Data Validation Time Series File: Description of File Format and Content

KSCI-19079-001
March 23, 2016

NASA Ames Research Center
Moffett Field, CA 94035

Prepared by: _____ Date: 3/23/2016
Susan E. Thompson, Science Office

Approved by: _____ Date: 3/23/16
Dwight Sanderfer, SOC Manager

Approved by: _____ Date: 3/23/16
Michael R. Haas, Science Office Director

Document Control

Ownership
This document is part of the Kepler Project Documentation that is controlled by the Kepler Project Office, NASA/Ames Research Center, Moffett Field, California.

Control Level
This document will be controlled under KPO @ Ames Configuration Management system. Changes to this document **shall** be controlled.

Physical Location
The physical location of this document will be in the KPO @ Ames Data Center.

Distribution Requests
To be placed on the distribution list for additional revisions of this document, please address your request to the Kepler Science Office:

> Michael R. Haas
> Kepler Science Office Director
> MS 244-30
> NASA Ames Research Center
> Moffett Field, CA 94035-1000
> Michael.R.Haas@nasa.gov

The correct citation for this document is: S. E. Thompson 2016, *Data Validation Time Series File: Description of the File Format and Content,* KSCI-19079-001.

DOCUMENT CHANGE LOG

CHANGE DATE	PAGES AFFECTED	CHANGES/NOTES
March 23, 2016		Original release

Contents

1   Introduction

The *Kepler* space mission searches its time series data for periodic, transit-like signatures. The ephemerides of these events, called Threshold Crossing Events (TCEs), are reported in the TCE tables at the NASA Exoplanet Archive (NExScI)[1]. Those TCEs are then further evaluated to create planet candidates and populate the Kepler Objects of Interest (KOI) table, also hosted at the Exoplanet Archive. The search, evaluation and export of TCEs is performed by two pipeline modules, TPS (Transit Planet Search) and DV (Data Validation). TPS searches for the strongest, believable signal and then sends that information to DV to fit a transit model, compute various statistics, and remove the transit events so that the light curve can be searched for other TCEs. More on how this search is done and on the creation of the TCE table can be found in Tenenbaum et al. (2012), Seader et al. (2015), Jenkins (2002). For each star with at least one TCE, the pipeline exports a file that contains the light curves used by TPS and DV to find and evaluate the TCE(s). This document describes the content of these DV time series files, and this introduction provides a bit of context for how the data in these files are used by the pipeline.

The input to TPS is the quarterly PDC (Pre-search Data Conditioning) flux time series and the data quality flags. The first task of TPS is to create a multi-quarter light curve that can be searched. While the details of this process are explained in Tenenbaum et al. (2013), the process can be summarized as follows:

1. Convert the flux from each quarter to a relative flux by dividing by the median flux of the quarter.

2. Fit and remove the strongest series of sine-waves in each quarter.

3. Fill gaps caused by poor or missing data.

TPS then takes this initial light curve (see LC_INIT §2.2) and applies a filter to account for the colored Gaussian noise (see §8.3.3 of Jenkins, 2011) to create a "whitened" light curve (see LC_WHITE §2.2). Basiclay, whitening scales the flux so the standard deviation is one and simplifies the transit search at the expense of distorting the shape of the transit in the data. The search is done by creating a Single Event Statistic (SES) time series: a time series which describes, for each cadence in the whitened time series, the significance (in σ) of a transit-like signal centered on that cadence. The SES time series is calculated by dividing the correlization by the normalization SES time series (see the columns SES_CORR and SES_NORM in §2.3). The normalization SES time series is proportional to the inverse of CDPP (Combined Differential Photometric Precision), the statistic used to evaluate the noise level of the light curve (Jenkins et al., 2010). For more information on the SES and CDPP time series and how they are calculated see Jenkins (2011) and Jenkins (2002). The SES time series is created at 14 different pulse durations ranging from 1.5 – 15 hours. They are folded at periods greater than 0.5 days to create a Multiple Event Statistic (MES) as a function of period and phase. Cadences that fall in gaps are given zero weight when calculating the MES, and cadences near

---

[1]http://exoplanetarchive.ipac.caltech.edu

events such as quarter boundaries are given less weight (see the DEWEIGHTS column in §2.3). The period and phase with the maximum MES value, if it passes other vetoes (Twicken et al., 2016; Seader et al., 2015; Tenenbaum et al., 2013), becomes a TCE. Finally, the TCE is sent to DV where the signal is fit with a transit model, the transits are removed, and the light curve is sent back to TPS to be searched for another periodic, transit-like signal.

