



Precision Light Curves from *TESS* Full-frame Images: A Different Imaging Approach

Ryan J. Oelkers^{1,2,4} and Keivan G. Stassun^{1,2,3}

¹ Vanderbilt University, Department of Physics and Astronomy, Nashville, TN 37235, USA; ryan.j.oelkers@vanderbilt.edu

² Vanderbilt Initiative in Data-Intensive Astrophysics (VIDA), USA

³ Fisk University, Department of Physics, Nashville, TN 37208, USA

Received 2018 March 6; revised 2018 July 12; accepted 2018 July 25; published 2018 August 29

Abstract

The *Transiting Exoplanet Survey Satellite* (*TESS*) will observe ~ 150 million stars brighter than $T_{\text{mag}} \approx 16$, with photometric precision from 60 ppm to 3%, enabling an array of exoplanet and stellar astrophysics investigations. While light curves will be provided for $\sim 400,000$ targets observed at 2 minute cadence, observations of most stars will only be provided as full-frame images (FFIs) at 30 minute cadence. The *TESS* image scale of $\sim 21''/\text{pix}$ is highly susceptible to crowding, blending, and source confusion, and the highly spatially variable point-spread function (PSF) will challenge traditional techniques, such as aperture and Gaussian-kernel PSF photometry. We use official “End-to-End 6” *TESS* simulated FFIs to demonstrate a difference image analysis pipeline, using a δ -function kernel, that achieves the mission specification noise floor of $60 \text{ ppm hr}^{-1/2}$. We show that the pipeline performance does not depend on position across the field, and only $\sim 2\%$ of stars appear to exhibit residual systematics at the level of ~ 5 ppt. We also demonstrate recoverability of planet transits, eclipsing binaries, and other variables. We provide the pipeline as an open-source tool at <https://github.com/ryanoelkers/DIA> in both IDL and PYTHON. We intend to extract light curves for all point sources in the *TESS* FFIs as soon as they become publicly available, and will provide the light curves through the *Filtergraph* data visualization service. An example data portal based on the simulated FFIs is available for inspection at https://filtergraph.com/tess_ffs.

Key words: catalogs – methods: data analysis – techniques: image processing

1. Introduction

The dramatic increase in efforts to search for transiting exoplanets at the turn of the millennium motivated the use of small-aperture, wide-field surveys to simultaneously monitor as many stars as possible (e.g., Bakos et al. 2002; Pepper et al. 2003; Pollacco et al. 2006). Coupled with technological and computing advances, which provided pathways to rapid reduction of large data sets, the resulting massive influx of time-series photometry has also helped to guide astronomy into an era of “big data.” The next wave of large astronomical surveys is expected to further drive the astronomy community’s need for powerful data reduction tools capable of providing large-scale data products with high precision and on a rapid timescale.

One such upcoming survey is the *Transiting Exoplanet Survey Satellite* (*TESS*; Ricker et al. 2014), which will be conducting a nearly all-sky photometric survey for two years, with a core mission to discover small transiting exoplanets orbiting nearby bright stars. One major consequence of the *TESS* survey will be the enormous amount of time-series data collected at high cadence and for a very large number of stars. While only the $\sim 400,000$ highest priority transiting-planet host stars will be observed with 2 minute cadence, nearly every star outside $\pm 6^\circ$ of the ecliptic plane will be observed with at least 30 minute cadence—a total of $\sim 423,000,000$ stars in the *TESS* Input Catalog (TIC; Stassun et al. 2018)—whose data will be released not as extracted light curves but in the form of full-frame images (hereafter, FFIs; Sullivan et al. 2015).

TESS is equipped with four 10 cm telescopes, each containing four charged couple devices (CCDs), with a total field of view of 24 deg^2 on the celestial sphere during a single pointing. Each telescope is aligned to cover different areas of a

longitudinal strip (called a sector), thus the four CCDs together in a given camera observe a total coverage of $96^\circ \times 24^\circ$. The satellite will observe each ecliptic hemisphere with 13 sectors over the course of a year, each sector will be observed continuously for 27 days. Additionally, the increasing areal overlap of sectors at higher ecliptic latitudes will combine in a way to create continuous viewing zones around the ecliptic poles, designed to (1) match the *James Webb Space Telescope* continuous viewing zone, to facilitate potential follow up, and (2) avoid the ecliptic plane to minimize contamination from solar system objects. Of particular consequence for handling the FFIs, the telescope is capable of observing such large areas of the sky because of the optical design, which maps a large 21 arcsec^2 area onto a single *TESS* pixel (Ricker et al. 2014).

The extraction of light curves from wide-field images with such large pixel scales is highly nontrivial. The effects of crowding and blending hinder methods such as aperture photometry, because contamination from nearby stars cannot be easily removed from even a small photometric aperture (Wang et al. 2011). Similarly, the point-spread functions (PSFs) of such large fields of view are notoriously highly spatially variable, which makes photometric extraction based on PSF-fitting extremely laborious, as individual PSFs need to be constructed for different parts of the image—particularly near the edges of the frame. Wide-field images are also typically undersampled, which has the tendency of causing the measured flux of a given star to be correlated with its position on the frame, and adds to the complexity of the reduction by requiring multiple frames to model the PSF. The combination of these effects can introduce systematics in the light curves that are difficult to interpret and correct, and can even mimic some astrophysical signals of interest, leading to false positives (Alard & Lupton 1998; Alard 2000; Bramich 2008; Miller et al. 2008).

⁴ Vanderbilt Data Science Fellow.

Light-curve extraction via difference imaging analysis (hereafter, DIA) is one approach used by some wide-field surveys because it mitigates many of the effects of crowded fields (Alard & Lupton 1998; Alard 2000; Bramich 2008; Miller et al. 2008; Oelkers et al. 2015). DIA involves the subtraction of two frames: one frame, typically the one with the higher signal-to-noise ratio (S/N), is blurred to appropriately match the seeing conditions of the other frame; then the two are subtracted to reveal any pixels whose flux has changed between the two frames. Stars with no inherent astrophysical variation in their flux will subtract to their Poisson noise level, while stars with true variations will leave a significant residual on the frame involving multiple correlated pixels within the PSF. DIA is useful in crowded environments because the majority of the stars on a given frame should subtract cleanly, and then simple aperture or PSF photometry can extract the residuals from the differenced frame without contamination from neighboring stars.

Such an approach has already proven successful with ground-based surveys such as the Kilodegree Extremely Little Telescope (KELT), whose imaging field of view and pixel scale are very similar to that of *TESS* (Pepper et al. 2007, 2012). Indeed, KELT light curves extracted via a customized DIA-based pipeline (Siverd et al. 2012) typically achieve a precision of $\lesssim 10$ mmag (e.g., Oelkers et al. 2018).

