



## TESS Data for Asteroseismology: Photometry

Rasmus Handberg<sup>1</sup>, Mikkel N. Lund<sup>1</sup>, Timothy R. White<sup>1,2,3</sup>, Oliver J. Hall<sup>1,4,5</sup>, Derek L. Buzasi<sup>6</sup>, Benjamin J. S. Pope<sup>7,8,14</sup>, Jonas S. Hansen<sup>1</sup>, Carolina von Essen<sup>1,9</sup>, Lindsey Carboneau<sup>1,5,6</sup>, Daniel Huber<sup>10</sup>, Roland K. Vanderspek<sup>11</sup>, Michael M. Fausnaugh<sup>11</sup>, Peter Tenenbaum<sup>12,13</sup>, and Jon M. Jenkins<sup>12</sup>

The T<sup>2</sup>DA Collaboration

- <sup>1</sup>Stellar Astrophysics Centre, Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, DK-8000 Aarhus C, Denmark; [rasmush@phys.au.dk](mailto:rasmush@phys.au.dk)  
<sup>2</sup>Sydney Institute for Astronomy (SfA), School of Physics, University of Sydney, 2006, Australia  
<sup>3</sup>Research School of Astronomy and Astrophysics, Mount Stromlo Observatory, The Australian National University, Canberra, ACT 2611, Australia  
<sup>4</sup>European Space Agency (ESA), European Space Research and Technology Centre (ESTEC), Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands  
<sup>5</sup>School of Physics and Astronomy, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK  
<sup>6</sup>Department of Chemistry and Physics, Florida Gulf Coast University, 10501 FGCU Boulevard S, Fort Myers, FL 33965, USA  
<sup>7</sup>Center for Cosmology and Particle Physics, Department of Physics, New York University, 726 Broadway, New York, NY 10003, USA  
<sup>8</sup>Center for Data Science, New York University, 60 Fifth Avenue, New York, NY 10011, USA  
<sup>9</sup>Astronomical Observatory, Institute of Theoretical Physics and Astronomy, Vilnius University, Sauletekio av. 3, 10257, Vilnius, Lithuania  
<sup>10</sup>Institute for Astronomy, University of Hawai'i, 2680 Woodlawn Drive, Honolulu, HI 96822, USA  
<sup>11</sup>Department of Physics and Kavli Institute for Astrophysics and Space Research, MIT, Cambridge, MA 02139, USA  
<sup>12</sup>NASA Ames Research Center, Moffett Field, CA 94035, USA  
<sup>13</sup>SETI Institute, Mountain View, CA 94043, USA

Received 2020 December 22; revised 2021 June 8; accepted 2021 June 9; published 2021 September 29

### Abstract

Over the last two decades, asteroseismology has increasingly proven to be the observational tool of choice for the study of stellar physics, aided by the high quality of data available from space-based missions such as CoRoT, *Kepler*, K2, and the Transiting Exoplanet Survey Satellite (TESS). TESS in particular will produce more than an order of magnitude more such data than has ever been available before. While the standard TESS mission products include light curves from 120 s observations suitable for both exoplanet and asteroseismic studies, they do not include light curves for the vastly larger number of targets observed by the mission at a longer 1800 s cadence in Full Frame Images (FFIs). To address this lack, the TESS Data for Asteroseismology (T<sup>2</sup>DA) group under the TESS Asteroseismic Science Consortium (TASC) has constructed an open-source pipeline focused on producing light curves for all stars observed by TESS at all cadences, currently including stars down to a TESS magnitude of 15. The pipeline includes target identification, background estimation and removal, correction of FFI time stamps, and a range of potential photometric extraction methodologies, though aperture photometry is currently the default approach. For the brightest targets, we transparently apply a halo photometry algorithm to construct a calibrated light curve from unsaturated pixels in the image. In this paper, we describe in detail the algorithms, functionality, and products of this pipeline and summarize the noise metrics for the light curves. Companion papers will address the removal of systematic noise sources from our light curves, and a stellar variability classification from these.

*Unified Astronomy Thesaurus concepts:* [Asteroseismology \(73\)](#); [Astronomy data analysis \(1858\)](#); [CCD photometry \(208\)](#); [Variable stars \(1761\)](#)

*Supporting material:* animation

### 1. Introduction

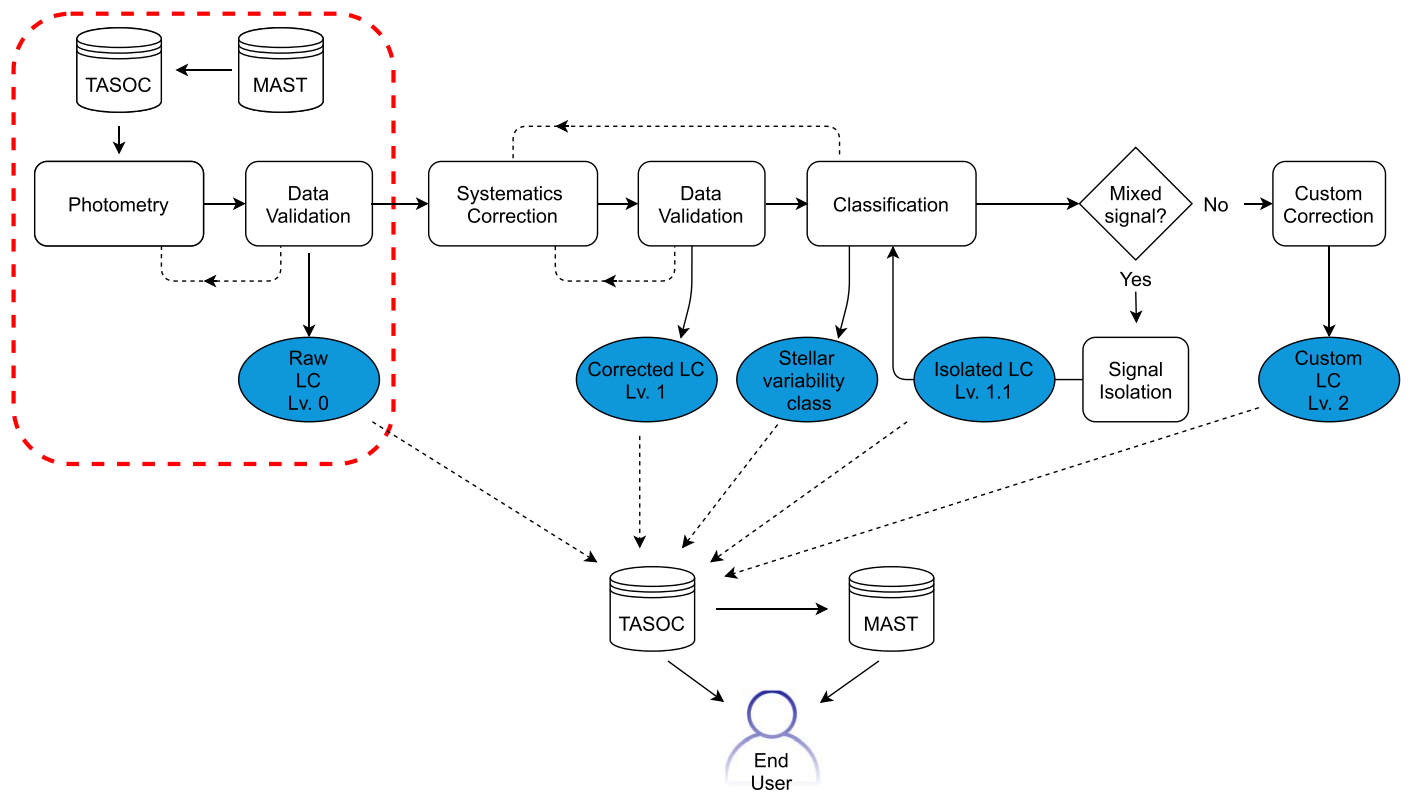
Asteroseismology, the study of stellar oscillations, has proven itself to be a powerful tool for accurate determination of global stellar properties like mass, radius, and age, utilized, e.g., in exoplanet characterization (see, e.g., Batalha et al. 2011), and for studies of the galaxy through galactic archeology (see, e.g., Miglio et al. 2013). Furthermore, in its own right, asteroseismology enables unique studies of the detailed internal structure and evolution of stars across the Hertzsprung–Russell diagram (HRD; see, e.g., Aerts et al. 2010).

During the last couple of decades asteroseismic investigations have undergone rapid development, fueled to a large extent by the large amount of high-quality data available. Starting with bespoke studies of single stars from the ground and space to large photometric surveys of thousands of stars from space-based missions like CoRoT, *Kepler*, and K2

(Baglin et al. 2002; Borucki et al. 2010; Howell et al. 2014), the move toward larger data sets has naturally meant a shift to a more pipeline-oriented processing of the data (see, e.g., Handberg & Lund 2014; Vanderburg & Johnson 2014; Lund et al. 2015; Luger et al. 2016; Jenkins 2017). Because stars exhibiting asteroseismic oscillations span the HRD, the oscillation signals themselves span orders of magnitudes in both frequencies and amplitudes, which on its own is a challenge for any processing.

With data from the Transiting Exoplanet Survey Satellite (TESS; Ricker et al. 2014; Vanderspek et al. 2018) the asteroseismic community is faced with a new set of challenges concerning the preparation of data. TESS gathers FFIs of  $96^\circ \times 24^\circ$  regions of the sky at 1800 s cadence, but no photometric extractions are performed, and hence no light curves are released as official data products. Existing community-built FFI photometry software includes *Eleanor* (Feinstein et al. 2019), the MIT Quick-look pipeline (Fausnaugh et al. 2018; Huang et al. 2020),

<sup>14</sup> NASA Sagan Fellow.



**Figure 1.** The overall structure of the full TASOC pipeline, with modules given as rectangular boxes, data products as ellipses, and “TASOC” and “MAST” indicating the databases hosting the data products. Dashed lines between modules indicate that an iteration might take place. The part enclosed by the red dashed line indicates the pipeline component described in this paper. The “systematics correction” part of the pipeline is described in Lund et al. (2021), while the “classification” is detailed in Audenaert et al. (2021).

the SPOC FFI pipeline (Caldwell et al. 2020), as well as difference imaging (Oelkers & Stassun 2018; Bouma et al. 2019; Montalto et al. 2020) and point-spread function (PSF) methods (Nardiello et al. 2019). However, neither of these techniques are optimized for asteroseismology, which is a large science driver for TESS photometry.

Therefore, the “TESS Data for Asteroseismology” (T<sup>2</sup>DA) coordinated activity within the TESS Asteroseismic Science Consortium (TASC; Lund et al. 2017) has been tasked with delivering light curves ready for asteroseismic analysis for each target observed by TESS, including those in the FFIs and from Target Pixel Files (TPFs). This is achieved using a custom-built, Open Source, pipeline the first steps of which we describe in this paper. Data from the pipeline are made available to the community via the database of the TESS Asteroseismic Science Operations Center (TASOC<sup>15</sup>) and the Mikulski Archive for Space Telescopes (MAST<sup>16</sup>). This collaborative effort is building on the combined experience in the TASC community gained from previous missions, various types of oscillating stars, and different analysis methods to process the TESS data in a way that yields data optimized for detailed scientific investigations in all subfields of asteroseismology.

We note that while the “TASOC pipeline” is developed with a focus on asteroseismology, the data will also be useful for, e.g., exoplanet science, eclipsing binaries, and the study of stellar activity (rotation, flares, etc.).

## 2. The TASOC Pipeline

The TASOC pipeline is developed by the T<sup>2</sup>DA group primarily to serve the asteroseismic community of TASC. Specifically, T<sup>2</sup>DA is responsible for delivering (1) raw photometric time series from FFIs (1800 s cadence) and TPFs (120 s cadence), (2) light curves corrected for systematic signals, ready for asteroseismic analysis, and (3) a classification of the variability/oscillation/pulsation type of each light curve from (1) to (2)—enabling each of the working groups of TASC, which are organized according to stellar type, to better target their analysis. We note that from TESS Cycle 3 onwards a new FFI cadence of 600 s will be adopted, and a new 20 s cadence data product will be introduced. The adoption of these new cadences poses no challenge for the pipeline.

The full TASOC pipeline is open-source and available on GitHub.<sup>17</sup> All data products from the pipeline are available via the TASOC database and as a “high-level science product” (HLSP) on MAST; see Section 7 for more details.

The first step in the TASOC pipeline is the photometric extraction of light curves for every target that falls on a TESS FFI or TPF, down to a TESS magnitude ( $T_{\text{mag}}$ ) of 15—this limit is set to reduce the processing time and because targets of interest to asteroseismology are typically brighter but could easily be changed if fainter targets are requested. This first step will be the focus of this paper. The following steps in the pipeline include the correction of systematics, described in Lund et al. (2021), and stellar classification, described in Audenaert et al. (2021). Figure 1 gives a general overview of

<sup>15</sup> <https://tasoc.dk>

<sup>16</sup> <https://archive.stsci.edu/hlsp/tasoc>

<sup>17</sup> <https://github.com/tasoc>

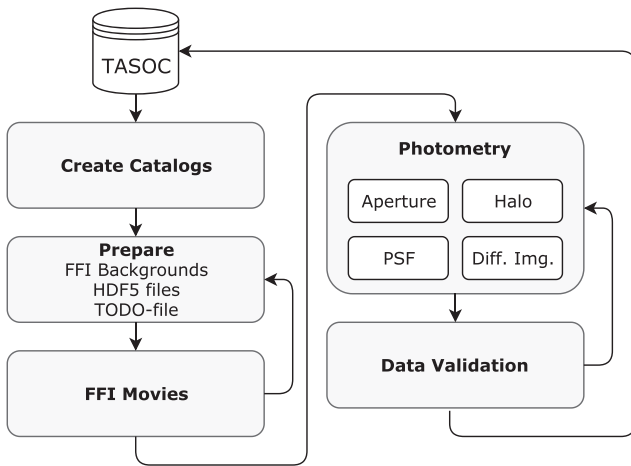


Figure 2. Structure of the TASOC Photometry pipeline.

the TASOC pipeline and the different data products produced. The red dashed line encloses the component parts described in this paper.

One very important design goal of the whole TASOC pipeline is that *all processing must be achievable in no more than 27 days of runtime*. This is due to the simple fact that the length of a single TESS observing Sector is 27 days, and we need to be able to process one Sector’s worth of data fully before the next Sector arrives. This necessitates the constraint that all the methods employed should not be unnecessarily computationally intensive. In any case, the whole pipeline has intentionally been constructed to be easily scaled up to larger computer systems should the need arise and more computational-intensive methods have to be employed. In that case, to limit total processing time the underlying computer hardware could simply be expanded without significant changes to the pipeline described here.

### 3. The Photometry Pipeline

The goal of the photometry pipeline is to extract light curves for all targets observed by TESS down to  $T_{\text{mag}} \leq 15$ . TESS is producing pixel-level data in two distinct modes: FFIs and TPFs. The first, as the name suggests, are the full  $2048 \times 2048$  pixel images for each of the 16 CCDs, stacked to 1800 s cadence (the base integration of data in TESS is 2 s). The TPFs consist of small postage stamps around preselected stars, usually only a few tens of pixels on each side, stacked to 120 s cadence. The basic input for the TASOC photometry pipeline are the calibrated TESS FFIs and TPFs, which are FITS (Flexible Image Transport System) files containing the individual pixels collected by TESS as a function of time. The basic pixel-level calibrations (e.g., dark, smear, flat field) are done by the TESS Science Processing Operations Center (SPOC; Jenkins et al. 2016), which means that each pixel we use as input has units of counts per second and therefore instrumental effects at the detector level will ideally have been sufficiently calibrated. They also contain a World Coordinate System (WCS) solution, which maps the pixel coordinates onto sky coordinates (R.A. and decl.). In all our processing we assume that these calibrations are sufficient, with the notable exception of the very brightest stars (see Section 4).

