



## *Kepler Data Release 5 Notes*

KSCI-19045-001

Data Analysis Working Group (DAWG)

*Jeffrey Van Cleve, Editor*

### Data Release 5 for Quarters Q0 and Q1

Q		First Cadence MJD midTime	Last Cadence MJD midTime	First Cadence UT midTime	Last Cadence UT midTime	Num Cads
0	LC	54953.038	54962.744	5/2/09 0:54	5/11/09 17:51	476
0	SC	54953.028	54962.754	5/2/09 0:40	5/11/09 18:05	14280
1	LC	54964.011	54997.481	5/13/09 0:15	6/15/09 11:32	1639
1	SC	54964.001	54997.491	5/13/09 0:01	6/15/09 11:47	49170

*The first Release of Planetary Search targets to the General Public*

Prepared by: \_\_\_\_\_ Date \_\_\_\_\_  
Jeffrey Van Cleve, Kepler Science Office, for the DAWG (next page)

Approved by: \_\_\_\_\_ Date \_\_\_\_\_  
Jon Jenkins, Co-I for Data Analysis & DAWG Lead

Approved by: \_\_\_\_\_ Date \_\_\_\_\_  
Michael R. Haas, Science Office Director

These Notes are the collective effort of the Data Analysis Working Group (DAWG), composed of Science Office (SO), Science Operations Center (SOC), and Science Team (ST) members as listed below:

Jon Jenkins\*, Chair

Doug Caldwell\*, Co-Chair

Allen, Christopher L.

Bryson, Stephen T.

Clarke, Bruce D.

Cote, Miles T.

Dotson, Jessie L.

Gilliland\*, Ron (STSci)

Girouard, Forrest

Haas, Michael R.

Hall, Jennifer

Ibrahim, Khadeejah

Klaus, Todd

Kolodziejczak, Jeff (MSFC)

Li, Jie

McCauliff, Sean D.

Middour, Christopher K.

Quintana, Elisa V.

Tenenbaum, Peter G.

Twicken, Joe

Uddin, Akm Kamal

Van Cleve, Jeffrey

Wohler, Bill

Wu, Hayley Y.

\*Science Team

Affiliations are Kepler Science Office or Science Operations Center unless otherwise noted.

## 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

## Table of Contents

<u><i>Prefatory Admonition to Users</i></u> .....	7
1. Introduction .....	8
2. Release Description .....	10
2.1 Summary of Contents .....	11
2.2 Pipeline Changes Since Previous Release .....	11
2.2.1 CAL: calibrated pixels .....	11
2.2.2 PA: uncorrected light curves and centroids .....	11
2.2.3 PDC: corrected light curves .....	11
2.3 Pipeline Changes Since the Previous Release of this Data Set .....	12
2.3.1 CAL: calibrated pixels .....	12
2.3.2 PA: uncorrected light curves and centroids .....	12
2.3.3 PDC: corrected light curves .....	12
3. Current Evaluation of Performance .....	13
3.1 Overall .....	13
3.2 Changes in Performance Since Previous Release .....	15
3.3 Known Calibration Issues .....	15
4. Data Delivered – Processing History .....	16
4.1 Overview .....	16
4.2 Pixel-Level Calibration (CAL) .....	17
4.3 Photometric Analysis (PA) .....	18
4.4 Pre-Search Data Conditioning (PDC) .....	19
4.4.1 Description .....	20
4.4.2 Performance .....	21
4.4.3 Removal of Astrophysical Signatures .....	25
5. Lost or Degraded Data .....	28
5.1 Momentum Desaturation .....	28
5.2 Reaction Wheel Zero Crossings .....	29
5.3 Data Anomalies .....	31
5.3.1 Safe Mode .....	31
5.3.2 Loss of Fine Point .....	31
5.3.3 Pointing Drift and Attitude Tweaks .....	31
5.3.4 Downlink Earth Point .....	32
5.3.5 Manually Excluded Cadences .....	32
5.3.6 Anomaly Summary Table .....	33

6.	Systematic Errors .....	34
6.1	Argabrightening .....	34
6.2	Variable FGS Guide Stars .....	37
6.3	Pixel Sensitivity Dropouts .....	39
6.4	Focus Drift and Jitter .....	40
6.5	Requantization Gaps .....	43
6.6	Spurious Frequencies in SC Data with Spacing of 1/LC .....	44
6.7	Known Erroneous FITS header keywords .....	45
7.	Data Delivered – Format .....	46
7.1	FFI .....	46
7.2	Light Curves .....	46
7.3	Pixels .....	47
7.4	Time and Time Stamps .....	47
7.4.1	Overview .....	47
7.4.2	Time Stamp Definitions for Release 5 .....	48
7.4.3	Caveats and Uncertainties .....	49
7.5	Future Formats Under Discussion .....	49
8.	References .....	50
9.	List of Acronyms and Abbreviations .....	51
10.	Contents of Supplement .....	54
10.1	Pipeline Instance Detail Reports .....	54
10.2	Thermal and Image Motion Data for Systematic Error Correction .....	54
10.2.1	Mod.out Central Motion .....	54
10.2.2	Average LDE board Temperature .....	55
10.2.3	Reaction Wheel Housing Temperature .....	55
10.2.4	Launch Vehicle Adapter Temperature .....	55
10.3	Background Time Series .....	55
10.4	Flight System Events .....	56
10.5	Calibration File READMEs .....	56

### **Prefatory Admonition to Users**

Kepler is a statistical mission dedicated to measuring the probability that Sun-like stars harbor Earth-like planets. Towards that end, the Pipeline is effective at removing instrumental signatures from the great majority of targets which are either astrophysically quiet or coherently variable. While it is not possible to perfectly preserve general stellar variability on long timescales with amplitudes comparable to or smaller than the instrumental systematics, significant effort has been expended to preserve the natural variability of targets in the corrected light curves in order to enable astrophysical exploitation of the Kepler data.

However, the current Pipeline is known to distort or remove astrophysically interesting signals or introduce excess noise for some targets. Therefore, users must understand the data analysis performed and ascertain its impact on their science before publishing results based on Data Release 5. **Users are strongly encouraged to compare the uncorrected ('raw') and corrected light curves, examine the centroid time series, and read these Release Notes to judge which products are best for their purposes.** Users considering a partial reanalysis of the uncorrected light curves will find ancillary engineering data and image motion time series provided for that purpose in the Supplement. The Science Office advises against publication of these Release 5 light curves without such careful consideration by the end user and dialog with the Science Office or Guest Observer Office as appropriate.

Users are encouraged to notice and document artifacts, either in the raw or processed data, and report them to the Science Office at [kepler-scienceoffice@lists.nasa.gov](mailto:kepler-scienceoffice@lists.nasa.gov).



*Users who neglect this Admonition risk seeing their works crumble into ruin before their time.*

## 1. Introduction

These notes have been prepared to give Kepler users of the Multimission Archive at STScI (MAST) a summary of flight system events occurring during the collection of the released data set which can impact the quality of the data, and a summary of the performance of the data processing pipeline version used on that data set in a particular Release. The Notes will be updated for each release of data to the public archive and placed on MAST along with other Kepler documentation, at [http://archive.stsci.edu/kepler/data\\_release.html](http://archive.stsci.edu/kepler/data_release.html). For the first time, these Notes describe the release of reprocessed data. These data were collected during the first 44 days of the science mission, and are now released to the general public after the expiration of the exclusive use period. The first processing of this data set was described in the Release 2 Notes (KSCI-19042).

