Draft version March 7, 2019
Typeset using LaTeX **twocolumn** style in AASTeX62

`eleanor`: A tool for extracting light curves from the *TESS* Full-Frame Images

Adina D. Feinstein,[1] Benjamin T. Montet,[1,*] Daniel Foreman-Mackey,[2] Megan E. Bedell,[2] Nicholas Saunders,[3,4] Jessie Christiansen,[5] Christina Hedges,[3,4] Rodrigo Luger,[2] Daniel Scolnic,[6] and José Vinícius de Miranda Cardoso[7,3]

[1] *Department of Astronomy and Astrophysics, University of Chicago, 5640 S. Ellis Ave, Chicago, IL 60637, USA*
[2] *Flatiron Institute, Simons Foundation, 162 Fifth Ave, New York, NY 10010, USA*
[3] *NASA Ames Research Center, Moffett Field, CA 94035, USA*
[4] *Bay Area Environmental Research Institute, 625 2nd St Ste. 209, Petaluma, CA 94952*
[5] *CalTech/IPAC-NASA Exoplanet Science Institute, M/S 100-22, 770 S. Wilson Ave, Pasadena, CA 91106 USA*
[6] *Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA*
[7] *Universidade Federal de Campina Grande, R. Aprgio Veloso, 882 Universitrio, Campina Grande, 58429, Brazil*

known

ABSTRACT

The *Transiting Exoplanet Survey Satellite* (*TESS*) observes a new 96° x 24° sector of the sky at 30 minute cadence as Full-Frame Images (FFIs), enabling the detection of thousands of previously undiscovered planets across 85% of the sky. We have created an open-source tool, `eleanor`, to produce light curves for objects observed by *TESS*. Here, we describe the methods used to produce light curves optimized for planet searches. We use both aperture and PSF photometry to recover known transiting exoplanets in the FFIs to validate the pipeline and perform a limited search for new planet candidates in Sector 1. Cadence stacked images, raw, and detrended `eleanor` light curves for each analyzed star will be hosted on MAST, with planet candidates on ExoFOP-TESS for broad community use.

Keywords: binaries: eclipsing, methods: data analysis, planets and satellites: detection, techniques: image processing, techniques: photometric

?

1. INTRODUCTION

The recently retired *Kepler* and *K2* missions (Borucki et al. 2010; Howell et al. 2014) revealed tremendous new insight into planet occurrence rates, planetary architecture, and planet formation (e.g. Huber et al. 2013a; Swift et al. 2013; Fabrycky et al. 2014). The *K2* mission ran out of fuel in October, 2018; it is therefore timely that *TESS*, the *Transiting Exoplanet Survey Satellite* (Ricker et al. 2014) is ready to start providing new insights into more nearby exoplanets than *Kepler* and *K2* detected. *TESS* is a two year survey observing 80% of the sky for exoplanet transits. TESS four cameras are aligned along a 96° × 24° degree sector of the sky and take observations of a sector for roughly 27 days. Within each sector, 20,000 stars, ~~which have been~~ pre-selected by TESS mission operators and through Guest Investigator (GI) proposals, are observed in a short, 2-minute cadence mode.

These targets have been selected not only for new exoplanet candidate searches, but also for asteroseismic studies, Solar System research, and additional galactic and extragalactic astrophysics.[1]

In addition to the short cadence targets, *TESS* obtains an image of the entirety of each sector every 30 minutes in what are known as Full-Frame Images (FFIs). The FFIs cover the entire 96° x 24° field of view. Within each sector of observations, there are roughly one million stars in the FFIs brighter than I=16. As such, the FFIs provide a huge data mining archive for many different sub-branches of astronomy, including the search for new transiting exoplanet candidates. Barclay et al. (2018) ran simulations which predict that within the FFIs, an additional 3,100 new exoplanets will be found orbiting bright ($T_{mag} < 11.0$) stars and 10,000 exoplanets will be found orbiting fainter stars. Within the FFIs, $\sim$ 1500 large planets ($> 4\ R_\oplus$) and $\sim$ 400 small planets ($\leq 4\ R_\oplus$) will be identified, where 67% of host stars are predicted to

Corresponding author: Adina D. Feinstein; @afeinstein20
afeinstein@uchicago.edu

* Sagan Fellow

[1] See, for example, *TESS* Guest Investigator Programs G011098, G011280, G011113, and G011163.