The TASOC Photometry pipeline is split into a few separate components as shown in Figure 2, which are run in succession. The three main components are “prepare”, “photometry”, and

“data validation”. We will go into the details of each of these in the following sections, but in general terms the process can be outlined as follows:

*Prepare*—The first step in the processing is to prepare the basic input files in such a way that the following photometric extraction of individual targets can be easily parallelized. The prime objective of this step is to perform all operations that need to have the information from the full FFIs available, e.g., estimate the background flux and flag problematic pixels (see Section 3.2) and to restructure the FITS files into a set of HDF5 files (Hierarchical Data Format). Among other things, the HDF5 file format allows us to extract arbitrary subsections of the FFI without having to load the entire FFIs into memory.

*Photometry*—The primary task is then to extract the brightness as a function of time for each star in the FFIs and TPFs, together with other relevant quantities, from the pixel-level data. This can be done in many different ways, several of which are implemented in the pipeline. The method that is used for the majority of stars in the TASOC pipeline is simple aperture photometry (SAP), but other methods are also used in special cases. The setup also allows us to easily add additional methods in the future. See Section 5 for details.

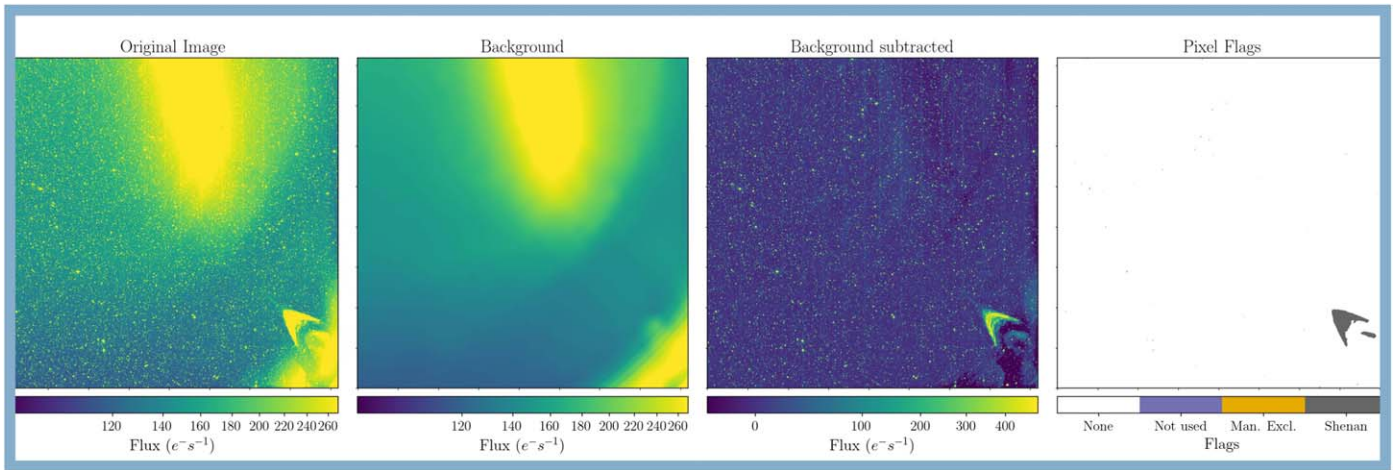
*Data Validation*—The objective of the data validation module is to check the final results of the photometry and the generated light curves. Various diagnostic metrics are checked to filter out erroneous light curves, so these will not be propagated to the following steps of the overall TASOC pipeline or released to the community. It will also flag spurious light curves, which are deemed to be of sufficient quality to avoid rejection but require warning users that potential issues may exist. See Section 8 for details.

#### 3.1. Catalogs and TODO-files

In TESS (unlike *Kepler/K2*) we can safely assume that we have a very good catalog of all stars of interest in the sky, due to TESS targets usually being much brighter. This catalog is provided in the form of the TESS Input Catalog (TIC; Oelkers et al. 2016; Stassun et al. 2018, 2019), which in its most recent incarnation (TIC-8) is based on the Gaia DR2 (Gaia Collaboration et al. 2018) and 2MASS (Skrutskie et al. 2006) catalogs. TASOC is storing a full copy of the TIC in a dedicated PostgreSQL database.

The first step in the photometry pipeline is to create full catalogs of all stars known to fall on or near the TESS detectors during a given observing Sector, to be used in the following steps in the pipeline. These catalogs are copied directly from the TIC but will only contain targets relevant for the observing Sector and CCD in question, and will be stored into local SQLite files. SQLite is a lightweight relational database management system that has the advantage of not requiring a database-server engine running but is simply embedded into a file system. Storing the catalogs into SQLite files gives the advantage that the program can run offline from the main TASOC database systems and will not be dependent on the response time of such a server or the network connecting to it.

In order to generate a catalog file for a given TESS observing Sector and CCD, we query the TIC using Q3C (Koposov & Bartunov 2006) for all targets falling within the footprint of the CCD with an additional  $0.1^\circ$  buffer around. We filter out any targets marked, e.g., as duplicates, and extract only the most relevant information to be stored into the SQLite files. This includes target position in R.A. and decl. (J2000), proper



**Figure 3.** Background extraction in FFIs, here for Sector 1, Camera 1, CCD 2. Panel 1 (leftmost): the raw calibrated FFI image. Panel 2: the estimated background including both the broad background and the corner glow. Panel 3: the background-corrected image. Panel 4 (rightmost): identification and flagging of the residual lens-flare “wineglass” feature in the background-corrected image. An animated version of this figure is available online, showing the change of the background during the full Sector [duration: 01:25].

(An animation of this figure is available.)

motions, TESS magnitudes ( $T_{\text{mag}}$ ), and effective temperature ( $T_{\text{eff}}$ ). We furthermore calculate the positions of the stars at the reference time of the Sector (usually around the apogee of the first orbit of the Sector) by projecting the coordinates forward in time with the known proper motions. These coordinates are the ones used in the rest of the pipeline as the reference coordinates of the targets. For one full TESS observing Sector a total of 16 catalog files are generated, one for each CCD.

Once we have created full catalogs of all targets, we can then create the list of targets to be processed in the photometry, as well as the following steps in the full TASOC pipeline. This file is called the TODO-file and, like the catalogs, this is also stored in an SQLite file. In this initial step, only the primary table of targets to be processed will be created, but at each step in the TASOC pipeline, more tables will be added to the TODO-file, containing results and diagnostic information. They are therefore the central repository for information in the TASOC pipeline.

For FFIs this means simply taking all targets that are brighter than the set threshold  $T_{\text{mag}} < 15$  and adding them to the TODO-list if they fall within the FOV of one of the cameras at the reference time of the Sector. For this we use the WCS transformation provided in the header of the FFIs, including the Simple Imaging Polynomial (SIP) image distortion correction.

For each TPF we add the central primary target to the list but furthermore also search for secondary targets, identified as other targets brighter than the cutoff that fall within the observed pixels in the postage stamp. If any such targets are found, these are added to the list. After all TPFs have been processed in this way, we filter the list by removing any secondary targets, which were also identified as primary targets. We also check secondary targets that were found to be present in more than one TPF, only keeping the one that is the farthest distance from the postage-stamp edge, because this is more likely to have captured more of the flux from the secondary target and not be truncated by the edge.

### 3.2. Background Estimation and Quality Flags

While FFIs provided by the TESS mission have been calibrated, nothing has been done to remove sky background

contributions to the images, originating from, e.g., scattered light from Earth, the Moon, other solar system objects, and zodiacal light. For FFIs, the first step in the processing is therefore to estimate the sky background level in each image. The challenge here is to estimate the changing large-scale background structures without including effects from variable stars (see Figure 3).

The first step is to mask out pixels in each image that are invalid or where very bright stars are present. Currently this is done simply by manually excluding regions of pixels with known problems (e.g., Mars in Sector 1) and making a threshold in the flux above which pixels are masked out. We do have plans in the future to use the known locations of bright stars from the catalog to improve the masking process, but this is not yet fully implemented.

The full  $2048 \times 2048$  pixel image is then cut into  $64 \times 64$  pixel tiles, wherein a  $3\sigma$  clipping from the median value is performed to further mask out stars. The mode of the image is then estimated using the same prescription as is used by SExtractor (SE; Bertin & Arnouts 1996):

$$\text{mode} = \begin{cases} 2.5 \cdot \text{median} - 1.5 \cdot \text{mean}, & \frac{\text{mean} - \text{median}}{\text{std}} \leq 0.3 \\ \text{median}, & \text{otherwise.} \end{cases} \quad (1)$$

These values are then run through a  $3 \times 3$  moving median filter and interpolated back onto the full image size. Other methods for estimating the sky backgrounds were also tested, and we concluded that the method described above performed the best in terms of both recovering the background and computational speed. See Appendix A for more details.

Another dominant feature of the TESS FFIs is the so-called “corner glow” (Vanderspek et al. 2018), which is believed to be caused by stray light from stars outside the camera field of view (FOV). The effect manifests itself in the four corners of the camera FOV farthest away from the camera center (in one corner per CCD) as a rise in the background level of  $\sim 1$ – $2$  times the overall sky background (see Figure 3). As this feature is only present in the corner of the CCD and seems primarily to

be a function of the radial distance from the camera center, the normal method using square tiles to estimate the background is suboptimal in capturing the corner glow. We mitigate this by reparameterizing the image by calculating the Euclidean distance from each pixel to the camera center. We then bin the pixels into 15 pixel wide bins starting from a distance of 2400 pixels and upwards, and in each bin we estimate the background in the same way as described above. This effectively creates “radial tiles” near the corner, where the background is estimated from pixels having the same radial distance. The bins are then interpolated back onto the full image and we are left with a background profile that rises smoothly toward the corner.

Our final estimate of the background for a single FFI will be a combination of the square tiles and radial components. We iteratively isolate the two components and create the final background from their sum. For a given FFI,  $\mathcal{I}$ , the two background components are iterated three times, first estimating the radial component,  $\mathcal{B}_{\text{radial}}$ , from  $\mathcal{I} - \mathcal{B}_{\text{square}}$ , and then estimating the square component,  $\mathcal{B}_{\text{square}}$ , from  $\mathcal{I} - \mathcal{B}_{\text{radial}}$ . The final background is  $\mathcal{B} = \mathcal{B}_{\text{square}} + \mathcal{B}_{\text{radial}}$ . The contribution from the radial component, compared to the pure square component, varies across the Sectors with the level of scattered light, raising the average estimated background level in the corners on the order of a couple of percent. In cases with additional features near the corners (e.g., in Figure 3) the largest differences can go up to  $\sim 25\%$ , due to the improved separation of the radial rise and nonradial contributions.

Once the background has been estimated for all images for a given CCD, the final background images are constructed where the estimated background for each pixel is smoothed in time using a three-point (1.5 hr) moving average filter. This avoids very rapid changes in the background, which could introduce high-frequency noise in the final photometry. At the same time, we construct an average image for each CCD, which is the co-added images after subtracting the final background, divided by the number of images. This sum image is used in some of the following steps for locating stars and is also included in the final output files.

Once we have constructed the background-subtracted FFIs, we also want to flag any areas of the images affected by additional features that are too sharp to be captured by the background estimation described above. This includes, e.g., the lens-flares seen in Sector 1 because of Mars entering the FOV of camera 4 (see Figure 3). These features move across the images and will impact different locations of the focal plane at different times. The goal is therefore to flag these features in each image by locating any pixels affected and propagate this information to the final light curves.

After estimating the backgrounds and average image as described above, we subtract the average image from each individual background-subtracted FFI, creating a difference image showing the departures from the “norm”. To avoid flagging variable pixels with small spatial scale structures, e.g., variable stars, the difference image is run through a  $15 \times 15$  pixel median filter. Any pixels departing more than a set threshold (nominally  $\sim 40$  counts per second) are flagged as irregular. As the final outcome of this process we produce a set of pixel-level flag images, which for each pixel and for each time stamp contains flags indicating the state of that pixel. This includes if the pixel was manually excluded, otherwise not used in the estimation of the background, or if it has been flagged using the procedure

described above. These pixel flag images are also stored internally in the HDF5 files for the given camera.

The next step in the “prepare” module is to do a matching between the data quality flags provided by the TESS team in the TPFs and FFIs (see Twicken et al. 2020) because the quality flags of the FFIs are generally not populated. For each CCD a small number of TPFs belonging to that CCD are randomly chosen, and their quality flags are binned onto the FFI cadence by enabling the flag if it is set in any of the corresponding TPF cadences. We then compare these quality flags from the TPFs with the ones provided with the FFIs, only including flags relevant for all targets (e.g., spacecraft attitude tweaks, course pointing, momentum dumps, etc.), and propagate any missing flags to the FFIs. We do not include the “Manual Exclude” flag, because a single 120 s data point marked as bad does not necessarily mean that the combined 1800 s FFI should be rejected. We also found that including this would discard  $\sim 20\%$  of all data in Sector 1. This step is included to ensure that quality flags present in the TPFs are propagated into the lower-level FFIs. In most cases, particularly in later Sectors, this has no effect on the final quality stamps, as the FFI quality stamps have been populated correctly.

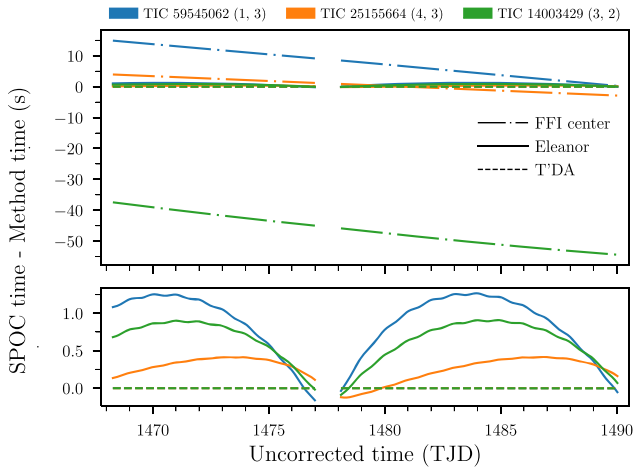
The final step in the “prepare” module is to create movies as a function of time of the original FFI images, estimated backgrounds, image-subtracted FFIs, and extracted pixel-level flags, as shown in Figure 3. These are used for visual inspection of the performance of the background estimation and the pixel-level flags. In cases where individual pixels need to be removed due to instrumental effects not identified in the automatic processing, these would be added and the “prepare” module would be rerun, using the new manually excluded pixels.

### 3.3. Time Stamps of FFIs

The FFIs provided by TESS on the MAST website all contain a time stamp in the header of each image. These time stamps are provided in TESS Barycentric Julian Date (TBJD) and might naively be anticipated to be of sufficient quality to use in any light curves that one would extract from the FFIs. However, these time stamps have been calculated for a fixed position in the center of the images, but because the barycentric correction depends on the exact coordinates of the target in question and each camera spans several degrees on sky, these time stamps actually need to be reprojected to the proper coordinates of the star being extracted.