The Notes are not meant to supplant the following documents, which are also needed for a complete understanding of the Kepler data:

1. **Kepler Instrument Handbook** (KIH, KSCI-19033) provides information about the design, performance, and operational constraints of the Kepler hardware, and an overview of the pixel data sets available. It was released on July 15, 2009, and is publicly available on MAST. Users will need to be familiar with the material in Sections 2 and 4.2-4.5 of the KIH to fully benefit from these Notes.
2. **Kepler Data Analysis Handbook** (KDAH) describes how these pixel data sets are transformed into photometric time series by the Kepler Science Pipeline, the theoretical basis of the algorithms used to reduce data, and a description of residual instrument artifacts after Pipeline processing. The initial release of the KDAH, focused on the needs of MAST users and Guest Observers (GOs), is expected about 9/1/2010. Until the KDAH is available, users seeking a discussion of Pipeline processing at a deeper level of detail than that provided in these Notes are directed to the SPIE papers (Refs. 3-5), which will be available from MAST on 6/15/2010 and from SPIE (<http://spie.org/>) on or about 8/1/2010.
3. **Kepler Archive Manual** (KDMC-10008) describes file formats and the availability of data through MAST. The Archive Manual is available on MAST
4. **Kepler Mission Special Issue of Astrophysical Journal Letters** (Volume 713, Number 2, 2010 April 20) contained several papers providing background on mission definition (Ref. 11), target selection (Ref. 12), science operations (Ref. 13), the Kepler point spread function (Ref. 14), instrument performance (Ref. 15), and the data processing pipeline (Ref. 9). Two papers discuss the characteristics of the Long Cadence data (Ref. 7), and Short Cadence data (Ref. 8) respectively. Numerous additional papers also provide early science results in both planet detections and asteroseismology, placing the use of Kepler data in context.

Users unfamiliar with the data processing pipeline should read Section 4 first, then the ApJ papers, then the SPIE papers. A list of acronyms and abbreviations appears in Section 9. Questions remaining after a close reading of these Notes and the Instrument Handbook may be addressed to [kepler-scienceoffice@lists.nasa.gov](mailto:kepler-scienceoffice@lists.nasa.gov).

A sentence at the start of a Section will indicate if it is "recycled" from earlier Notes. If the Notes pertaining to a given data Release are revised, they will be reapproved for release and given an incremented document number KSCI-190XX-00n. n starts at 1 for the original version of the notes for a data Release. Reference to Release Notes will refer to the most recent version (highest n) unless otherwise stated.

Data which would be unwieldy to print in this document format are included in a tar file, the Data Release Notes #5 Supplement, which will be released with this document. Supplement files are called out in the text, and a README file in the tar file also gives a brief description of the files contained. All supplement files are either ASCII or FITS format, though some are also provided



as MATLAB \*.mat files for the convenience of MATLAB users. The contents of the Supplement are described in Section 10.

**Dates, Cadence numbers, and units:** Each set of coadded and stored pixels is called a *Cadence*, while the total amount of time over which the data in a Cadence is coadded is the *Cadence period*, which in the case of the flight default operating parameters is 1766 s = 0.49 h, or 270 frame times for Long Cadence, and 58.85 s or 9 frame times for Short Cadence. Cadences are absolutely and uniquely enumerated with *Cadence interval numbers* (CIN), which increment even when no Cadences are being collected, such as during downlinks and safe modes. The *relative Cadence index* (RCI) is the Cadence number counted from the beginning of a quarter (LC) or month (SC). For example, the first LC of Q1 would have an RCI = 1 and CIN = 1105 while the last LC of Q1 has RCI = 1639 and CIN = 2743. Figures, tables, and supplement files will present results in CIN, RCI, or MJD, since MJD is the preferred time base of the Flight System and Pipeline, and can be mapped one-to-one onto CIN or RCI. On the other hand, the preferred time base for scientific results is Barycentric Julian Date (BJD); the correction to form BJD is done on a target-by-target basis in the files users download from MAST, as described in detail in Section 7.4. Unless otherwise specified, the MJD of a Cadence refers to the time at the midpoint of the Cadence. Data shown will be for Q0 and Q1, unless otherwise indicated in the caption. Flux time series units are always the number of detected electrons per Long or Short Cadence.

## 2. Release Description

A *data set* refers to the data type and observation interval during which the data were collected. The observation interval is usually a *quarter*, indicated by Q[n], though Q0 and Q1 are 10 days and one month, respectively, instead of 3 months as will be the case for the rest of the mission. The *data processing* descriptor is the internal Kepler Science Operations (KSOP) ticket used to request the data processing. The KSOP ticket contains a "Pipeline Instance Report," included in the Supplement, which describes the version of the software used to process the data, and a list of parameter values used. Released software has both a release label, typically of the form m.n, and a revision number (preceded by "r") precisely identifying which revision of the code corresponds to that label. For example, the code used to produce Data Release 5 has the release label "SOC Pipeline 6.1" and the revision number r36439. Unreleased software will, in general, have only a revision number for identification.

The same data set will, in general, be reprocessed as the software improves, and will hence be the subject of multiple releases. The combination of data set and data processing description defines a *data product*, and a set of data products simultaneously delivered to MAST for either public or proprietary (Science Team or GO) access is called a *data release*. The first release of data products for a given set of data is referred to as "new," while subsequent releases are referred to as "reprocessed." Release 5 is a reprocessing of the same data set as Release 2.

Data products are made available to MAST users as FITS files, described in the Kepler Archive Manual and Section 7 of these Notes. While the Kepler Archive Manual refers to light curves which have not been corrected for systematic errors as 'raw', in these Notes they will be referred to as 'uncorrected' since the uncorrected light curves are formed from calibrated pixels, and 'raw' will refer only to the pixel values for which only decompression has been performed. The relationship of Pipeline outputs and MAST files is shown in Figure 2. The keywords `DATA_REL = 5` has been added to the FITS headers so users can unambiguously associate Release 5 FITS files with these Notes. In addition, the keyword `QUARTER` has been added; in the case of Release 5, all light curve files have `QUARTER = 0` or `1`.

Data Release 5 was produced with released code, with formal verification and validation of the pipeline and the resulting data products. The resulting data products are a substantial improvement over Release 2, which used pipeline code that was developed before launch. While the Kepler data analysis Pipeline continues to evolve to adapt to the performance of the flight system and our understanding of the data, the rate of evolution is expected to be slower in the second year of the mission than in the first, with major upgrades on a roughly annual basis.

## 2.1 Summary of Contents

**Table 1: Contents of Release 5. CIN is the Cadence interval number described in Section 1. The Pipeline Instance ID (PID) for CAL, PA, and PDC is shown. Refer to Section 7.4 for a discussion of time and time stamps.**

Q	PipeIDs	PipeIDs			First Cadence MJD midTime	Last Cadence MJD midTime	CIN Start	CIN end	Num Cads
		CAL	PA	PDC					
0	LC	1336	1356	1400	54953.038	54962.744	568	1043	476
0	SC	1396	1396	1396	54953.028	54962.754	5500	19779	14280
1	LC	1257	1257	1416	54964.011	54997.481	1105	2743	1639
1	SC	1376	1376	1437	54964.001	54997.491	21610	70779	49170

All Release 5 Cadence data were processed under KSOP-435 with SOC Pipeline 6.1, revision number r36439, and were first released 9/15/2009 as Release 2 using Pipeline 5 software. Other previous Releases include the first Releases of Q2 (Release 3) and Q3 (Release 4) data.

