TESS Data for Asteroseismology: Light-curve Systematics Correction

Mikkel N. Lund1, Rasmus Handberg1, Derek L. Buzasi2, Lindsey Carboneau1,2,3, Oliver J. Hall1,3,4, Filipe Pereira5,6, Daniel Huber7, Daniel Hey8, and Timothy Van Reeth9
The T’DA Collaboration

1 Stellar Astrophysics Centre, Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, DK-8000 Aarhus C, Denmark; mikkelnl@phys.au.dk
2 Department of Chemistry and Physics, Florida Gulf Coast University, 10501 FGCU Blvd. S., Fort Myers, FL 33965, USA
3 School of Physics and Astronomy, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK
4 European Space Agency (ESA), European Space Research and Technology Centre (ESTEC), Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands
5 Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, 4150-762 Porto, Portugal
6 Departamento de Física e Astronomia, Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre, s/n, PT4169-007 Porto, Portugal
7 Institute for Astronomy, University of Hawai‘i, 2680 Woodlawn Dr., Honolulu, HI 96822, USA
8 School of Physics, Sydney Institute for Astronomy (SIfA), The University of Sydney, NSW 2006, Australia
9 Institute of Astronomy, KU Leuven, Celestijnenlaan 200D, B-3001, Leuven, Belgium

Received 2020 December 22; revised 2021 August 24; accepted 2021 August 24; published 2021 December 1

Abstract

Data from the Transiting Exoplanet Survey Satellite (TESS) have produced of the order of one million light curves at cadences of 120 s and especially 1800 s for every ~27 day observing sector during its two-year nominal mission. These data constitute a treasure trove for the study of stellar variability and exoplanets. However, to fully utilize the data in such studies a proper removal of systematic-noise sources must be performed before any analysis. The TESS Data for Asteroseismology group is tasked with providing analysis-ready data for the TESS Asteroseismic Science Consortium, which covers the full spectrum of stellar variability types, including stellar oscillations and pulsations, spanning a wide range of variability timescales and amplitudes. We present here the two current implementations for co-trending of light curves from TESS, which cover different regimes of variability to serve the entire seismic community. We find performance in terms of commonly used noise statistics meets expectations and is applicable to a wide range of intrinsic variability types. Further, we find that the correction of light curves from a full sector of data can be completed well within a few days, meaning that when running in steady state our routines are able to process one sector before data from the next arrives. Our pipeline is open-source and all processed data will be made available on the websites of the TESS Asteroseismic Science Operations Center and the Mikulski Archive for Space Telescopes.

Unified Astronomy Thesaurus concepts: Astronomy data analysis (1858); Asteroseismology (73); CCD photometry (208); Variable stars (1761); Exoplanets (498)

1. Introduction

The space-based photometric missions CoRoT (Baglin et al. 2009), Kepler (Gilliland et al. 2010), and K2 (Howell et al. 2014) have over the last decade and a half have revolutionized the field of asteroseismology (e.g., Chaplin & Miglio 2013; Bowman 2017; Hekker & Christensen-Dalsgaard 2017; García & Ballot 2019). The success of these missions can be attributed to their ability to deliver high-quality photometric observations at high cadence and spanning long continuous baselines. The Transiting Exoplanet Survey Satellite (TESS; Ricker et al. 2014) is the most recent mission to satisfy these criteria, and it promises to continue the advancement of asteroseismology (Campanile et al. 2016; Schofield et al. 2019). As with Kepler, the TESS asteroseismic investigation is organized within the TESS Asteroseismic Science Consortium (TASC; see Lund et al. 2017), and the data for asteroseismic analysis is hosted at the TESS Asteroseismic Science Operations Center (TASOC10) and the Mikulski Archive for Space Telescopes (MAST11; 10.17909/t9-4smn-dx89).

The primary science goal of TESS is to find exoplanets smaller than Neptune that can be characterized in detail from follow-up observations. The main data product obtained to meet the science goal is observations of a preselected list of targets at a 120 s cadence. In addition to being suited for exoplanet studies, these data are required for the asteroseismic study of main-sequence solar-like oscillators. Of particular importance for the asteroseismic study of evolved stars, such as red giants for galactic archeology studies (Miglio et al. 2013; Silva Aguirre et al. 2020), or classical pulsators (Antoci et al. 2019; Holdsworth et al. 2021), are the targets observed in the full-frame images (FFIs; Jenkins et al. 2016), which were obtained every 1800 s during the nominal mission.

A key distinction between the processing of data from Kepler versus that for TESS is that TESS will neither reduce the FFIs nor deliver light curves for all targets. Targets observed at a 120 s cadence will be processed to produce light curves, while FFIs will be made available to the community in a calibrated form and light curves will be delivered for a subset of targets (Caldwell et al. 2020). In the extended TESS mission the FFI cadence is reduced to 600 s, and an additional cadence of 20 s has been introduced. Hence, to make full use of all the data from TESS there is a need for a flexible, extensible pipeline that can produce analysis-ready data for a variety of astrophysics applications, including exoplanet science, stellar activity, and of course asteroseismology. Such a pipeline, the TASOC pipeline, is provided and run by the coordinated activity TESS Data for Asteroseismology (T’DA) within TASC (see

10 https://tasoc.dk/
11 https://archive.stsci.edu/hlsp/tasoc
Section 2). In this paper we describe components of the TASOC pipeline dedicated to removing instrumental effects from the raw light curves extracted from data delivered by TESS. Other pipelines for TESS FFI data reduction exist (see, e.g., Oelkers & Stassun 2018; Feinstein et al. 2019; Huang et al. 2020), but these are mostly focused on exoplanet searches, while the TASOC pipeline seeks to preserve intrinsic stellar variability covering a range of timescales and amplitudes.

The paper is structured as follows. In Section 2 we give an overall introduction to the full TASOC pipeline, and then focus on the details of the light-curve correction component in Section 3. Section 4 details the performance of the light-curve correction pipeline, while Section 5 describes the format of the data products. We conclude, provide an outlook, and discuss potential improvements to the pipeline in Section 6.

2. The TASOC Pipeline

The TASOC pipeline is developed by the T’DA group to serve the asteroseismic community of TASC. Specifically, T’DA is responsible for delivering (1) raw photometric time series from FFIs (1800 s/600 s cadence) and target pixel files (TPFs; 120 s/20 s cadence); (2) light curves corrected for systematic signals, ready for asteroseismic analysis; and (3) a preliminary classification of the variability/oscillation/pulsation type of each light curve from (1) to (2), enabling each of the working groups (WGs) of TASC, which are organized according to stellar type, to better target its analysis.

The TASOC pipeline is open-source and available on GitHub12 (Handberg et al. 2021a, 2021b). All data products from the pipeline are available via the TASOC database and on MAST (10.17909/t9-4smn-dx89); see Section 5 for more details.

12 https://github.com/tasoc/photometry, https://github.com/tasoc/corrections

The photometry module of the pipeline is described in Handberg et al. (2021, hereafter Paper I), and the stellar classification is described in Audenaert et al. (2021, hereafter Paper III). The current paper deals with the correction of systematics of raw light curves. Figure 1 gives a general overview of the TASOC pipeline, and its different data products. The components enclosed by the red dashed line are the ones described in this paper.