However, most DIA pipelines, including that used by KELT, utilize a Gaussian basis to solve for the kernel that is used to blur the reference frame before subtraction. These kernels are advantageous because they require a solution for only a small number of basis vectors, and, in general, the PSFs of many astronomical images are approximately Gaussian in shape. However, when the PSF of the image deviates significantly from Gaussian, as in very wide-field images like the *TESS* FFIs, these methods tend to produce lower-quality subtractions and begin to suffer from similar problems of standard PSF photometry when the model PSF is poorly matched to the true PSF⁵ (see, e.g., Bramich 2008; Miller et al. 2008; Oelkers et al. 2015).

This manuscript describes a DIA approach, which uses a Dirac δ -function kernel to solve for non-Gaussian, arbitrarily shaped PSFs. It is a heavily modified version of the pipeline from Oelkers et al. (2015), optimized for *TESS* FFIs. We apply the adapted pipeline to official NASA “End-to-End 6” (Jenkins et al. 2004; Bryson et al. 2010; Jenkins et al. 2018, hereafter, ETE-6) *TESS* simulated FFIs to analyze its performance relative to nominal *TESS* specifications.

The remainder of this manuscript is organized as follows: Section 2 describes the simulated data; Section 3 details our data reduction pipeline; Section 4 presents the performance of the pipeline in terms of photometric precision relative to the expected *TESS* noise model; and Section 5 summarizes our results, including intended improvements to the pipeline prior to the first official *TESS* data release near the end of 2018.

2. Data: Simulated *TESS* Full Frame Images

We tested our pipeline using the ETE-6 FFIs provided by NASA (Jenkins et al. 2004; Bryson et al. 2010; Jenkins et al. 2018). These images were built using the expected photometric precision of the satellite, PSF shapes from lab

testing, and were designed to have a cadence similar to that expected for the mission. The PSF shape appears to vary significantly as a function of position on the detector (we measured a range of 1.13–2.76 pixels on a typical image, see Figure 1), but, in general, it has a typical full-width at half-maximum of 1.88 pixels. The location and intensity of stars on each frame were modeled after real stars found in the TIC (Stassun et al. 2018), and several hundred of the stars were injected with signals mimicking real astrophysical variations observed by the *Kepler* satellite (Borucki et al. 2010). Spacecraft jitter was also introduced into the data to simulate systematics not yet included in the official noise model (see Sullivan et al. 2015, updated by J. Pepper et al. 2018, in preparation), but which are expected to occur during the mission.

NASA released the ETE-6 FFIs on the Mikulski Archive for Space Telescopes (MAST) on 2018 February 15 in both uncalibrated and calibrated forms. Currently, the FFIs are available for all 16 CCDs and simulate a single observing sector for 27 days. To simplify the testing process, we selected the calibrated FFIs from a single CCD camera (in this case, CCD #2 on Camera 2, centered on $\alpha \sim 16.1$, $\delta \sim +28$).

The reduction of all 1348 images completed in ~ 4.5 days on 2 cores, and 106,399 light curves were extracted with *TESS* magnitudes $7 < T < 16$. The tested pipeline can later be applied to all 16 CCDs in parallel, thus we expect to be able to extract light curves from the eventual real FFI stars of a given sector within ~ 1 week of their public release.

3. Methods: A Difference Imaging Pipeline for Wide-field Images

The pipeline presented in this paper is a heavily modified version of the pipeline from Oelkers et al. (2015), which uses the mathematical precepts of Alard & Lupton (1998), Bramich (2008), and Miller et al. (2008), and has already been tested on numerous wide-field telescope systems during the past five years. The pipeline has reached the predicted precision floor⁶ for each system from which data have been fully reduced and published (see Figure 1, top). These systems have ranged in pixel scales from $0''.24/\text{pix}$ (Díaz et al. 2016) to $15''/\text{pix}$ (Oelkers et al. 2015) for published data, and up to $21''/\text{pix}$ in private testing with unofficial *TESS* FFIs (Z. Berta-Thompson 2016, private communication, SPyFFI pipeline).

The pipeline was originally designed to reduce data from the Chinese Small Telescope ARray (hereafter, CSTAR; Zhou et al. 2010; Oelkers et al. 2015). CSTAR was a small-aperture (~ 14.5 cm), wide-field (~ 20 deg²) telescope that was deployed to Dome-A in Antarctica during the Antarctic winters of 2008, 2009, and 2010. The telescope collected more than 10^6 images of nearly continuous time-series photometry of the south celestial pole during each ~ 6 month observing season. Data reductions of each season identified more than 100 new variable stars and transiting-planet candidates (Zhou et al. 2010; Wang et al. 2011, 2013; Oelkers et al. 2015, 2016b). While CSTAR represented a number of advances in its design and scientific results, the system was not without its limitations. The 2009 observing season had a number of technical issues that greatly stressed the data reduction (Oelkers et al. 2015). The most prominent feature of

⁵ We acknowledge that, in practice, many Gaussian kernels truncate to a δ -function kernel, if the proper combination of Gaussians is selected.

⁶ Based on the expected noise from stellar photon counts, sky background, and the scintillation limit for each system.

