\}, \{y\})$, and any other parameters that are expected⁵ 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 5). 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.
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

⁴ We used the temperature from the on-chip aluminum-copper sensor measurements included in the engineering data: archive.stsci.edu/missions/teess/engineering/.

⁵ 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.

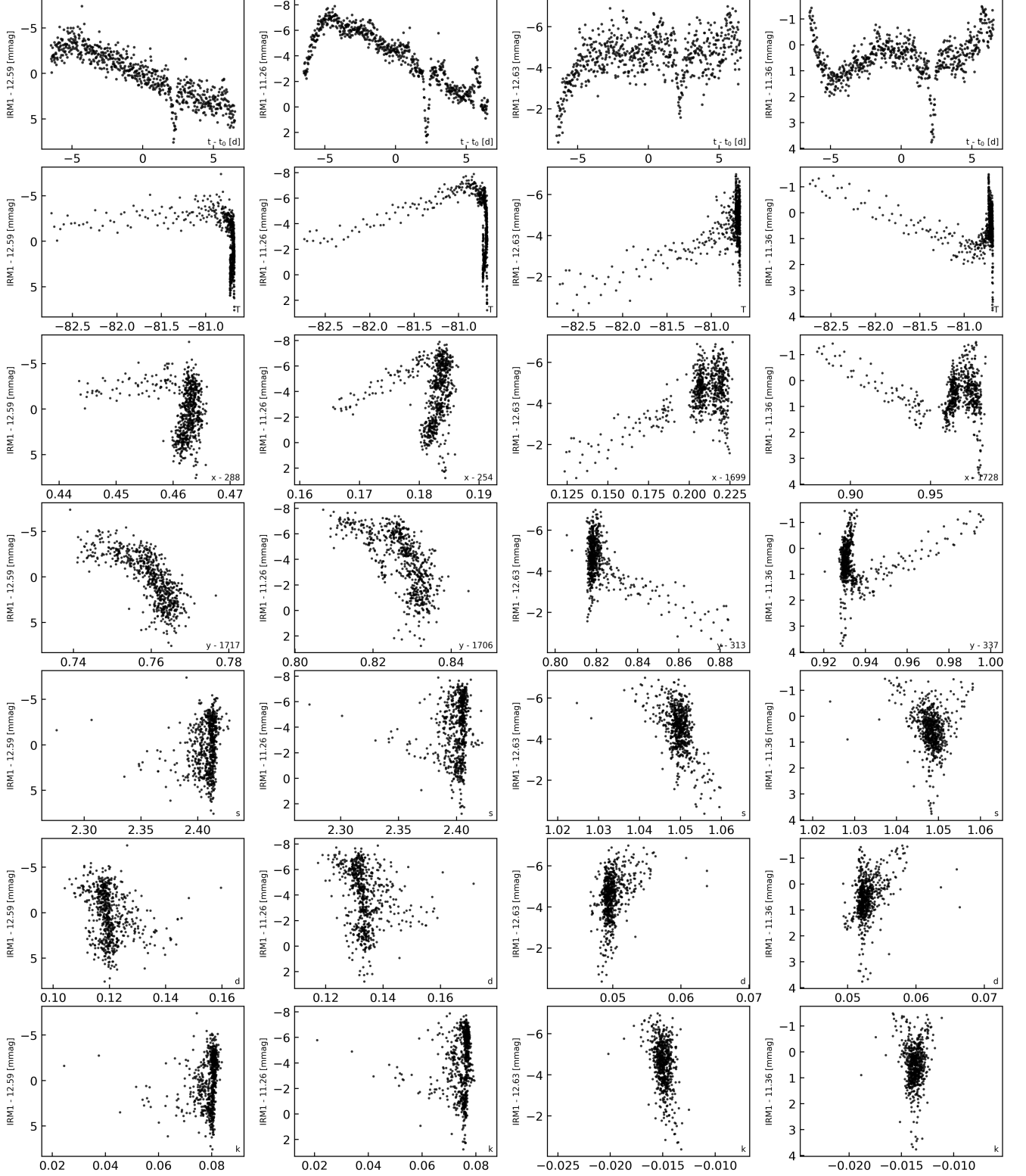


Figure 5. Flux as a function of “external parameters” for four representative stars. The left two columns are stars at the corner of a camera’s field; the right two columns are from the centers. Each row shows a different parameter along the x -axis, given in text at the bottom right of each subplot. “Hooks” are common features in flux as a function of temperature and centroid position. § 3.5.1 gives a verbose description.

