

Homogeneous Light Curves for Stars in Clusters from TESS: Method & Sector 1-4 Data Release

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 February 1, 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

Lorem ipsum.

In what follows, § 2 presents BLAH, and § 3 describes BLAH BLAH. § 4 discusses, and § 5 concludes.

2. METHOD

2.1. Overview

We reduced the TESS images to lightcurves by performing a sequence of steps using stand-alone programs. A conceptual overview of our “pipeline” is given in Figure 2. Our overall method is in the spirit of the reduction approaches used by Pál (2009), Soares-Furtado et al. (2017) and Oelkers & Stassun (2018).

We begin with the calibrated full frame images produced by SPOC (§ 2.2). We then perform a collection of preparatory steps, including source extraction of bright stars, astrometry using the resulting positions, and coarse simple aperture photometry (§ 2.3). Using the shape values from the initial astrometry, we select an astrometric reference frame and transform all of the calibrated images to it. We construct a photometric reference by stacking a collection of frames, and then convolve all the transformed frames to the photometric reference, and subtract (§ 2.4). We perform aperture photometry on the subtracted images using positions projected onto the frame from Gaia DR2. We detrend the resulting lightcurves (§ 2.5), and then assess the resulting lightcurves using a variety of statistics for their white noise and red noise properties (§ 3).

2.2. 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 in ecliptic longitude every lunar month. Data acquired throughout each “sector” is 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 reads out every 2 seconds. To produce a manageable telemetric load, the resulting pixel values are averaged by the onboard computer into 30 minute exposures. (An on-board cosmic ray mitigation algorithm is applied.) 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 dividing out a flat field. The details are given 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.

2.3. Preparatory Steps

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

The first is to trim and mask the images. We convert the calibrated image from MAST into a `fits`-compatible image by retrieving the image data, and omitting the uncertainty map. We then trim the image to remove virtual rows and columns using the `SCIROWS`, `SCIROWE`, `SCCSA`, `SCCED`, header values. We mask out saturated stars using the ‘`fiqn`’ program with a fixed saturation level of 2^{16} ADU. As described by Pál (2009), the mask is metadata to the image, and is not ever actually applied to the pixel values, except

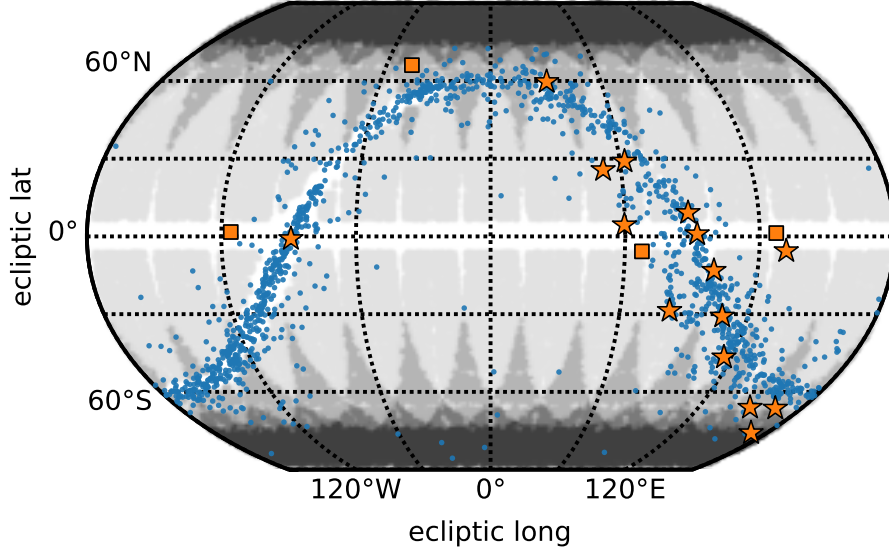


Figure 1. There are many clusters; TESS looks at them. Placeholder sky map.