The four TESS cameras
or TESS's four cameras

I would just link to the
page of selected proposals,
rather than call some out

Maybe clarify here that the Kepler/K2 LC data had 30-minute cadence explicitly

be F and G type stars (Barclay et al. 2018). In addition to the new stars observed every sector, due to spacecraft orientation, TESS has a continuous viewing zone (CVZ) at each ecliptic pole, where stars are observed for up to 351 days. The CVZ is focused on both the Northern and Southern poles and will be the ideal location for finding longer period planets than detectable at lower ecliptic latitudes.

Targets in the CVZ will prove useful to community members who are interested in measuring rotation periods of stars and asteroseismology. Long baseline photometry will allow for better tracking of spots along the surface of a star. This will prove useful for young stars with many spots and could be promising for determining ages of older stars with slower rotation periods (Barnes 2003).

Asteroseismology, the study of stellar oscillations to probe the internal structure of stars, will additionally benefit from the TESS FFIs. The ability to measure stellar oscillations with 30-minute cadence data has been previously demonstrated with the Kepler long-cadence data (Campante et al. 2016). Transiting planets identified with TESS will benefit from accurate stellar parameters, and thus calculated planet parameters, ~~with the addition of being able to constrain fundamental stellar properties with asteroseismology~~. This technique for determining stellar parameters will prove useful when spectral typing new stellar hosts, particularly looking at main sequence versus subgiant and giant stars (Huber et al. 2013b). ~~Obtaining high-accuracy planet parameters will prove essential when determining ideal candidates for follow-up via the TESS Working Groups.~~

redundant w/ "accurate stellar parameters"

at the FOP stage don't need super accurate parameters yet

Within the FFIs, there is also the possibility to study Solar System objects, and galactic and extragalactic sources. Using a difference imaging approach and K2 long-cadence data, Dimitriadis et al. (2018) were able to obtain a light curve for a supernova, SN2018oh, at an approximate distance of 52.7 Mpc away. The light curve begins 18 days before peak brightness, which is a feat advanced surveys that are triggered by supernova events could not have achieved. In addition to supernovae, teams, such as Molnar et al. (2015), were able to detect extragalactic RR Lyrae stars in Leo IV, a dwarf galaxy at a distance of ∼ 154 kpc. Using K2 observations, they were able to observe the farthest measurement of the Blazhko effect, or long-period modulations in the period and amplitude of the light curve. The discovery of this effect was the first discovery outside of the Milky Way and the Magellanic Clouds.

if you don't add the note suggested above, clarify here that K2 LC = 30-min.

The TESS mission delivers calibrated FFIs. However, the formatting of the FFIs is not conducive for the entire community to be able to access. The FFIs are approximately 35 MB each. If each sector is observed for roughly 27 days every half hour, this ~~means all data for a single target in a single sector is stored in~~ ∼ 1300 separate files, totaling ∼ 45 GB in size. Moreover, systematic effects such as background scattered light from the Earth and Moon exist in the photometry and can overwhelm extrasolar signals, especially when the telescope is near perigee.

what does this mean? which calibration steps have happened?

this implies that, even for analyses concerned with only a single target, the data would be spread across

I don't really understand this "Moreover". Isn't it better to have the large files with all the pixels so you can deal with the systematic effects?

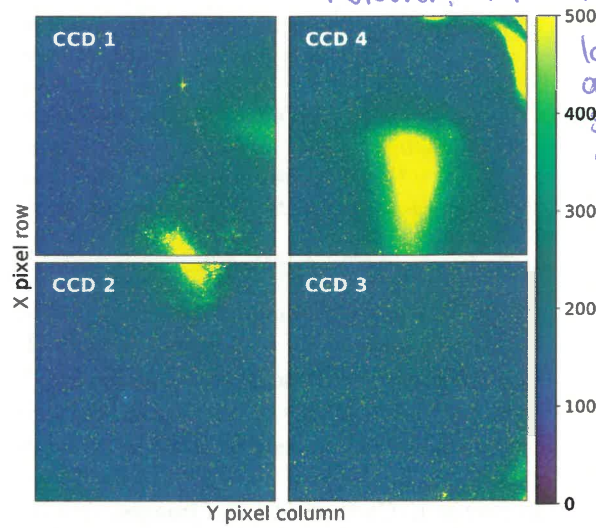

Figure 1. An example FFI from Sector 1, Camera 4. There is noticeable structured background in the corner of the CCD as well as in the center. There are 44 columns of non-science on either side of the FFI in the y-direction and 30 virtual, smear, and buffer rows at the top in the x-direction.