We start by reverting the barycentric correction of the provided FFI time stamps back to Julian Dates in TDB (Barycentric Dynamical Time) in the TESS frame of reference. We then use SPICE kernels provided by the TESS Team to calculate the accurate position of the TESS spacecraft at the time of observation and calculate the light-travel time correction resulting from the different light paths from the position of the target to the TESS spacecraft versus the solar system barycenter. We have compared the new time stamps generated in this way with the time stamps given in the TPFs provided by the SPOC. In all cases we arrive at time stamps within less than a millisecond from the ones in the TPFs. For TPFs we do not use this procedure, as these already have time stamps calculated for the target in question.

Figure 4 gives a comparison between the times computed using different reference points for three targets in Sector 6 at different locations in the TESS FOV against the times provided by SPOC in the TPF data. The “FFI center” method refers to the times



**Figure 4.** Comparison between time stamps provided in SPOC TPF data against times obtained from corresponding FFI data using different reference points for the time calculation. Differences are plotted against the uncorrected times. The comparison is shown for three targets observed in Sector 6 with different colors according to the top legend. The numbers in parentheses after the TIC ID indicate the position of the star in the FOV in terms of “camera” and “CCD”. The targets were selected to cover a large portion of the FOV. Line styles indicate the method used for the calculation of times (see Section 3.3). The bottom panel provides a zoomed-in view of the y-range close to zero to better show the time differences obtained by the method adopted in this work and by the *Eleanor* pipeline (Feinstein et al. 2019).

provided in the FFI files, and used, e.g., by *TESSCut* (Brasseur et al. 2019) (which in turn is used internally by several other photometry pipelines), where the times can be off by several tens of seconds. Shown is also the reference used in *Eleanor* (Feinstein et al. 2019), where the reference point is defined to be the Greenwich observatory—this reduces the error on the times, but the resulting differences now have oscillations on timescales of both the *TESS* orbit and the Earth rotation (Greenwich is positioned on the Earth surface, hence periodically varying its distance to the observed stars). As mentioned above, the differences obtained in the *TASOC* pipeline from the *SPICE* kernels are consistently at the sub-millisecond level.

As can be seen in Figure 4, the timing differences are generally within a few seconds for *Eleanor* and can be on the order of minutes for *TESSCut*. Such offsets will likely have only a small effect on scientific analysis for most targets. However, in some cases, such as studies relying on the phase stability for high-amplitude coherent oscillators or transit timing variations for exoplanets, the analysis could be affected by such offsets (von Essen et al. 2020), which in the case of *Eleanor* would also have periodic components besides the yearly periodicity. We note also that any potential impact will be increased with the reduction in the FFI cadence from 1800 s to 600 s in the extended mission.

From *TASOC* data release 5 and onward, the time stamps for both FFIs and TPFs are corrected for the known constant time offsets (see Fausnaugh et al. 2019, 2020a, 2020b), including earlier *TESS* data where this was not yet taken into account (before *TESS* data release 30). This includes effects from both “staggered readout” (FFIs) and detailed correction to start and end times (FFIs and TPFs).

#### 4. Very Bright Stars

The brightest stars observed by *TESS* provide unique challenges. Stars brighter than a  $T_{\text{mag}} \sim 6$  saturate the detector,

with excess charge spilling along the CCD columns. As these bleed columns become longer for very bright stars, larger target pixel files are required to capture the flux. Problems can be encountered when the bleed columns become excessively long.

The most significant problem arises when the bleed column extends beyond the region covered by the TPF. This can arise either because the stamp selected for the target was too small, or because the star is too close to the edge of the CCD. In these cases, the light curves cannot be constructed by simple aperture photometry. However, they may still be constructed from the “halo” photometry of scattered light and the wings of the PSF (Section 5.2).

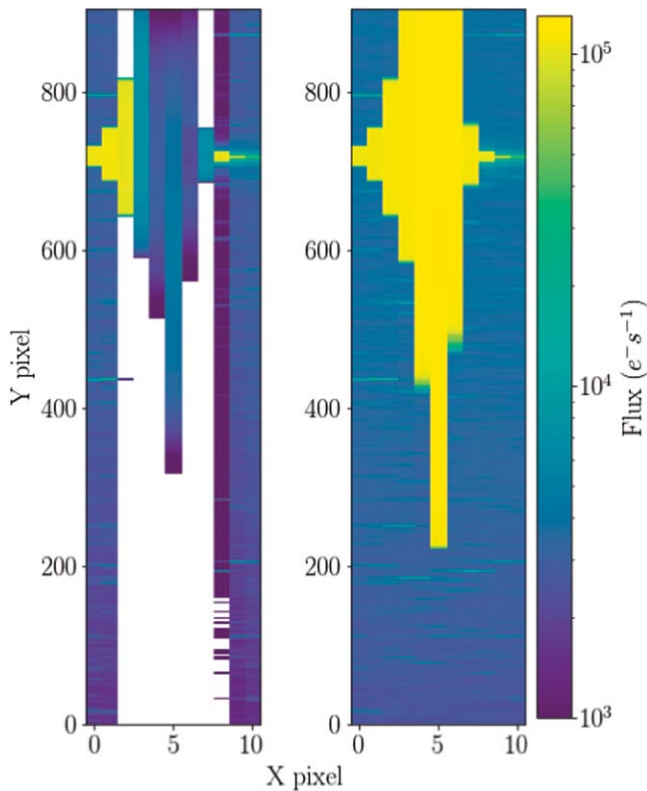
For particularly bright stars near the edge of the CCD, the saturation can cause calibration problems. The smear correction, intended to remove the photometric smear that occurs during the shutterless readout, is calculated as the robust mean across virtual rows 2059–2068. If these pixels become saturated, as has happened with, e.g., *Procyon* and  $\alpha$  Cen, the smear can be significantly overestimated along these columns. To rectify this, the TPF needs to be recalibrated with a better smear correction. The recalibration is done starting from the raw counts. In addition to the smear correction, it also differs from the pixel-level calibration applied in the SPOC pipeline by including an undershoot correction. All other calibration steps are performed as normal (see Vanderspek et al. 2018; Clarke et al. 2020).

The new smear correction is estimated from the science pixels themselves, selecting background pixels that do not contain light from any stars, and calculating a robust mean for each column from these pixels. Because this includes both smear and background flux, we estimate the background from the difference between this and the standard smear correction for columns unaffected by saturation and remove it from our new smear correction. This new smear correction can then be applied.

Undershoot is a signal distortion that appears as a slight depression of the signal in a pixel that is read out after a bright pixel. This is generally a small correction for *TESS* and is subsequently not applied in the SPOC pipeline (see Vanderspek et al. 2018). However, because it is most significant for light-to-dark transitions as those that occur around bright stars, we have included the correction in our recalibration.

Figure 5 shows a single frame from the Sector 11 TPF of  $\alpha$  Cen before and after this recalibration procedure has been applied. For stars such as  $\alpha$  Cen, or any star that happens to share a column with these stars, this procedure makes previously unusable data viable. To compare how well this procedure performs relative to the standard calibration of the SPOC pipeline, we can apply it to a bright star that is not affected by an overestimated smear correction, such as  $\beta$  Hyi in Sector 1, and see how the flux of each pixel at each cadence differs. Despite the smear and undershoot calibration differences, the median ratio of the recalibrated flux to the SPOC flux for all pixels in the Sector 1 TPF of  $\beta$  Hyi is 0.999, with a 16th percentile of 0.986 and an 84th percentile of 1.006.

An additional issue for bright stars is that in each Sector the brightest stars (typically with  $T_{\text{mag}} \lesssim 2$ ) with large pixel stamps are not processed by the photometric module of the SPOC pipeline. As a result the TPFs provided by SPOC are without flux uncertainties, background fluxes, or an aperture calculated or even basic pixel calibrations (Sectors 1 and 2 only). Because of this, these targets are currently not processed



**Figure 5.**  $\alpha$  Centauri ( $T_{\text{mag}} = -0.1$ ) as observed by TESS in Sector 11. Left: original flux with improper smear-column correction. Right: recovered flux after recalibration.

by the TASOC photometry pipeline either. We are working on the implementation of basic pixel-level calibrations on these targets, enabling high-quality photometry for these as well.

Over the large photometric apertures used for bright stars, the background calculated from the SPOC photometric pipeline can be a poor match for the spatially varying background, because it is uniform across the image. This can have the effect of introducing spurious signals, which can be especially severe in the case of halo photometry (Section 5.2), where the fainter parts of the PSF can be of comparable flux density to the background (Eisner et al. 2019). We are also considering if the background calculated from the SPOC photometric pipeline is appropriate for less bright stars in terms of capturing spatial and temporal variations. Therefore, we have experimented with allowing for more spatial structure in the background used for TPF files, which is also allowed to vary in time. The strategy is to utilize the full background structure estimated from the FFIs (Section 3.2), interpolating it onto the higher temporal cadence, and then scale it to have the same mean as estimated from the background pixels in the TPF. Thereby, we can utilize the structural information from the FFIs, combined with the temporal resolution from the TPFs.

We found that this helped partially remove spurious signals from scattered light in some cases, but was not sufficient in other cases. To improve this we also added the possibility of additionally fitting a low-order 2D polynomial light map on top of the FFI structure, which of course comes at the risk of overfitting. Additional work is still needed to find the optimal strategy, and this feature is not yet enabled for the current data releases.

## 5. Photometric Methods

The central module in the pipeline is the “Photometry” module, which is responsible for the extraction of light curves from the FFIs and TPFs, after the initial preparations and restructuring of data. The pipeline is built in a modular fashion, so different methods for photometry can be implemented into the same framework without duplicating functionality that is of a more general concern to CCD photometry. This is implemented by having a general class, `BasePhotometry`, which defines the general concept of “a photometry of a single target”, that implements all common functionality between the photometric methods, described in the following sections. All specific photometric methods inherit all methods and properties from the `BasePhotometry` class, thus defining an abstraction layer between the data and the photometric methods. In this way, the individual methods do not need to worry about low-level data input/output but can focus on the task of extracting the brightness of the target star as a function of time from the provided postage-stamp images provided by this underlying system.

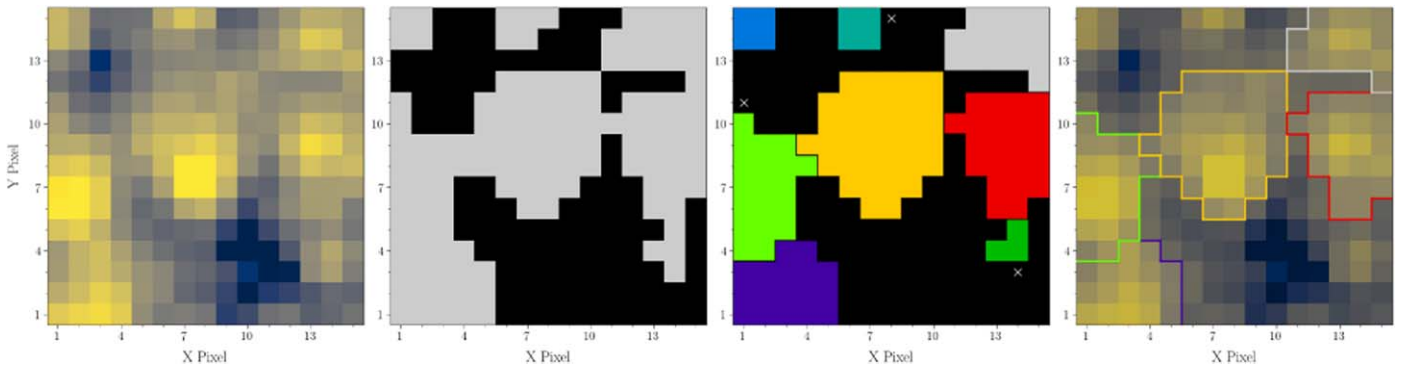
When the photometry of a given target is started, the underlying `BasePhotometry` class will automatically create a “postage-stamp” cutout of the FFI or TPF in question, centered on the target, and load the necessary data into memory. The size of the stamp is scaled by the  $T_{\text{mag}}$  of the target, creating larger stamps for bright targets (potentially thousands of pixels, as can be seen in Figure 5), and scaling down to a minimum stamp size of  $11 \times 11$  pixels for fainter targets (see Figure 13).

`BasePhotometry` provides functionality for accessing the postage-stamp images, uncertainty images, extracted background images, and average image in several ways, always returning only the pixels within the postage stamp, i.e., without having to load the entire FFI into memory. It also defines functionality for changing the size of the postage stamp, should any photometric method require it.

An interface for obtaining lists of all stars that are present within the current postage stamp is also available. This is created using the WCS solution to map the footprint of the postage stamp onto the sky and querying the catalog files (see Section 3.1) for all targets that fall within these coordinates (plus/minus a small buffer) at the time of observation. The resulting list contains both the TIC identifiers, sky and pixel coordinates, and the  $T_{\text{mag}}$  of the targets. Furthermore, it can provide the same catalog of targets where the pixel coordinates take into account the spacecraft pointing jitter, which depends on the time and position on the FOV, meaning the catalog is time dependent. We will return to this in Section 5.3.

When starting the photometric reduction, `BasePhotometry` will also define a light curve table, which will contain the time stamps (see Section 3.3), quality flags, cadence numbers, and any other relevant information corresponding to each postage-stamp image as a function of time. Columns for measured flux and corresponding uncertainty are also created, but initially left empty. It is then up to the specific photometric algorithms to populate these columns of the table.

Once the photometry has populated the light curve table with measured flux and uncertainties, `BasePhotometry` also provides the functionality for saving the extracted light curves to FITS files, with all the appropriate headers and meta-information. The format of the output files will be described in Section 7. At the same time, a set of diagnostic information



**Figure 6.** Example of apertures set using the described procedures for a target in a relatively crowded region, here for the star TIC 410446575 in Sector 3 (central yellow aperture). Going from the left, the first panel shows the summed image of the pixel cutout used for the star; the second panel gives in gray the pixels with flux levels above the adopted threshold; the third panel shows the apertures found from applying the clustering and watershed algorithms, and pixels identified as noise are marked with a white “x”; and the fourth panel shows the outlines of the final aperture for the cutout after discarding apertures with 4 pixels or less.

about the extracted light curve is calculated and stored into the TODO-file for use in the following steps of the pipeline. This includes, for instance, the median flux level, rms- and point-to-point noise, and a variability metric (see Lund et al. 2021, for details).

### 5.1. Aperture Photometry

The primary method for extracting light curves in the pipeline is simple aperture photometry, where the total flux of the target is summed up within a pixel mask containing, ideally, only the target in question. The approach taken builds heavily on the K2P<sup>2</sup> (Lund et al. 2015; Handberg & Lund 2017) method, which was developed for the K2 mission, but with a few modifications.

The method defines apertures by means of DBSCAN (Density-Based Spatial Clustering of Applications with Noise; Ester et al. 1996), as implemented in `scikit-learn` (Pedregosa et al. 2011). This is applied to pixels with a flux level above a given threshold, given by the mode plus 0.8 times the standardized MAD of the flux in the pixel stamp. A constraint of a minimum of 4 pixels is placed on the size of the clusters such that, e.g., single high-flux pixels are discarded as being noise. In the case where several distinct clusters are identified, the position of the target in question is used to identify to which pixel cluster the target belongs.