### Notes

The “golden” FFIs collected just before Q0 have not been reprocessed since Release 2. FFIs are available from the MAST public ftp site: <http://archive.stsci.edu/pub/kepler/ffi/>

## 2.2 Pipeline Changes Since Previous Release

This Section describes changes in the Pipeline code and parameters since the previous release (the Q3 data in Release 4), while Section 2.3 summarizes the changes in the Pipeline since the previous release of this data set (Release 2). Changes are listed by Pipeline module outputs, and the corresponding data products on MAST. The software modules comprising the science data analysis Pipeline are described briefly in Section 4. Users unfamiliar with the Pipeline should read Section 4 before reading this Section. The PA and PDC versions and input parameters can be unambiguously referenced by their Pipeline Instance Identifier (PID), shown in Table 1.

### 2.2.1 CAL: calibrated pixels

No significant changes to CAL were made between Release 4 and Release 5.

### 2.2.2 PA: uncorrected light curves and centroids

There were no significant changes to PA between Release 4 and Release 5.

### 2.2.3 PDC: corrected light curves

There were no significant changes to PDC itself between Release 4 and Release 5. However, the following parameters were changed:

1. No LDE board temperatures were used for cotrending, only motion polynomials. The board temperatures were not used since there are no thermal transients to be corrected, because the short time scale heater cycle (Section 6.4) is generally very well corrected by cotrending with the motion polynomials, and noise results indicate that the corrected light curves were previously being overfitted.
2. The center-to-peak variability detection threshold was reduced to 0.25% from 0.5%

3. Cotrending is allowed to only add 5% to the noise instead of 20%, before being rejected.
4. Cotrending performance is assessed over 3 day intervals instead of 1 day.

### **2.3 Pipeline Changes Since the Previous Release of this Data Set**

This Section gives a summary of changes in the Pipeline since the Q0 and Q1 data were released as Release 2. Since the changes are substantial, however, users are advised to read Section 4 and Refs 3-5.

#### **2.3.1 CAL: calibrated pixels**

No major changes were made to CAL between Release 2 and Release 5.

#### **2.3.2 PA: uncorrected light curves and centroids**

1. Barycentric times (Section 7.4) calculated by PA are now included in the FITS files exported to MAST.
2. Optimal apertures were recalculated using improved knowledge of the PRF and focal plane geometry.
3. All exported centroids are flux-weighted instead of a mix of flux-weighted and PRF-fit centroids.

#### **2.3.3 PDC: corrected light curves**

1. Data is flagged for flight system anomalies before processing.
2. Discontinuities in individual light curves are detected and repaired.
3. Variable stars are identified and a harmonic fit and removal attempted before cotrending. The harmonic fit is restored to the corrected light curve before export to MAST.

### 3. Current Evaluation of Performance

#### 3.1 Overall

The Combined Differential Photometric Precision (CDPP) of a photometric time series is the effective white noise standard deviation over a specified time interval, typically the duration of a transit or other phenomenon that is searched for in the time series. In the case of a transit, CDPP can be used to calculate the S/N of a transit of specified duration and depth. For example, a 6.5 hr CDPP of 20 ppm for a star with a planet exhibiting 84 ppm transits lasting 6.5 hours leads to a single transit S/N of  $4.1 \sigma$ .

The CDPP performance has been discussed by Borucki et al. (Ref. 2) and Jenkins et al. (Ref. 7). Jenkins et al. examine the 33.5-day long Quarter 1 (Q1) observations that ended 2009 June 15, and find that the lower envelope of the photometric precision on transit timescales is consistent with expected random noise sources. The reprocessed Q1 corrected light curves discussed in these Notes have the same properties, as shown in Figure 1. Nonetheless, the following cautions apply for interpreting data at this point in our understanding of the Instrument's performance:

1. Many stars remain unclassified until Kepler and other data can be used to ascertain whether they are giants or otherwise peculiar. Since giant stars are intrinsically variable at the level of Kepler's precision, they must be excluded from calculations of CDPP performance. A simple, but not foolproof, way to do this is to include only stars with high surface gravity ( $\log g > 4$ ).
2. Given the instrument artifacts discussed in detail in the KIH, it is not generally possible to extrapolate noise as  $1/\sqrt{\text{time}}$  for those channels afflicted by artifacts which are presently not corrected or flagged by the Pipeline.
3. Stellar variability and other instrumental effects are not, in general, white noise processes.
4. There is evidence from the noise statistics of Q0 and Q1 (see below) that the Pipeline is overfitting the data for shorter data sets (a month or less of LC data) and fainter stars, so users are urged to compare uncorrected and corrected light curves for evidence of signal distortion or attenuation. The problem is less evident in the Q2 and Q3 data in Releases 3 and 4.

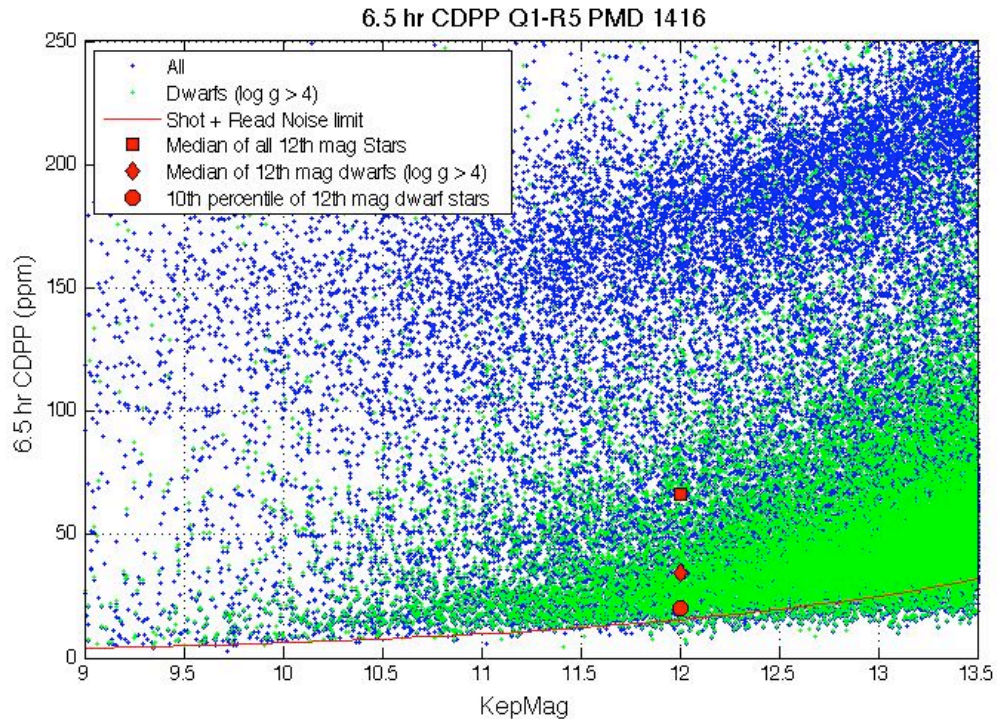
Example published data is shown in [2] and [10].