The recently ended K2 mission lead to the ~~create~~ creation of several community driven pipelines for data reduction. There is a significant benefit to having multiple pipelines using different methods for data reduction for the same data sets. For example, one could find a new planet candidate in the `EVEREST` light curves (Luger et al. 2018) and compare to the `K2SFF` light curves (Vanderburg & Johnson 2014) to determine if the signal is real. Additionally, Shaya et al. (2015) created the *Kepler* Extra-Galactic Survey (KEGS), with the goal of producing light curves for extra-galactic sources. These pipelines were especially useful due to the largest systematics found within the *K2* data. Other pipelines, including `K2VARCAT` (Armstrong et al. 2015), K2SC (Aigrain et al. 2015), and POLAR (Barros et al. 2015) created light curves for the public to use as well, each with their own methods for removing systematics and therefore their own strengths and weaknesses.

choose a consistent "package" font

In this article, we present the **eleanor** pipeline and data products for light curve extraction from the *TESS* FFIs. In Section 2 we describe the methods used to can you quantify the typical duty cycle / observation efficiency of your FFI postcards as a result of this conservatism?? on average, how many end up w/ flags set??

create our light curves. In Section 3, we demonstrate the photometric capabilities of `eleanor`by presenting early science results, including recovering known transiting planets, new planet candidates, and other stellar variability, including supernovae in the *TESS* field of view. In Section 4, we discuss the light curve data products, availability, and how to download the open-source software package.

2. CREATING LIGHT CURVES

2.1. *Handling Raw Data*

The `eleanor` pipeline first manipulates the large format of the FFIs to be able to efficiently create light curves. That is, before extracting light curves, we create an intermediate data product called "postcards. Postcards are 148 × 104 pixel cut-out regions of the FFIs, and are created with a 50 pixel overlap between each postcard to avoid edge effects for individual stars. Unlike FFIs, the postcards are time-stacked, including all cadences for which observations are available, and are background-subtracted. As the background of the FFIs is highly structured ~~and varies greatly~~ across the detector (Figure 1), the localized scaling of the postcards provides a sufficient region for initial background subtraction.

does this mean spatially?

you haven't described what these are, or are going to be yet

Within each postcard, the World Coordinate System (WCS) from the FFIs are conserved. The postcards also contain quality flags to highlight potentially suspect cadences. We follow a two-step process for assigning the quality flags. ~~The short-cadence targets have quality flags, some of which indicate issues at the FFI level, such as the spacecraft~~ being ~~is in coarse pointing mode rather than fine pointing, there is a reaction wheel desaturation event, or the spacecraft pointing towards the Earth rather than its field (see Tenenbaum & Jenkins (2018) for more information).~~

redundant w/ below

The *TESS* mission assigns twelve different quality flags to the short cadence data, of which eight are applicable to the FFIs. The quality issues are: attitude tweaks; when the spacecraft is in coarse or Earth point; an argabrightening event occurs; reaction wheel desaturation event; a cosmic ray is detected on a collateral pixel row or column; and there is straylight from the Earth or Moon in the camera field-of-view (see Table 28 in Tenenbaum & Jenkins (2018) for more information). Additionally, there is a "manual exclude" quality flag set in the processing of short cadence data.By identifying short-cadence targets that fall on each camera-CCD pairing for a given sector, we can copy these quality flags into our postcards. There are roughly 15 short-cadence observations for every 1 FFI observation, and we follow the most conservative procedure: any applicable quality flag that falls during an individual FFI exposure is copied into the postcards.

In addition, we are able to introduce our own quality ~~flag based on the pointing model.~~ We fit a line to each orbit in a sector to the centroid positions of each target on a given CCD after the pointing model was applied. If the centroid position is $\geq 2\sigma$ from the line, we introduce a quality flag (value = 4096) in our produced light curves. The creation of the pointing model is described below.

expand! Maybe show a figure?

there is not enough detail / this is confusing. clarify. 2 orbits/sectors: why a line? fit line to what?