We note that while the TASOC pipeline is primarily developed with a focus on asteroseismology, the data will also be useful for a wider range of science applications, including exoplanet science, eclipsing binaries, and the study of activity (rotation, flares, etc.).

3. Light-curve Correction

Starting from the raw light curves extracted as detailed in Paper I, the light-curve corrections described in this paper are of a co-trending nature, i.e., using information shared among many stars to estimate and account for any systematic signals. For specific types of stars additional detrending may be required to isolate specific signals—an example could be the filtering of transits for an oscillating exoplanet host, or the high-pass filtering of low-frequency variability from stellar rotation. Depending on the identified variability type, based on the stellar classification module, the TASOC pipeline will at a later stage include a “custom correction” making use of classification results to improve the performance of the detrending algorithms and thus produce improved light curves for the different TASC WGs. The modular nature of the pipeline is intended to simplify incorporation of such future improvements and extensions.

Because T’DA serves the entire community within TASC, the light-curve correction needs to be able to deal with stars with variability ranging from stochastic solar-like oscillations with a lifetime of minutes, to more coherent pulsators, such as...
δ-Sct/γ-Dor type, RR Lyrae stars, and long-period variables, such as Miras and Cepheids. In addition to the large range of typical timescales, the variability amplitudes range from a few parts per million to several magnitudes. The range of timescales and amplitudes (and phase stabilities) poses a challenge in terms of fully preserving the astrophysical signal, as the level of overlap with the systematic contribution will vary with stellar type.

In the first releases of the correction module, we employ two independent co-trending methodologies. As the quality of the co-trending from the different methods varies with stellar type, we will release both versions. It will then be up to the user to identify which correction is best suited for their specific science case. In a future release we are hopeful that an informed decision can be made on which correction to use for a given star by iterating with the stellar classification module (Paper III).

The two methods incorporated and used for the current release are the co-trending basis vector (CBV; Section 3.1) and ensemble methods (Section 3.2). In Section 4 we will provide a comparison of the two methods to show typical use cases where a given method dominates in quality.

3.1. CBV Fitting Method

The first of the correction methods employs so-called co-trending basis vectors for the correction (Smith et al. 2017). CBVs are a set of orthonormal vectors describing the dominant sources of variability shared among many stars. The idea is that such shared variability can likely be attributed to systematic signals induced by external sources, rather than representing variability intrinsic to the individual stars.

Below we go through the steps included in generating CBVs and using them to co-trend the raw TESS light curves.

3.1.1. CBV Generation

Many of the procedures used for generating CBVs are inspired by the pipeline used to process Kepler data. For the sake of completeness we go through the steps here, but refer the reader to Smith et al. (2017) for more details on the specifics of the Kepler pipeline.

The key difference from the approach used for Kepler is in the definition of the spatial regions for which the CBVs are computed. In Kepler a set of CBVs was computed for each full CCD, and this approach has been carried on to the correction of 120 s data from TESS by the Science Processing Operations Center (SPOC; Jenkins et al. 2016). We introduce a subdivision of the TESS CCDs into CBV areas and compute a set of CBVs for each of these. The areas are currently defined by three concentric circles centered on the camera centers—in this manner each of the four CCDs per camera will be divided into four areas, giving a total of 16 sets of CBVs per camera. Figure 2 gives an illustration of the segmentation for camera 4 (here for observations in Sector 4). The area is given by a three-digit number, where the first digit gives the camera number, the second gives the CCD number, starting in the top-left quadrant and increasing in the counterclockwise direction, and the third gives the region within the CCD, ranging from 1 toward the camera center to 4 at the camera corner.

The reasoning behind this segmentation is that some systematics, such as focus changes, are expected to depend on the distance to the camera center. By having areas with a difference in distance to the camera center some of the variation in the systematics should be better captured, allowing for a reduced number of CBVs required in the co-trending and hence reducing the risk of overfitting.

The procedure for generating CBVs is as follows:

1. The sample of light curves from a given CCD area is first pruned for targets of high variability, because these can contaminate the identification of shared variability for CBV generation. A cut is made that deselects targets with a variability larger than 1.3 times the median variability of all targets in the area. This cutoff value is found to provide a good compromise between deselecting the most highly variable targets and retaining a large sample of stars for CBV generation. We note that the CBV correction will ultimately be applied to all targets. For the measurement of variability we first subtract a third-order polynomial fitted using a weighted least-squares routine—the variability is then given by the standard deviation of the residual normalized by the median flux error (as propagated from the photometry pipeline). The third-order polynomial is found to provide a good removal of most long-term variability with a likely systematic origin. Consequently, this variability is not used to prune targets for the generation of CBVs.

2. Data points with certain quality flags or with NaN values in the raw flux time series are removed from the following analysis. Specifically, we remove data points with TESS quality flags 2 (safemode), 8 (Earthpoint), 32 (desaturation event), and 128 (manual exclude), and potential TASOC quality flag 2 (manual exclude). We note that, at the time of writing, many of the quality flags are not yet populated by the TESS team. Currently we retain data obtained during coarse pointing, as spike CBVs (see below) have been found to be able to correct for much of the coarse-pointing jitter. We also do not propagate the “manual exclude” flags from the 120 s data.
—in Sector 1 it was found that by propagating these flags about ∼20% of the 1800 s data of the FFIs would have been discarded.

3. From this reduced low-variability sample we compute the correlation matrix between the median-normalized light curves. Based on the correlation matrix we retain the 50% of targets with the highest median absolute correlation to other targets. The idea behind this selection of targets is that if a light curve is dominated by a systematic variability shared among many stars, then their light curves should also show a high level of correlation. For the calculation of the correlation matrix we use Einstein summation in Python, which allows for a very rapid derivation of correlations even when dealing with several hundred thousand 27 day light curves.

4. The light curves from the above step are interpolated using a piecewise cubic Hermite-interpolating polynomial (as implemented in SciPy: Virtanen et al. 2020) to fill gaps from masked quality flags or bad data points (e.g., NaN flux from the photometry module) that are individual to a given target—gaps from shared quality flags are not interpolated. We note that such individual gaps rarely occur; hence generally very few or no points are interpolated. This step ensures that all light curves have the same number of data points, which is required for the following procedures. After the interpolation the light curves are mean-normalized. The interpolated points are removed in the final light curves.

5. The CBVs are given by the orthonormal basis vectors of a principal-component analysis (PCA). The PCA is performed with singular-value decomposition using the Python package scikit-learn (Pedregosa et al. 2011). We currently include 16 components in the PCA.

6. The calculation of CBVs in the above step is iterated with pruning of single targets with a large weight in a given CBV. From the PCA we can assess the weight of each light curve in the construction of a given basis vector. If a given CBV truly represents a shared variability, then a given light curve should not have a weight significantly larger than that of the other light curves. To identify such potential targets we compute the entropy of the weights and compare it to expectations from an assumed Gaussian distribution with a width similar to the standard deviation of the weights (see Smith et al. 2017 for details). A negative difference indicates a poor entropy with a dominant contribution from a single or a few stars. Based on testing different limits we find that a value of −0.5 provides a good upper limit on the difference—if this criterion is not met we remove the star with the largest relative weight and recompute the CBVs. This process is iterated until the calculated entropy meets expectations. Given the pruning of highly variable targets in a previous step, this step typically only removes a few stars, if any at all, from contributing to the CBVs.