After the identification of pixel clusters, the watershed image segmentation from van der Walt et al. (2014) is run to split clusters that may contain two or more close targets. In a departure from the approach of Lund et al. (2015), we ensure that all basins used for the watershed, which are found from a local peak-finding algorithm, can be associated with a target from the TIC that lies within the pixel stamp. In the end, the set of pixels from the clustering and watershed segmentation that contain the position of the target star is returned as the aperture to use for the photometric extraction. If no pixels are found above the flux threshold, or if the known position of the target star does not lie within any of the identified sets of pixels, a minimum  $2 \times 2$  pixel aperture is returned at the target position and a warning flag is set (see Table 2). Figure 6 provides two examples of the definition of apertures, one for an isolated star and one for a star in a crowded field. As seen, the clustering combined with the image segmentation allows apertures to be defined even in relatively crowded regions. We note that for 120 s data, where the adopted pixel stamp is taken from the TESS mission, we also perform photometric extraction for potential secondary targets found in the stamp. So, while we currently only include primary targets with

$T_{\text{mag}} < 15$  in the processing, some secondary targets in the 120 s stamps will have magnitudes beyond this limit. These additional 120 s targets are most clearly seen as the sparse points in Figures 8–12 beyond  $T_{\text{mag}} = 15$ .

When the aperture belonging to the target star has been identified it is checked if any of the aperture pixels lie at the edge of the pixel stamp (only for 1800 s data). If this is the case, the stamp is extended by 5 pixels in the direction of the edge containing aperture pixels, and the aperture definition is run anew. The initial size of the stamp is determined by the TESS magnitude of the target star, with a relation optimized to return the smallest stamp without the need for resizing (Figure 13). So, while the resizing is allowed to take place five times, this rarely happens. If the resizing limit is reached, an error flag is set for the star.

When the final aperture has been defined, a contamination metric is computed if the catalog indicates that stars other than the target star lie within the aperture. The contamination metric,  $C$ , is given by one minus the flux ratio of the target star to all stars in the aperture, including the target. In practice it is calculated from the sum of catalog magnitudes of stars located within the defined aperture:

$$C = 1 - \frac{f_{\text{target}}}{f_{\text{total}}} = 1 - 10^{0.4(T_{\text{mag, total}} - T_{\text{mag, target}})}. \quad (2)$$

In this manner the contamination will tend to 1 if the flux from secondary stars equals that of the target star. See Section 8 and Figure 11 for the cuts made in the data validation based on the contamination metric. We note that the metric will not include flux from stars just outside the aperture, so the metric can be seen as a minimum value for the flux contamination. We do not currently attempt to correct the flux level of the target star for crowding effects.

### 5.2. Halo Photometry

For the very brightest stars, saturation produces a bleed column that is so long that it reaches the edge of the CCD chip, or which is otherwise too saturated to include in a photometric aperture. For these stars we employ the halo photometry method (`halophot`: White et al. 2017; Pope et al. 2019), which constructs a calibrated light curve from unsaturated pixels in the deep PSF (the “halo”).

The light curve is constructed as a linear combination of the time series of the individual unsaturated pixels in the vicinity of

the bright star. The weights are determined by gradient-descent optimization to minimize an objective function of the resulting light curve, the generalized “lagged total variation” (Pope et al. 2019). This is a convex objective function (“TV-min”), and has analytic derivatives which can be computed with automatic differentiation, for which we use `autograd` (Maclaurin et al. 2015).

`halophot` has been shown on K2 data to produce light curves with comparable noise to conventionally observed unsaturated stars, without significant loss of sensitivity to planets or stellar variability (Pope et al. 2019).

In our TESS implementation, we first adopt a default circular aperture with a 20 pixel radius centered on the position of the star. We then discard saturated pixels by an iterative process: using a conservative range of estimates of the number of saturated pixels, we apply the TV-min algorithm to pixel sets with one more pixel discarded at a time. Pixels that are saturated contribute very little total variation, and therefore when we have cut out all saturated pixels there is a significant increase in the standard deviation of the resulting light curve. Next, background stars are identified using `DBSCAN`, similarly as done in aperture photometry (Section 5.1), and masked out. We then apply the TV-min algorithm to this censored array of pixels. The resulting light curve is scaled to absolute flux using the relation described in Section 8. The resulting weight maps, indicating the final weights given to each pixel, are stored as an extra extension in the final output FITS files (see Section 7).

### 5.3. PSF Photometry

The TESS PSFs are provided by the TESS team and give the two-dimensional subpixel distribution of light from a point source as measured by the TESS spacecraft. The PSFs were created in flight by microdithered observations of bright isolated stars, resulting in PSFs measured at 25 positions spanning each CCD (see Vanderspek et al. 2018). These PSFs can be linearly interpolated to any position in the FOV and integrated onto the pixel grid to yield the Pixel Response Function (PRF), which is the “pixelated” version of the PSF.

The objective of PSF photometry is to fit this inferred shape of a star to the measured images, assuming a good catalog of stars in the image. In the standard PSF photometry, this means we have three free parameters per star in the image: position in pixels and flux. A nonlinear optimization algorithm is used to optimize all the free parameters by, at each iteration, re-integrate the underlying PSF to a new PRF for each star and match this to the observed image.

However, the TIC (see Section 3.1) contains the positions of stars with very small uncertainties, especially when compared to the 21” pixels of TESS. It is therefore a reasonable assumption that we can fix the positions of the stars for a given image using the TIC. In that case the PSF-fitting problem becomes linear and can be written in the following way:

$$\begin{bmatrix} \vdots & \vdots \\ \text{PRF}_{i,1} & \cdots & \text{PRF}_{i,N} \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} \vdots \\ f_j \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ p_i \\ \vdots \end{bmatrix}, \quad (3)$$

where  $f_j$  is the desired flux of the  $j$ th star,  $p_i$  is the measured flux in pixel  $i$ , and  $\text{PRF}_{i,j}$  is the PRF of the  $j$ th star in pixel  $i$ . This problem can be solved much faster, using standard linear algebra routines, which significantly speeds up the calculation.

As a final step, the flux is adjusted by performing aperture photometry on the flux in the residual image in a small aperture ( $2 \times 2$  pixels) centered on the target star and adding this to the extracted flux. This is inspired by Kjeldsen & Frandsen (1992) and will correct the measured flux for small deviations near the PSF center. Preliminary tests show that this correction can be on the order of  $\sim 1\%$  in the case of a  $T_{\text{mag}} = 7$  star, which is near the saturation limit of the detector.

Both the standard PSF fitting and the linear PSF fitting are implemented in the pipeline but are currently not used. One shortcoming of PSF photometry is its inability to handle bright saturated stars, which often are the ones of most interest to asteroseismology. This is due to the PSF model not being a valid representation of the observed image due to saturated pixels and bleed trails causing deviations unique to the brightness and position of the star on the detector. In future data releases, PSF photometry could be used for nonsaturated stars in crowded regions, where aperture photometry becomes inefficient and prone to errors from contamination.

### 5.4. Difference Imaging

The last method of photometry currently implemented in the TASOC Photometry pipeline is difference imaging, where aperture photometry is performed on the difference between the image in questions and a reference image. The method is well established and has been demonstrated to work on TESS data in crowded fields (see Bouma et al. 2019). The implementation in the pipeline is currently not at a stage where it can be run in a routine manner and is therefore currently not in use. We are, however, planning on including this in future releases for use in crowded fields.

## 6. Parallel Processing

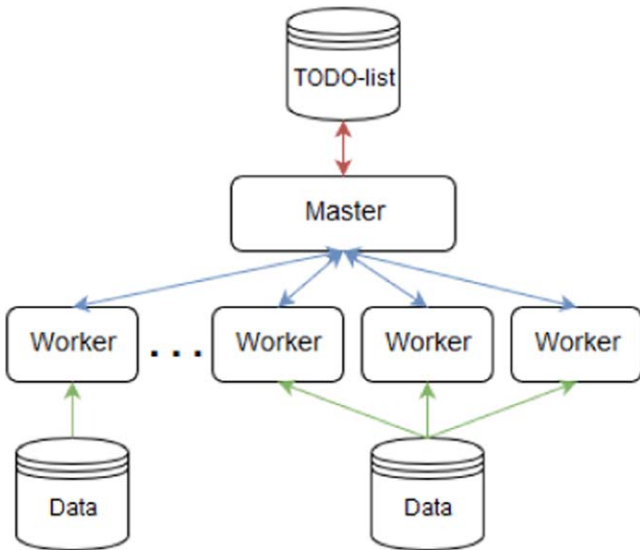
In most of the TASOC pipeline setup, the same basic setup for processing the many tasks in a parallel scalable way is used. The setup is illustrated in Figure 7 and is also used in the corrections and classifications parts of the full TASOC pipeline. This consists of a master process that hands out tasks to workers, which upon completion return the results of that task back to the master. This is an efficient way to perform a large number of independent tasks when there are more tasks than workers, especially when the run times can vary for each task. In practice this is done by utilizing a combination of Python’s multiprocessing module, Message Passing Interface (MPI), and the Slurm workload manager.<sup>18</sup>

Specifically for the photometry part of the pipeline, as discussed in this paper, splitting the work into independent subtasks is done in two different ways:

The parallelization is fairly trivial in the initial “prepare” step, because each FFI can be treated independently in the estimation of backgrounds and pixel-level flags (Section 3.2). Each worker can therefore handle all the loading of the FFI and all calculations independently as a single “task” and the master process only has to receive the output of each worker, perform any time-dependent calculations, and save the results to disk.

For the photometry itself, we define a “task” as being the photometry of a single star. The master is responsible for finding the next task to be processed from the TODO-list (see Section 3.1) and passing this to an idle worker. The worker, upon receiving a task, will perform the required photometry

<sup>18</sup> <https://slurm.schedmd.com>



**Figure 7.** Schematic overview of the computational setup used for the majority of the TASOC Pipeline. The Master process keeps track of the tasks to be processed using the TODO-list and communicates the tasks to the workers. Each worker has direct access to the input data through a shared or distributed file system.

(see Section 5), save the resulting light curve to file (see Section 7), and return the status and additional diagnostic information from the photometry to the master process. Based on the returned information from the worker, together with the information in the TODO-list, the master can make decisions on which task to process next, if any other tasks need to be skipped, or if additional processes need to be added. An example where this is used is if fainter stars are found to fall within the derived aperture of a brighter star. The master will mark this in the TODO-list, so photometry is not subsequently attempted on these fainter stars, because that would result in the same aperture as found for the bright star. Complete records of which stars were skipped, and which brighter stars were responsible, are also stored in the TODO-list.

Using this overall computational setup, we are able to use the power of a distributed computing system, while still being able to dynamically handle dependencies between tasks, like the ones described above. Currently this capability is only moderately exploited but could be used to dynamically change the tasks to be performed on the fly based on the information extracted up until that point, e.g., changing the type of photometry based on the crowding of a particular region of the FOV.

## 7. Light-curve Data Formats

The primary data format for extracted and corrected light curves is FITS (Flexible Image Transport System<sup>19</sup>) and is provided in a compressed Gzip format. A FITS light-curve file produced by T<sup>2</sup>DA and stored on TASOC will be named following the structure:

```
tess{TICID} - s{sector} - {camera}
  - {ccd} - c{cadence}
-dr{data release} - v{version}
-tasoc_lc.fits.gz
```

<sup>19</sup> [https://fits.gsfc.nasa.gov/fits\\_standard.html](https://fits.gsfc.nasa.gov/fits_standard.html)

**Table 1**  
TASOC “QUALITY” Flags

Bit	Value	Description
0	0	All is OK.
1	1	Data point flagged as bad based on the quality flag by TESS team (their bits 1, 2, 4, 8, 32, 64, 128, and/or 4096).
2	2	Manually excluded by the TASOC team.
3	4	Data point has been sigma clipped.
4	8	An additive constant jump correction has been applied.
5	16	An additive linear jump correction has been applied.
6	32	A multiplicative constant jump correction has been applied.
7	64	A multiplicative linear jump correction has been applied.
8	128	Data point has been interpolated.
9	256	Data point has been rejected in processing.

The “TIC ID” (TESS Input Catalog identifier) of the star is zero (pre) padded to 11 digits, the “sector” is zero prepadded to 3 digits, the “cadence” is in seconds and zero prepadded to 4 digits, the “data release” is zero prepadded to 2 digits and refers to the official release of the data from the mission, the “version” is zero prepadded to 2 digits and refers to the TASOC data release (counting from 1). As an example, the star TIC 62483237, observed in Sector 1 on camera 1, CCD 2, in 120 s cadence and part of the first data release and fifth TASOC processing will have the name:

```
tess00062483237 - s001 - 1 - 2 - c0120
-dr01 - v05 - tasoc_lc.fits.gz
```

Note that files released prior to TASOC data release 5 did not include the camera and CCD in the file names.

Each light-curve FITS file has four extensions: a “PRIMARY” header with general information on the star and the observations; a “LIGHTCURVE” table with time, raw flux, corrected flux, etc.; a “SUMIMAGE” with an image given by the time-averaged pixel data; and an “APERTURE” image. The information provided in the FITS file is intended to mimic that provided in the official TESS products—please consult the “TESS Science Data Products Description”<sup>20</sup> for more information.

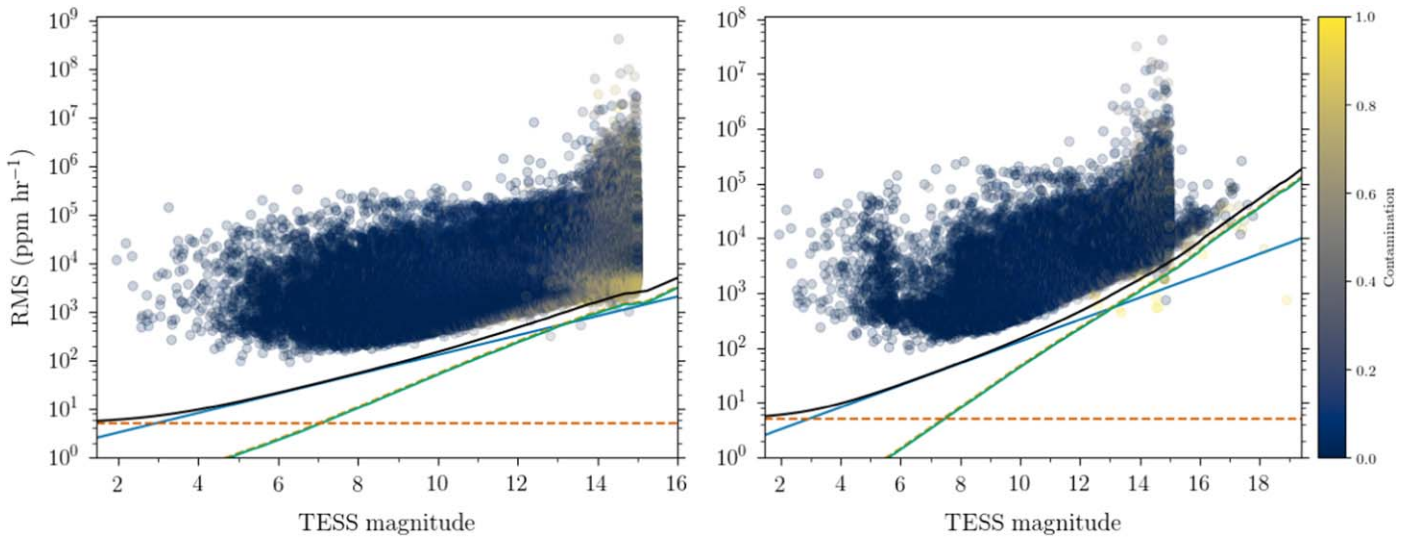
Note, targets processed with the Halo photometry option (see the PHOTMET key in primary FITS header for the adopted photometry method) have the additional extension “WEIGHT-MAP” in their FITS file, which gives the weight assigned to each pixel in the Halo photometry method.