2.2. *Pointing Model*

We build a pointing model to ensure, regardless of any possible issues with the WCS, that our stars are well-centered in our $\boxed{\text{TPF cutouts}}$ at each cadence. For each CCD, we complete a search of targets in the TIC with $7.5 \leq T_{mag} \leq 12.5$ and $0 \leq \text{contam} \leq 5 \times 10^{-3}$, leaving us with bright, but not saturated, uncrowded stars to calibrate our pointing model. Due to spacecraft ~~observatory~~ motion, the true position of the star and the position of the star on the detector may not be the same, even after applying the FFI WCS.

We then determine the affine transformation that minimizes the square of the differences between the predicted and observed positions for all stars on the detector at each cadence. By applying a pointing model, we improve the precision of placing each target in or very near the center of each target pixel file (TPF) image.

do you want to quantify these statements at all?

2.3. *From TPFs to Light curves*

how big a cutout do you usually make?

2.3.1. *Aperture selection*

Once the TPF has been extracted, `eleanor` uses a predefined library of apertures of various shapes and sizes (see Figure 2) for aperture extraction. It initially tests apertures that were shown to work well for *Kepler* photometric extraction, including 2 × 1 rectangles and an L shape, both in four different orientations about the center. We also test standard circular and square apertures defined using the `photutils` package. For completion, we use both binary (pixel values of either 0 or 1) and weighted (pixel values ranging from 0 to 1) masks when testing. Circular apertures have radii of 1.25, 2.5, 3.5, and 4 pixels. Squares have lengths and widths of 3, 4.1, and 5 pixels as well as a 3 pixel length and width that is rotated 45°. We chose these orientations to maximize diversity in the aperture testing. All apertures and associated extracted light curves are saved in the `eleanor` data product.

weighted by what?

A user of the software has the option to pass in their own aperture masks as well, in the `TargetData.get_lightcurve()` function. Masks are required to be 2D and the same length and height as the final pro-

Show some real TESS FFI targets to show these are reasonable starting points, e.g. what happens at the very edges of the field to the PSF?

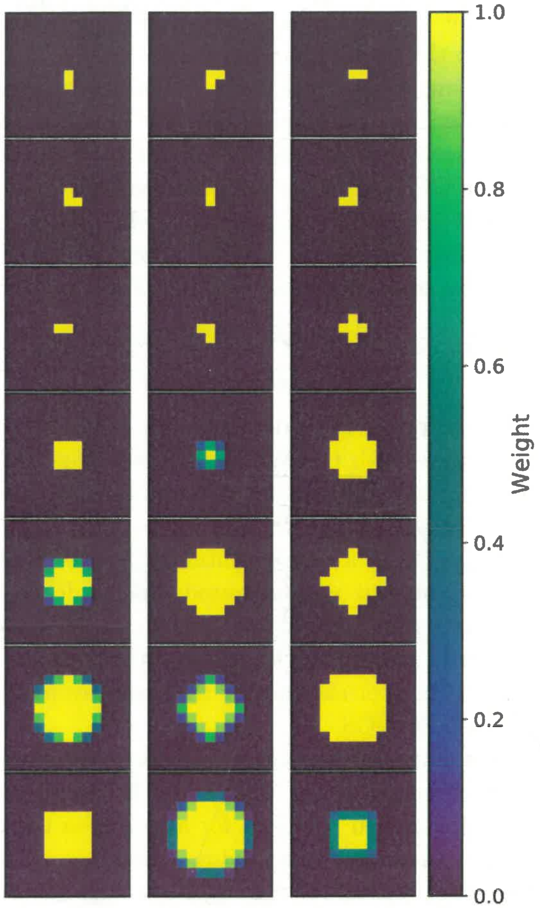

Figure 2. The default library of apertures **eleanor** uses to test and create light curves. We try a range of shapes and sizes as well as assigning varying weights to each pixel. The binary apertures have pixel values of 0 and 1; the weighted apertures have pixel values from 0 to 1. All apertures are tested unless `crowded_field = True`, in which case only apertures with less than 9 pixels are used to extract light curves.

duced target pixel file. There is an additional options for users to customize a `photutils` aperture using `TargetData.custom_aperture()`. This is a basic function that allows the user to customize the radius or length and width and angle of a circular or rectangular aperture. Users can also pass in a position in TPF pixel space to offset the aperture from the center, where it is placed by default. The aperture can additionally be defined as binary or weighted. What kind of weights?