The Photometer Performance Assessment (PPA) tool formally calculates CDPP on 6 hr timescales as a function of Cadence  $c$  for each target  $k$ . The temporal median of the CDPP for each target is  $\text{TMCDPP}_k = M_{\{c\}}(\text{CDPP}_{ck})$ , where  $M_{\{c\}}$  denotes the median over the set of Cadences  $\{c\}$ , which in this case is all the Cadences of a Quarter.  $\text{TMCDPP}_k$  is then divided by  $\sqrt{13/12}$  to approximate the results on the 6.5 hr benchmark time scale. The results (Figure 1) separate into two branches, mostly corresponding to giants with  $\log g < 4$  and dwarfs with  $\log g > 4$ .

$\text{TMCDPP}_k$  can be further summarized by aggregate statistics, which are calculated over all targets which are members of a set  $K$  satisfying certain criteria, such as the aggregate median  $\text{CDPP}_K = M_{\{k \in K\}}(\text{TMCDPP}_k)$ , where  $M_{\{k \in K\}}$  indicates the median calculated over all targets  $k$  satisfying the criteria for membership in  $K$ . For example,  $K$  can be  $\{\text{all targets with magnitude between } 11.75 \text{ and } 12.25 \text{ and } \log g > 4\}$ , loosely referred to as "12th magnitude dwarfs." Aggregate percentiles can be defined in the same way. Table 2 summarizes the results for Q1. Note that the median CDPP over  $K = \{\text{all stars in a given magnitude bin}\}$  actually decreases as stars get fainter beyond 10th magnitude since the proportion of all stars which are (quiet) dwarfs increases considerably as the stars get fainter.

The Jenkins et al. (Ref. 7) expression for the lower noise envelope for stars brighter than 14th magnitude may be thought of as the RSS sum of shot noise and an effective read noise

corresponding to a source extracted from 2.6 pixels (given a per-frame, per pixel read noise of 100 e-). Extending this expression to the benchmark 6.5 hr transit time gives the results shown in Figure 1 and Table 2.



**Figure 1: 6.5 hr Temporal Median (TM) of the per-Cadence CDPP calculated for Quarter 1 by PPA for stars between 9th and 13.5th magnitude. The 6 hr TMCDDPs have been divided by  $\sqrt{13/12} = 1.041$  to approximate 6.5 hr TMCDDPs. Stars on the planetary target list with Kepler Magnitude < 13.5 and  $\log g > 4$ , which are almost certainly dwarf stars, are shown as green '+'s; other stars are marked with blue '+'s. The red line is the CDPP calculated from a simple shot and effective read noise model derived from Jenkins et al. (Ref. 7).**

**Table 2: Aggregate Statistics for the TMCDDPs plotted in Figure 1. Column Definitions: (1) Kepler Magnitude at center of bin. Bins are +/- 0.25 mag, for a bin of width 0.5 mag centered on this value. (2) Number of dwarfs ( $\log g > 4$ ) in bin. (3) 10th percentile TMCDDP for dwarfs in bin. (4) Median TMCDDP for dwarfs in bin. (5) Number of all stars in bin. (6) 10th percentile TMCDDP of all observed stars in bin. (7) Median TMCDDP for all stars in bin. (8) Simplified noise model CDPP. (9) Percentage of all observed stars with TMCDDP < noise model. TMCDDP is in units of ppm.**

center mag	number of dwarfs in bin	10th prc CDPP, dwarfs	median CDPP, dwarfs	number of all stars in bin	10th prc CDPP, all stars	median CDPP, all stars	simple Model CDPP	% of stars below model
9	30	5.8	31.6	229	10	95.1	3.8	0
10	170	9.7	34.6	708	12.6	108.4	6.0	0.56
11	655	14.8	29.9	2003	18.6	100.5	9.5	0.50
12	2311	19.9	34.5	4808	22.2	66.5	15.2	1.44
13	7261	28.2	43.7	11503	30.3	55.5	24.4	3.28

The Q0 data are not shown in this Section, since the overfitting problem is worse for this 10-day data set and the CDPP results produced by PPA are questionable. For example, 10<sup>th</sup> percentile 12<sup>th</sup> magnitude dwarf has a CDPP of 13.2 ppm, substantially less than the read + shot noise limit of 15.2 ppm. This problem will be worked and resolved before the next Release of Q0 data.

### **3.2 Changes in Performance Since Previous Release**

In Q1 Release 2, the median 12<sup>th</sup> magnitude dwarf had a CDPP of 43 ppm, while in Q1 Release 5 this benchmark is 34.5 ppm. Even discounting for some overfitting, CDPP has improved as a result of improved processing. Just as important, but more difficult to quantify in some aggregate sense, is a greatly improved ability to detect and remove systematic errors in the presence of serious complicating factors such as flight system anomalies, pixel response dropouts, and stellar variability. Additional issues have already been identified and will be addressed by future releases as promptly as possible.

### **3.3 Known Calibration Issues**

Topics under consideration by the DAWG which may change future calibration parameters or methods include:

1. Find a set of ancillary engineering data (AED) and Pipeline-generated metrics which more effectively remove systematics errors without overfitting the data (and hence distorting the astrophysical signal). Correlations between the light curves of different targets suggest the existence of unrepresented systematic errors. The unrealistically low CDPP (below the read + shot noise limit) for some stars suggests overfitting.
2. Improve the characterization of stellar variability to represent weaker and more complex waveforms, so cotrending can be more effective when the stellar variability is temporarily removed from the light curve.
3. Characterize and correct for in-orbit change of focus (Section 6.4).
4. Identify particular light curves that are poorly corrected, and understand why generally effective remedies do not work in these cases. Feedback from users to [kepler-scienceoffice@lists.nasa.gov](mailto:kepler-scienceoffice@lists.nasa.gov) is essential for the SO and SOC to identify, flag, and fix all such "hard cases."
5. Mitigate or at least identify the Artifacts described in the KIH, Section 6.7.
6. Assess and improve the focal plane characterization models which are inputs to CAL.

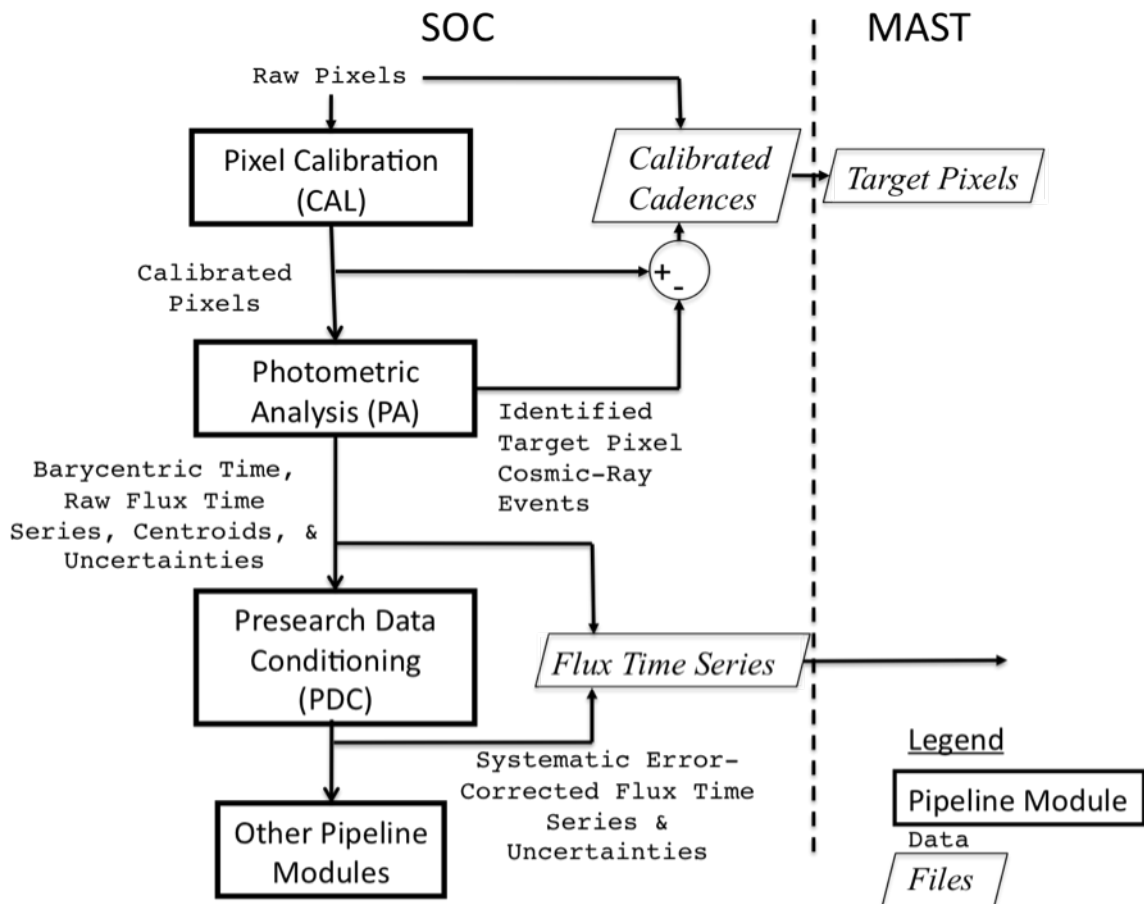
Calibration and data analysis issues related to the focal plane and its electronics are discussed in the Instrument Handbook.