From file version 1.3, additional columns have been added to the “LIGHTCURVE” table containing quality flags. One of these, “PIXEL\_QUALITY”, contains the quality flag provided by the TESS team. For an explanation of the bit values used here see the TESS Archive Manual.<sup>21</sup> The column “QUALITY” gives the quality flags set by the TASOC pipeline, which have the meanings as shown in Table 1.

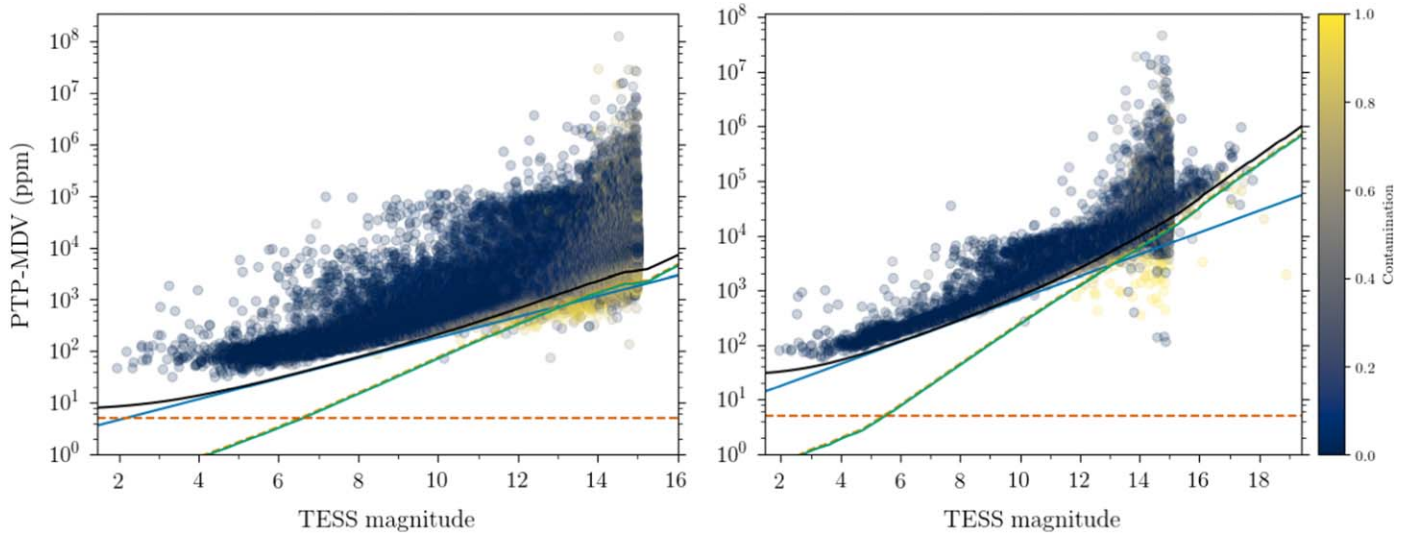
With file version 1.4 a few additional keywords have been added to the “LIGHTCURVE” header. Of particular note is the “XPOSURE” key, which gives the actual exposure of the observations, taking into account dead time from readout and

<sup>20</sup> <https://archive.stsci.edu/missions/tess/doc/EXP-TESS-ARC-ICD-TM-0014.pdf>

<sup>21</sup> <https://outerspace.stsci.edu/display/TESS/2.0+-+Data+Product+Overview>



**Figure 8.** rms noise on 1 hr timescale, here for stars observed in Sector 3. Left: targets extracted from 1800 s FFI data. Right: targets extracted from 120 s TPF data. The lines give the predicted noise estimates following Sullivan et al. (2015; blue full: shot noise; green full: read noise; yellow dashed: zodiacal noise; red dashed: adopted systematic noise; black full: total noise).



**Figure 9.** Point-to-point median-differential variability (MDV), here for stars observed in Sector 3. Left: targets extracted from 1800 s FFI data. Right: targets extracted from 120 s TPF data. The lines give the predicted noise estimates following Sullivan et al. (2015; blue full: shot noise; green full: read noise; yellow dashed: zodiacal noise; red dashed: adopted systematic noise; black full: total noise).

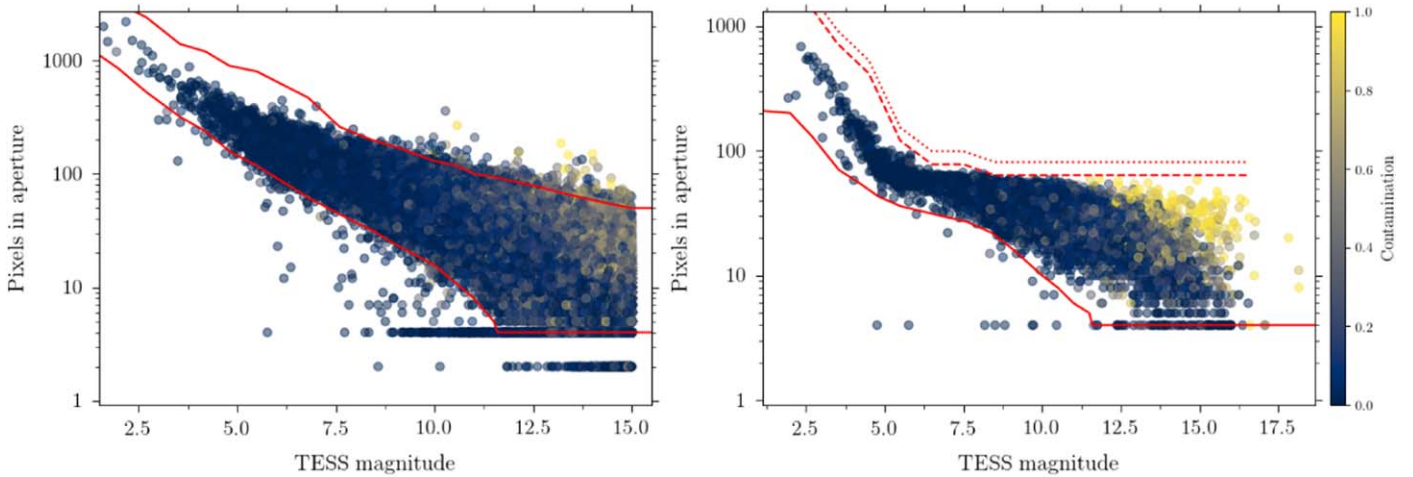
from the cosmic-ray mitigation (see Berta-Thompson et al. 2015; Vanderspek et al. 2018). Using this value for the integration of measured flux will, for instance, be important for the calculation of signal apodization.

With file version 5 onwards, a photometric data validation (DATAVAL) flag has also been added to the “PRIMARY” header. These flags have the following meanings:

Bits marked with a “\*” in Table 2 have been used internally to identify targets to hold back from being released—these targets will be scrutinized further and may be made available with a subsequent release. Therefore, only bits not marked with “\*” will actually appear in the released data. The boundaries used for the flags are given in Figures 8–12. We note that a target may have an aperture of 4 pixels (i.e., the minimum allowed aperture) without being flagged with bit 5, because this bit is only set when the aperture definition has failed in some manner and has defaulted to the minimum  $2 \times 2$ .

**Table 2**  
TASOC “DATAVAL” Flags

Bit	Value	Description
0	0	All is OK.
1	1	Not in use.
2	2	Star has lower flux than given by magnitude relation.
3	4	Not in use.
4	8	Not in use.
5*	16	Star has minimum $2 \times 2$ mask.
6*	32	Star has smaller mask than general relation.
7*	64	Star has larger mask than general relation.
8	128	Not in use.
9	256	PTP-MDV lower than theoretical.
10	512	rms lower than theoretical.
11*	1024	Invalid contamination.
12	2048	Contamination high.
13*	4096	Invalid mean flux.
14*	8192	Invalid noise.



**Figure 10.** Pixel in apertures as a function of TESS magnitude, here for Sector 1 data. The left panels show apertures for 1800 s targets, while the right panels show apertures for 120 s targets. The individual points are color-coded by the contamination. The full red lines give the boundaries for the data validation. The dotted line in the right panel gives the median relation for the maximum allowed number of pixels given the stamp size, while the dashed line shows the corresponding relation under the assumption that the aperture is circular.

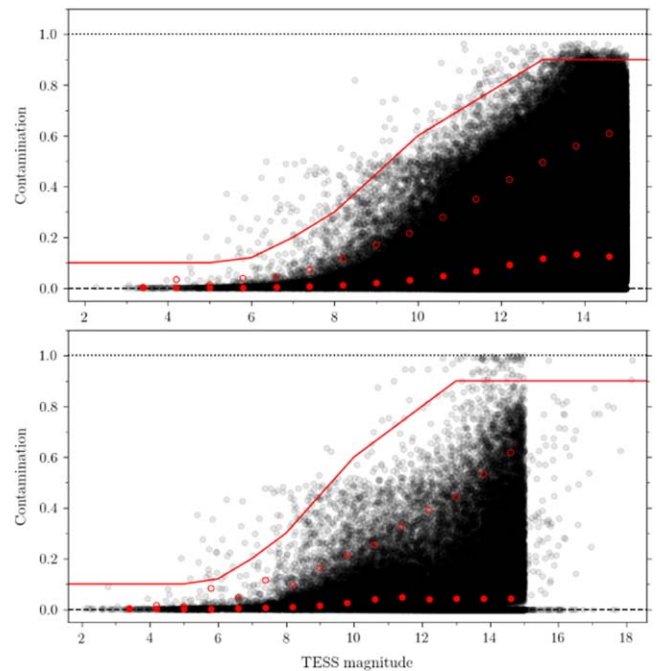
Users should make sure to check the photometric data validation (DATAVAL) flag of any specific star under study, as well as the aperture and sum images.

## 8. Data Validation and Performance

Below we describe in turn the different metrics used in the validation of the extracted photometry before release and for our assessment of the pipeline performance. Information on how the data validation results are reported can be found in Section 7 and Table 2.

*Noise metrics*—The photometric quality of the reduced (raw) light curves is summarized in Figures 8–9. Figure 8 shows the 1 hr rms noise in parts per million (ppm) as a function of Tmag; Figure 9 gives the point-to-point median-differential variability (MDV) (corresponding to rms on the timescale of the observing cadence). For the expected noise curves (Sullivan et al. 2015) we used relations for mean flux (Figure 12) and number of aperture pixels (Figure 10) as a function of TESS magnitude derived from the processed data. As seen the raw photometry generally follows the expected noise characteristics. A comparison of the noise metrics for systematics-corrected data will be provided in Lund et al. (2021).

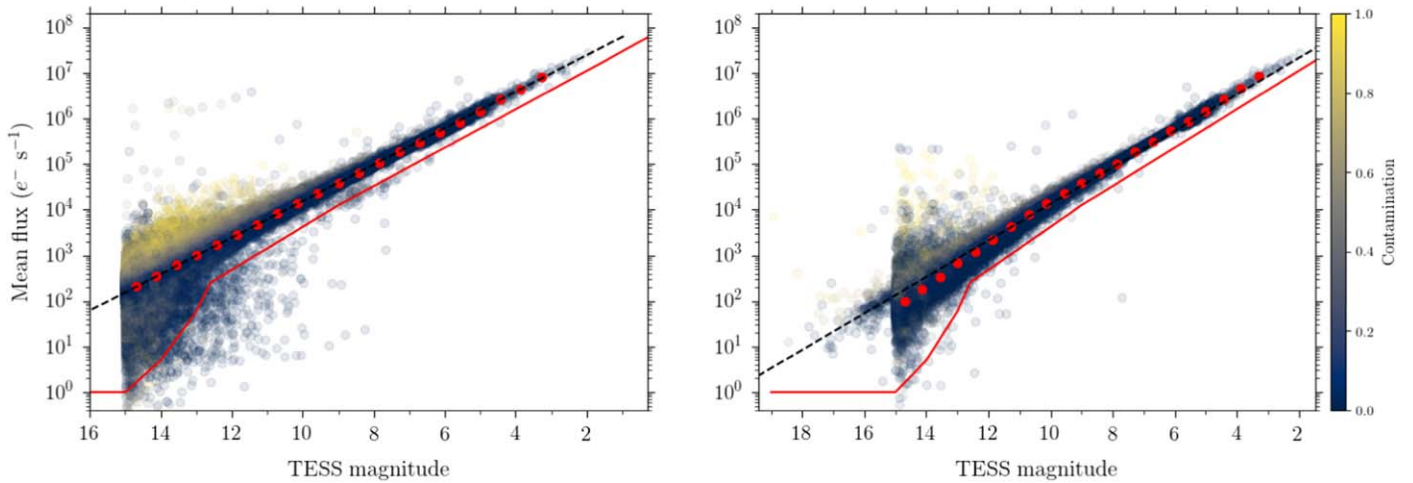
*Aperture size*—Figure 10 shows the sizes of the defined apertures as a function of Tmag. A minimum aperture of 4 pixels has been adopted for the TASOC processing—targets with smaller apertures in Figure 10 are situated on CCD edges and have not been released (see Table 2). The full red lines give the boundaries used in the data validation, defined by visual inspection of the pixel stamps at different magnitudes. For 120 s cadence targets, only a lower bound is used because the upper aperture limit is typically set by the downloaded stamp size. This restriction is clearly seen in Figure 10 (right panel) where the dotted line indicates the maximum number of pixels given the downloaded stamp if requiring that the aperture does not include the edge pixels of the stamp—the dashed line gives the corresponding aperture size assuming the pixel stamp is square and that the aperture is circular. As seen, this latter limit nicely follows the upper boundary of the obtained aperture sizes, implying that in the future the TESS team might consider increasing the size of the stamp selected for 120 s stars to reduce the risk of missing parts of the stellar flux. We caution



**Figure 11.** Contamination metric as a function of Tmag, here for Sector 5 data. The top (bottom) panel gives contamination for 1800 s (120 s) data. The red full curve gives the boundary used in the photometry data validation. The red full points give the median-binned values, while the empty red points give the binned 95th percentile.

also that one should be aware of contamination (see below), especially at high magnitudes. As seen from Figure 10 the faint targets with larger-than-average apertures are typically significantly contaminated.

*Contamination*—Figure 11 shows the contamination metric for each star as a function of Tmag (Section 5.1; provided in the FITS light-curve header as AP\_CONT). We encourage users of the data to keep this value in mind when interpreting signals extracted for a given star—the metric gives the fraction of flux in the light curve contributed from stars other than the main one (see Equation (2)). Note therefore that flux in the aperture from a neighboring star that does not lie within the aperture is not



**Figure 12.** Relation between extracted flux from simple aperture photometry and the TESS magnitude, color-coded by contamination, here for Sector 3 data. The left (right) panels show values for 1800 s (120 s) data. The black dashed line gives the relations obtained following the prescription in Equation (4), which has been fitted to the red median-binned points. The full red line gives the adopted boundary for the data validation.

taken into account. The WCS provided with the “APERTURE” extension in the FITS file can be used to identify which other stars fall within the aperture of the main star.