After the raw ~~photometry~~ is extracted, we correct for possible systematic effects on an orbit-by-orbit basis, creating a flux time series called "CORR_FLUX". We regress the raw flux time series against the x pixel position, y pixel position (both taken from our pointing model), measured background at the location of the

How do you measure the background? maybe describe in the postcard section

TPC, and time, effectively removing any signals correlated with ~~any of~~ these parameters. We note that very long signals, such as starspot-induced stellar variability for slowly rotating stars, or transients like supernovae could produce an approximately linear signal over a single orbit. This signal would then be removed by "CORR_FLUX" in a similar way as it was removed by the *Kepler* pipeline in the generation of their pre-search data conditioning (PDCSAP) flux.

Although the extracted TPF is already background subtracted at the postcard level, we test if additional background subtraction on the TPF level will increase the precision of the extracted light curve. The amount of background subtraction is marked in the header by the flag "BKG_LVL", which reads as either "PC_ONLY" indicating the background subtracted postcard was enough or "PC_TPF" indicating an additional background subtraction was completed on the TPF level. The background is defined using the `photutils` function, `MMMBackground`, which calculates a background using a mode estimator of a the form $3\bar{X} - 2\mu$. We chose not to complete the background subtraction on the FFI level due to significant changes across the detector.

I don't know what $\bar{X}$ and μ here are

Within each individual FFI, there is significant "background" emission, largely driven by reflected light from the Earth and Moon, although other bright Solar System objects contribute as well. As seen in Figure 1, it is clear that the background varies greatly across the detector. Significant background will affect both the ability to detect planets, to characterize signals with variability similar to the background, and to accurately measure planet parameters derived from the light curves. To date, there are minimal quality flags within the FFIs to help users identify which observations are ~~too~~ affected by background (e.g. Mars came into the field of view in Sector 1 ~~one~~ and saturated a majority of the detector) or which observations may have been affected by spacecraft corrections or errors, leaving the user to mask regions that may provide less-than-optimal photometry.

After a light curve has been extracted using every aperture shape and size, and the light curve has received background treatment on both the postcard ~~and postcard~~ and TPF level, an ideal aperture is chosen for that star. Here, we are primarily interested in creating the best light curve possible for planet search, preserving the sharp features on short timescales induced by planet transits. The corrected light curve light curve is flattened using `lightkurve.flatten()`, which applies a 2nd order Savitzky-Golay filter to the light curve. The flattened light curve with the smallest measured CDPP value on one-hour timescales is then chosen. The asso-

You haven't said what this is yet, or how you calculate it.

ciated aperture and level of background subtraction are flagged in the header of the TPF.

2.3.2. *Principal Component Analysis*

you run PCA *after* flattening with Savitzky-Golay ?!?

We use the python package **sklearn** to perform principal component analysis (PCA). PCA is a machine learning technique which calculates orthogonal eigenvectors between a set of input vectors, such as light curves. The purpose of conducting a PCA analysis is to remove any additional systematics that are still potentially present and shared between nearby stars on the detector. Removing commonalities between the **eleanor** light curves could prove useful when separating astrophysical from instrumental effects. PCA is the basis of the development of "PDCSAP" light curves in the *Kepler* and *K2* missions (Smith et al. 2012; Stumpe et al. 2012).

We develop our eigenvectors by collecting 12,000 stars across each camera. We use the "CORR_FLUX" flux from each star as inputs, identifying the 16 principal components. The components are visually inspected as a quality check. The **eleanor** light curve products include a "PCA_flux", which is created by subtracting the first 8 principal components for the appropriate camera to minimize the possibility of astrophysical variability being imprinted in an eigenvector. All 16 components are available for more artisinal analyses of light curves; the option to use more or ~~less~~ fewer components is built into `eleanor.TargetData().pca()`.

which part of the preceding sentence is relevant here?