The DV time series files contain light curves and other time series used by TPS to find and evaluate TCEs. It is one of many products created for each TCE from the results of the TPS/DV run. The full set of products for each TCE are:

- TCE table, including many metrics to evaluate the quality of the TCE

- DV report (one per target that has a TCE)

- DV one-page summary (one per TCE)

- DV time series file (one per target that has a TCE)

- TCERT (Threshold Crossing Event Review Team) vetting report

2   File Format and Content

The project exports one DV time series Flexible Image Transport System (FITS) file for each unique Kepler target that produces at least one TCE. The file contains a primary header (§2.1), one extension per TCE found on the specified target (§2.2), and a statistics extension (§2.3). The TCE extensions are numbered according to the order in which the TCEs were found. Each run of DV will contain a new set of TCEs and thus a new set of time series files and reports.

The file name of the DV time series files is formatted as follows:

$$\mathrm{kplr}{<}KeplerId{>}\text{-}{<}yyyymmddhhmmss{>}\_\mathrm{dvt.fits}$$

where *KeplerId* is the Kepler Input Catalog (KIC) Identification number of the searched star, except when a TCE is found on a custom aperture, and *yyyymmddhhmmss* is the date that the DV exporter was run.

2.1   Primary Header

The primary header contains keywords that describe the overall properties of the target that produced the TCEs. Those not standard to the FITS format, or not easily explained in the keyword's comment field, are listed below. All stellar information, such as effective temperature and gravity, are those DV used when performing the transit fit to the TCE.

- CREATOR – The program name and pipeline job number used to run DV.

- OBJECT – The string version of the Kepler Identification number. For all cases, except where a TCE was found on a custom aperture, this value is the KIC Identification number. The value matches the *KeplerId* found in the file name.

- KEPLERID – The object's Kepler Identification number formatted as an integer.

- SKYGROUP – The skygroup containing the star. This is the same as the channel number during season 2.

- DATA_REL – A unique number specifying the data release. This number matches the data release number of the light curve and target pixel data used to find these TCEs. It also matches the data release number of the TCE table and the associated KOIs at NExScI.

2.2   TCE Extensions

One binary FITS extension is provided for each TCE found on the specified target. The header contains information about the transit model fit performed at the TCE's period, epoch and depth. The data columns contain the light curves at different stages of the planet search or evaluation. For instance, the quarter-stitched light curve used to search for planets is available in its whitened (LC_WHITE) and un-whitened form (LC_INIT). The median detrended light curve shown in the DV one-page reports is also available (LC_DETREND). Each extension contains the following columns:

TCE Column Descriptions

1. TIME – [float] The Barycentric Julian date of each cadence minus the offset 2454833. This value is available for all cadences in the file, even if no data is available at those times.

2. TIMECORR – [float] The Barycentric correction plus the time slice correction applied to each cadence in order to convert between time reported by the Kepler spacecraft and the Barycentric time available in the TIME column (Thompson et al., 2016). This value is available for all cadences in the file, even if no data is available at those times.

3. CADENCENO – [integer] The cadence number, a unique value that is incremented by one with each cadence. This value is available for all cadences in the file, even if no data is available at those times.

4. PHASE – [float] The phase of each cadence given the period and epoch of the identified TCE (see they keywords TPERIOD and TEPOCH). The phases span $-\frac{1}{4}P$ to $+\frac{3}{4}P$, where P is the period in days listed under the keyword TPERIOD. The transit occurs at a value of zero days.

5. LC_INIT – [float] The initial light curve discussed in the introduction. This light curve has been quarter-stitched, harmonic-removed and has a mean of zero.

6. LC_INIT_ERR – [float] The estimated error for LC_INIT.

7. LC_WHITE – [float] The whitened light curve that was searched by TPS to find this TCE. This is LC_INIT after the whitening filter is applied.

8. LC_DETREND – [float] The quarter-stitched, median detrended, relative-flux light curve, as shown in the one page summary reports.