7. The derived CBVs are then tested for having a high signal-to-noise ratio (S/N). Some CBVs, typically the ones with the least information content, will have a large and apparently random scatter. While some of this variability will be truly systematic, the risk of introducing high-frequency noise in the light curve when fitting the CBV outweighs the benefit of including the CBV in the co-trending. We require that in order for a CBV to be considered in the co-trending the power level between the standard deviation of the CBV ($\sigma_{\text{signal}}$) and the standard deviation of the first differences of the CBV ($\sigma_{\text{noise}}$) should be less than 5 dB, that is,

$$10 \log_{10} \left( \frac{\sigma_{\text{signal}}^2}{\sigma_{\text{noise}}^2} \right) \text{dB} < 5. \quad (1)$$

8. The high-S/N CBVs are then split into components showing low-frequency and/or low-amplitude variations (we shall refer to these as normal CBVs) and spike CBVs, which, as the name suggests, contain large single-bin spikes. The spikes are typically remnants of reaction wheel desaturation events (where currently only the central cadence is flagged in 1800 s data; see Figure 11) or times when TESS observed in coarse pointing. The normal and spike CBVs are fitted independently, but always in pairs. To identify and separate out the spikes of a given CBV we first apply a Savitzky–Golay filter to the CBV. This low-pass-filtered version of the CBV is then subtracted and the absolute value is taken of the residuals. In these values we then identify peaks that are larger than three times the standard deviation of the absolute residuals (computed using the median absolute deviation). We note that propagating the quality flags from the 120 s data (especially the “manual exclude” flag) could mitigate a lot of the spikes, though this would be at the cost of discarding many data points.

Figure 3 gives an example of the computed CBVs from Sector 6 data, showing both the normal and spike CBVs. As seen, the normal CBVs typically exhibit a 2 week variation, corresponding to the orbit of TESS. The most dominant contribution to the 1800 s spike CBVs is the reaction wheel desaturation events every ~3.125 days. Figure 4 illustrates the variation in the primary CBV component with the position on the camera, showing the importance of segmenting the camera into areas with their own sets of CBVs. If only a single set of CBVs were calculated for a camera, several components would be needed to capture the variability of the primary CBV in the segmented setup.

### 3.1.2. CBV Fitting

The correction of the raw photometry with the CBVs is done by representing the systematic signal to be removed as a linear combination of the CBVs:

$$F_{\text{corr}} = F_{\text{raw}} - \sum_{i=0}^{N} \hat{c}_i \times \text{CBV}_i, \quad (2)$$

where $i$ runs over the $N$ CBVs included in the correction, and $\hat{c}_i$ is the corresponding scaling coefficient. The coefficients are given by the ordinary least-squares solution as

$$\hat{c} = (A^T A)^{-1} A^T f, \quad (3)$$

where $A$ is a matrix with the normal and spike CBVs as columns and $f$ is the raw flux light curve. In $A$ we also add a column with ones to allow for an offset.

The correction with Equation (2) and fitting with Equation (3) is done iteratively with a sigma-clipping of the residual corrected flux ($\text{Flux}_{\text{corr}}$), where points more than $4\sigma$ away are removed (with $\sigma = 1.4826 \text{MAD}(|\text{Flux}_{\text{corr}}|)$). This iteration continues until the fitting coefficients from...
Equation (3) remain unchanged, or a maximum of 50 iterations has been reached—the maximum number of iterations is rarely reached, but in the odd case it is reached a warning flag is set for the processing of the target. This robust sigma-clipping ensures that outliers that have not been flagged by the quality flags or transient signals (flares, eclipses, transits, etc.) do not influence the fit of the CBVs to the raw light curve. Finally, we use the Bayesian information criterion (BIC) to decide on the number \( N \) of calculated CBVs (16) to include in the correction.

For the correction of 120 s cadence targets we do not compute separate CBVs, but use an interpolated (cubic spline) version of the 1800 s cadence CBVs. The reason for not generating CBVs based on 120 s cadence data is the much reduced number of targets available in a given CBV area, resulting in noisy CBVs that have a tendency to add significant amounts of white noise to the co-trended light curves.

3.2. Ensemble Fitting Method

The second correction method employs ensemble photometry (Gilliland & Brown 1988) to characterize shared variability by constructing an average standard star using a combination of light curves from stars nearby in the field. The approach focuses on characterizing and removing local systematic effects. Below we describe the steps involved in the ensemble correction algorithm.

We note that currently the ensemble method is currently only applied to 1800 s data because of a lack of sufficient stars for constructing a proper ensemble from 120 s data. Using synthetic data, we have conducted numerical experiments interpolating a 120 s ensemble from the 1800 s data, and results are promising, but not yet ready for inclusion in our final product. Further development of the method to allow the inclusion of 120 s stars is underway.

3.2.1. Ensemble Generation

The most important parts of the ensemble correction method are the determination of appropriate members of the stellar ensemble and the creation of the “average” light curve from that group. Here we outline the steps included in that process, which is illustrated in Figure 5.

1. Data points in the raw flux time series with NaN values or nonzero quality flags as listed for the CBV method in Section 3.1.1 are removed from the succeeding analysis. This typically removes less than 2% of points, primarily from cadences associated with spacecraft anomalies and scattered light. For the resulting time series we determine three basic statistical parameters for later use in the process. Two of these are the median value and the variability characterized by the range between the 5th and 95th percentiles of the flux values \( R_{\text{var}} \); see, e.g., Basri et al. 2011). We also characterize the noise level of the time series using the standard deviation of the differenced light curve \( \sigma_D \), a technique that accounts for the nonstationary nature of the typical stellar light curve (Nason 2006). We then return a median-divided light curve to use for detrending purposes.

2. Potential ensemble members are identified based on the Euclidean distance from the target star. We begin with 100 candidates, and further require that they lie on the same camera and CCD as the target. For each potential ensemble member, we calculate the same basic statistical parameters as done for the target, and normalize by the light curve’s median value.

3. We then loop through the ensemble candidates in order of increasing Euclidean distance from the target, adding each member to the ensemble list if it meets the following criteria, the values for which were determined from numerical experiments with the targets in TESS Sector 1:
The differenced standard deviation of the ensemble candidate does not exceed a maximum value, \( \sigma_D < \sigma_{D,\text{max}} \). The maximum allowed value is \( \sigma_{D,\text{max}} = 1 \), which ensures that the ensemble is not contaminated by very noisy light curves (typically from faint stars).

(b) The differenced standard deviation of the ensemble candidate is less than 10 times that of the target.

(c) We require that the range parameter be \( R_{\text{var}} < 0.4 \), to ensure that highly variable stars, such as classical variable stars, are not included in the ensemble.

These limiting values for these parameters are recorded in the header to allow for future flexibility. As mentioned above the values for the listed criteria are based on Sector 1 data. Given the improvements of TESS operations from Sector 4 these numbers will possibly be updated in later versions of the pipeline based on new tests. Any such changes will be commented on in the associated data release notes.