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 4):

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 4).

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 quarternion 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

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.

4. RESULTS

4.1. Light Curve Statistics

RMS vs mag plots.

ACF statistics.

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

As an initial search for transiting planets, strong stellar rotators, and eclipsing binaries, we performed a few steps of post-processing. For simplicity, we chose a single aperture size – “aperture 2” – with a radius of 1.5 pixels, to perform the period search.

First, 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 the `tls` implementation⁶ are slightly more efficient than in most BLS implementations, since it uses the cubic-in-frequency sampling advocated for by Ofir (2014), rather than standard linear-in-frequency sampling. During 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 shrinks the data volume by about 5%, but also lowers the number of systematic false positives in subsequent vetting. We also ran a generalized Lomb-Scargle periodogram (CITE), simply to have additional periodicity information.

We then performed a cut on the TLS signal detection efficiency: $SDE > 12$. This yielded XXX light curves, and we then performed TFA signal reconstruction on these lightcurves using `varertools`. 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. Generally speaking, this helps improve the quality of the resulting light curve.

We then make a multi-page `.pdf` document with the information necessary to make classifications for vetting. These documents are released along with the light curves, as 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 ..., 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` `astroquery.simbad` `astroquery.mast` `astroquery.nasaexoplanetarchive` `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), `scipy` (Jones et al. 2001), `TagSpaces` (CITE), `tesscut` (CITE)

⁶ <https://github.com/hippke/tls>

REFERENCES

- Alard, C., & Lupton, R. H. 1998, *ApJ*, 503, 325
- Bakos, G. A., Torres, G., Pál, A., et al. 2010, *The Astrophysical Journal*, 710, 1724
- Bell, C. P. M., Murphy, S. J., & Mamajek, E. E. 2017, *Monthly Notices of the Royal Astronomical Society*, 468, 1198
- Bhatti, W., Bouma, L. G., & Wallace, J. 2018, *astrobase*
- 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
- Cantat-Gaudin, T., Jordi, C., Vallenari, A., et al. 2018, *Astronomy & Astrophysics*, 618, A93
- Carrasco, J. M., Evans, D. W., Montegriffo, P., et al. 2016, *Astronomy and Astrophysics*, 595, A7
- 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*
- 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
- Eastman, J., Siverd, R., & Gaudi, B. S. 2010, *Publications of the Astronomical Society of the Pacific*, 122, 935, *arXiv: 1005.4415*
- 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
- 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
- Hippke, M., & Heller, R. 2019, *arXiv:1901.02015 [astro-ph]*, *arXiv: 1901.02015*
- Huang, C. X., Penev, K., Hartman, J. D., et al. 2015, *Monthly Notices of the Royal Astronomical Society*, 454, 4159
- Hunter, J. D. 2007, *Computing in Science & Engineering*, 9, 90
- Jaffe, T. J., & Barclay, T. 2017, *Tessgi/Ticgen: V1.0.0*, DOI: 10.5281/zenodo.888217
- 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
- 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
- Miller, J. P., Pennypacker, C. R., & White, G. L. 2008, *Publications of the Astronomical Society of the Pacific*, 120, 449
- 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
- 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
- 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
- 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
- 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>
- Vanderspek, R., Doty, J., Fausnaugh, M., et al. 2018, *TESS Science Document*
- Walt, S. v. d., Colbert, S. C., & Varoquaux, G. 2011, *Computing in Science & Engineering*, **13**, 22
- Wang, D., Hogg, D. W., Foreman-Mackey, D., & SchÅlkopf, B. 2017, [arXiv:1710.02428 \[astro-ph\]](https://arxiv.org/abs/1710.02428), arXiv: 1710.02428
- 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

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.*