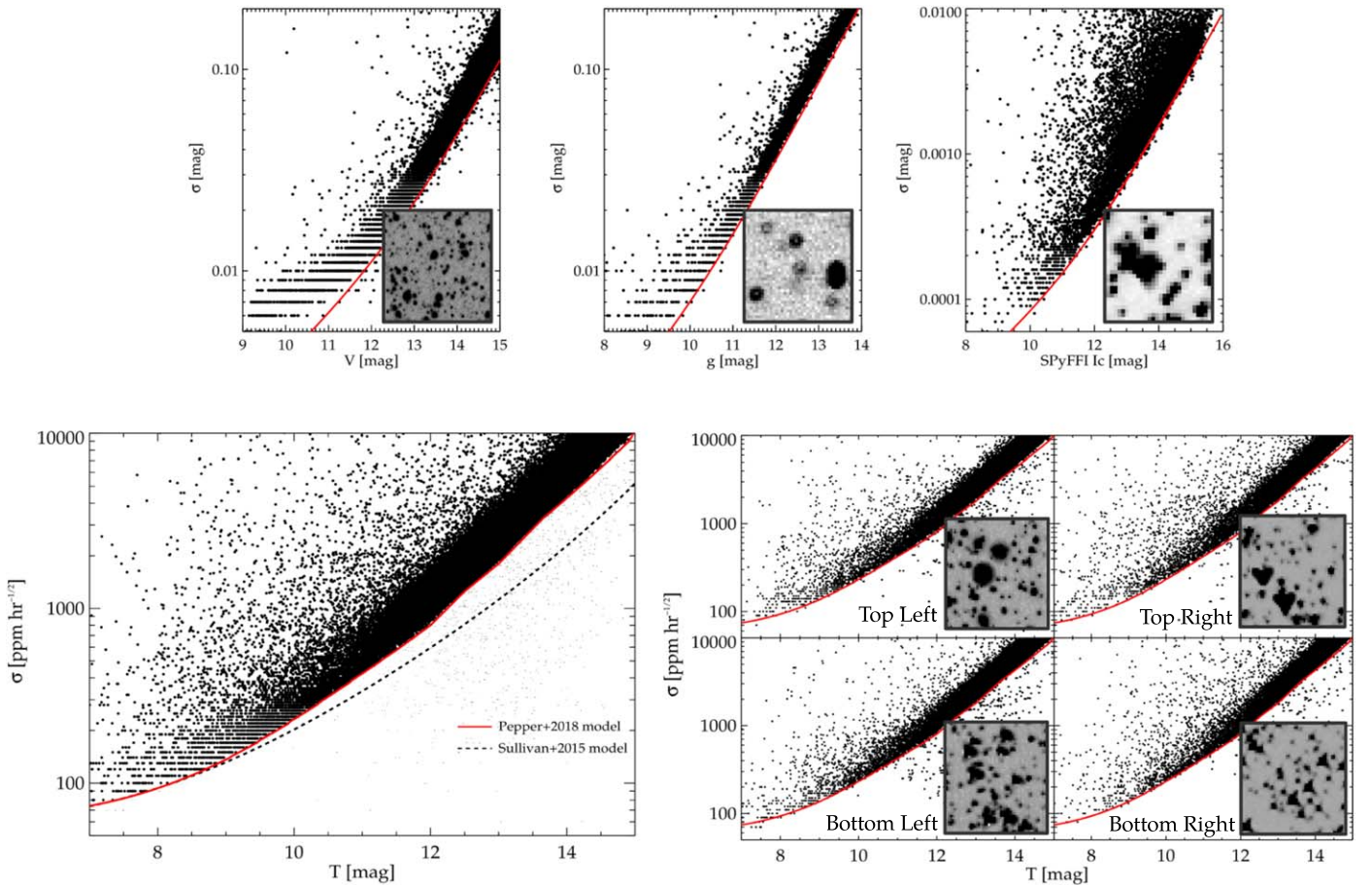


Figure 1. (Top) The achieved rms with the pipeline for $6''/4/\text{pix}$ (Oelkers et al. 2016a); the 2009-CSTAR data set with a $15''/\text{pix}$ with a defocused PSF Oelkers et al. (2015, 2016b); and initial testing with the SPyFFI images with $\sim 21''/\text{pix}$ (Z. Berta-Thompson 2016, private communication, SPyFFI data). (Bottom left) The achieved rms for the pipeline described in this work with ETE-6 simulated *TESS* data. The red line shows the expected *TESS* photometric precision from J. Pepper et al. (2018, in preparation) and the dashed line shows the expected *TESS* photometric precision from Sullivan et al. (2015). A small fraction ($\sim 2\%$) of stars, which were over-fit by the detrending routine, are plotted as small black dots under the red line. (Bottom right) The achieved rms as a function of position on the detector. An example PSF is shown in the inset of each panel to demonstrate the changing PSF shape. No discernible difference can be seen between the precision of the light curves as a function of position on the frame.

the data set was the defocused PSF that persisted through all three working cameras, which led Oelkers et al. (2015) to modify the Alard & Lupton (1998) and build upon the Bramich (2008) and Miller et al. (2008) DIA routines using a Dirac δ -function kernel to compensate for the changing, highly irregular PSF.

The current version of the pipeline is composed of eight routines designed to extract light curves from the FFIs in three steps: (1) background subtraction and image alignment; (2) master frame creation; and (3) image subtraction, fixed-aperture photometry, and trend removal. All routines for the pipeline are written in IDL and PYTHON (with the exception of the differencing step, which is written in C with IDL and PYTHON wrappers) to allow increased user flexibility.

We describe the basic routines below and identify specific IDL and PYTHON routines where appropriate. The pipeline typically fully reduces a single *TESS* FFI (from background subtraction to light-curve extraction) in 3 minutes using Intel Quad-Core Xeon 2.33 GHz/2.8 GHz processors, and the parameters described in this work.

3.1. Background Subtraction and Image Alignment

The fully calibrated FFIs exhibited a low-frequency residual sky background, which is meant to represent the zodiacal background and faint contamination from galaxies and unresolved stars (Stassun et al. 2018). We applied a residual background subtraction following the approach of Wang et al. (2013) and Oelkers et al. (2015). The residual background model is constructed by sampling the sky background every 32×32 pixels over the entire CCD, and replacing pixels with values of ADU $> 25,000$, with the median sky background in each box. A model sky is then generated by interpolating between all boxes to make a “thin plate spline” (Duchon 1976). We used the IDL implementation GRID_TPS and the PYTHON implementation of Rbf to make the splines that are then subtracted from every frame.

Difference imaging requires precise frame alignment in order to produce a proper subtraction. Slight variations caused by improperly aligned frames can contribute to poorly measured fluxes, which will introduce additional dispersion in the extracted light curves. While the *TESS* pointing is expected to be much better than 1 pixel, spacecraft jitter is likely and has been introduced into the simulated FFIs. We use the world coordinate solution (WCS) from the image headers to align the

images using the IDL implementation of HASTROM and PYTHON implementation of HCONGRID with cubic-spline interpolation.⁷

3.2. Master Frame Combination

A high-quality master frame, which will be convolved and subtracted from each science frame, is required to preserve the precision of the extracted photometry. Typically, this frame is generated by median combining many individual frames with high S/N and the best seeing, obtained throughout the observing campaign. We created our master frame by median combining all 1348 images. To save machine memory, images are median combined in sets of 50, producing 26 temporary master frames, which are then combined into a single final master frame. This procedure does not apply any pixel weighting or treatment for cosmic rays, but could be improved with such measures in future reductions.

Next, we identified all 106,399 stars in the TIC with *TESS* magnitude (T) $T < 16$ that were within the master frame field of view. We then calculated the curve of growth for a variety of aperture sizes, and selected the aperture radius that optimizes the flux calculation for the majority of the stars on the frame. The flux within the aperture is calculated using a polygon approximation of the intersection between a circular aperture and a square pixel, normalized by the total area of the sum of the pixels to match the circular aperture precisely. We found this aperture radius size to be 2.5 pixels. This is consistent with the measurements of the typical PSF FWHM of 1.88 pixels, which would transcribe >99.7% of a typical star’s flux within 2.39 pixels.