The minimum number of stars in the ensemble is five. Once that number is reached, a fit of the ensemble light curve to the target star is done. We then add the next candidate star to the ensemble and perform a new ensemble fit (see Section 3.2.2) generating the corrected flux light curve \( F_{\text{corr}} \). We track the quality of the corrected light curve as a function of the number of ensemble members using the normalized first-order total variation (TV; White et al. 2017), or equivalently the geometric length of the time series (Buzasi et al. 2015; Prša et al. 2019), as our figure of merit:

\[
TV = \sum_{i=1}^{N} \frac{|F_{\text{corr},i} - F_{\text{corr},i-1}|}{M(F_{\text{corr}})}.
\]

where \( M \) refers to the median operator. Additional members are added to the ensemble until the quality of the correction, as traced by the TV, no longer increases; the typical resulting ensemble sizes are \( \sim 10 \) stars.

Practically, we have noted that stars with identical TESS magnitudes do not always produce light curves with similar numbers of counts, and that these observed offsets are correlated with local background levels, implying that background subtraction is (unsurprisingly) imperfect. We refer the reader to Paper I for details on the background estimation. For bright stars, minor errors in background estimation produce insignificant offsets, but the impact is more significant for fainter stars. Accordingly, we adjust the background flux level \( B \) of each ensemble member to minimize the difference between the flux of the median-filtered ensemble \( F_E \) and the target \( F_t \) light curves, via the statistic \( \zeta \):

\[
\zeta = \sum_{i=0}^{N} \left[ \frac{F_{t,i}}{M(F_t)} - \frac{F_{E,i} + B}{M(F_E + B)} \right]^2.
\]

We then apply the optimal offset value \( B \), typically a few counts per pixel, to the ensemble light curve. Figure 6 illustrates the results arising from the application of this algorithm to a few members of a typical ensemble.

3.2.2 Ensemble Application

To apply the ensemble fit to the target star we remove significant outliers, both instrumental and astrophysical, at each cadence by clipping points that are more than \( 2\sigma \) from the mean value at that cadence. The ensemble light curves are then median-filtered at each time step to produce a single ensemble light curve \( F_E \) that can be used for correction purposes. Before application, we allow for a modest flux offset \( K \) of \( F_E \) through
minimization of the scaled Euclidean length of the time series:

$$F_2 = \frac{\sum_{i=1}^{N} \Delta_i \frac{E_i}{M(F_E + K)}}{M(F_E + K)}$$

where $\Delta_i$ represents the $i$th first difference; the first difference is the difference between one time-series value and the next (Nason 2006). The resulting offset is applied to $F_E$ and the result is divided into the target light curve to produce a corrected light curve $F_{corr}$.

### 4. Performance

#### 4.1. Photometric Performance

As is customary in assessing the performance of photometric correction methods, we look first at the noise metrics of the resulting light curve. Figure 7 shows the impact of the systematics correction for Sector 6 1800 s cadence targets on the 1 hr rms variability, the point-to-point median differential variability (PTP-MDV), and the overall variance of the light curve. The differently colored contours/points refer to the values for the different correction methods and the raw photometry.

From Figure 7 we find that (1) the largest reduction happens for the overall variance, followed by the 1 hr rms and lastly the PTP-MDV, where the reduction is minimal, and (2) in all cases the reduction is overall larger for the CBV method. The first point indicates that the co-trending most aggressively removes long-term trends, while the point-to-point scatter is largely unaffected, and the second point shows that the CBV method is more aggressive than the ensemble method. This behavior is expected and intended, because shared systematic trends, often originating from variations attributed to the spacecraft, are expected to be somewhat secular and because no priors are used on the fitting of CBV components, increasing the risk of overfitting, especially for stars with a variability longer than the observational baseline. We do, however, see that the lower envelope of the noise metrics follows the expected noise level from Sullivan et al. (2015). The relatively few cases (values shown as points, which are all below the 1 per million level) at high magnitudes, where the noise falls below the expected level, are typically highly contaminated stars (see Paper I).

Beyond the agreement between the lower envelope and the expected noise it is difficult to use the noise metrics to evaluate the performance of the co-trending, because our goal is to preserve intrinsic variability not simply to minimize scatter on a given timescale, as is often done for planet searches using, e.g., the 1 hr combined differential photometric precision.

Concerning the number of components $N$ included in the corrections of the CBV method, a dependence is mainly seen on magnitude—the brighter the star, and hence the less dominated by photometric noise, the greater the number of CBV components typically included. At magnitudes $T_{mag} > 12$ the distribution for $N$ is fairly uniform between 3 and 16 components.

We note that it is impossible to provide a single figure of merit that summarizes the quality of the corrections performed, as this will depend on the type of star, the magnitude, the position on the CCD, etc. A user of these light curves will have to make their best judgment of the quality of the corrections conditioned on the type of object being observed.

---

**Figure 5.** A high-level summary of the ensemble correction algorithm. The ensemble is initially constructed from light curves for nearby stars that satisfy basic statistical criteria, and ensemble membership is then increased and refined to maximize the light curve figure of merit described in the text. Along with the final correction, the number and identity (TESS Input Catalog identifiers (TIC IDs)) of the members of the final ensemble are returned in the light-curve header. The end of the algorithm is represented by the black filled shape.
4.2. Example Light Curves

In order to properly evaluate performance we would need to simulate light curves with all different types of expected intrinsic variability, covering a range of magnitudes and crowding values, and include known systematic variability covering a range of timescales and relative amplitudes. Such an exercise is beyond the scope of the current paper, but we are planning a comparison for a future paper, with the intent of including all methods currently available for the reduction of TESS photometry. We can, however, qualitatively evaluate the performance and compare strengths and complementarity between the CBV and ensemble methods by examples of corrected light curves with identifiable intrinsic variability.

Figures 8–10 show examples of corrected light curves for different types of variables. These cases are selected from a randomly generated set of 200 targets, to uniformly cover the focal plane and a spread in magnitudes.

In Figure 8 we show four cases highlighting the particular strengths of the ensemble method—panels (a) and (b) give examples of long-period variable stars where the ensemble method is able to preserve the astrophysical signal while also correcting drops in flux from momentum wheel desaturation events (see Figure 11). In comparison, the CBV method performs poorly for these types of stars from overfitting the variability. Panels (a), (c), and (d) give examples of the local nature of the ensemble correction method, i.e., building the systematics correction from stars in close proximity on the detector, which enables a correction of the high levels of spatially localized noise seen in these light curves. The local nature of the noise means that it is not well represented in the CBVs, and thus the CBV method struggles to properly correct the noise in these cases.

Figure 9 shows examples of stellar variability where both methods are able to preserve the astrophysical signal. However, the CBV method here and in general better corrects small-scale (as compared to the intrinsic variability) systematic trends, most prominently seen in panels (b) and (c). In general, the CBV method also appears better at correcting the residuals from reaction wheel desaturation events, which are not captured by the ensemble method for the light curves in panels (b) and (d), and systematics from scattered light near the data downlink gap.

Figure 10 shows examples of two oscillating red giants found among the random set of targets. The oscillations are evident from the corrected data of both methods, but the CBV method is seen to perform slightly better in correcting systematics from scattered light near the data downlink gap and desaturation events, which remain in the corrected ensemble data in panel (b). We also note that the CBV method has much lower high-frequency noise in the power density spectra (PDSs) compared to the raw data in panel (a) and both the raw and ensemble-corrected data in panel (b).