*Flux relation*—Figure 12 shows the relation between the extracted mean flux for a star and its TESS magnitude. This relation can be described well by the relation:

$$Tmag \approx zp - 2.5 \log_{10}(S), \quad (4)$$

where  $S$  gives the mean flux in  $e^- s^{-1}$ , and  $zp$  is the zero point for the relation. The fit of Equation (4) is performed to median-binned values for a given Sector and cadence (see Figure 12), and considers only targets with a contamination below 0.15. As expected, targets with a flux level above the median relation are typically found to have a significant contamination. From Sectors 1–6 and considering both relations for 1800 s and 120 s data, we find  $zp$  values in the range 20.401–20.506. In general we find slightly lower  $zp$  for 120 s data as compared to 1800 s data, which likely can be attributed to the difference in the background correction, and possibly the fact that the 120 s aperture size is limited by the downloaded pixel stamp (see “aperture size” above). It is also expected that the  $zp$  will vary slightly from Sector to Sector due to differences in the overall crowding, and hence the relative importance of the sky background. Equation (4), adopting a mean value of  $zp = 20.451$ , is used for stars with photometry extracted using the Halo method (Section 5.2), in order to obtain the correct relative amplitudes.

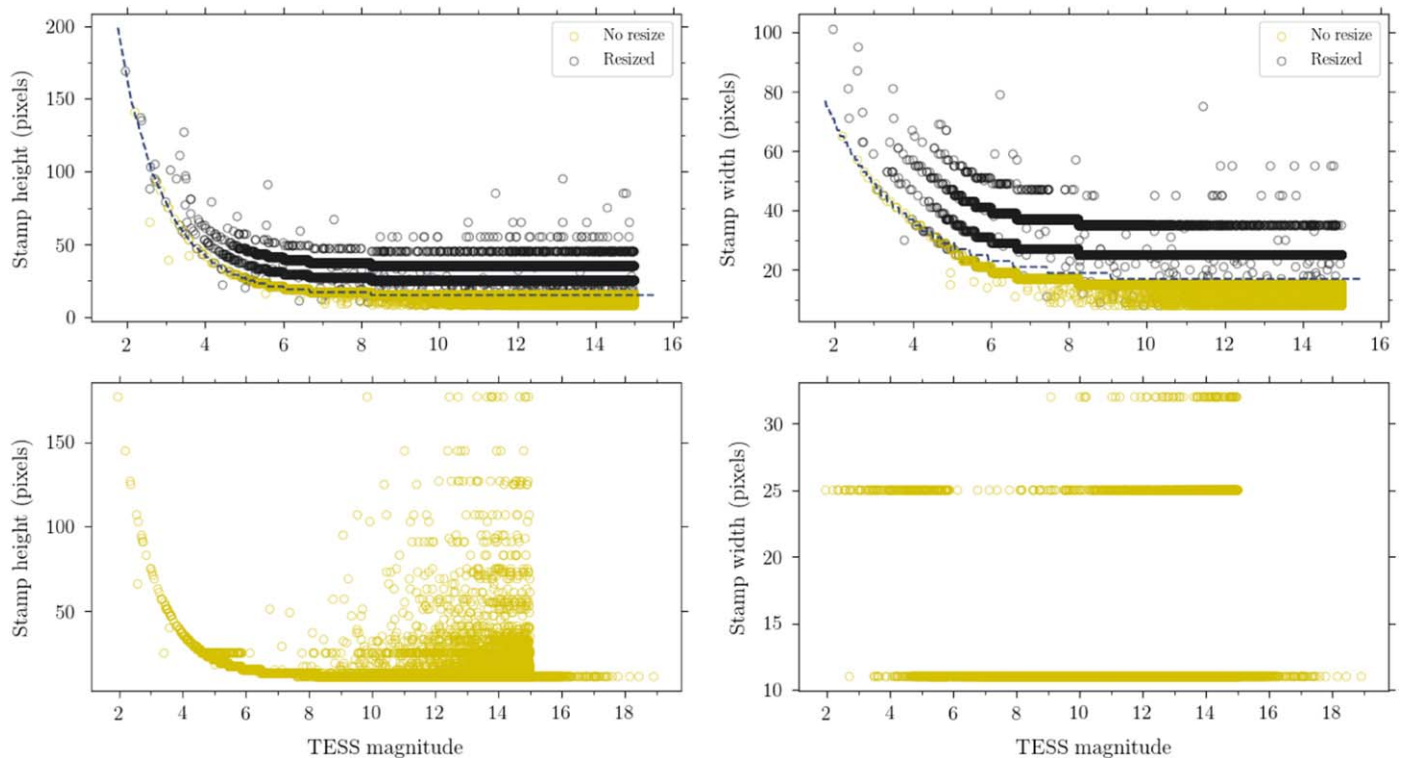
*Pixel stamp size*—Figure 13 shows the stamp sizes for the cutouts made around each processed target. For 120 s data the stamp provided by the TESS mission is always used. In cases where a defined aperture touches the edge of the pixel stamp (for 1800 s data), the stamp is allowed to resize in one or both directions by a fixed step of 5 pixels, and the aperture is defined anew (see also Section 5.1). The starting guess for the stamp size (width and height) has been optimized to reduce the number of required resizes and thereby also processing time. The maximum number of allowed resizes is currently set to 5, at which state the photometry would stop, but this is almost never reached in practice. FFI targets seen to have heights/widths falling below the well-defined relation are found on the

edges of the CCDs and hence have a limited possible height/width. For secondary targets identified in TPF stamps, a fixed aperture of  $11 \times 11$  pixels is used.

## 9. Conclusions and Outlook

We have developed an automated, flexible, modular, and robust TESS pipeline, optimized for producing light curves for asteroseismology from FFIs and TPFs, but useful for a wide variety of astrophysical applications. The pipeline runs in parallel at the Centre for Scientific Computing, Aarhus, and is easily capable of keeping pace with monthly TESS data deliveries. Initial inputs are TESS FFIs and TPFs after low-level calibration has been performed by the SPOC. We use the TIC to identify all stars down to approximately  $Tmag = 15$ , estimate and remove time-resolved sky background from the images in a way that captures the complexity of the observed TESS backgrounds, and calculate precise timing information, including barycentric corrections, for all our targets.

After these processing steps, we then perform photometric extractions for all targets. The pipeline is designed for flexibility and is able to transparently make use of multiple extraction algorithms including aperture photometry, two different PSF-fitting implementations, difference imaging, and halo photometry. At present the default for most stars is aperture photometry, using an approach modeled on the K2P<sup>2</sup> algorithm developed for K2 data (Lund et al. 2015). This methodology uses a clustering approach to assign pixels to groups of nearby targets and then segments the clusters to construct noncircular apertures for each individual star, allowing successful application of aperture photometry to even quite crowded fields. For the brightest stars (typically with  $Tmag \lesssim 4$ ), we apply an implementation of halo photometry (White et al. 2017; Pope et al. 2019) to construct a calibrated light curve from unsaturated pixels in the outer portions of the PSF of a saturated stellar image. The modular nature of our approach allows us to easily implement additional photometric algorithms, and we plan to include PSF fitting and difference imaging for crowded fields at a later stage. We have designed data formats for maximum compatibility with existing TESS products hosted at MAST and software widely used by the community such as Lightkurve (Lightkurve Collaboration et al. 2018). Noise metrics for the



**Figure 13.** Relation between stamp height (left) and width (right) as a function of Tmag, here for Sector 2 data. The top (bottom) panels show the values for 1800 s (120 s) data. Yellow points indicate stamps that have not been resized, while black points show values for resized stamps. The dashed line in the top panels shows the starting values of the stamp size.

resulting light curves, whether measured using rms or MDV, are generally consistent with expectations from the relations shown in Sullivan et al. (2015).

As noted above, we have designed the TASOC photometric pipeline with flexibility in mind. During the extended mission, the long cadence will be reduced to 600 s, while an additional short cadence mode of 20 s will be available on a limited shared-risk basis. Our pipeline is already able to successfully process these higher-cadence frames with limited impact on processing time, though we are still understanding the impacts of cosmic rays and potentially more limited calibration information available at these cadences. We anticipate continuing to make improvements to the pipeline and its products throughout the lifetime of the mission and beyond in response both to shortcomings we identify and to community input.

Funding for the Stellar Astrophysics Centre is provided by The Danish National Research Foundation (grant agreement No. DNR106). R.H. and M.N.L. acknowledge the ESA PRODEX program. This research was supported by the National Aeronautics and Space Administration (80NSSC18K1585 and 80NSSC19K0379) awarded through the TESS Guest Investigator Program. O.J.H. acknowledges the support of the UK Science and Technology Facilities Council (STFC). This work was performed in part under contract with the Jet Propulsion Laboratory (JPL) funded by NASA through the Sagan Fellowship Program executed by the NASA Exoplanet Science Institute.

The numerical results presented in this work were obtained at the Centre for Scientific Computing, Aarhus.<sup>22</sup>

Resources supporting this work were provided by the NASA High-End Computing (HEC) Program through the NASA Advanced Supercomputing (NAS) Division at Ames Research Center for the production of the SPOC data products.

This research made use of Astropy,<sup>23</sup> a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013; Price-Whelan et al. 2018).

This paper includes data collected by the TESS mission, which are publicly available from the Mikulski Archive for Space Telescopes (MAST) and described in Jenkins et al. (2016). Funding for the TESS mission is provided by NASA’s Science Mission directorate. This research has made use of NASA’s Astrophysics Data System, as well as the NASA/IPAC Extragalactic Database (NED) which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

Funding for the TESS Asteroseismic Science Operations Centre is provided by the Danish National Research Foundation (grant agreement No. DNR106), ESA PRODEX (PEA 4000119301), and Stellar Astrophysics Centre (SAC) at Aarhus University. We thank the TESS team and staff and TASC/TASOC for their support of the present work.

*Software:* Astropy (Astropy Collaboration et al. 2013; Price-Whelan et al. 2018), Photutils (Bradley et al. 2019), NumPy (Harris et al. 2020), SciPy (Virtanen et al. 2020), matplotlib (Hunter 2007), scikit-learn (Pedregosa et al. 2011), scikitimage (van der Walt et al. 2014), Bottleneck,<sup>24</sup> SpiceyPy (Annex & Carcich 2019), h5py,<sup>25</sup> Q3C (Koposov & Bartunov 2006).

<sup>23</sup> <https://www.astropy.org>

<sup>24</sup> <https://bottleneck.readthedocs.io>

<sup>25</sup> <https://www.h5py.org>

<sup>22</sup> <https://phys.au.dk/forskning/faciliteter/cscaa/>

## Appendix A Background Tests

TESS FFIs are contaminated by light from background sources (e.g., zodiacal light, scattered light from solar system objects). This is of particular concern for TESS as it lies on an orbit that takes it close to earth periodically, causing large-scale changes in background flux.

Before settling on the methodology presented in Section 3.2, a series of background estimation algorithms were developed and tested on TESS data, both real and simulated (e.g., by Lund et al. 2017), and assessed on their ability to remove the background. The main difficulty for these methods is dealing with the corner glow, due to scattered light from stars off the CCD. Because the corner glow does not flow continuously into the rest of the background, methods relying on smoothing or interpolation tend to struggle slightly more.

### A.1. Row-by-column Estimate

The background continuum was estimated individually along each row of pixels in the FFI, and consequently along each column. Each column or row of data was divided into 60 bins. The 10th percentile value in the bins was then smoothed using a Hamming filter to provide a background continuum for a single row or column. The full FFI background was then obtained by averaging the two. This estimate performs well on the center of the CCD but undercorrects at the edges.

### A.2. Binned Mode Estimate

First, the data were subdivided into tiles of  $128 \times 128$  pixels. The continuum pixel fluxes in each tile were fit using a RANSAC algorithm (Pedregosa et al. 2011), which provides outlier flags under the assumption that there are more background pixels than source pixels in each tile. The modes of the RANSAC-qualified values are taken as representative of that tile, resulting in 256 modes across the full FFI. These are then again fit with RANSAC. Finally, to obtain the background continuum, a 2D polynomial function is fit to the 256 modes using the RANSAC inlier probabilities as weights, thus minimizing the impact of tiles with increased density of stars resulting in anomalously high mode values. This method performs most poorly, as the RANSAC fitting is computationally expensive, and the spatial resolution does not effectively treat corner glow.

### A.3. Kepler-style Estimate

The *Kepler* mission (Borucki et al. 2010) estimated the sky background across its camera by measuring flux in small  $2 \times 2$  pixel tiles across each readout channel, in order to minimize contamination from recognizable stars, and then interpolating between them (Bryson et al. 2010, 2020).

This method attempts to do the same by taking tiles of  $9 \times 9$  pixels equally distributed across the CCD and calculating the mode of the pixels inside these tiles, before interpolating between these data to obtain the background estimate for the full FFI. Interpolation tends to fail near the edges of the measured data, so in order to provide an accurate measure of

the background, the density of the points is doubled near the edges and corners of the FFI. The points are taken as the intersections of an irregular Cartesian grid, as was the original method for the *Kepler* background. The modes were sigma clipped once at  $1\sigma$ , and then interpolated across the full FFI to obtain the background estimate. This method fails to consistently return a background measure, as the location of the pixels used to measure the background flux does not account for the location of any known stars, causing some stellar flux to be included in the background, and the interpolation can fail in frames with particularly bad corner glow (see Figure 14). It should be stressed that this method is not the same as was used by the *Kepler* mission, but merely inspired by it. In the *Kepler* mission, the FFIs from *Kepler* and the catalog were analyzed in order to define the locations of the  $2 \times 2$  background apertures so that they accurately reflected the dimmest, normal pixels in each scene. In the version used here, this was not done.

### A.4. SE-style Estimate

The full  $2048 \times 2048$  pixel image was cut into  $64 \times 64$  pixel tiles, wherein a  $3\sigma$  clipping is performed to further mask out stars. The mode of the image was estimated using the same prescription as is used by SE (Bertin & Arnouts 1996) (see Equation (1), Section 3.2). It treats the corner glow by reparameterizing the image into radial coordinates and performing the background fit a second time, but binned in radius instead of in Cartesian coordinates. The two measurements are performed iteratively. These values were run through a  $3 \times 3$  moving median filter and interpolated back onto the full image size.

