Kepler Data Release 21 Notes
Q0–Q14

KSCI-19061-001
Data Analysis Working Group (DAWG)
Susan E. Thompson (Editor)

Prepared by: Susan E. Thompson Date: 5/7/13
Susan E. Thompson, Kepler Science Office, for the DAWG (page 4)

Approved by: JChristiansen Date: 5/7/13
Jessie L. Christiansen, Editor-in-Chief

Approved by: J Jenkins Date: 5-7-13
Jon Jenkins, Co-I for Data Analysis & DAWG Lead

Approved by: Michael R. Haas Date: 5/7/13
Michael R. Haas, Kepler 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 place 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 Data Characteristics Handbook and accompanying Data Release Notes are the collective effort of the Data Analysis Working Group (DAWG), composed of Science Office (SO), Science Operations Center (SOC), and Guest Observer (GO) Office members as listed below:

Jon Jenkins, Chair
Doug Caldwell, Co-Chair
Barclay, Thomas
Bryson, Stephen T.
Burke, Christopher J.
Campbell, Jennifer
Catanzarite, Joseph
Christiansen, Jessie L.
Clarke, Bruce D.
Coughlin, Jeffrey L.
Girouard, Forrest
Haas, Michael R.
Ibrahim, Khadeejah
Klaus, Todd
Kolodziejczak, Jeffery (MSFC)
Li, Jie
McCauliff, Sean D.
Morris, Rob
Mullally, Fergal
Quintana, Elisa V.
Rowe, Jason
Sabale, Anima
Seader, Shawn
Smith, Jeffrey Claiborne
Still, Martin D.
Tenenbaum, Peter G.
Thompson, Susan E.
Twicken, Joe
Uddin, Akm Kamal

The correct citation for this document is: S. E. Thompson, J. L. Christiansen, J. M. Jenkins, D. A. Caldwell, T. Barclay, S. T. Bryson, C. J. Burke, J. R. Campbell, J. Catanzarite, B. D. Clarke, J. L. Coughlin, F. Girouard, M. R. Haas, K. Ibrahim, T. C. Klaus, J. J. Kolodziejczak, J. Li, S. D. McCauliff, R. Morris, F. Mullally, E. V. Quintana, J. Rowe, A. Sabale, S. Seader, J. C. Smith, M. D. Still, P. G. Tenenbaum, J. D. Twicken, and A. K. Uddin, 2013, Kepler Data Release 21 Notes (KSCI-19061-001).

Contents

1 Introduction

These Data Release Notes provide information specific to the current reprocessing and re-export of Q0–Q14 data. Specifically, this data release provides an improved version of the PDC light curves and corrects the barycentric times so they include the conversion to TDB (Barycentric Dynamical Time). As such, these Notes are an addendum to the Data Release Notes previously provided for each of these quarters (see Table 1). Specific data events relevant to each quarter, i.e. Safe Modes, Corona Mass Ejections, Attitude Tweaks etc., may be found in the previous Data Release Notes. Previous notes relevant to the PDC module or the archive FITS formats have been superseded by these Data Release Notes.

The data products included in this data release include light curve files, FFIs, and CBVs. The target pixel files, background, ARP and collateral data files have not been reprocessed or re-exported. As a result, the barycentric times in the target pixel and background files for Q0–Q14 will remain incorrect until these data are re-exported in the future. (The collateral and ARP data files do not contain barycentric time.)

Table 1: Previous Data Release Numbers and SOC Pipeline version numbers by Quarter

Quarter(s)	DRN number	CAL and PA SOC Pipeline version
0–4	14	8.0
5–8	16	8.1
9	12	8.0
10	13	8.0
11	15	8.0
12	17	8.1
13	18	8.2
14	19	8.3

1.1 The SOC Pipeline for Q0–Q14

Data Release 21 was processed with the SOC Pipeline 8.3 for the PDC module and 9.0 for the exporter module. The SOC Pipeline number for the CAL and PA modules are listed in Table 1. The data has not been reprocessed through the CAL and PA modules, so the SAP_FLUX data are unchanged from the previous deliveries listed in Table 1. For details on how Kepler processes the data through the front-end of the pipeline (modules CAL, PA, PDC), please see the Data Processing Handbook (Fanelli et al., 2011). Notable changes and improvements to the pipeline for this data release include the following:

- The PDC (Presearch Data Conditioning) module of the pipeline has implemented multi-scale MAP (msMAP) for long cadence data and quickMAP for short cadence data. For more information on how these affect the PDC light curves, see Section 3.1.
- The export module (AR) of the pipeline has been updated to ensure that all barycentric times are exported in the TDB time system. See Section 3.2 for more information.
- Since PDC was processed with SOC Pipeline 8.3 for this data release, the PDC related keywords described in DRN 20 (Q15) are not populated as part of this data release. This includes the PDC method, earth point goodness metric, and sudden pixel sensitivity dropout keywords. However, data release 21 data products were created with SOC pipeline 9.0, so the PDC keywords are present but do not contain any information.