Fluxes for all stars were then extracted from the master frame using fixed-aperture photometry. We defined the zeropoint offset between the instrumental magnitude scale and the TIC magnitude scale as the median difference between the instrumental magnitude and the TIC T magnitude for all stars. We found the zeropoint to be 4.825 mag. We note that for a small number of stars we found a different median zeropoint offset that was ~ 0.75 mag brighter than the offset above. This suggests that a second star of roughly equal brightness is present in the simulated FFIs at the same location, even though only one TIC object exists at that location. This 0.75 mag different offset was found only around some, but not all, bright stars, $T < 11$. These “duplicate” stars are a known issue in the TIC for version 5 and earlier, and were caused by ghost artifacts in the SDSS extended source catalog that occur near bright stars. These objects have been removed from future TIC versions. Therefore, we accepted the zeropoint offset that was consistent with the majority of stars (Stassun et al. 2018).

3.3. Image Subtraction and Aperture Photometry

Typically, DIA routines use an adaptive kernel, $K(x, y)$, defined as the combination of 2 or more Gaussians. While effective at modeling well-defined, circular PSFs, this kernel has difficulty properly fitting other PSF shapes, particularly for highly distorted stars near the edge of the frame (Bramich 2008; Miller et al. 2008). The *TESS* PSF varies quite significantly

across the frame. Therefore, we used a Dirac δ -function kernel to compensate for the noncircular, irregular PSF shape. We defined the kernel as

$$K(x, y) = \sum_{\alpha=-w}^w \sum_{\beta=-w}^w c_{\alpha,\beta}(x, y) K_{\alpha,\beta}(u, v), \quad (1)$$

where $K_{\alpha,\beta}$ is a combination of $(2w + 1)^2$ δ -function basis vectors and $K_{0,0}$ is the centered δ function (Bramich 2008; Miller et al. 2008). We defined our basis vectors to ensure a constant photometric flux ratio between images (Alard 2000; Bramich 2008; Miller et al. 2008). In the case of $\alpha \neq 0$ and $\beta \neq 0$,

$$K_{\alpha,\beta}(u, v) = \delta(u - \alpha, v - \beta) - \delta(u, v), \quad (2)$$

while for $\alpha = 0$ and $\beta = 0$,

$$K_{0,0}(u, v) = \delta(u, v). \quad (3)$$

Stamps are taken around at most 500 bright, isolated⁸ stars to solve for the coefficients $c_{\alpha,\beta}(x, y)$ using the least-squares method. We allowed $c_{0,0}(x, y)$ to be spatially variable to compensate for variations that may correlate with position on the detector: such as poor sky subtractions, blending, or imperfections in the flat-field. Stars are identified as suitable candidates for stamps provided their measured photometric error was < 50 mmag and there were no other, brighter stars within 3 pixels of the target star. We found that we could use a 5×5 pixel kernel across the frame without appreciably affecting the quality of the subtraction, but significantly reducing the runtime.

We set the photometric extraction aperture at a 2.5 pixel radius ($52''$) and set the image background to be the value determined in Section 3.1. The differential flux was then combined with the flux from the reference frame, and zero-pointed using the offset described above. Finally, the photometric errors were rescaled following a methodology similar to Lupton et al. (1989) and Kaluzny et al. (1998).

3.4. Light-curve Detrending

Some low-level systematics persisted in the extracted light curves, even with the care taken to properly preprocess and difference the images. While the exact source of these systematics is unknown, the likely explanations include improper image alignment, imperfect stellar stamp selection, minor variations in the calibration process, or a combination of these and other factors.

We opted to remove (detrend) these systematics from the photometry using an ensemble of light curves to craft a model of nonastrophysical signals. First, for each star, we identified the 1000 closest stars of similar magnitude ($\sim \pm 0.1$ mag). We have found that stars of similar magnitude tend to show similar systematics for a variety of reasons, including response to similar pixels, optical effects on a particular area of the CCD, blending effects from bright stars, and local background from sources such as zodiacal light or the Galactic plane. Atmospheric effects cause similar systematics in nearby stars in ground-based surveys, but this would not affect stars observed by *TESS*. Then we identify a subset of stars that, when subtracted from the light curve, decrease the root-mean-square

⁷ We assume that the noise in the interpolated images is nearly identical to the original image. Given the typical FWHM of the *TESS* mock-FFIs is near 1.88 pixels, suggesting the images are undersampled, it is possible that additional noise is being added to the images. However, given we reach the noise floor (as shown in Figure 1), we do not believe appreciable levels of noise are being added in this step.

⁸ The stars need not be isolated, but in this pipeline, isolated stars are preferred.