### A.5. Method Comparison

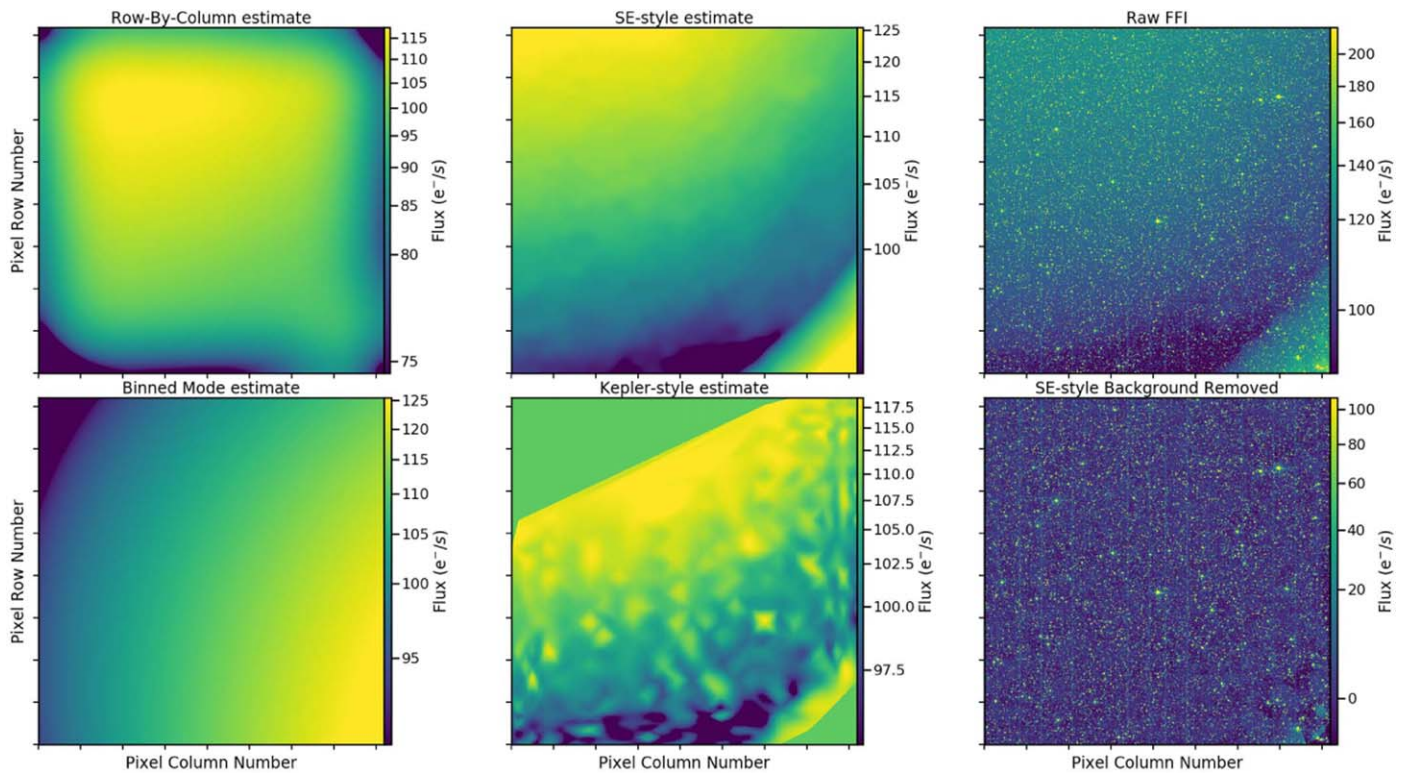
All four methods were used to estimate a simulated background. The “binned mode” estimate performed well in general but struggled in cases of significant corner glow and had a large computational time due to the RANSAC fitting process.

The “row-by-column” estimate performed quickly and estimated the background well overall but struggled near the edges due to smoothing methods. In some cases, it introduced a regular grid-like structure due to the nature of the estimation method.

The “Kepler-style” method failed to consistently produce smooth interpolations. Increasing the grid density increased the computational time to impractical length, and replacing the interpolation with a 2D polynomial fit failed to capture the global structure as well.

The “SE-style” estimate also performed quickly and estimated the background well overall, and is the only method that sufficiently treats the corner glow. Due to the robustness and simplicity of the underlying code and methodology, this method was applied in the main pipeline.

The four background estimates, along with raw and SE-style background-subtracted FFIs, are shown in Figure 14.



**Figure 14.** Background estimates of an FFI from TESS Camera 1, CCD 2 in Sector 3, for each of the four estimation methods. The raw and background-subtracted images are shown on the right. The SE-style estimate removes the background most robustly.

## Appendix B FITS Headers of Output Light Curves

We here include an example of the full FITS headers of a typical FITS light curve file produced by the TASOC Photometry Pipeline as described in Section 7. In extensions 3 and 4 (SUMIMAGE and APERTURE), several headers pertaining to the World Coordinate System solution and SIP distortions have been removed to save space.

### HDU-0 (PRIMARY)

```

SIMPLE = T/ conforms to FITS standard
BITPIX = 8/ array data type
NAXIS = 0/ number of array
          dimensions
EXTEND = T
NEXTEND = 3/ number of standard
          extensions
EXTNAME = 'PRIMARY' / name of extension
ORIGIN = 'TASOC/Aarhus' / institution responsible
          for creating this file
DATE = '2020-06-06' / date the file was created
TELESCOP = 'TESS' / telescope
INSTRUME = 'TESS Photometer' / detector type
FILTER = 'TESS' / Photometric bandpass
          filter
OBJECT = 'TIC 444953480' / string version of TICID
TICID = 444953480/ unique TESS target
          identifier
CAMERA = 2/ Camera number
CCD = 4/ CCD number
SECTOR = 6/ Observing sector
PROCVER = 'master-v4.5' / Version of photometry
          pipeline

```

```

FILEVER = '1.4' / File format version
DATA_REL = 8/ Data release number
VERSION = 5/ Version of the processing
PHOTMET = 'aperture' / Photometric method used
RADESYS = 'ICRS' / reference frame of
          celestial coordinates
EQUINOX = 2000.0/ equinox of celestial
          coordinate system
RA_OBJ = 91.263592/ [deg] Right ascension
DEC_OBJ = -12.035933/ [deg] Declination
PMRA = 0.9823499999999999/ [mas/yr] RA proper motion
PMDEC = -0.199674/ [mas/yr] Dec proper
          motion
PMTOTAL = 1.002437643335484/ [mas/yr] total proper
          motion
TESSMAG = 13.668/ [mag] TESS magnitude
TEFF = 4878.0/ [K] Effective
          temperature
TICVER = 7/ TESS Input Catalog
          version
CRMITEN = T/ spacecraft cosmic ray
          mitigation enabled
CRBLKSZ = 10/ [exposures] s/c cosmic
          ray mitigation block siz
CRSPOC = F/ SPOC cosmic ray cleaning
          enabled
KP_THRES = 0.8/ K2P2 sum-image threshold
KP_MIPIX = 4/ K2P2 min pixels in mask
KP_MICLS = 4/ K2P2 min pix. for cluster
KP_CLSRA = 1.414213562373095/ K2P2 cluster radius
KP_WS = T/ K2P2 watershed
          segmentation
KP_WSBLLR = 0.5/ K2P2 watershed blur
KP_WSTHR = 0/ K2P2 watershed threshold
KP_WSFOT = 3/ K2P2 watershed footprint

```

KP\_EX = T/ K2P2 extend overflow  
 AP\_CONT = 0.5487687205693473/ AP contamination  
 DATAVAL = 0/ Data validation flags  
 CHECKSUM = 'geOejdNegdNegdNe' / HDU checksum updated  
 2020-06-06T16:44:04  
 DATASUM = '0' / data unit checksum updated  
 2020-06-06T16:44:04

TUNIT7 = 'ppm' / column units: rel. flux in  
 parts-per-million  
 TDISP7 = 'E26.17' / column display format  
 TTYPE8 = 'FLUX\_CORR\_ERR' / column title: corrected  
 photometric flux error  
 TFORM8 = 'D' / column format: 64-bit  
 floating point

*HDU-1 (LIGHTCURVE)*

XTENSION = 'BINTABLE' / binary table extension  
 BITPIX = 8/ array data type  
 NAXIS = 2/ number of array  
 dimensions  
 NAXIS1 = 96/ length of dimension 1  
 NAXIS2 = 993/ length of dimension 2  
 PCOUNT = 0/ number of group  
 parameters  
 GCOUNT = 1/ number of groups  
 TFIELDS = 14/ number of table fields  
 INHERIT = T/ inherit the primary  
 header  
 TTYPE1 = 'TIME' / column title: data time  
 stamps  
 TFORM1 = 'D' / column format: 64-bit  
 floating point  
 TUNIT1 = 'BJD-2457000, days' / column units: Bar-  
 ycenter corrected TESS  
 Julian  
 TDISP1 = 'D14.7' / column display format  
 TTYPE2 = 'TIMECORR' / column title: bar-  
 ycenter-timeslice  
 correction  
 TFORM2 = 'E' / column format: 32-bit  
 floating point  
 TUNIT2 = 'd' / column units: day  
 TDISP2 = 'E13.6' / column display format  
 TTYPE3 = 'CADENCENO' / column title: unique  
 cadence number  
 TFORM3 = 'J' / column format: signed  
 32-bit integer  
 TDISP3 = 'I10' / column display format  
 TTYPE4 = 'FLUX\_RAW' / column title: photo-  
 metric flux  
 TFORM4 = 'D' / column format: 64-bit  
 floating point  
 TUNIT4 = 'e- / s' / column units: electrons  
 per second  
 TDISP4 = 'E26.17' / column display format  
 TTYPE5 = 'FLUX\_RAW\_ERR' / column title: photo-  
 metric flux error  
 TFORM5 = 'D' / column format: 64-bit  
 floating point  
 TUNIT5 = 'e- / s' / column units: electrons  
 per second  
 TDISP5 = 'E26.17' / column display format  
 TTYPE6 = 'FLUX\_BKG' / column title: photo-  
 metric background flux  
 TFORM6 = 'D' / column format: 64-bit  
 floating point  
 TUNIT6 = 'e- / s' / column units: electrons  
 per second  
 TDISP6 = 'E26.17' / column display format  
 TTYPE7 = 'FLUX\_CORR' / column title: corrected  
 photometric flux  
 TFORM7 = 'D' / column format: 64-bit  
 floating point

TUNIT8 = 'ppm' / column units: parts-  
 per-million  
 TDISP8 = 'E26.17' / column display format  
 TTYPE9 = 'QUALITY' / column title: photo-  
 metric quality flags  
 TFORM9 = 'J' / column format: signed  
 32-bit integer  
 TDISP9 = 'B16.16' / column display format  
 TTYPE10 = 'PIXEL\_QUALITY' / column title: pixel  
 quality flags  
 TFORM10 = 'J' / column format: signed  
 32-bit integer  
 TDISP10 = 'B16.16' / column display format  
 TTYPE11 = 'MOM\_CENTR1' / column title: moment-  
 derived column centroid  
 TFORM11 = 'D' / column format: 64-bit  
 floating point  
 TUNIT11 = 'pixels' / column units: pixels  
 TDISP11 = 'F10.5' / column display format  
 TTYPE12 = 'MOM\_CENTR2' / column title: moment-  
 derived row centroid  
 TFORM12 = 'D' / column format: 64-bit  
 floating point  
 TUNIT12 = 'pixels' / column units: pixels  
 TDISP12 = 'F10.5' / column display format  
 TTYPE13 = 'POS\_CORR1' / column title: column  
 position correction  
 TFORM13 = 'D' / column format: 64-bit  
 floating point  
 TUNIT13 = 'pixels' / column units: pixels  
 TDISP13 = 'F14.7' / column display format  
 TTYPE14 = 'POS\_CORR2' / column title: row posi-  
 tion correction  
 TFORM14 = 'D' / column format: 64-bit  
 floating point  
 TUNIT14 = 'pixels' / column units: pixels  
 TDISP14 = 'F14.7' / column display format  
 EXTNAME = 'LIGHTCURVE' / extension name  
 TIMEREFF = 'SOLARSYSTEM' / barycentric correction  
 applied to times  
 TIMESYS = 'TDB' / time system is Bary-  
 centric Dynamical  
 Time (TDB)  
 BJDREFI = 2457000/ integer part of BTJD  
 reference date  
 BJDREFF = 0.0/ fraction of the day in  
 BTJD reference date  
 TIMEUNIT = 'd' / time unit for TIME,  
 TSTART and TSTOP  
 TSTART = 1468.276025411983/ observation start time  
 in BTJD  
 TSTOP = 1490.046743795897/ observation stop time  
 in BTJD  
 DATE-OBS = '2018-12-15T18:36:19.412' / TSTART as UTC calen-  
 dar date  
 DATE-END = '2019-01-06T13:06:09.480' / TSTOP as UTC calen-  
 dar date  
 MJD-BEG = 58467.77602541199/ observation start time  
 in MJD  
 MJD-END = 58489.54674379589/ observation start time  
 in MJD

TELAPSE = 21.77071838391339/ [d] TSTOP-TSTART	/ [deg] Coordinate value
LIVETIME = 4.310602240014852/ [d] TELAPSE multiplied	at reference point
	by DEADC
DEADC = 0.198/ deadtime correction	LONPOLE = 180.0/ [deg] Native longitude
EXPOSURE = 4.310602240014852/ [d] time on source	of celestial pole
XPOSURE = 356.4/ [s] Duration of exposure	LATPOLE = -11.930165289899/ [deg] Native latitude of
TIMEPIXR = 0.5/ bin time beginning=0	celestial pole
	TIMESYS = 'TDB' / Timescale
	middle=0.5 end=1
TIMEDEL = 0.02083333333333333/ [d] time resolution	TIMEUNIT = 'd' / Time units
	of data
	DATEREF = '1858-11-17' / ISO-8601 fiducial time
INT_TIME = 1.98/ [s] photon accumulation	MJDREFI = 0.0/ [d] MJD of fiducial time,
	integer part
	time per frame
READTIME = 0.02/ [s] readout time per	MJDREFF = 0.0/ [d] MJD of fiducial time,
	fractional part
	frame
FRAMETIM = 2.0/ [s] frame time (INT_TIME	DATE-OBS = '2018-12-19T00:06:23.366Z' / ISO-8601 time of
+ READTIME)	observation
NUM_FRM = 900/ number of frames per time	MJD-OBS = 58471.004437106/ [d] MJD of observation
	MJD-OBS = 58471.004437106/ [d] MJD at start of
	stamp
	observation
NREADOUT = 720/ number of read per	TSTART = 1471.5052378466/ [d] Time elapsed since
	fiducial time at start
	cadence
CHECKSUM = 'eeSbfZSbedSbeZSb' / HDU checksum updated	DATE-END = '2018-12-19T00:36:23.366Z' / ISO-8601 time at end of
	2020-06-06T16:44:04
	observation
DATASUM = '3537856224' / data unit checksum	MJD-END = 58471.02527044/ [d] MJD at end of
	observation
	updated 2020-06-
	06T16:44:04

**HDU-2 (SUMIMAGE)**

XTENSION = 'IMAGE' / Image extension
BITPIX = -64/ array data type
NAXIS = 2/ number of array
NAXIS1 = 15
NAXIS2 = 15
PCOUNT = 0/ number of parameters
GCOUNT = 1/ number of groups
INHERIT = T/ inherit the primary
WCSAXES = 2/ Number of coordi-
CRPIX1 = 863.0/ Pixel coordinate of
CRPIX2 = 18.0/ Pixel coordinate of
PC1_1 = -0.0057098640987826/ Coordinate transforma-
PC1_2 = -0.0003675563609303/ Coordinate transforma-
PC2_1 = 0.00022773533217156/ Coordinate transforma-
PC2_2 = -0.005696929639529/ Coordinate transforma-
CDEL1 = 1.0/ [deg] Coordinate incre-
CDEL2 = 1.0/ [deg] Coordinate incre-
CUNIT1 = 'deg' / Units of coordinate
CUNIT2 = 'deg' / Units of coordinate
CTYPE1 = 'RA-TAN-SIP' / TAN (gnomonic) projec-
CTYPE2 = 'DEC-TAN-SIP' / TAN (gnomonic) projec-
CRVAL1 = 86.20939522998/ [deg] Coordinate value
CRVAL2 = -11.930165289899