Concerning the reaction wheel desaturation events, Figure 11 shows the representative impact on the light curves, and the correction of this by our methods. The 200 random 1800 s light curves are first detrended by a 2 day moving median filter and then phase-folded at a period of ~3.1250 days, corresponding to the time between the reaction wheel desaturations. The light-curve segments before and after the data downlink are treated separately as the phase of the desaturations is shifted at the gap. The folded light curves are then median-binned, with a bin width corresponding to the 1800 s cadence. The central cadence of the desaturation (marked in red) is impacted the most and this point will be removed using the pixel quality flags (bit 32). However, the neighboring cadences (marked in green) are also heavily affected, but are not tagged with a quality flag in the photometry. As seen, the CBV method is able to correct for the impact at these cadences, resulting in near-zero median-binned flux (note that the ordinate in Figure 11 is linear in the ±100 ppm region). The ensemble method often leaves a residual, but still reduces the impact by an order of magnitude. Because the affected cadences are often perfectly corrected they are left in the final light curves, but the user should (as always) make sure to assess the quality of the corrections before further analysis. Removing a single cadence on both sides of the flagged desaturations should in any case alleviate issues from uncorrected residuals. We plan in a future version of the pipeline to include, for both methods, a check of the residuals around flagged desaturations. If the points are found to be outliers because of an insufficient correction they will be flagged as such in the QUALITY extension of the final data product (see Section 4.5 and Paper I, Table 1). The inclusion of such an additional check will be noted in the data release notes associated with the given version of the pipeline.
In Figure 12 we show examples of light curves for known exoplanet systems observed during Sector 6, including the systems CoRoT-4 (Aigrain et al. 2008), CoRoT-5, CoRoT-12, CoRoT-18 (Hébrard et al. 2011), HATS-4 (Jordan et al. 2014), HATS-6 (Hartman et al. 2015), and HATS-45 (Brahm et al. 2018). For these 120 s targets we show in addition to the CBV light curves the ones produced by SPOC (the raw light curves are SAP—simple aperture photometry; the corrected light curves are PDCSAP—presearch data conditioning SAP). We omit showing light curves from the eleanor pipeline (Feinstein et al. 2019) because the flux extraction and correction here requires user input on a star-by-star basis, and data from the method of Oelkers & Stassun (2018) is currently unavailable for Sector 6.

We see that in all cases the planetary transits are clearly visible when the data are phase-folded at the planetary orbital period. When comparing the transits from the CBV method with those from SPOC, we see that the transit depth from the SPOC data is often significantly larger. In most cases this is a consequence of the correction of SPOC SAP (and therefore also PDCSAP) data for both the estimated crowding and the target star flux fraction captured in the optimum aperture (see Stumpe et al. 2012). If we revert the applied corrections to the SPOC data we generally see excellent agreement in the transit shapes, which highlights the validity of using T’DA data for exoplanet searches and analysis—and importantly, T’DA also offers data for all identified 1800 s FFI targets. Remaining differences in the transit shapes, specifically in the depths, can be traced to the raw photometry from different aperture sizes (Paper I), most prominently seen for HATS-6 (panel (a)) and CoRoT-18 (panel (e)). It is clear, however, that with a correction for crowding and fractional flux contained in the aperture, which should be applied before an analysis of the transit, most of the differences between the CBV- and SPOC-corrected data should be remedied. We leave this task to the expert user—as seen from Figure 12 the effect of the correction by SPOC can be quite significant, and appears in some cases to be overestimated based on a comparison with transit depths from the literature. We see also that for CoRoT-18 (panel (f)) we are able to preserve the rotational modulation of the star in addition to the exoplanet transit (panel (e)). We note that all the known exoplanet hosts shown are relatively faint ($T_{\text{mag}} > 12.75$) and crowded.

4.3. Correction Yield

Also worth comparing are the numbers of corrected light curves for the ensemble and CBV methods. Due to the architecture of the ensemble code, there is a requirement of a certain number of stars in the vicinity of the target star with similar magnitudes. As the target star gets brighter it becomes increasingly difficult to find suitable nearby comparison stars for the ensemble. In Figure 13 we compare the number of corrected light curves with the number of available light curves as a function of magnitude. As seen, the ensemble method will for 1800 s targets generally return only a high percentage of stars with TESS magnitudes $\gtrsim 7–8$. The numbers do of course differ between campaigns from variations in field density, for example, but the user of ensemble light curves should not be surprised if a corrected light curve is unavailable for relatively bright targets.

4.4. Processing Time

In terms of calculation time the co-trending of targets with the CBV approach proceeds at a pace of $\sim 0.34$ s per star, while the ensemble method takes $\sim 0.88$ s per star. For the CBV approach, this calculation time assumes that the CBVs have been generated, a process that in itself takes of the order of $\sim 0.5–1$ days. Figure 14 shows the total processing time for all 1800 s cadence targets in Sector 6 with our current setup with a single 36-CPU-core node on the Grendel Slurm cluster at the Centre for Scientific Computing Aarhus.

The total processing time for $\sim 1,650,000$ stars amounts to $\sim 0.3$ and $\sim 0.7$ days for the CBV and ensemble methods, respectively. A secondary time component entering the total processing time comes from idle time, when a process waits to

---

13 Obtained from the TOI catalog: https://tev.mit.edu/data/collection/193/.
14 The uncorrected flux $F_0$ is obtained from the PDCSAP flux $F$ as $F_0 = f(F + M(F)^{1/(1 - \epsilon)})$, where $\epsilon$ is the crowding factor (CROWDSAP), $f$ is the flux fraction (FLIRCDSAP), and $M$ is the median operator.

15 http://www.csaa.dk/grendel-s/overview/
start because output is being written by other processes to a file shared among the processes on the node. While further optimizations are being pursued the processing is at present fast enough to keep up with the extraction of raw photometry, and the processing can be parallelized over several nodes if required.

4.5. Light-curve Recommendations

Detailed recommendations on best practices are still being investigated, but in general we suggest that users consult corrected data from both methods to see if they are compatible. In cases where there are large differences the underlying stellar variability will typically be of the long-period kind, and in these cases then ensemble data should generally be preferred. When the corrected light curves are compatible (differences below or at the level of the variability range of the raw light curve) the CBV method will generally do a better correction of the light-curve systematics, including reaction wheel momentum desaturations (Figure 11) and the strong systematics from scattered light around the data downlink gaps. We recommend that users always consult the quality flags accompanying the raw and corrected light curves, and the release notes associated with the data. As detailed in Paper I the corrected light curves contain both a column of PIXEL_QUALITY, which gives the quality flags provided by the TESS team (see the TESS Archive Manual\(^\text{[16]}\)), and a column of QUALITY, which gives the quality flags set by the TASOC pipeline (see Paper I, Table 1). We also encourage users to specifically check the corrected light curves around the downlink gaps—while the strong systematics typically found here from the scattered light of the Earth are properly corrected in many cases, especially for the CBV method, residuals can occur and could hamper a subsequent analysis if not removed or detrended from the light curve. We note that the corrected light curves are not gap-filled. It is up to the user of the data to

---

\(^{16}\) https://outerspace.stsci.edu/display/TESS/2.0+-+Data+Product+Overview
decide if gap filling/inpainting of the data is useful for the specific science case and in that case what method should be adopted (see, e.g., García et al. 2014; Pascual-Granado et al. 2018).