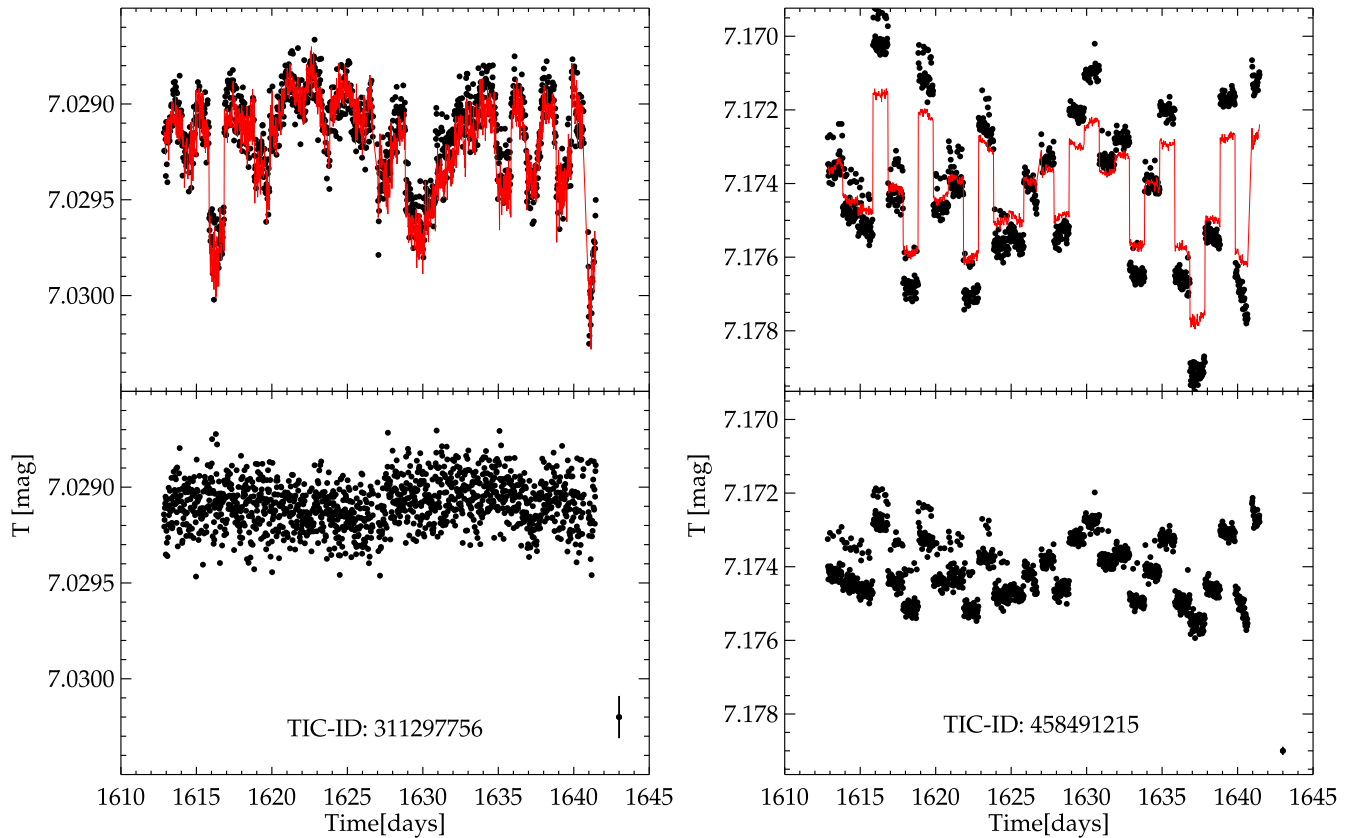


Figure 2. Two light curves from the simulated data set, before (top) and after (bottom) undergoing the detrending process. Typical photometric uncertainties are shown on the bottom right of each bottom panel. The model trend (red line) is built from a set of stars with similar magnitude that produce at least a 10% improvement in the star’s rms. While the star on the left is visibly cleaned by the detrending process, the star on the right still suffers from systematics; we estimate such systematics persist in $\sim 2\%$ of light curves.

(rms) variations of the light curve by at least 10%. Stars that do not decrease the rms are discarded from the model. If no comparison stars were found to decrease the rms of the target star, then no detrending was applied. If multiple trend stars were found to decrease the rms of the target light curve, then they were median combined into a master model.

During this process, we noticed that many stars exhibited abrupt changes in magnitude, typically after 48 observations—which is 1 day given the 30 minute FFI candence. These changes in magnitude are likely a recovery of how the simulated images were generated. While a selection of stars were fully rendered throughout the entire observing season, most stars were regenerated once per day. This effect created a visual break in the light curves every 48 data points (Jenkins et al. 2018). We removed these offsets by subtracting subsequent data points in a given trend model, and identifying observations where the deviation between data points was larger than 5σ of the mean deviation between data points. Each subsection of the trend model was then scaled to match the median magnitude of the light curve during this subset. However, if subtracting the scaled trend did not improve the rms by more than 5%, the trend during this subsection of the light curve was not scaled. This was done to ensure that true astrophysical signals would not be removed in the detrending routine.

Figure 2 shows an example of this detrending and its proficiency at removing nonastrophysical signals. While the detrending routine was proficient at reducing the rms of many stars, there were still stars that failed the detrending process

(see Figure 2, right). We mention these artifact signals as a caution, but note that these events are uncommon in our tests and likely affect $\sim 2\%$ of stars.

Of course, systematics typically vary greatly between systems, and the systematics of the simulated FFIs analyzed here may not fully capture the full systematics of the final, true FFIs after *TESS* launch. Therefore, we propose several methods that users can implement and that we plan to include with the final release of the FFI data products. These methods include:

1. Position based detrending: While the pointing of the telescope is expected to be exceptional, it will not be immune to slight corrections for drift. One can use the correlation of time and a star’s x , y position to remove trends that appear as the star’s centroid shifts from pixel to pixel. While in theory, many of these correlations should be accounted for within the δ -function kernel, in practice, we have found that for very wide-field images, the δ -function kernel is not a perfect representation across the entire image and small systematics remain on millimag levels. Given the light curves have a precision of parts-per-million, we expect there to be variations that cannot be adequately fit for by the kernel, and instead we suggest one may be able to remove these systematics with a “residual-flat-field” as shown in Wang et al. (2013) and Vanderburg & Johnson (2014).
2. Magnitude-based detrending: This method loosely follows the plan we outline above. Stars of similar magnitude can be combined to make a model of

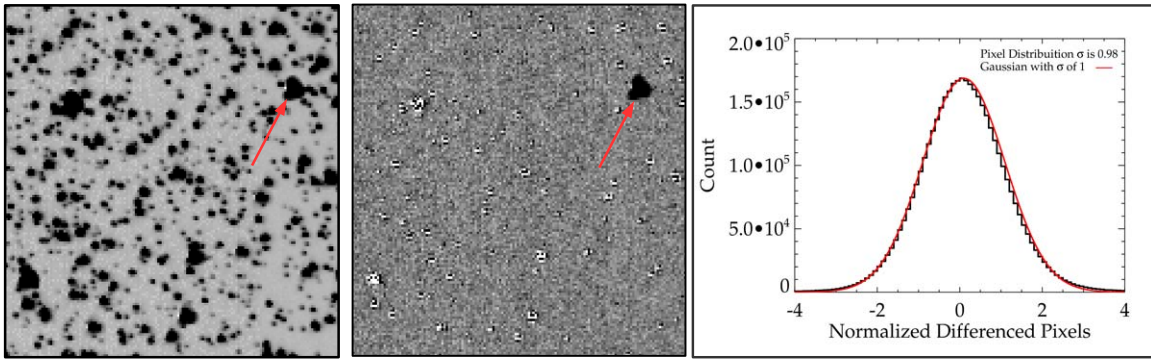


Figure 3. 150×150 pixel ($52/5 \times 52/5$) cutouts of a typical science frame (left) and differenced frame (center). The color has been inverted on each frame for clarity. The majority of the stars on the differenced frame either subtract cleanly or show uncorrelated residuals. The red arrows point to an area of correlated positive residuals suggesting a possible variable object. (Right) The normalized histogram of differenced pixel values on the differenced frame (black line) with an overplotted Gaussian of $\sigma = 1$ (red line). The normalized distribution and Gaussian match quite well, suggesting the noise in the differenced frame is meeting the expectation.

variations that are not astrophysical in nature. Typically, two stars of similar magnitude should not vary in identical astrophysical patterns between images, even if they were the same spectral type or variable type.