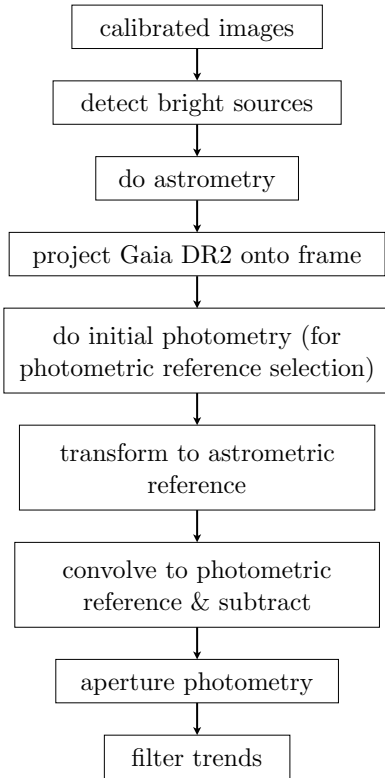


Figure 2. Pipeline. This is what we do.

during specific image processing steps that follow. During masking, we also use the algorithm described by Pál (2009) to extend masks beyond purely saturated pixels to “bloomed” pixels horizontally and vertically adjacent to the saturated pixels. Finally, for frames with the `DQUALITY` bit-flag set to 32 — corresponding to the “momentum dumps” described

by Vanderspek et al. (2018) — we mask out 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.

Next, we perform some initial analysis steps to produce metadata needed during image subtraction. To obtain an astrometric solution (independent from the WCS data packaged 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 (§ 2.4). 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.

With the resulting WCS information, we then project a source catalog from Gaia-DR2 onto the frame, down to a pre-selected magnitude cutoff (Gaia Collaboration et al. 2018). We use these expected positions to center the apertures in our photometry, rather than attempting to detect the positions. Such “forced-aperture photometry” is preferable to performing source extraction because of the large TESS pixels, and the accurate Gaia positions. 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 and catalog-indexing is performed using `gaia2read`¹ (Kim 2018).

Finally, we 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 (e.g., 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.

2.4. 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. 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 $D_{xy} = I_{xy} - M_{xy}$.

We then select two “reference frames” for image subtraction. The first is the astrometric reference; the second is the photometric reference. To choose the astrometric reference, we use the following heuristics:

1. The frame must have large and round stars (largest “S”, smallest “D” and “K” values).
2. The frame must minimal background noise, as measured in annuli around the bright stars selected in § 2.3.
3. The frame must have, relative to the other frames being considered, a large number of detected sources.

We sort the frames using the above metrics, and then select the astrometric reference from successive intersections of each sorted list. Using the algorithm presented by Pál & Bakos (2006), we then calculate a spatial transformation that maps the X and Y coordinates from each calibrated frame onto the the astrometric reference.

“Registering” the images onto a common astrometric reference frame requires some care. In particular, the transformation must preserve the fluxes of the sources. Since the transformation is affine (a combination of e.g., translation, rotation, dilation, and shear) standard bilinear or bicubic interpolation would not achieve flux conservations. We therefore use the flux-conserving interpolation scheme described by Pál (2009) which is based on analytic integration of the surfaces determined by the pixel values. The largest component of the transformation is a translation, of order 2 arcseconds, or about 0.1 TESS pixels.

The second reference frame needed for image subtraction is the photometric reference. Generically, if one were to simply subtract two arbitrary frames, the noise level in the resulting difference image would be $\sqrt{2}$ times greater than that in the original target frame. To minimize this effect, we create a high-SNR median average of N selected frames, so that the noise level in the resulting subtracted images goes as $\sqrt{1 + 1/N}$. This “photometric reference” is used both to calculate the convolution kernel during image subtraction, and to measure the reference flux of each star. To select viable frames for inclusion in the photometric reference, we use the results of our initial photometry step (§ 2.3), and sort frames for each sector by:

1. Lowest median scatter in photometry
2. Lowest median error in photometry
3. Lowest median background measurement
4. Lowest median absolute deviation in background measurements
5. Largest number of stars detected by `fiphot` with good flags.

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.

Some details of, and motivation for, the convolution process are warranted. The general idea of the image subtraction problem is to...

This has some benefits for crowded regions. Most notably, it enables a measurement of the background flux level.

Our approach follows from the algorithm presented by Alard & Lupton (1998), and explained in detail by Miller et al. (2008). The process is:

1. Select a set of isolated stars (“stamps”).
2. Choose basis vectors K_n for the kernel, $K = \sum_{n=1}^N a_n K_n$. Our goal is to determine the best-fitting coefficients a_n .

¹ github.com/samuelyeewl/gaia2read

3. Using these stamps and the least-squares method, ...

Finally, we convolve and subtract all the frames that have been translated from the photometric reference (`grcombine`). We use the same kernel formalism as described by Soares-Furtado et al. (2017). To compute the convolution kernel, we minimize the difference between each image and the photometric reference (or something like this), using the Alard & Lupton 1998 method (CITE, also CITE Miller’s reference). The software implementation is that by Pál (2009), in the `ficonv` program. The implemented math is as follows.

$$XXX = XXX. \quad (5)$$

We explored a variety of kernel options, including X, Y, and Z. We wound up including an “identity” term (MEANING?), a “background term” (MEANING?) and a delta function term (MEANING? Give math for all of these). There are also spatially-varying terms of order XX. (MEANING?)