5. Data Format and Access

We refer the reader to Paper I for a general overview of the data formatting, but describe here the details specific to systematics-corrected light curves.

Similar to that for raw photometry, the foremost data format for corrected light curves is the Flexible Image Transport System (FITS\textsuperscript{17}), and the FITS data is provided in a compressed Gzip format. A FITS light-curve file produced by T’DA for a corrected target will be named following the structure

$$\text{tess}[^{\text{TIC ID}}] - \text{s[sector]} - \{\text{camera}\} - \{\text{ccd}\} - \{\text{cadence}\} - \text{dr[data release]} - \{\text{version}\} - \text{tasoc} - \{\text{method}\} \_\text{lc.fits.gz}$$

The TIC ID of the star is zero (pre-)padded to eleven digits, the sector is zero (pre-)padded to three digits, the cadence is in seconds and zero (pre-)padded to four digits, the data release is zero (pre-)padded to two digits and refers to the official release of the data from the mission, and the version is zero (pre-)padded to two digits and refers to the TASOC data release (counting from 1). The file naming is thus identical to that of the raw photometry, with the exception of the method, which refers to the co-trending method used. The method can either be “cbv” for the CBV approach (Section 3.1) or “ens” for the ensemble approach (Section 3.2).

\textsuperscript{17} https://fits.gsfc.nasa.gov/fits_standard.html

\begin{figure}
\centering
\includegraphics[width=\textwidth]{example_light_curves.png}
\caption{Example light curves for stars observed in 1800 s during Sector 6, showing cases where both methods are able to preserve the intrinsic stellar variability. The TIC number of the targets is given at the top of the panels, including in parentheses the target camera, CCD, and TESS magnitude. The color of the points/lines indicates the source of the data. The raw light curves have been offset vertically from the corrected light curves by the difference between the horizontal dashed and dotted lines. The vertical red lines in the left panels mark the position in time for the reaction wheel desaturations.}
\end{figure}
When accessing the data files via MAST (10.17909/t9-4snm-dx89) as a “high-level science product” the filename will be slightly different—we refer the reader to MAST for a description of the naming convention.

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 Document for more information.

The header of the LIGHTCURVE extension will provide some details on the co-trending. In addition to the filename, the key CORRMET will provide the adopted correction method, while CORRVER will refer to the version used of the correction module, according to the version tag of the code on GitHub.

In the case of a CBV correction (Section 3.1) the header will include values for the weights of the fitted CBVs as CBV_c# (normal) and CBVS_c# (spike), where # refers to the number of the CBV; the CBV_AREA the star belongs to; the number of fitted CBVs (CBV_COMP); whether BIC was used to decide on the best number of CBVs (CBV_BIC); and the number of CBVs allowed in the fit (CBV_MAX), i.e., the number of CBVs passing the S/N test (Section 3.1). Information will also be provided on the method used to fit the CBVs (CBV_MET).

In the case of an ensemble correction (Section 3.2) the LIGHTCURVE extension header will include values for the number of targets included in the ensemble (ENS_NUM). An additional ENSEMBLE extension will list the TIC IDs of the ensemble stars, and the associated value for the background correction parameter $B$ (Equation (5) as BZETA). We refer the reader to the Appendix for examples of the full headers of the LIGHTCURVE and ENSEMBLE extensions.
Figure 11. Median-binned relative flux values of 200 random 1800 s light curves, phase-folded at the period of the reaction wheel desaturation events of ~3.1250 days. The color of the points indicates the data source, and within a cadence the points have been offset horizontally for better distinction. The light curves have been rectified by a moving median filter with a width of 2 days before the phase folding. The central cadence marked by the red region is captured in the pixel quality flags and hence removed for the CBV-corrected and ensemble-corrected (ENS) data. The green regions mark the two cadences flanking the flagged central cadence that are also significantly affected by the desaturation events. The ordinate has a linear scale in the ±100 ppm region and is logarithmic beyond this.

5.1. CBV Files

The CBVs generated for the processing will also be made available for full reproducibility of the co-trending. The CBVs will be available in FITS format and named following the structure

tess - s{sector} - c{cadence} - a{area} - v{version} - tasoc_cnv.fits

The sector is zero (pre-)padded to four digits, the cadence is in seconds and zero (pre-)padded to four digits, the area is the three-digit CBV area introduced in Section 3.1.1, and the version is zero (pre-)padded to two digits and refers to the TASOC data release (counting from 1). The file currently has two extensions, one for the normal CBVs named CBV.SINGLE-SCALE.{area}, and one for the spike CBVs named CBV.SPIKE.{area}. Each extension gives a column with time and cadence number, and then the respective CBV vectors for the area. If the pipeline, as envisioned, is later extended to use multiscale CBVs, these will be added in a separate extension.

As mentioned in Section 3.1.2, 120 s cadence targets are fitted using interpolated 1800 s cadence CBVs; hence CBV files will only be available at a 1800 s cadence.

5.2. Availability and Data Policy

Data for processed targets can first and foremost be obtained via the TASOC database. The TASOC data releases are described, including the data release notes, under the Data Releases tab. Individual target search is accessed via the Data Search tab, and here specifying the “Data types” to include the “TASOC Lightcurves.”

Data will also be available at MAST as a high-level science product (https://doi.org/10.17909/r9-4smn-dx8919). We note that in addition to a traditional search on MAST, the data can also be accessed programmatically using the MAST search option in Astroquery20 and is directly accessible via the Lightkurve (Lightkurve Collaboration et al. 2018) package using the author=’TASOC’ input parameter when searching for light curves.

6. Conclusions and Outlook

We have presented a module for the TASOC pipeline for producing systematics-corrected light curves from FFIs and TPF data. The two methods currently included, the CBV (Section 3.1) and ensemble methods (Section 3.2), allow for the processing of stars with a wide range of variability timescales and amplitudes. The ensemble method is found to better preserve the astrophysical signal for stars with long-period variability, while the CBV method is better in general at removing secular and often small-scale systematic trends from the TESS orbit, and imprints from the regular reaction wheel desaturation events.

We find that the processing of photometry from a full sector, typically comprising ~1.6 million stars, can be completed within a few days with our current setup at the Centre for Scientific Computing Aarhus.

We have several envisioned improvements and extensions to the current pipeline. The goal is to have a pipeline with only one method implemented that can properly treat all cases of variability. As seen in Section 4 the advantage of having both the CBV and the ensemble method is their strengths for different timescales of variability, where the CBV method for long-period variability has a tendency to overfit. This behavior is expected given the lack of constraints on the CBV scaling coefficients.

However, if constraints could be placed on the coefficients based on the cohort of nearby stars in the form of a prior, it should be possible to gain the benefits of the ensemble method (which utilizes information from nearby stars) but still using orthonormal basis vectors. To indicate the potential of using shared information for the construction of fitting priors we show in Figure 15 a median-binned map of the scaling coefficients for CBV number 1 (i = 1 in Equation (3)) for targets in Sector 6, camera 4. For the run used to generate this plot all possible CBVs are included in the fits, i.e., the BIC is not used for selecting the number of CBVs to include. As seen there are clear structures in the scaling coefficients within a given CBV area and in some cases across different areas if CBV number 1 happens to capture the same component of the variability. One also sees some apparent residuals from the corner glows (see Paper I) that are captured by the CBVs. In the current testing we build for a given star and a given CBV a prior by a weighted kernel density estimate for the scaling coefficients. The weighting is given by the distance between the target star and those contributing to the prior in the space of positions on the CCD and in TESS magnitudes. This approach is very similar to that adopted in the Kepler mission for the PDC-MAP data product (Smith et al. 2017). We are still in the testing phase of the implementation of priors. Similar to Smith et al. (2017) we plan to add a multiscale component to the CBVs, where sets of CBVs are created for different scales of variability.