3. Fourier-based detrending: By investigating the power spectrum of the light curve, multiple stars may show similar periods, or aliases of similar periods. These stars could be combined to create a trend pattern, if the period is known a priori to be spurious.

4. Results: Pipeline Demonstration with the Simulated Full Frame Images

4.1. Performance of the δ Function Kernel

The quality of a differenced image can be quantified using stars that cleanly subtract in an image. Specifically, the quality of a differenced frame can be described by assessing the degree to which the deviations left in the differenced frame match the expectations for the noise from the science frame and the master frame (Alard 2000).

To show that our differenced frames match the expectations for the noise, we select a frame and normalize each pixel value by the combination of the noise from the science frame and the master frame. We defined the expected noise⁹ as $\delta = \sqrt{I_N + R_N}$, where I_N is the photon counts in the science image and R_N are the photon counts in the master frame, both normalized by the number of frames used to create each image. If the subtraction generally matches the expectations from Poisson noise, the normalized pixels should show a Gaussian distribution centered near 0 with a standard deviation of 1. Figure 3 shows the histogram of pixel values for a typical differenced frame normalized by δ . The normalized residuals show a mean of 0.132 and a standard deviation of 0.98. We believe this bias is likely due to the sky background, given each pixel is $21''$ in size, and we only account for sky background during the preprocessing of the images as described in Section 3.1. It is possible that the procedure could be improved with an additional sky subtraction after the differencing step, but scatter in the normalized pixels is nearly 1, which suggests only minimal levels of noise are being introduced by this bias.

⁹ This definition differs slightly from the Alard (2000) model, which uses the *convolved* master frame noise. We found that the two models do not produce significantly different results, and because our master frame is a high S/N template, we therefore accepted the simpler of the two for our analysis.

We accept this as sufficient evidence that the residuals in the differenced frame are closely matching the expected Poisson noise.

4.2. Light-curve Precision

Next, we compared our resulting light curves to the current models of expected *TESS* precision as a function of stellar magnitude. In particular, we use the original noise model from Ricker et al. (2014) and Sullivan et al. (2015) and the updated noise model from J. Pepper et al. (2018, in preparation). These models incorporate shot noise from the star, sky background (in the case of *TESS*, contamination from zodiacal light, unresolved stars, and background galaxies), the read noise of the detector, and a mission-specified noise floor of 60 ppm on 1 hr timescales.¹⁰ J. Pepper et al. (2018, in preparation) included an additional estimation of the contamination by nearby stars, which we have excluded from our model because this effect is present in the simulated FFI data through physical contamination, and is manifested in our analysis by stars that deviate from the expected noise floor. These noise models differ quite substantially, particularly in the sky-dominated regime. This is principally because Ricker et al. (2014) and Sullivan et al. (2015) used an initial optimal-aperture size estimate, while J. Pepper et al. (2018, in preparation) use an updated optimal-aperture model based on lab testing of the cameras. We adopt the J. Pepper et al. (2018, in preparation) model, as a more up-to-date estimation of the photometric precision,¹¹ but we include both models for comparison.

As shown in Figure 1 (bottom left), our pipeline very satisfactorily reproduces the expected noise floor of J. Pepper et al. (2018, in preparation); the noise floor model matches the lower envelope of stars, representing nonvariable objects. Stars above the lower envelope noise floor are interpreted to represent variable objects of various types (see below).

Wide-field imagers typically have a PSF that is largely dependent on the location of the stars on the detector, and the severity of the noncircular PSF shape can be exacerbated near

¹⁰ While not necessarily useful for the purposes of testing our differencing routine, we have included the systematic floor in our model because it is standard in *TESS* yield predictions.

¹¹ We note that we smoothed the fourth-order estimate of the optimal aperture from J. Pepper et al. (2018, in preparation) for our analysis, because we wanted to remove the structure it created in the final model. We also assume a maximum aperture no greater than what is used for a $T = 4$ star, and a minimum aperture of 1 pixel following the work of Stassun et al. (2018).