We perform forced-aperture photometry on the subtracted frames, using the locations from the Gaia projection, to measure the resulting difference fluxes. We use circular apertures, with the appropriate weights along the boundaries (`fiphot`; Pál (2009) Section FIXME).

Note that to measure the fluxes on the differenced images, a slightly different procedure from Eq. 2 is relevant. (Pal 2009 Section 2.9). Specifically, the subtracted image S is given by $I - B - R * K$, for I the original image, B the background, R the reference image, and K the convolution kernel. The flux in the original image is given by

$$f = \sum_{x,y} w_{x,y} I_{x,y} \quad (6)$$

$$= \frac{1}{\|K\|_1^2} \sum_{x,y} S_{x,y} (w * K)_{x,y} + \sum_{x,y} R_{x,y} w_{x,y}, \quad (7)$$

Finally, to 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.

2.5. Lightcurve Detrending

The preceding steps produce lightcurves that include both instrumental systematics as well as astrophysical variability. The detrending philosophy adopted by the HAT group typically proceeds in two sequential stages (see discussions from e.g. Bakos et al. 2010; Huang et al. 2015; Zhang et al. 2016). The first stage is to decorrelate against external parameters that are known to affect the stellar flux measurements (EPD, Bakos et al. 2010). The second stage is to filter remaining systematic trends caused by unknown parameters by decorrelating the measured brightnesses of stars against each other (TFA, Kovács et al. 2005).

In general though, one must adapt detrending procedures to the data at hand. In § 2.5.1, 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. We go on to show (§ 2.5.2) that a plausible detrending approach for the purpose of transit discovery is to simply decorrelate against other nearby stars with standard TFA.

2.5.1. Flux versus external parameters

The traditional approach to EPD (Pál 2009; Bakos et al. 2010; Zhang et al. 2016) is to fit and subtract a model for the magnitudes m of the form

$$m = \text{const.} + \sum_i c_i \theta_i, \quad (8)$$

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 environment², the centroid positions (x, y) , the fractional part of the centroid positions ($\{x\}, \{y\}$), and any other parameters that are expected⁴ to correlate strongly with the observed flux. The coefficients c_i are calculated through linear least-squares, and subtracted to produce “EPD” lightcurves.

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 reasonable 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 lightcurves 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 lightcurves 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 4). 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 lightcurves 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.

² We used the temperature from the on-chip aluminum-copper sensor measurements included in the engineering data³.

⁴ 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.