2.3.3. *Point-Spread Function Modeling*

Although we do not provide PSF modeled flux as part of our deliverable, it is included as an option using the **eleanor** software package. *BEN* TO DO: It'd be super useful if you wrote this section :D

Move this info earlier where you talk about construction

2.4. *TPF data product*

The final light curve is stored as a FITS file. Each FITS file contains the cadence-stacked, background-subtracted flux pixels for a 13 × 13 region, centered on the source, and the cadence-stacked flux error pixels for the same region. The files contain all 21 aperture masks tested in the light curve extraction process as well as the raw and corrected fluxes for these apertures. For the automatically selected best aperture, three light curves are available: raw flux, corrected flux, and PCA flux. Users have the option to create a point-spread function (PSF) modeled light curve with the **eleanor** software package. However, the PSF flux is not included in the default files. An example of each type of light curve can be seen in Figure 3, which additionally shows the recovery the known WASP-100b planet (Hellier et al. 2014; Stassun et al. 2017).

lower case previously

In addition to photometric information, the produced files contain the X and Y centroid positions, as inferred by our pointing model, and quality flags, based on the two-minute cadence ~~targets~~ ~~and~~ our own quality flag for cadences where the X and Y centroid has jumped 2σ away from the mean, as discussed previously. We create light curve files for sources in the TIC with $I \leq 16$. Members of the community can use the **eleanor** package to create light curves for fainter or extragalactic objects, or for a more detailed or optimized analysis of individual objects.

3. RESULTS

We calculated the combined differential photometric precision (CDPP) for 32,000 light curves in each of *TESS*'s four cameras. The CDPP is a metric defined for *Kepler* that assesses by the ability to detect a weak terrestrial planet transit signal in a light curve (Christiansen et al. 2012). To calculate the CDPP, we remove any long term trend again using `lightkurve.flatten()`. The data is binned on one hour timescales and the CDPP is evaluated. The CDPP as a function of *TESS* magnitude is shown in Figure 4.

Each camera is color-coded by the four CCDs on each. The solid lines represent the lower-limit of the CDPP for each CCD. The CDPP remains fairly consistent for all CCDs in a given camera, with the exception of Camera 4. CCD 4 experiences noticeably more systematics (see Figure 1) than the other CCDs, leading to an overall increase in CDPP values. Note that the presence of the Large Magellanic Cloud in CCDs 1 and 2 does not lead to a significant difference in CDPP values when compared to other cameras: while diffuse light from the LMC is obvious, it is stable.

We additionally compare the **eleanor** light curve CDPP to that of Oelkers & Stassun (2019) and the MIT Quick-Look Pipeline (QLP). QLP light curves are not available to the public except for confirmed planets; for comparison we use the published light curve of TOI-172 b (Rodriguez et al. 2019). We apply the **eleanor** quality flags across all light curves for a uniform comparison. We apply additional masks for times when the observatory went out of its fine-pointing mode ($1338 <$ Time < 1339 and $1346.8 <$ Time < 1348.6) and for the 3 transits of TOI-172 b, marked in Fig. 5 by the vertical orange lines (Rodriguez et al. 2019). We leave the transits in Figure 5 to demonstrate **eleanor**'s transit detection capabilities. Each light curve has been flattened using identical processes.

For TOI-172, a $T_{mag} = 10.711$ host star, the Oelkers & Stassun (2019) light curve reaches a CDPP=577.3; the QLP reaches a CDPP=376.3; **eleanor** reaches a