## 4. Data Delivered – Processing History

### 4.1 Overview

*This Section is unchanged from the Release 4 Notes.*

The delivered FITS files were processed as shown in simplified form in Figure 2. What is referred to as “raw” flux time series in the Kepler Archive Manual is the result of calibrating pixels, estimating and removing sky background, and extracting a time series from a photometric aperture, and is referred to in these notes as “uncorrected” flux time series. The “corrected” flux time series has been decorrelated against known system state variables, such as pointing. In these Notes, we refer to “detrending” as an operation that removes low-frequency features of a light curve, using only the light curve data itself – such as subtracting the results of a median boxcar or centered polynomial (Savitzky-Golay) fit from the data. “Cotrending,” on the other hand, removes features correlated between the light curve and ancillary data, with some loss of low-frequency information and consequent signal distortion. Cotrending is also referred to as “systematic error removal.”



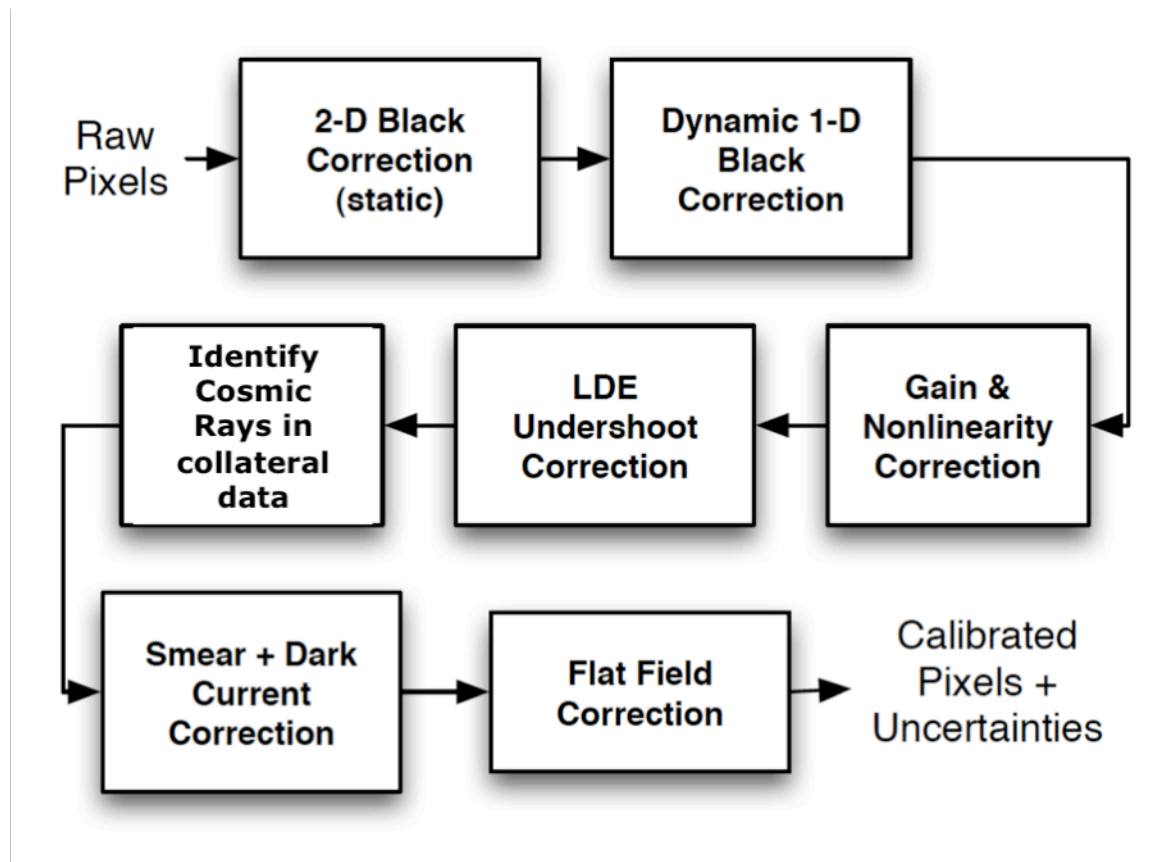
**Figure 2: Processing of data from raw pixels to flux time series and target pixel files archived at MAST. The target pixel files generated at MAST from the calibrated Cadence files delivered by the SOC have identified cosmic-ray events removed. The corrected flux time series delivered to MAST contain stellar variability and have -Infs for bad or missing data. In the corrected light curves used internally in the SOC for detecting planets, outliers are identified, and bad or missing data are filled by an autoregressive (AR)**



algorithm. See Section 7 and the MAST Kepler Archive Manual for details of MAST file contents.

#### 4.2 Pixel-Level Calibration (CAL)

The first step, pixel calibration (software module CAL), performs the pixel level calibrations shown in Figure 3. The SOC receives raw pixel data from each Kepler CCD, including collateral pixel data that is collected primarily for calibration. These collateral pixels include serial register elements used to estimate the black level (voltage bias), and masked and over-clocked rows used to measure the dark current and estimate the smear that results from the lack of a shutter on the spacecraft. Detailed models of each CCD have been developed from pre-flight hardware tests, along with full-frame images (FFIs) taken during commissioning prior to the dust cover ejection. These models are applied within CAL to correct for 2D bias structure, gain and nonlinearity of the ADU-to-photoelectron conversion, the electronic undershoot discussed in KIH Section 6.6, and flat field. CAL operates on long (30 min) and short (1 min) Cadence data, as well as FFIs [3, 9].



**Figure 3: Pixel Level Calibrations Performed in CAL.** See the Instrument Handbook for a discussion of signal features and image contents processed in CAL.