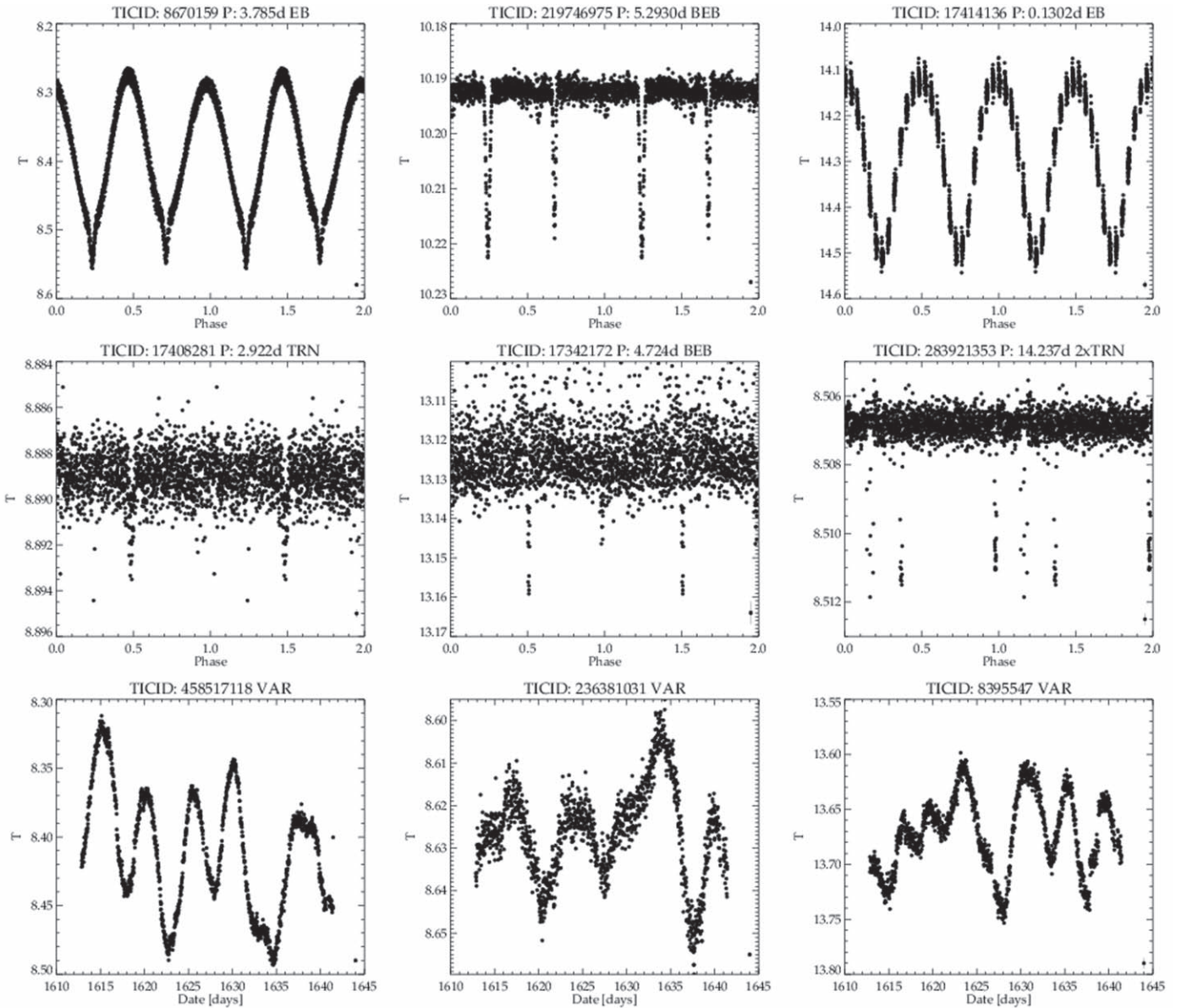


Figure 4. Nine representative light curves from the simulated data set: two simulated eclipsing binaries and one background eclipsing binary (top row); two simulated transits and one background eclipsing binary (middle row); and three simulated large amplitude variable stars (bottom row). Each object is identified by its injected “Ground-Truth” signature: TRN for transit; EB for eclipsing binary; VAR for variables; and BEB for background eclipsing binary. TICID 283921353 is a system with two transiting signals injected. Phased light curves have been phase-folded on the “Ground-Truth” period and plotted twice for clarity, but not binned. Representative photometric error bars can be seen at the bottom right of each panel.

the edges of the frame (Pepper et al. 2007). Since the shape of the PSF is largely position-dependent on the simulated FFIs, we checked 1024×1024 pixel subimages in the four CCD corners to see if the precision changed as a function of detector position. As shown in Figure 1 (lower right) there is no appreciable difference on any part of the frame.

4.3. Identifying Variable Stars and Transiting-planet Candidates

The data release notes for the simulated FFIs indicate that several hundred variable-star and planet-transit signatures were injected into the FFIs to enable checks of signal recoverability. We searched for stars in the simulated data that showed stellar variability and/or transit candidates using the variability metrics of Oelkers et al. (2018), a basic Lomb–Scargle

periodicity search (Lomb 1976; Scargle 1982, LS;), and the box-least-square (BLS; Kovács et al. 2002) algorithm to identify transit-like events.

The variability metrics of Oelkers et al. (2018) were shown to work well to identify stars with large amplitude variability in time-series data. We calculated the metrics (Stetson 1996; Wang et al. 2013, rms, $\Delta 90$, Welch–Stetson J and L) for every star in the frame, and each star has their specific metric values compared to stars of similar magnitude (for rms and $\Delta 90$ metrics) or to the entire field (for J and L metrics). Stars that have metric values larger than a threshold (typically $+2\sigma$ for rms and $\Delta 90$; and $+3\sigma$ for J and L) are typically considered variable. These metric values are calculated for every star, and are included in our data release.

We provide metrics for periodic signals using two methods. First, we ran a basic LS search (Lomb 1976; Scargle 1982) for

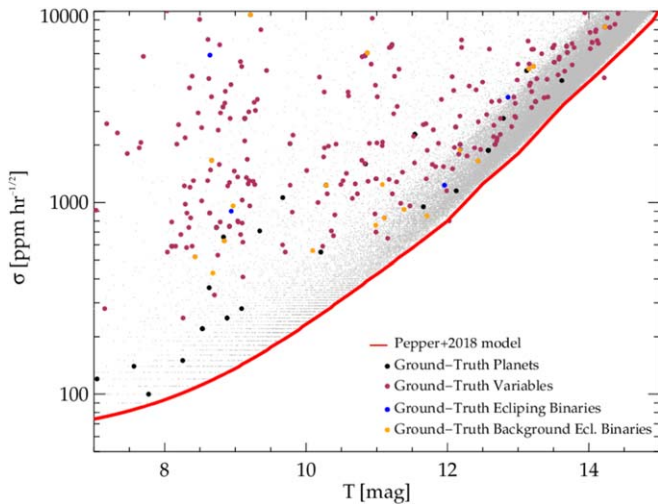


Figure 5. Reproduction of the bottom left panel of Figure 1 showing the achieved precision for “ground-truth” objects in our analysis. Injected planets are shown as black dots, variable stars are shown as maroon dots, eclipsing binaries are shown as blue dots, background eclipsing binaries are shown as orange points, and all other stars are shown as gray dots. The noise model is also shown as a red line.

periods between 0.01 and 27 days. Second, we ran a BLS search (Kovács et al. 2002) for periods between 0.3 and 27 days with 10,000 frequency steps and 100 phase bins. The LS and BLS results are included in our data release.

Typically, heuristic cutoffs will be identified using the data from the entire frame to identify the most likely astrophysical events. In the future, we plan to detect events using a methodology similar to previous work (Wang et al. 2011, 2013; Rodriguez et al. 2017; Ansdell et al. 2018; Oelkers et al. 2018). We visually inspected light curves of several stars identified in the official “Ground-Truth” lists released by NASA (Jenkins et al. 2018), and we have plotted example stars in Figure 4. Additionally, we show the precision achieved for each “Ground-Truth” object observed in our analysis in Figure 5. Some objects with large rms values are clearly not “Ground-Truth” objects, but instead were identified as poorly detrended stars (<2% of all objects). In any case, we include all calculated metrics in our data release portal (see below) to allow users to experiment with the data using different significance thresholds.

5. Summary

We have presented a data reduction pipeline that has been adapted for *TESS* FFIs. The pipeline includes routines for background subtraction, image alignment, master frame combination, difference imaging, aperture photometry, and magnitude-based trend removal. The final precision of the produced light curves has been shown to match the expected photometric precision of the detector. Finally, a variety of injected astrophysical variables and transit signatures have been identified.

The full pipeline described in this paper is currently available for download through a GITHUB repository at the URL <https://github.com/ryanoelkers/DIA/> (Oelkers et al. 2018). We have tested the routines on multiple computers with multiple data sets, and we believe the code can be readily used as is or adapted. Some external libraries are required, and we list them with the README file at the repository. The pipeline

is currently in a “version zero” state, and users should expect the pipeline to be updated and improved prior to the public release of *TESS* data from the first observed sector, expected for late 2018.