9. MODEL_INIT – [float] The transit model calculated for the initial light curve.

10. MODEL_WHITE – [float] The whitened transit model calculated for the whitened light curve.

Keywords

The keywords added to each TCE extension describe the TCE and the transit fit model. The header also provides information about the observation times available in this extension. The time related keywords will be the same in all TCE extensions, however the transit fit model and TCE ephemeris will be different for each TCE extension. We describe below those keywords that are unique to the DV time series files and not well described by the comments in the header. Other keywords are described in the Kepler Archive Manual (Thompson et al., 2016) or are FITS standard.

- TPERIOD – The period of the detected TCE in days.

- TEPOCH – The epoch of the detected TCE in BJD minus 2454833.0.

- TDEPTH – The depth of the best fit transit model in parts per million.

- TSNR – The signal-to-noise of the best fit transit model.

- TDUR – The duration in hours of the best fit transit model.

- INDUR – The ingress duration in hours of the best fit transit model.

- DRRATIO – The ratio of the planet distance and the stellar radius.

- RADRATIO – The ratio of the planet radius and the stellar radius.

- PRADIUS – The planet radius in Earth radii of the best fit transit model.

- MAXMES – The MES value for the TCE.

- MAXSES – The maximum SES value for the TCE.

- NTRANS – The number of transits with non-gapped data.

- CONVRGE – A true/false (T/F) flag indicating whether the full planet model converged.

- MEDDETR – The length scale used by the median detrender, in hours, to create the LC_DETREND light curve.

2.3   Statistics Extension

The statistics extension contains times series data that apply to the object in general and are not specific to a particular TCE. The extension contains the original PDC light curve for all quarters, as is available in the *Kepler* light curve FITS files. It also contains the SES and CDPP time series, for all pulse durations. The CDPP, SES_CORR, and SES_NORM columns reflect the state of the respective time series after all TCEs are removed from the data. The Bootstrap Statistics, available in the TCE Table at NExScI, are calculated using these final SES time series.

Statistics Column Descriptions

1. TIME – [float] The Barycentric Julian date of each cadence minus the offset 2454833. This is the same as in the TCE extension, see §2.2.

2. TIMECORR – [float] The Barycentric correction plus the time slice correction. This is the same as in the TCE extension, see §2.2.

3. CADENCENO – [integer] The cadence number. This is the same as in the TCE extension, see §2.2.

4. PDCSAP_FLUX – [float] The PDC flux measurement for every cadence in units of electrons per second. This matches exactly what is available in the light curve files on a per quarter basis.

5. PDCSAPFLUX_ERR – [float] The error in the PDC flux measurement. This matches exactly what is available in the light curve files on a per quarter basis.

6. RESIDUAL_LC – [float] The residual light curve. This light curve is the LC_INIT time series with the transits for all TCEs removed.

7. DEWEIGHTS – [float] A value between zero and one indicating the weight of the data in that cadence when searching for the first TCE event.

8. QUALITY – [integer] A bit flag indicating when certain events occured that may impact the quality of the data. This matches exactly what is available in the light curve files on a per quarter basis. See the *Kepler* Archive Manual (Thompson et al., 2016) for a detailed list of the events and which bit is set for each.

9–22. SES_CORR_*d*ur – [float] The correlation SES time series for each of the 14 searched pulse durations. The duration is encoded in the name in hours using an underscore for the decimal point.

23–36. SES_NORM_*d*ur – [float] The normalization SES time series for each of the 14 searched pulse durations. The duration is encoded in the name in hours using an underscore for the decimal point.

37–50 CDPP_*dur* – [float] The CDPP time series statistic as measured for each of the 14 searched pulse durations. The duration is encoded in the name in hours using an underscore for the decimal point.

References

Jenkins, J. 2011, in "Kepler Data Processing Handbook, KSCI-19081-001", ed. M. N. Fanelli

Jenkins, J. M. 2002, ApJ, 575, 493

Jenkins, J. M., Chandrasekaran, H., McCauliff, S. D., et al. 2010, in Proceedings of the SPIE, Vol. 7740, Software and Cyberinfrastructure for Astronomy, 77400D

Seader, S., Jenkins, J. M., Tenenbaum, P., et al. 2015, ApJS, 217, 18

Tenenbaum, P., Christiansen, J. L., Jenkins, J. M., et al. 2012, ApJS, 199, 24

Tenenbaum, P., Jenkins, J. M., Seader, S., et al. 2013, ApJS, 206, 5

Thompson, S. E., Fraquelli, D., van Cleve, J., & Caldwell, D. 2016, Kepler Archive Manual, KDMC-10008-006

Twicken, J. D., et al. 2016, in prep