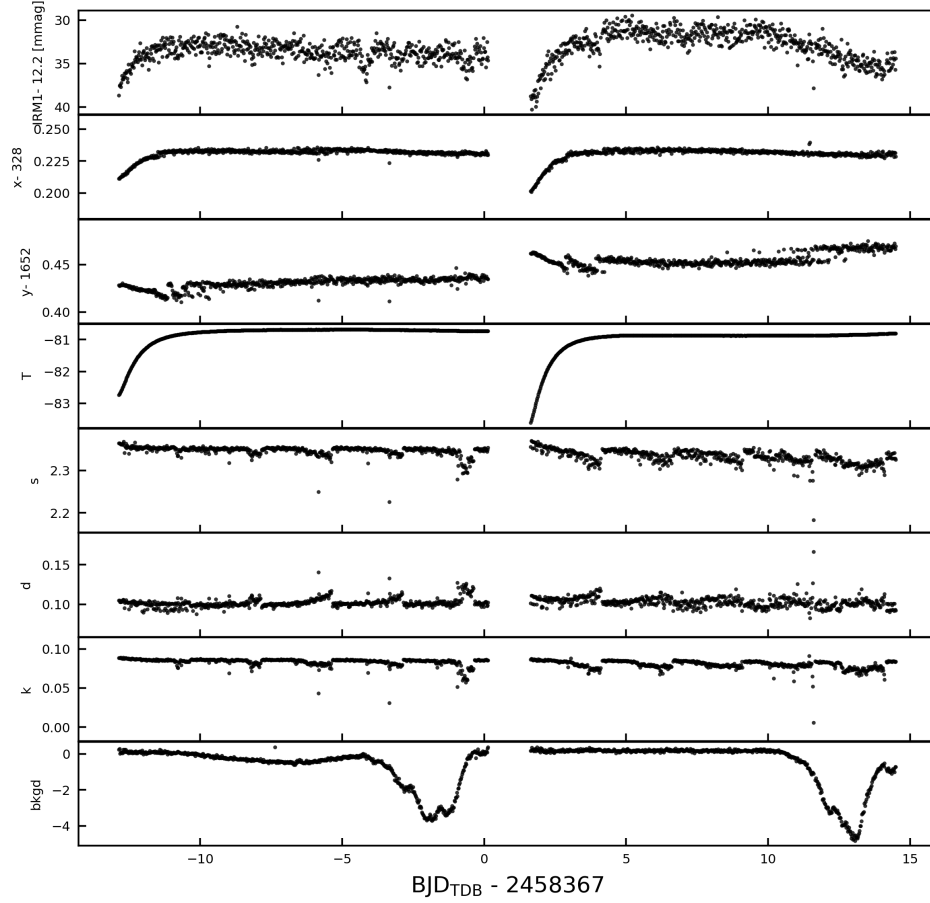


Figure 3. Timeseries of “external parameters” for a representative star. *Top:* Instrumental raw magnitude (with a particular aperture size), as a function of time. *Second and third from top:* x and y centroid positions as a function of time. Continuing in order are the CCD temperature, the (s, d, k) shape parameters, and the measured background value. Most of the apparent variability is instrumental: see § 2.5.1.

4. *Flux vs. shape parameters.* For lightcurves 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.

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 3).

Considering the timeseries of parameters other than flux (Figure 3):

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.

Given the characteristics of the variability, a linear model of the form given in Equation 8 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 lightcurves, and also distorted transits.

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 lightcurves.