There may be rare cases where the bleeding smear value exceeds the 23 bits of the Science Data Accumulator, and the bleeding star is also variable. Then the difference between the masked and virtual smear will toggle between positive and negative values, and CAL will produce incorrect results. Users noticing repeated large step-like transitions between two discrete flux

levels should consult the FFIs to see whether either the masked or virtual smear regions contain bleeding charge. If so, please contact the Science Office for further investigation.

### 4.3 Photometric Analysis (PA)

*This Section is unchanged from the Release 4 Notes, except to reference the mod.out center motion time series now provided in the Supplement.*

The primary tasks of this module are to compute the photometric flux and photocenters (centroids) for up to 170,000 Long Cadence (thirty minute) and 512 Short Cadence (one minute) targets across the focal plane array from the calibrated pixels in each target's aperture, and to compute barycentric corrected timestamps per target and Cadence [4].

The tasks performed by Photometric Analysis (PA) are

1. Calculation of barycentric time correction, obviating the need for manual correction discussed in the Release 2 Notes (KSCI-19042).
2. Detection of Argabrightening events (Section 6.1). Argabrightening detection and associated Cadence gapping take place before CR detection; otherwise, many of the Argabrightening events would be cleaned as cosmic rays in the respective pixels, and not effectively detected and marked as data gaps.
3. Cosmic ray (CR) cleaning of background and target pixels, logging of detected CRs, and calculation of CR metrics such as hit rate and mean energy. The CR's are corrected by subtracting the residual differences after median filtering is performed on the detrended pixel time series. The same method and parameters are used for LC and SC. (Note 5)
4. Robust 2-D polynomial fitting to calibrated background pixels
5. Background removal from calibrated target pixels
6. Aperture photometry. In release 5, the flux is the sum of pixels in the optimal aperture after background removal (Simple Aperture Photometry, SAP).
7. Computation of flux-weighted (first moment) centroids. (Note 1).
8. Fitting of 2-D motion polynomials to target row and column centroids, which smoothly maps (RA,DEC) to (row, column) for a given output channel. (Note 4).
9. Setting gap indicators for Cadences with Argabrightening (Section 6.1). The gapped Cadences have all  $-\text{Inf}$  values in the FITS light curve files, except for the first two columns: time and Cadence number.

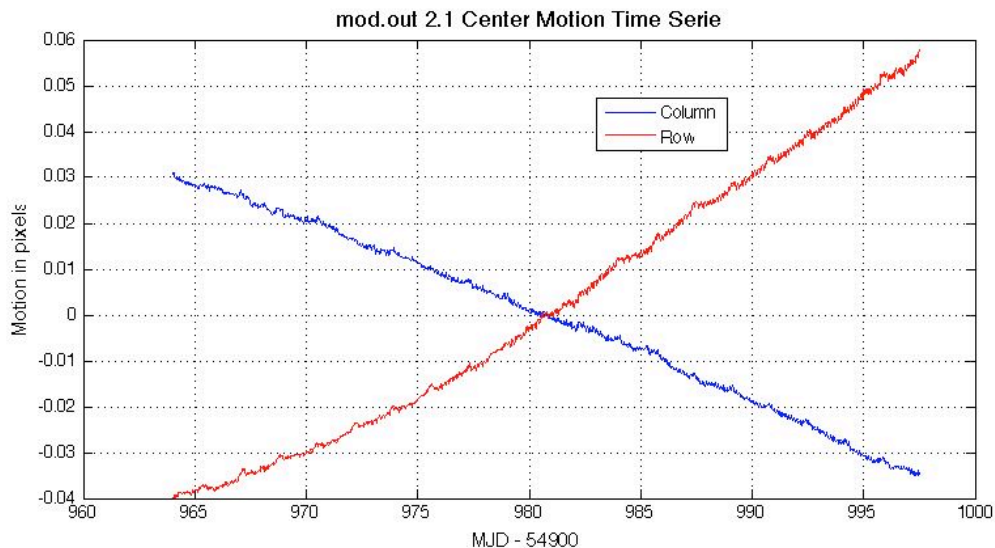
#### Notes

1. Flux-weighted (first moment) centroids are calculated for all targets. PRF centroids are also computed, but only for a small subset (PPA\_STELLAR) of the Long Cadence targets due to the heavy computational requirements of the PRF centroiding algorithm. The PRF fitting does not necessarily converge for targets which are faint, or located in complex fields. While only flux-weighted centroids are exported to MAST in Release 5, they are suitable for precision astrometry [6] in uncrowded apertures. Users wishing to improve on the flux-weighted centroids need to consider the distribution of flux from non-target sources in the optimal aperture pixels or use the PRFs provided in the KIH Supplement to do their own fits.
2. There is no identification of bad pixels in PA in Release 5, nor is there any exclusion, gapping or other treatment of known bad pixels. Bad pixels may be identified in future releases. The treatment of bad pixels is TBD, and may depend on how the pixel is bad (high read noise, unstable photoresponse, low photoresponse, etc.) and its location in the target aperture. While the Pipeline flags bad data on a per mod.out, per Cadence basis, bad pixels affect individual targets, and users are cautioned to carefully inspect the target pixels before believing peculiar light curves.
3. The output of PA is called 'raw' in the light curve FITS file, even though it is the sum of 'calibrated' pixels, because systematic errors have not been removed by PDC.

4. Motion polynomials are a means of estimating local image motion, and do not assume rigid body motion of the entire focal plane. They thus account for changes in plate scale, rotation, image distortion, and differential velocity aberration (DVA) on a channel-by-channel and Cadence-by-Cadence basis (Figure 4). The simplified mod.out center motion time series provided in the Supplement are the row and column of the nominal center (in RA and DEC) of the mod.out as calculated from these motion polynomials.

5. Data which are greater than 12 median absolute deviations (MAD), after the removal of a trend formed by a quadratic fit followed by a 5 Cadence wide median filter, are identified as CRs. The MAD is calculated over a sliding window 145 Cadences wide after the trend is removed. The amplitude, Cadence, and location of the removed CR will be made available to users in future Releases, either as Cadence-to-Cadence cosmic ray correction tables, or integrated into the target pixel files (Section 7.5), so that users may restore the CRs and use their own methods of CR detection and removal if desired.

Astrophysical phenomena of only a single Cadence duration cannot be distinguished from CRs, as the Pipeline does not check for correlated outliers on adjacent pixels.



**Figure 4: Mod.out 2.1 Center Motion Time Series** calculated from motion polynomials. The median row and column values have been subtracted. Since this mod.out is at the edge of the field, it shows large differential velocity aberration (DVA) with respect to the center of the field, as well as a higher sensitivity to focus jitter and drift. The 1.7 d period of an eclipsing binary FGS guide star (Section 6.2) can also be seen.

#### 4.4 Pre-Search Data Conditioning (PDC)

The primary tasks of PDC for MAST users are to correct systematic errors and remove excess flux in target apertures due to crowding. PDC also identifies outliers, and fills gaps in uncorrected flux light curves before passing the conditioned light curve to the Transiting Planet Search (TPS) component of the Pipeline for internal SOC use, but does not provide the outlier and gap fill results to MAST users. PDC was designed to remove systematic errors that are correlated with ancillary engineering or Pipeline generated metrics (such as motion polynomials), and also to condition Long Cadence light curves for the transiting planet search (TPS). Significant effort has been expended to preserve the natural variability of targets, though further effort is still required to strike the right balance between preserving stellar variability signals and emphasizing transit signals. Users will therefore need to be cautious when their phenomena of interest are much shorter (<1 h) or much longer (>5 d) than a transit, or have complex light curves with multiple

extrema on transit time scales (such as eclipsing and contact binaries). Examples of astrophysical features removed or significantly distorted by PDC are shown in Section 4.4.3.