We encourage adaptation of the code, and we encourage users to cite this work as well as Alard (2000), Alard & Lupton (1998), Bramich (2008), Miller et al. (2008), and Oelkers et al. (2015) given the large contributions of those works to the formulation of the pipeline presented here.

We plan to use our pipeline to reduce *TESS* FFIs and we will release all data products to the public on a rapid timescale.¹² All data products will be released through the *Filtergraph* visualization portal at the URL https://filtergraph.com/teess_ffi. Users will be able to use the portal to access TIC information for the extracted stars, light-curve data files, image files of light curves, and links to Simbad and Aladin for each star. We will also provide basic variability information calculated for each light curve using the metrics from Oelkers et al. (2018), and periodicity information using a Lomb–Scargle analysis (Lomb 1976; Scargle 1982), as well as a best-fit box-least-square period (Kovács et al. 2002). We encourage readers to visit the *Filtergraph* URL even in advance of the first sector data release to see an example of the data release using the data from this work, and welcome suggestions for how the data release can be improved in the future.

The authors would like to thank the referee, Robert Lupton, for his valuable critiques and comments, which greatly improved the quality of this manuscript. The authors would also like to thank the *TESS* Science Processing Operations Center for making the simulated images available for the community to download and experiment with. R.J.O. would like to thank the Stassun group for their useful discussion and suggestions. R.J.O. acknowledges support through the NASA 17-XRP-17 2-0024 grant. This work has made extensive use of the *Filtergraph* data visualization service (Burger et al. 2013) at <https://filtergraph.com>. This research has made use of the VizieR catalogue access tool, CDS, Strasbourg, France. This work was conducted in part using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University, Nashville. We acknowledge use of the ADS bibliographic service. This work has made use of the TIC and CTL, through the *TESS* Science Office’s target selection working group (architects K. Stassun, J. Pepper, N. De Lee, M. Paegert, R. Oelkers).

ORCID iDs

Ryan J. Oelkers <https://orcid.org/0000-0002-0582-1751>

Keivan G. Stassun <https://orcid.org/0000-0002-3481-9052>

References

- Alard, C. 2000, *A&AS*, **144**, 363
- Alard, C., & Lupton, R. H. 1998, *ApJ*, **503**, 325
- Ansdell, M., Oelkers, R. J., Rodriguez, J. E., et al. 2018, *MNRAS*, **473**, 1231
- Bakos, G. Á, Lázár, J., Papp, I., Sári, P., & Green, E. M. 2002, *PASP*, **114**, 974
- Borucki, W. J., Koch, D., Basri, G., et al. 2010, *Sci*, **327**, 977
- Bramich, D. M. 2008, *MNRAS*, **386**, L77
- Bryson, S. T., Jenkins, J. M., Peters, D. J., et al. 2010, *Proc. SPIE*, **7738**, 773808

¹² Currently, we estimate this will be approximately 1–2 weeks after a *TESS* public data release given our current computing resources.

- Burger, D., Stassun, K. G., Pepper, J. A., et al. 2013, in ASP Conf. Ser. 475, *Astronomical Data Analysis Software and Systems XXII*, ed. D. N. Friedel (San Francisco, CA: ASP), 399
- Díaz, M. C., Beroiz, M., Peñuela, T., et al. 2016, *ApJL*, **828**, L16
- Duchon, J. 1976, *ESAIM: Mathematical Modelling and Numerical Analysis—Modélisation Mathématique et Analyse Numérique*, 10, 5
- Jenkins, J. M., Peters, D. J., & Murphy, D. W. 2004, *Proc. SPIE*, **5497**, 202
- Jenkins, J. M., Tenenbaum, P., Caldwell, D. A., et al. 2018, *RNAAS*, **2**, 47
- Kaluzny, J., Stanek, K. Z., Krockenberger, M., et al. 1998, *AJ*, **115**, 1016
- Kovács, G., Zucker, S., & Mazeh, T. 2002, *A&A*, **391**, 369
- Lomb, N. R. 1976, *Ap&SS*, **39**, 447
- Lupton, R. H., Fall, S. M., Freeman, K. C., & Elson, R. A. W. 1989, *ApJ*, **347**, 201
- Miller, J. P., Pennypacker, C. R., & White, G. L. 2008, *PASP*, **120**, 449
- Oelkers, R. J., Macri, L. M., Marshall, J. L., et al. 2016a, *AJ*, **152**, 75
- Oelkers, R. J., Macri, L. M., Wang, L., et al. 2015, *AJ*, **149**, 50
- Oelkers, R. J., Macri, L. M., Wang, L., et al. 2016b, *AJ*, **151**, 166
- Oelkers, R. J., Rodriguez, J. E., Stassun, K. G., et al. 2018, *AJ*, **155**, 39
- Oelkers, R. J., & Stassun, K. G. 2018, DIA: Delta Function Difference Imaging Code, v0.1, Zenodo, doi:[10.5281/zenodo.1309975](https://doi.org/10.5281/zenodo.1309975)
- Pepper, J., Gould, A., & Depoy, D. L. 2003, *AcA*, **53**, 213
- Pepper, J., Kuhn, R. B., Siverd, R., James, D., & Stassun, K. 2012, *PASP*, **124**, 230
- Pepper, J., Pogge, R. W., DePoy, D. L., et al. 2007, *PASP*, **119**, 923
- Pollacco, D. L., Skillen, I., Collier Cameron, A., et al. 2006, *PASP*, **118**, 1407
- Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2014, *Proc. SPIE*, **9143**, 914320
- Rodriguez, J. E., Ansdell, M., Oelkers, R. J., et al. 2017, *ApJ*, **848**, 97
- Scargle, J. D. 1982, *ApJ*, **263**, 835
- Siverd, R. J., Beatty, T. G., Pepper, J., et al. 2012, *ApJ*, **761**, 123
- Stassun, K. G., Oelkers, R. J., Pepper, J., et al. 2018, *AJ*, **156**, 102
- Stetson, P. B. 1996, *PASP*, **108**, 851
- Sullivan, P. W., Winn, J. N., Berta-Thompson, Z. K., et al. 2015, *ApJ*, **809**, 77
- Vanderburg, A., & Johnson, J. A. 2014, *PASP*, **126**, 948
- Wang, L., Macri, L. M., Krisciunas, K., et al. 2011, *AJ*, **142**, 155
- Wang, L., Macri, L. M., Wang, L., et al. 2013, *AJ*, **146**, 139
- Zhou, X., Fan, Z., Jiang, Z., et al. 2010, *PASP*, **122**, 347