An important addition will also be a potential iteration of the co-trending with the variability classification of the pipeline (Paper III). If the stellar type is confidently identified from the classification this can be used to guide which correction method is best to use, or possibly a weighting of different

---

19 https://archive.stsci.edu/hlsp/

20 https://astroquery.readthedocs.io/en/latest/mast/mast.html
Figure 12. Examples of phase-folded systematics-corrected light curves for known exoplanets observed during Sector 6. The light curves have been smoothed by a moving median filter with a width corresponding to 5% of the transit duration (for panel (f) the filter width is 1% of the stellar rotation period). The names of the individual systems along with the periods used are indicated in the panels. The colors of the lines indicate the different corrections from the CBV method and the PDCSAP from SPOC. SPOC\textsubscript{uncorr} gives the version of the SPOC data after reverting the correction for the crowding and flux fraction captured in the aperture. Panels (e) and (f) both show data for CoRoT-18, displaying the transit in panel (e) and the rotational modulation in panel (f) (see Hebrard et al. 2011; Raetz et al. 2019).
scales of CBVs. Tests of such an iteration will be started once the current versions of the correction and classification components of the pipeline are running in a steady state.

We are in the process of improving the ensemble method for dealing with 120 s data. We note, however, that the 120 s targets for asteroseismology are unlikely to benefit from adopting the ensemble approach over the CBV one, as these are generally stochastic solar-like oscillators.

As mentioned in Section 4 we are planning a larger comparison of TESS FFI reduction methods currently available. This should serve as a guide to the community on the use of different publicly available data for specific use cases.

We thank the two anonymous referees for comments and suggestions that helped improve the paper. Funding for the Stellar Astrophysics Centre is provided by the Danish National Research Foundation (grant agreement No. DNRF106). The numerical results presented in this work were obtained at the Centre for Scientific Computing Aarhus. This research was partially conducted during the Exostar19 program at the Kavli Institute for Theoretical Physics at UC Santa Barbara, which was supported in part by the National Science Foundation under grant No. NSF PHY-1748958. M.N.L. and R.H. acknowledge support from the ESA PRODEX program. D.L.B., L.C., and D.H. acknowledge support from the TESS GI Program under grant 80NSSC18K1585. L.C. acknowledges support from the

---

Figure 13. Left: difference between number of corrected targets and number of targets with available photometry as a function of TESS magnitude. Right: percentage ratio of corrected light curves as a function of TESS magnitude.

Figure 14. Cumulative processing time in days for 1800 s cadence targets from Sector 6 with our current setup of a 36-CPU-core node. The different colors refer to the different components of the total (green) processing time, namely the actual calculation time (blue) and the idle wait time (yellow) for writing results to a shared output file. The solid lines give the values for the CBV method and the dashed lines give the corresponding values for the ensemble method (ENS).

Figure 15. Median-binned map of scaling coefficients from the fit of CBV number 1 ($i = 1$ in Equation (3)) for targets in Sector 6, camera 4. The pixel coordinates are given relative to the camera center. Note that the variability captured by the specific CBV can vary for the different CBV areas (see Figure 2)—this is clearly visible for CCDs 3–4, where the segmentation into concentric areas is obvious. The vertical/horizontal stripes show the positions of the conducting straps, whose known positions are marked by vertical red lines in CCD 2.

---

21 http://phys.au.dk/forskning/cscaa/
European Research Council under the European Union’s Horizon 2020 research and innovation program (CartographY GA, 804752). O.J.H. acknowledges the support of the UK Science and Technology Facilities Council. F.P. acknowledges support from fellowship PD/BD/135227/2017 funded by Fundação para a Ciência e Tecnologia (Portugal) and Programa Operacional Potencial Humano/FSE (EC). D.H. acknowledges support from the Alfred P. Sloan Foundation. The research leading to these results has received funding from the KU Leuven Research Council (grant C16/18/005: PARADISE), and from the Research Foundation Flanders (FWO) by means of a junior postdoctoral fellowship with grant agreement No. 12ZB620N. This paper includes data collected by the TESS mission, which are publicly available from the Barbara A. Mikulska Archive for Space Telescopes. Funding for the TESS mission is provided by NASA’s Science Mission Directorate. Funding for the TESS Astroseismic Science Operations Center is provided by the Danish National Research Foundation (grant agreement No. DNRF106), ESA PRODEX (PEA 4000119301), and the Stellar Astrophysics Centre at Aarhus University. We thank the TESS team and staff and TASC/TASOC for their support of the present work.

Facilities: TESS, MAST (TESS).

Software: Astropy (Astropy Collaboration et al. 2013; Price-Whelan et al. 2018), scikit-learn (Pedregosa et al. 2011), NumPy (Harris et al. 2020), SciPy (Virtanen et al. 2020), Matplotlib (Hunter 2011), Bottleneck, 

Lightcurve (Lightcurve Collaboration et al. 2018), Seaborn (Waskom 2021).

Appendix

FITS Headers of Output Light Curves

We here include an example of the full headers of the LIGHTCURVE extension of a typical FITS light-curve file from both the CBV and ensemble methods, and the header of the ENSEMBLE extension for the ensemble method (see Section 5).

A.1. CBV 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 |
| EXTNNAME = ‘LIGHTCURVE’ / extension name |
| TIMEREF = ‘SCLABS’ / barycentric correction |
| applied to times |
| TIMESYS = ‘TDB’ / time system is Barycentric 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 = 1468.27663914153 / observation start time in BTJD |
| TSTOP = 1490.04694595576 / observation stop time in BTJD |
| DATE-OBS = ‘2018-12-15T18:37:12.438’ / TSTART as UTC calendar date |

(Continued)

| DATE-END = ‘2019-01-06T13:06:26.946’ / TSTOP as UTC calendar date |
| MJD-DEG = 58467.77663914146 / observation start time in MJD |
| MJD-END = 58489.54694595576 / observation start time in MJD |
| TELAPSE = 21.7703068430688 / [d] TSTOP − TSTART |
| LIVETIME = 4.310520749232762 / [d] TELAPSE multiplied by DEADC |
| DEADC = 0.198 / deadtime correction |
| EXPOSURE = 4.310520749232762 / [d] time on source |
| XPOSURE = 356.4 / [s] Duration of exposure |
| TIMEPIXR = 0.5 / bin time beginning = 0 |
| middle = 0.5 end = 1 |
| TIMEDEL = 0.02083333333333333 / [d] time resolution of data |
| INT_TIME = 1.98 / [s] photon accumulation time per frame |
| READTIME = 0.02 / [s] readout time per frame |
| FRAMETIM = 2.0 / [s] frame time (INT_TIME + READTIME) |
| NUM_FRM = 900 / number of frames per time stamp |
| NREADOUT = 720 / number of read per cadence |
| CHECKSUM = ‘cNOAfTM3cTM9cTM9’ / HDU checksum updated |
| 2020-08-05T02:43:40 |
| DATASUM = ‘645682208’ / data unit checksum updated 2020-08-05T02:43:40 |