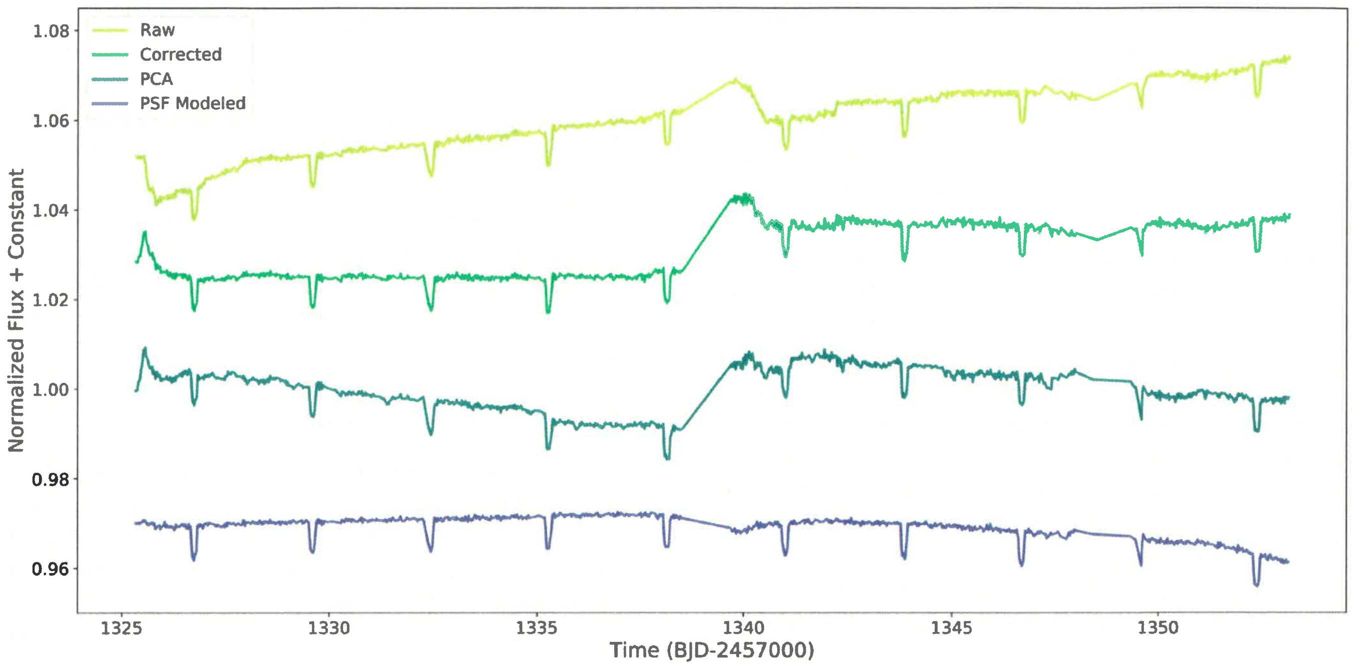

Figure 3. The four available light curves available using `eleanor`. Raw flux is the sum of pixels in the aperture. Corrected flux is raw flux with a linear regression as a function of pixel location, background, and time. PCA flux is a Principal Component Analysis subtracted flux, to remove common systematics between targets on the same camera. PSF modeled flux is the 2D Gaussian point-spread function modeled flux. The 10 transits for WASP-100 are clearly seen in all four light curves.

CDPP=320.0. The smaller scatter in the `eleanor` light curve will maximize community ability to detect small planet transit signals in the FFIs.

3.1. *Recovery of Known WASP Planets*

In order to demonstrate the capabilities of the `eleanor` light curves, we have recovered known transiting WASP planets and show three of those results in Figure 6. Using these results, we use the `batman` (Kreidberg 2015) transit fitting software package, following the methods from Mandel & Agol (2002), to derive planet parameters from the `eleanor` light curves and compare our results with the known parameters. The planet parameters are consistent with values quoted in Stassun et al. (2017).

We complete a more extensive Markov Chain Monte Carlo (MCMC) analysis for WASP-126b (Maxted et al. 2016) using `emcee` (Foreman-Mackey et al. 2013), an implementation of the affine-invariant ensemble sampler of Goodman & Weare (2010). We initialized our MCMC run with the best-fit parameters from a single `BATMAN` fit and ran it for 200 steps, to complete a burn-in. The parameters from the 200$^{\text{th}}$ run were then used as the starting point for the next run of 1500 steps. The parameters and uncertainties are quoted in Table 1. All system parameters derived with the `eleanor` light curve fall within 1σ of the accepted parameters from Maxted et al. (2016), with the exception of a/R_*, which agrees within 1.5σ. Overall, the derived parameters for WASP-126 correspond well to the quoted values.

3.2. *New Science with* ***eleanor*** *Light Curves*

We demonstrate the potential of new science to come out of the *TESS* FFIs using `eleanor` by performing a limited search for periodic signals in Sector 1 data. Using the 12,000 stars generated for the PCA components, we performed a limited search using the box-least squares (BLS; `bls.py`) module in `astropy` to phase fold light curves and identify new periodic candidates, ranging from new planet candidates to RR Lyraes. We searched a limited period range of 0.5 to 8 days and saved the TICs of candidates where the maximum peak in the periodogram (period vs. log likelihood) was above 9σ above the mean.

The TICs were then vetted by-eye. The resulting new science candidates can be found in Table 2. We present the properties from the BLS fits and uncertainties for new planet candidates and eclipsing binaries. This is an incomplete list of new exoplanet candidates identified with `eleanor`. A full list of candidate signals will be included in future work. An example of the new candidate light curves can be seen in Figure 7.

We note that Sullivan et al. (2015) concluded that there will be roughly 1000 false positives within the 2-