Tuning the parameters of PDC requires assessing the relative merits of removing instrumental artifacts, preserving transits and their shapes, and preserving other astrophysical phenomena, and it is not likely that any single choice can give satisfactory results for all observing conditions, targets, and phenomena of interest. Hence, PDC is discussed in greater detail in these Notes than is CAL or PA. Users concerned about the impact of PDC on their signals of interest are invited to use the ancillary engineering data (AED) and motion time series in the Supplement to perform their own systematic error correction.

#### 4.4.1 Description

The tasks performed by PDC are:

1. Accept data anomaly flags for Cadences which are known to be lost or degraded (Section 4.4.2). These Cadences and their corresponding data anomalies are shown in Section 5.3.6
2. Resampling of AED to match the sampling rate of LC and SC data.
3. Identification and correction of unexplained discontinuities (i.e. unrelated to known anomalies), an iterative process.
4. Cotrend target flux time series against AED and motion polynomials derived by PA (Section 4.3 item 8) to remove correlated linear and nonlinear deterministic trends. Singular Value Decomposition (SVD) is used to orthogonalize the set of basis vectors and numerically stabilize the model fit. While thermal engineering data were not used for cotrending in Release 5, the motion polynomials contain similar information in this case.
5. Identify variable stars (>0.25% center-peak variability).
6. For variable stars only, perform coarse systematic error correction with the following steps:
  - a. Correct discontinuities due to attitude tweaks.
  - b. Compare phase-shifting harmonic fitting to simple polynomial fitting, and select the method which gives the smallest error for initial detrending.
  - c. Correct thermal recovery transients with a polynomial fit for each target.
  - d. Remove a low-order polynomial trend from the transient-corrected light curve
  - e. Repeat the harmonic fit first done in Step 6.b, and save this improved harmonic fit for later restoration.
  - f. Subtract the improved harmonic fit from the light curve resulting from Step 3, and cotrend the harmonic-removed light curve as in Step 4.
7. For stars initially identified as variable, if the standard cotrended result is not variable and cotrending has reduced the noise then the identification of the star as variable is considered mistaken. Then the result of the standard cotrending is retained. Otherwise, the result of the harmonic free cotrending is retained. The harmonic content is restored later in PDC.
8. Assess results of cotrending. If cotrending has increased the noise by >5%, restore the uncorrected light curve at this point.
9. Correction for the excess flux in the optimal aperture for each target due to crowding, as calculated over the optimal aperture.
10. Identification and removal of impulsive outliers after masking off astrophysical events such as giant transits, flares, and microlensing. A median filter is applied to the time series after the removal of obvious astrophysics, and the residual is determined by subtracting the median series from the target flux series. A robust running mean and standard deviation of the residual is calculated and points more than  $12 \sigma$  (LC) or  $8 \sigma$  (SC) from the mean are excluded. Not all astrophysical events are successfully masked, and hence may be falsely identified as outliers or may unnecessarily increase the noise threshold for outliers. The masked events are restored to the MAST light curves.

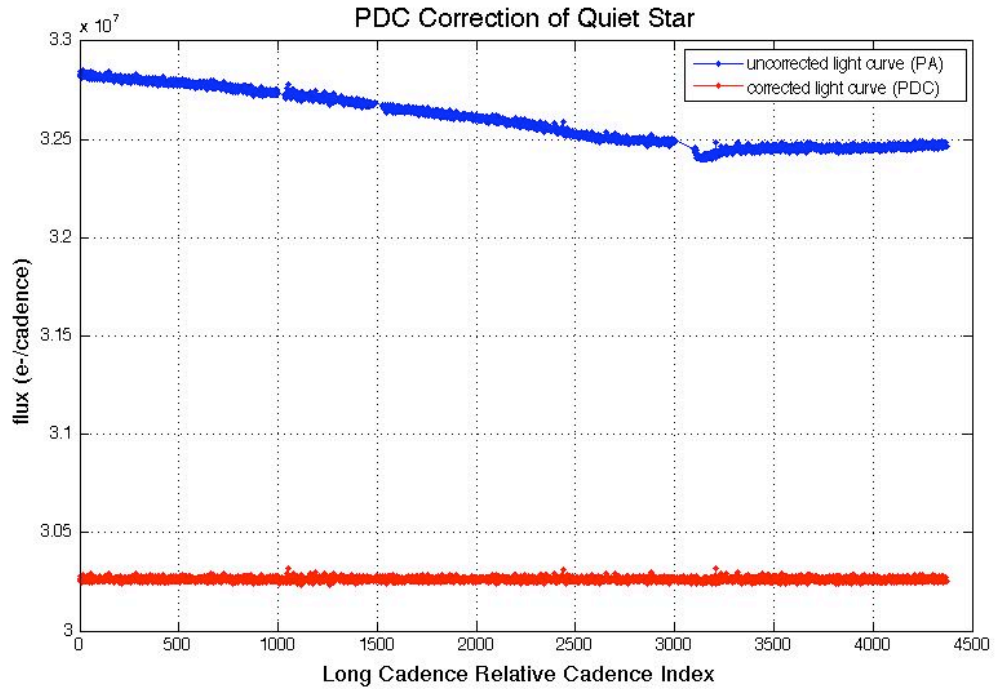
### Notes

1. The crowding metric is the fraction of starlight in an aperture which comes from the target star. For example, a crowding metric of 1 means that all the light in an aperture comes from the target, so the light curve needs no correction. A crowding metric of 0.5 means that half the light is from the target and half from other sources, so the flux must be decreased by half before the correct light curve for the target is obtained. Note that the uncorrected flux time series are *not* corrected for crowding. The crowding metric is based on the Kepler Input Catalog (KIC) star locations and brightnesses, the local PRF of the target star and its neighbors, and the optimal aperture. It is averaged over a Quarter, and neglects seasonal and secular changes in the PRF compared to the model established by observations during Commissioning. A given star will move to different parts of the Kepler focal plane from Quarter to Quarter as Kepler rolls, so the PRF, aperture, and crowding metric will also vary from Quarter to Quarter.
2. Gaps are not filled in the MAST files, and are represented as -Infs. Intermediate data products generated by PDC and internal to the SOC do have gaps filled, before being passed to planetary search parts of the Pipeline.
3. Different frequencies of AED will physically couple to the photometric light curve with different strengths. PDC represents this by decomposing AED time series into low and high bandpass components, and allowing the coefficients of the components to vary independently in the fit.
4. The output of PDC is referred to as 'corrected' data in the delivered files. Users are cautioned that systematic errors remain, and their removal is the subject of ongoing effort as described in Section 3.3.