22 https://bottleneck.readthedocs.io
### A.2. ENSEMBLE HDU-1 (LIGHTCURVE)

| TYPET = | 'XTENSION' | /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 |
| EXTNAME = | 'LIGHTCURVE' | /extension name |
| TMEKEF = | 'SOLAR SYSTEM' | /barycentric correction |
| TIMESYS = | 'TDB' | /time system is Barycentric 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 |
A.3. ENSEMBLE HDU-4 (ENSEMBLE)

XTENSION = 'BINTABLE' / binary table extension
BITPIX = 8 / array data type
NAXIS = 2 / number of array dimensions
NAXIS1 = 12 / length of dimension 1 NAXIS2 = 7 / length of dimension 2
PCOUNT = 0 / number of group parameters
GCOUNT = 1 / number of groups
TFIELDS = 2 / number of table fields
TTYPE1 = 'TIC' / column title: TIC identifier
TFORM1 = 'K' / column format: signed 64-bit integer
TTYPE2 = 'BZETA' / column title: background scale
TFORM2 = 'E' / column format: 32-bit floating point
EXTNAME = 'ENSEMBLE' / extension name
TDISP1 = '110' / column display format
TDISP2 = 'E' / column display format
END

References

Aigrain, S., Collier Cameron, A., Ollivier, M., et al. 2008, A&A, 488, L43
Antoci, V., Cunha, M. S., Bowman, D. M., et al. 2019, MNRAS, 490, 4040
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Audenaert, J., Kuszelewicz, J. S., Handberg, R., et al. 2021, AJ, 162, 209
Baglin, A., Auvergne, M., Barge, P., et al. 2009, in IAU Symp. 253, Transiting Planets, ed. F. Pont, D. Sasselov, & M. J. Holman (Cambridge: Cambridge Univ. Press), 71
Basri, G., Walkowicz, L. M., Batalha, N., et al. 2011, AJ, 141, 20
Bowman, D. M. 2017, Amplitude Modulation of Pulsation Modes in Delta Scuti Stars (Berlin: Springer)
Brahm, R., Hartman, J. D., Jordán, A., et al. 2018, AJ, 155, 112
Buzasi, D. L., Carboneau, L., Hessler, C., Leznano, A., & Preston, H. 2015, IAU, 29, 2256843
Caldwell, D. A., Tenenbaum, P., Twicken, J. D., et al. 2020, RNAAS, 4, 201
Campante, T. L., Schofield, M., Kuszelewicz, J. S., et al. 2016, ApJ, 830, 138
Chaplin, W. J., & Miglio, A. 2013, ARA&A, 51, 353
Feinstein, A. D., Monet, B. T., Foreman-Mackey, D., et al. 2019, PASP, 131, 094042
Garcia, R. A., & Ballot, J. 2019, LRSP, 16, 4
García, R. A., Mathur, S., Fries, S., et al. 2014, A&A, 568, A10
Gililland, R. L., & Brown, T. M. 1988, PASP, 100, 754
Gililand, R. L., Brown, T. M., Christensen-Dalsgaard, J., et al. 2010, PASP, 122, 131
Handberg, R., Hansen, J. S., Pope, B., & Lund, M. N. 2021a, tasoc/photometry: v6.2.5, Zenodo, doi:10.5281/zenodo.5153073
Handberg, R., Lund, M. N., Carboneau, L., Pereira, F., & Reeth, T. V. 2021b, tasoc/corrections: v2.0.1, Zenodo, doi:10.5281/zenodo.5154027
Handberg, R., Lund, M. N., White, T. R., et al. 2021, AJ, 162, 170
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Natur, 585, 357
Hartman, J. D., Bayliss, D., Brahm, R., et al. 2015, AJ, 149, 166
Hebrard, G., Evans, T. M., Alonso, R., et al. 2011, A&A, 533, A130
Hekker, S., & Christensen-Dalsgaard, J. 2017, A&ARv, 25, 1
Holdsworth, D. L., Cunha, M. S., Kurtz, D. W., et al. 2021, MNRAS, 506, 1073
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, 206
Huber, D., Bedding, T. R., Stello, D., et al. 2011, ApJ, 743, 143
Hunter, J. D. 2011, CSE, 9, 90
Jenkins, J. M., Twicken, J. D., McCauliff, S., et al. 2016, Proc. SPIE, 9913, 99132E
Jordán, Á., Brahm, R., Bakos, G. Á., et al. 2014, AJ, 148, 29
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
Lund, M. N., Handberg, R., Kjeldsen, H., Chaplin, W. J., & Christensen-Dalsgaard, J. 2017, EPJDC, 160, 01005
Miglio, A., Chiappini, C., Morel, T., et al. 2013, MNRAS, 429, 423
Nason, G. 2006, Statistics in Volcanology (London: Geological Society of London), 129
Oelkers, R. J., & Stassun, K. G. 2018, AJ, 156, 132
Pascual-Granado, J., Suárez, J. C., Garrido, R., et al. 2018, A&A, 614, A40
Pedrosa, F., Varouguas, G., Gramfort, A., et al. 2011, J. Mach. Learn. Res., 12, 2825
Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., et al. 2018, AJ, 156, 123 Prsa, A., Zhang, M., & Wells, M. 2019, PASP, 131, 068001
Raetz, S., Heras, A. M., Fernández, M., Casanueva, V., & Marka, C. 2019, MNRAS, 483, 824
Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2014, Proc. SPIE, 9143, 20
Schofield, M., Chaplin, W. J., Huber, D., et al. 2019, ApJS, 241, 12
Silva Aguirre, V., Stello, D., Stokholm, A., et al. 2020, ApJL, 889, L34
Smith, J. C., Stumpe, M. C., Smith, J. C., et al. 2009, in IAU Symp. 253, Transiting Planets, ed. F. Pont, D. Sasselov, & M. J. Holman (Cambridge: Cambridge Univ. Press), 71
Stumpe, M. C., Smith, J. C., Van Cleve, J. E., et al. 2012, PASP, 124, 985
Sullivan, P. W., Winn, J. N., Berta-Thompson, Z. K., et al. 2015, ApJ, 809, 77
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261
Waskom, M. 2021, JOSS, 6, 3021
White, T. R., Pope, B. J. S., Antoci, V., et al. 2017, MNRAS, 471, 2882

ORCID iDs

Mikkel N. Lund @ https://orcid.org/0000-0001-9214-5642
Rasmus Handberg @ https://orcid.org/0000-0001-8725-4502
Derek L. Buzasi @ https://orcid.org/0000-0002-1988-143X
Lindsey Carboneau @ https://orcid.org/0000-0003-1001-5137
Oliver J. Hall @ https://orcid.org/0000-0002-0468-4775
Filipe Pereira @ https://orcid.org/0000-0002-2157-7146
Daniel Huber @ https://orcid.org/0000-0001-8832-4488
Daniel Hey @ https://orcid.org/0000-0003-3244-5357
Timothy Van Reeth @ https://orcid.org/0000-0003-2771-1745