2 Data Quality in Q0–Q14

We present no new CDPP statistics or data anomaly flags for these data release notes. See previous data release notes for these statistics based on the previous PDC processing.

The CDPP metrics (3, 6 and 12 hour) for individual targets may be found in the headers of the light curve files. When comparing values with previous data releases, remember that the CDPP calculation was performed with a new CDPP algorithm, introduced in SOC 8.3 (see Data Release 19). This algorithm generally has the effect of slightly reducing the CDPP metric.

3 Notable Features of the Q0–Q14 Data

In this section we only discuss features of the data that have changed as a result of this reprocessing and re-export. For all other details, see the DRN listed in Table 1 and the Data Characteristics Handbook (Christiansen et al., 2012).

3.1 A New Implementation of PDC and its Effects on the Data

3.1.1 Multi-Scale MAP

The major improvement to PDC for Long Cadence data is multi-scale MAP (msMAP). msMAP is a wavelet-based band-splitting framework for removing systematics from the light curves. It decomposes each light curve into three characteristic length-scales (or bands). This reduces the chance that an astrophysical signal will be accidentally removed and that systematic affects will be preserved. The Cotrending Basis Vectors (CBVs) are calculated separately for each band, but then each of the three bands is handled differently. The longest band ($\geq$ 1024 cadences, i.e., $\geq$21 days) performs a simple robust fit to the CBVs because it is difficult to distinguish between astrophysical signals and systematics on length scales similar in length to a quarter. The middle band performs a MAP fit. The shortest band ($<$ 4 cadences, i.e., $<$ 2 hours) simply preserves all signals to output.

Multi-scale MAP does not always perform better than regular MAP (Smith et al., 2012; Stumpe et al., 2012). PDC therefore automatically compares the performance of msMAP to the regular MAP algorithm, and chooses the reduction with the better performance based on the Goodness Metric calculated by PDC (see Smith et al., 2012; Stumpe et al., 2012). Multi-scale MAP is preferred for approximately 90% of targets. In future releases, the keywords in the headers of the FITS files will indicate which PDC light curve was chosen for individual targets, however this feature is not available in this data release.

CBVs for the different bands are not provided to the end user. The CBVs provided with this data release are calculated using regular MAP, as they have been in the past.

3.1.2 Application of MAP to Short Cadence Data (quickMAP)

We now apply a version of MAP, called quickMAP, to all short cadence light curves. quickMAP performs in much the same way as regular MAP does for long cadence light curves (Smith et al., 2012). The primary difference arises from the small number of short cadence targets, making it impossible to construct useful basis vectors and priors directly from the short cadence targets. Instead, quickMAP interpolates the long cadence basis vectors to the short cadence exposure time and uses the prior information calculated for the target's long cadence light curve. As a result of the interpolation, no systematics with frequencies higher than the long cadence Nyquist frequency are removed (see Christiansen et al., 2012, §5.12). There is also the potential that signals could be introduced above the long cadence Nyquist Frequency; however, Fourier analysis has shown there to be virtually no introduced signals above this frequency.

Short cadence PDC light curves will be more similar to long cadence PDC light curves than in previous data releases. However, the algorithms and data sets are not identical, so there will be noticeable differences for individual targets. For example the long period attenuation discussed below occurs at shorter periods for short cadence data. Also, msMAP for long cadence data does a better job of removing Earth point recoveries.

3.1.3 Stellar Preservation in PDC Data – Attenuation of Long Timescale Signals

Users interested in events with durations comparable to the length of a data set (3 months for long cadence, 1 month for short cadence) should not use the PDC light curves. All PDC algorithms have difficulty distinguishing between astrophysical and instrumental signals on these timescales. msMAP assumes all long period signals that are correlated with the lowest band's basis vectors are systematic and removes them. We quantified the effect on an astrophysical signal by injecting sinusoidal signals of different periods and amplitudes equal to 0.1% times the median flux level into real data. We then measured the attenuating effect of the PDC processing. We show the results in Figure 1 below. Signals with periods less than ~3 days are very well preserved, while periods greater than 20 days are almost entirely removed or shifted and distributed to other frequencies. Note that transits, or any sharp, periodic features, are not affected by this corruption. Though the interval between transits is long, their durations are always short. A more detailed analysis of how well stellar signals are preserved is currently in development.