	ISO-8601 time of
	observation
	[d] MJD of observation
	[d] MJD at start of
	observation
	[d] Time elapsed since
	fiducial time at start
	ISO-8601 time at end of
	observation
	[d] MJD at end of
	observation
	[d] Time elapsed since
	fiducial time at end
	[d] Elapsed time (start
	to stop)
	[d] Time resolution
	Reference position of
	timestamp in bin-
	ned data
	/ Equatorial coordinate
	system
	/ extension name
	HDU checksum updated
	2020-06-06T16:44:04
	/ data unit checksum
	updated 2020-06-
	06T16:44:04

**HDU-3 (APERTURE)**

XTENSION = 'IMAGE' / Image extension
BITPIX = 32/ array data type
NAXIS = 2/ number of array
NAXIS1 = 15
NAXIS2 = 15
PCOUNT = 0/ number of parameters
GCOUNT = 1/ number of groups
INHERIT = T/ inherit the primary
WCSAXES = 2/ Number of coordi-
CRPIX1 = 863.0/ Pixel coordinate of
CRPIX2 = 18.0/ Pixel coordinate of
PC1_1 = -0.0057098640987826/ Coordinate transforma-
PC1_2 = -0.0003675563609303/ Coordinate transforma-
PC2_1 = 0.00022773533217156/ Coordinate transforma-
PC2_2 = -0.005696929639529/ Coordinate transforma-

CDELTA1 = 1.0/ [deg] Coordinate increment at reference point  
 CDELTA2 = 1.0/ [deg] Coordinate increment at reference point  
 CUNIT1 = 'deg' / Units of coordinate increment and value  
 CUNIT2 = 'deg' / Units of coordinate increment and value  
 CTYPER1 = 'RA-TAN-SIP' / TAN (gnomonic) projection + SIP distortions  
 CTYPER2 = 'DEC-TAN-SIP' / TAN (gnomonic) projection + SIP distortions  
 CRVAL1 = 86.20939522998/ [deg] Coordinate value at reference point  
 CRVAL2 = -11.930165289899/ [deg] Coordinate value at reference point  
 LONPOLE = 180.0/ [deg] Native longitude of celestial pole  
 LATPOLE = -11.930165289899/ [deg] Native latitude of celestial pole  
 TIMESYS = 'TDB' / Timescale  
 TIMEUNIT = 'd' / Time units  
 DATEREFF = '1858-11-17' / ISO-8601 fiducial time  
 MJDREFI = 0.0/ [d] MJD of fiducial time, integer part  
 MJDREFF = 0.0/ [d] MJD of fiducial time, fractional part  
 DATE-OBS = '2018-12-19T00:06:23.366Z' / ISO-8601 time of observation  
 MJD-OBS = 58471.004437106/ [d] MJD of observation  
 MJD-OBS = 58471.004437106/ [d] MJD at start of observation  
 TSTART = 1471.5052378466/ [d] Time elapsed since fiducial time at start  
 DATE-END = '2018-12-19T00:36:23.366Z' / ISO-8601 time at end of observation  
 MJD-END = 58471.02527044/ [d] MJD at end of observation  
 TSTOP = 1471.5260711749/ [d] Time elapsed since fiducial time at end  
 TELAPSE = 0.020833328235/ [d] Elapsed time (start to stop)  
 TIMEDEL = 0.020833333333333/ [d] Time resolution  
 TIMEPIXR = 0.5/ Reference position of timestamp in binned data  
 RADESYS = 'ICRS' / Equatorial coordinate system  
 EXTNAME = 'APERTURE' / extension name  
 CHECKSUM = 'U6QAW4P5U4PAU4P5' / HDU checksum updated 2020-06-06T16:44:04  
 DATASUM = '8545' / data unit checksum updated 2020-06-06T16:44:04

### ORCID iDs

Rasmus Handberg  <https://orcid.org/0000-0001-8725-4502>  
 Mikkel N. Lund  <https://orcid.org/0000-0001-9214-5642>  
 Timothy R. White  <https://orcid.org/0000-0002-6980-3392>  
 Oliver J. Hall  <https://orcid.org/0000-0002-0468-4775>  
 Derek L. Buzasi  <https://orcid.org/0000-0002-1988-143X>  
 Benjamin J. S. Pope  <https://orcid.org/0000-0003-2595-9114>  
 Jonas S. Hansen  <https://orcid.org/0000-0002-4165-8592>  
 Carolina von Essen  <https://orcid.org/0000-0002-6956-1725>  
 Lindsey Carboneau  <https://orcid.org/0000-0003-1001-5137>

Daniel Huber  <https://orcid.org/0000-0001-8832-4488>  
 Roland K. Vanderspek  <https://orcid.org/0000-0001-6763-6562>  
 Michael M. Fausnaugh  <https://orcid.org/0000-0002-9113-7162>  
 Peter Tenenbaum  <https://orcid.org/0000-0002-1949-4720>  
 Jon M. Jenkins  <https://orcid.org/0000-0002-4715-9460>

### References

- Aerts, C., Christensen-Dalsgaard, J., & Kurtz, D. W. 2010, *Asteroseismology* (Berlin: Springer)
- Annex, A., & Carcich, B. 2019, AndrewAnnex/SpiceyPy: SpiceyPy 2.2.0, Zenodo, doi:[10.5281/zenodo.2576445](https://doi.org/10.5281/zenodo.2576445)
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, **558**, A33
- Audenaert, J., Kuzlewicz, J. S., Handberg, H., et al. 2021, *AJ*, in press (arXiv:[2107.06301](https://arxiv.org/abs/2107.06301))
- Baglin, A., Auvergne, M., Barge, P., et al. 2002, in Proc. First Eddington Workshop on Stellar Structure and Habitable Planet Finding, ESA Special Publication, 485, Noordwijk, ed. B. Battrick et al. (Noordwijk: ESA Publications), 17
- Batalha, N. M., Borucki, W. J., Bryson, S. T., et al. 2011, *ApJ*, **729**, 27
- Berta-Thompson, Z. K., Levine, A., & Sullivan, P. 2015, Cosmic Ray Rejection Strategies for TESS, Tech. Rep., Kavli Institute for Astrophysics and Space Science, Massachusetts Institute of Technology
- Bertin, E., & Arnouts, S. 1996, *A&AS*, **117**, 393
- Borucki, W. J., Koch, D., Basri, G., et al. 2010, *Sci*, **327**, 977
- Bouma, L. G., Hartman, J. D., Bhatti, W., Winn, J. N., & Bakos, G. Á. 2019, *ApJS*, **245**, 13
- Bradley, L., Sipocz, B., Robitaille, T., et al. 2019, *astropy/photutils*: v0.6, Zenodo, doi:[10.5281/zenodo.2533376](https://doi.org/10.5281/zenodo.2533376)
- Brasseur, C. E., Phillip, C., Fleming, S. W., Mullally, S. E., & White, R. L. 2019, *Astrocute: Tools for Creating Cutouts of TESS images*, ascl:[1905.007](https://doi.org/10.5281/zenodo.45007)
- Bryson, S. T., Jenkins, J. M., Klaus, T. C., et al. 2010, *Proc. SPIE*, **7740**, 77401D
- Bryson, S. T., Jenkins, J. M., Klaus, T. C., et al. 2020, *Kepler Data Processing Handbook: KSCI-19081-003* (Moffett Field, CA: NASA Ames Research Center), 3, 27, [https://archive.stsci.edu/files/live/sites/mast/files/home/missions-and-data/kepler/\\_documents/KSCI-19081-003-KDPH.pdf](https://archive.stsci.edu/files/live/sites/mast/files/home/missions-and-data/kepler/_documents/KSCI-19081-003-KDPH.pdf)
- Caldwell, D. A., Tenenbaum, P., Twicken, J. D., et al. 2020, *RNAAS*, **4**, 201
- Clarke, B. D., Caldwell, D. A., Quintana, E. V., et al. 2020, in *Kepler Data Processing Handbook: KSCI-19081-003*, ed. J. M. Jenkins (Moffett Field, CA: NASA Ames Research Center), 5, 65, [https://archive.stsci.edu/files/live/sites/mast/files/home/missions-and-data/kepler/\\_documents/KSCI-19081-003-KDPH.pdf](https://archive.stsci.edu/files/live/sites/mast/files/home/missions-and-data/kepler/_documents/KSCI-19081-003-KDPH.pdf)
- Eisner, N. L., Pope, B. J. S., Aigrain, S., et al. 2019, *RNAAS*, **3**, 145
- Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. 1996, in Proc. Second Int. Conf. on Knowledge Discovery and Data Mining, KDD'96, ed. E. Simoudis, J. Han, & U. Fayyad (Palo Alto, CA: AAAI Press), 226
- Fausnaugh, M., Huang, X., Glidden, A., & Guerrero, N. 2018, AAS Meeting Abstracts, 231, 439.09
- Fausnaugh, M. M., Burke, C. J., Caldwell, D. A., et al. 2019, TESS Data Release Notes: Sector 18, DR25, NASA, [https://archive.stsci.edu/missions/tess/doc/tess\\_drn/tess\\_sector\\_18\\_dr25\\_v02.pdf](https://archive.stsci.edu/missions/tess/doc/tess_drn/tess_sector_18_dr25_v02.pdf)
- Fausnaugh, M. M., Burke, C. J., Caldwell, D. A., et al. 2020a, TESS Data Release Notes 27 Memo: Updates to Sector 20, Data Release 27 Products, NASA, [https://archive.stsci.edu/missions/tess/doc/tess\\_drn/tess\\_s20\\_dr27\\_data\\_product\\_revision\\_memo\\_v02.pdf](https://archive.stsci.edu/missions/tess/doc/tess_drn/tess_s20_dr27_data_product_revision_memo_v02.pdf)
- Fausnaugh, M. M., Burke, C. J., Caldwell, D. A., et al. 2020b, TESS Data Release Notes 29 Memo: Updates to Sector 21, Data Release 29 Products, NASA, [https://archive.stsci.edu/missions/tess/doc/tess\\_drn/tess\\_s21\\_dr29\\_data\\_product\\_revision\\_memo\\_v02.pdf](https://archive.stsci.edu/missions/tess/doc/tess_drn/tess_s21_dr29_data_product_revision_memo_v02.pdf)
- Feinstein, A. D., Montet, B. T., Foreman-Mackey, D., et al. 2019, *PASP*, **131**, 094502
- Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, *A&A*, **616**, A1
- Handberg, R., & Lund, M. N. 2014, *MNRAS*, **445**, 2698
- Handberg, R., & Lund, M. N. 2017, *A&A*, **597**, A36
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Natur*, **585**, 357
- Howell, S. B., Sobeck, C., Haas, M., et al. 2014, *PASP*, **126**, 398
- Huang, C. X., Vanderburg, A., Pál, A., et al. 2020, *RNAAS*, **4**, 204
- Hunter, J. D. 2007, *CSE*, **9**, 90
- Jenkins, J. M. 2017, *Kepler Data Processing Handbook: KSCI-19081-002* (Moffett Field, CA: NASA Ames Research Center), [https://archive.stsci.edu/files/live/sites/mast/files/home/missions-and-data/kepler/\\_documents/KSCI-19081-002-KDPH.pdf](https://archive.stsci.edu/files/live/sites/mast/files/home/missions-and-data/kepler/_documents/KSCI-19081-002-KDPH.pdf)

- Jenkins, J. M., Twicken, J. D., McCauliff, S., et al. 2016, *Proc. SPIE*, **9913**, 99133E
- Kjeldsen, H., & Frandsen, S. 1992, *PASP*, **104**, 413
- Koposov, S., & Bartunov, O. 2006, *adass XV*, **351**, 735
- Lightkurve Collaboration, Cardoso, J. V. d. M., Hedges, C., et al. 2018, Lightkurve: Kepler and TESS Time Series Analysis in Python, Astrophysics Source Code Library, ascl:[1812.013](https://ascl.net/1812.013)
- Luger, R., Agol, E., Kruse, E., et al. 2016, *AJ*, **152**, 100
- Lund, M. N., Handberg, R., Buzasi, D. L., et al. 2021, [arXiv:2108.11780](https://arxiv.org/abs/2108.11780)
- Lund, M. N., Handberg, R., Davies, G. R., Chaplin, W. J., & Jones, C. D. 2015, *ApJ*, **806**, 30
- Lund, M. N., Handberg, R., Kjeldsen, H., Chaplin, W. J., & Christensen-Dalsgaard, J. 2017, *EPJ Web Conf.*, **160**, 01005
- Maclaurin, D., Duvenaud, D., & Adams, R. P. 2015, in ICML 2015 AutoML Workshop, <https://indico.ijclab.in2p3.fr/event/2914/contributions/6483/subcontributions/180/attachments/6060/7185/automl-short.pdf>
- Miglio, A., Chiappini, C., Morel, T., et al. 2013, *MNRAS*, **429**, 423
- Montalto, M., Borsato, L., Granata, V., et al. 2020, *MNRAS*, **498**, 1726
- Nardiello, D., Borsato, L., Piotto, G., et al. 2019, *MNRAS*, **490**, 3806
- Oelkers, R. J., & Stassun, K. G. 2018, *AJ*, **156**, 132
- Oelkers, R. J., Stassun, K. G., Pepper, J., De Lee, N., & Paegert, M. 2016, in 2016 IEEE Int. Conf. on Big Data (Big Data) (Piscataway, NJ: IEEE), 3204
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *Journal of Machine Learning Research*, **12**, 2825
- Pope, B. J. S., White, T. R., Farr, W. M., et al. 2019, *ApJS*, **245**, 8
- Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., et al. 2018, *AJ*, **156**, 123
- Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2014, *JATIS*, **1**, 014003
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *AJ*, **131**, 1163
- Stassun, K. G., Oelkers, R. J., Paegert, M., et al. 2019, *AJ*, **158**, 138
- Stassun, K. G., Oelkers, R. J., Pepper, J., et al. 2018, *AJ*, **156**, 102
- Sullivan, P. W., Winn, J. N., Berta-Thompson, Z. K., et al. 2015, *ApJ*, **809**, 77
- Twicken, J. D., Caldwell, D. A., Jenkins, J. M., et al. 2020, TESS Science Data Products Description Document EXP-TESS-ARC-ICD-0014 Rev F, NASA Ames Research Center, <https://archive.stsci.edu/missions/tess/doc/EXP-TESS-ARC-ICD-TM-0014-Rev-F.pdf>
- van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., et al. 2014, *PeerJ*, **2**, e453
- Vanderburg, A., & Johnson, J. A. 2014, *PASP*, **126**, 948
- Vanderspek, R., Doty, J. P., Fausnaugh, M., et al. 2018, TESS Instrument Handbook, Tech. Rep., Kavli Institute for Astrophysics and Space Science, Massachusetts Institute of Technology, [https://archive.stsci.edu/missions/tess/doc/TESS\\_Instrument\\_Handbook\\_v0.1.pdf](https://archive.stsci.edu/missions/tess/doc/TESS_Instrument_Handbook_v0.1.pdf)
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nature Methods*, **17**, 261
- von Essen, C., Lund, M. N., Handberg, R., et al. 2020, *AJ*, **160**, 34
- White, T. R., Pope, B. J. S., Antoci, V., et al. 2017, *MNRAS*, **471**, 2882