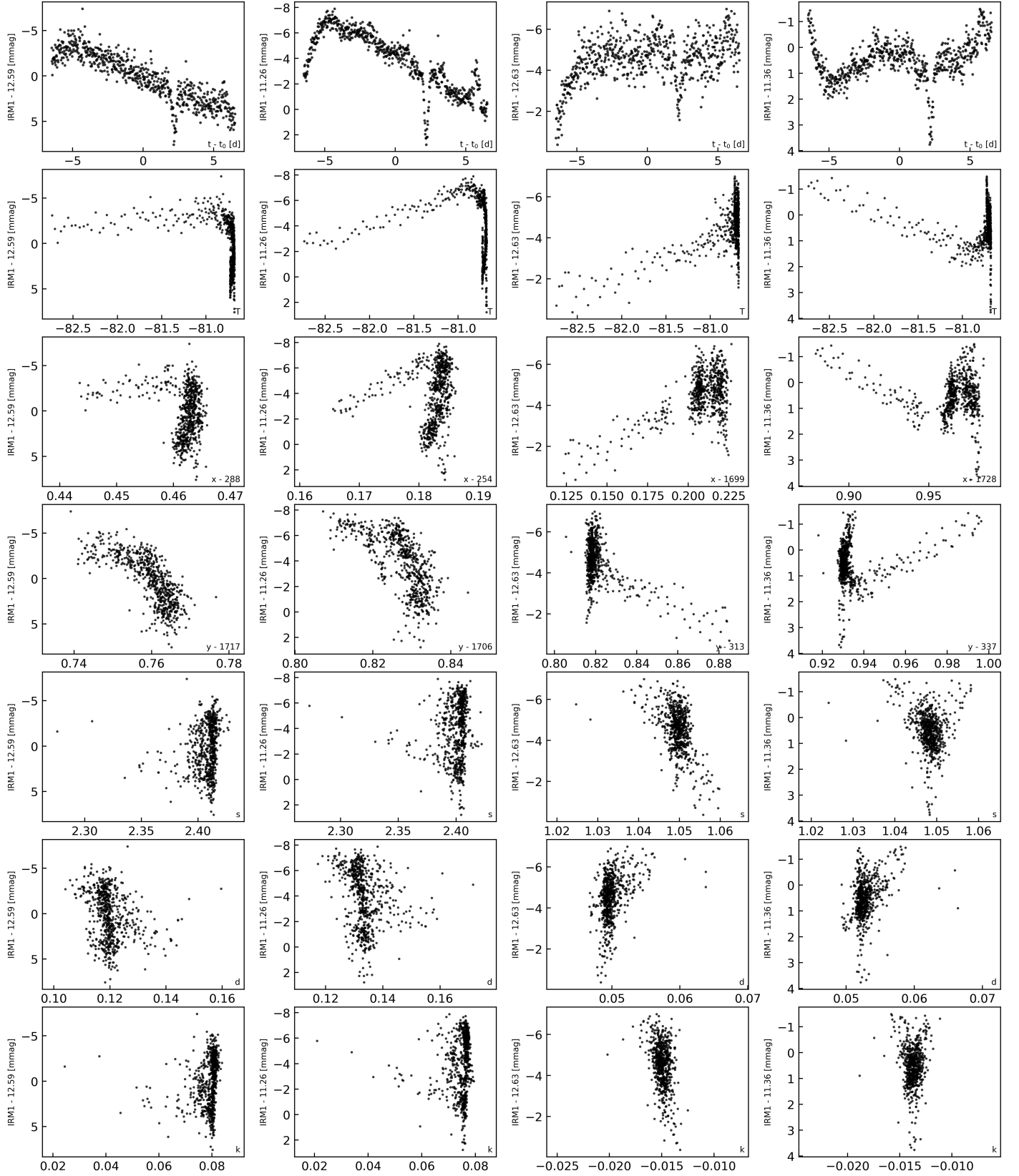


Figure 4. Flux as a function of “external parameters” for four representative stars. The left two columns are stars at the corner of a camera’s field; the right two columns are from the centers. Each row shows a different parameter along the x -axis, given in text at the bottom right of each subplot. “Hooks” are common features in flux as a function of temperature and centroid position. § 2.5.1 gives a verbose description.

2.5.2. 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 (TFA, Kovács et al. 2005).

This requires selecting “template stars”, which are a subsample of star that are supposed to represent all the types of systematics across the dataset. We select 500 template stars randomly from the stars on-chip with G_{Rp} magnitudes between 8.5 and 13 (TODO: rather bright?). We additionally require these stars to not have excessive variability, by fitting a parabola in the RMS-magnitude plane, and discarding stars more than 4σ away from the prediction of this fit.

Once the template stars are selected, we use the secret TFA implementation from the `HATpipe` source code, that you and nobody else is allowed to see.

To select template stars, we impose a

The details of this selection procedure are

3. RESULTS

RMS vs mag plots.

SNR of retrieved HJs.

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

Some focus on actual cluster fields.

4. DISCUSSION

Lorem ipsum.

5. 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, 2018)

Software: `astrobases` (Bhatti et al. 2018), `astropy` (Collaboration et al. 2018), `astroquery` (Ginsburg et al. 2018), `BATMAN` (Kreidberg 2015), `corner` (Foreman-Mackey 2016), `emcee` (Foreman-Mackey et al. 2013), `fitsh` (Pál 2012), `IPython` (Pérez & Granger 2007), `matplotlib` (Hunter 2007), `numpy` (Walt et al. 2011), `pandas` (McKinney 2010), `scipy` (Jones et al. 2001).

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
- Bhatti, W., Bouma, L. G., & Wallace, J. 2018, *astrobases*
- Clarke, B. D., Caldwell, D. A., Quintana, E. V., et al. 2017, *Kepler Science Document*, 5
- Collaboration, T. A., Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *arXiv:1801.02634 [astro-ph]*, *arXiv: 1801.02634*
- Dierckx, P. 1996, *Curve and surface fitting with splines*, repr edn., Monographs on Numerical Analysis (Oxford: Clarendon Press), oCLC: 245719230
- 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
- Gaia* Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, *Astronomy and Astrophysics*, 595, A1
- Gaia* Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, *Astronomy and Astrophysics*, 616, A1
- Ginsburg, A., Sipőcz, B., Madhura Parikh, et al. 2018, *Astropy/Astroquery: V0.3.7 Release*
- 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
- Jones, E., Oliphant, T., Peterson, P., et al. 2001, *Open source scientific tools for Python*
- Kim, J. 2018, *Querying Gaia for Wide Binary Companions to Exoplanet Hosts*, Princeton Junior Thesis (Unpublished)
- Kovács, G., Bakos, G., & Noyes, R. W. 2005, *Monthly Notices of the Royal Astronomical Society*, 356, 557
- 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
- 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
- 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
- Soares-Furtado, M., Hartman, J. D., Bakos, G. Á., et al. 2017, *Publications of the Astronomical Society of the Pacific*, **129**, 044501
- 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
- Zhang, M., Bakos, G. Á., Penev, K., et al. 2016, *Publications of the Astronomical Society of the Pacific*, **128**, 035001