3.1.4 Outliers and SPSDs Detected During Transits

There are a small number of instances where the outlier or SPSD (Sudden Pixel Sensitivity Dropouts) detector will incorrectly trigger on deep transits. This can cause transit depths to decrease or be corrupted. We are working on a solution to this problem for a future release.

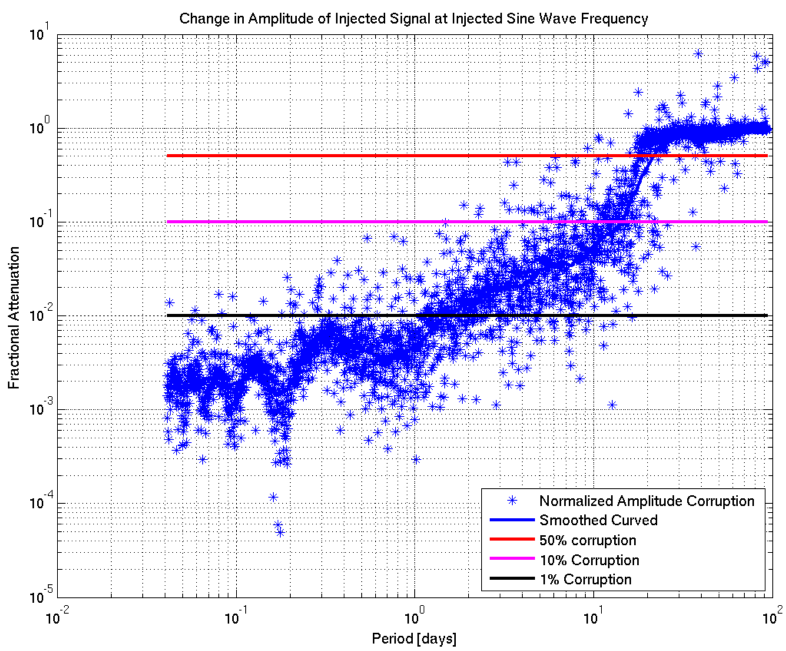

Figure 1: Signal attenuation as a function of period in msMAP. Sine waves of different periods were injected into long cadence light curves before being processed with PDC. Their amplitudes before and after processing were compared. Each blue symbol represents the fractional change in amplitude of a given injected sine wave. Short period signals were attenuated by less than 1%, while long period signals were entirely removed.

3.2 Time and Time Stamps

This data release has corrected the ∼66 s timing error present in the barycentric times previously reported by Kepler. For a more detailed description of the issue, please see Data Release Notes 19. Once ingest is complete, all Q0-Q14 light curve files and FFIs will have barycentric times reported in terrestrial dynamic time (TDB).

To ensure that you have data with times reported in TDB, check the keywords in the primary header of the FITS files. The times have the TDB correction if the FILEVER keyword is ≥ 5.0 in the light curve files and ≥ 3.0 for the calibrated FFIs. From Q15 onward, all data products will have barycentric times reported in the TDB time system.

WARNING: The barycentric times in the target pixel files will not be re-exported as part of this data release. Because future target pixel files will have the correct times, there will be a discontinuity in time between the Q0–Q14 target pixel files and future quarters until the Q0–Q14 target pixel files are re-exported. Users of the target pixel files are encouraged to use the time stamps from the light curve files until this time correction is propagated to these files.

References

Christiansen, J. L., Van Cleve, J. E., Jenkins, J. M., Caldwell, D. A., Barclay, T., Bryson, S., Burke, C. J., Twicken, J. D., & Uddin, A. K. (2012). Kepler Data Characteristics Handbook. *KSCI-19040-003*.

Fanelli, M. N., Jenkins, J. M., Bryson, S. T., Quintana, E. V., Twicken, J. D., Wu, H. W., Tenenbaum, P., Allen, C. L., Caldwell, D. A., Chandrasekaran, H., Christansen, B. D., & Uddin, A. K. (2011). Kepler Data Processing Handbook. *KSCI-19081-001*.

Smith, J. C., Stumpe, M. C., Van Cleve, J. E., Jenkins, J. M., Barclay, T. S., Fanelli, M. N., Girouard, F. R., Kolodziejczak, J. J., McCauliff, S. D., Morris, R. L., & Twicken, J. D. (2012). Kepler Presearch Data Conditioning II - A Bayesian Approach to Systematic Error Correction. *PASP*, *124*, 1000–1014.

Stumpe, M. C., Smith, J. C., Van Cleve, J. E., Twicken, J. D., Barclay, T. S., Fanelli, M. N., Girouard, F. R., Jenkins, J. M., Kolodziejczak, J. J., McCauliff, S. D., & Morris, R. L. (2012). Kepler Presearch Data Conditioning I - Architecture and Algorithms for Error Correction in Kepler Light Curves. *PASP*, *124*, 985–999.