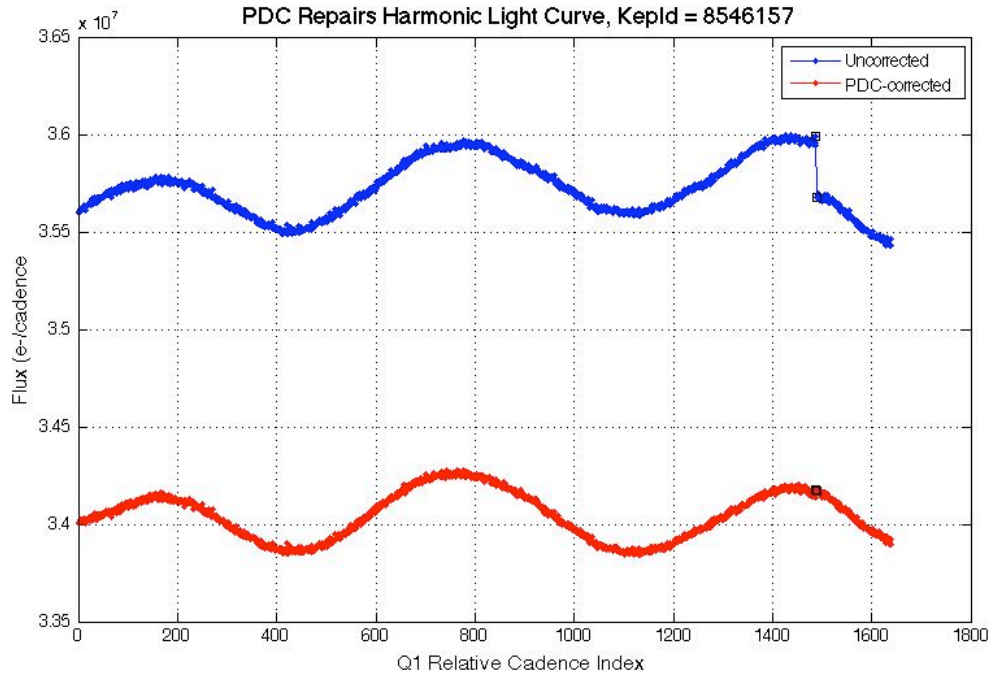
#### **4.4.2 Performance**

*This Section contains some Figures reproduced from Release 4 Notes (Quarter 3) since Q0 and Q1 were in some ways too peaceful to demonstrate PDC's performance. Captions clearly indicate whether a Q3 example is shown.*

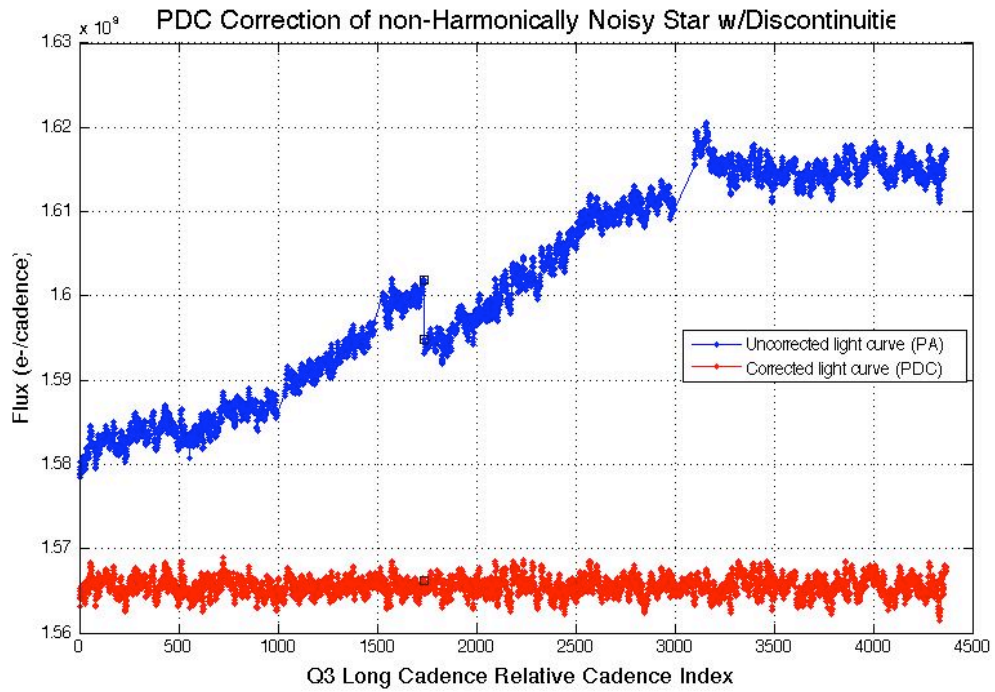
PDC gives satisfactory results on most stars which are either intrinsically quiet (Figure 5), or have well-defined harmonic light curves above the detection threshold (Figure 6); in most of these cases, the standard deviation of the corrected flux is within a factor of 2 of the noise expected from read and shot noise in the calibrated pixels summed to form the uncorrected light curve. It also performs well in many cases where the star is variable, but without a dominant harmonic term (Figure 7). However, PDC will sometimes not identify a target-specific discontinuity (Figure 8), and will sometimes introduce noise into complex lightcurves (Figure 9). Conversely, PDC sometimes identifies eclipses as discontinuities and introduces a discontinuity in an attempt to correct the false discontinuity (Figure 10).



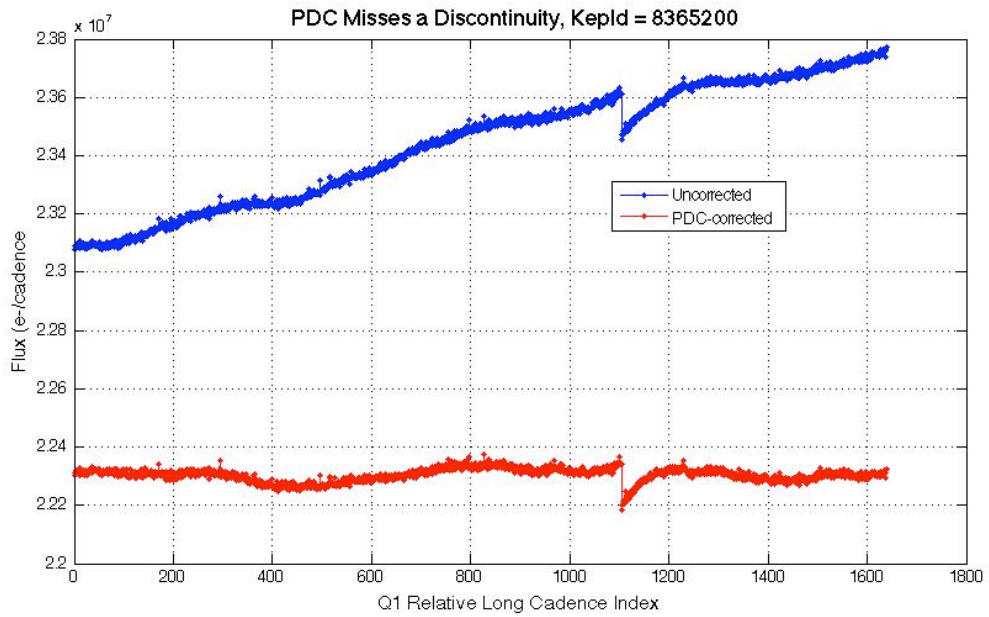
**Figure 5: Q3 example of PDC removal of slopes and discontinuities from the light curve of a quiet star of Kepler magnitude 14.6. The noise in the corrected light curve is only 20% greater than the noise expected from the calibrated pixels, a considerable improvement over the uncorrected light curve. The gap between RCI = 3000 and 3100 is a Safe Mode event (Section 5.3.1) followed by a monthly download and subsequent thermal transient.**



**Figure 6: Q1 example of PDC correction of a harmonically variable star with a target-specific discontinuity. MAST users receive the light curve corrected for systematic errors, with the harmonic variability restored and gaps in the data represented by -Infs. The corrected light curve delivered to MAST is shown in red in this Figure.**

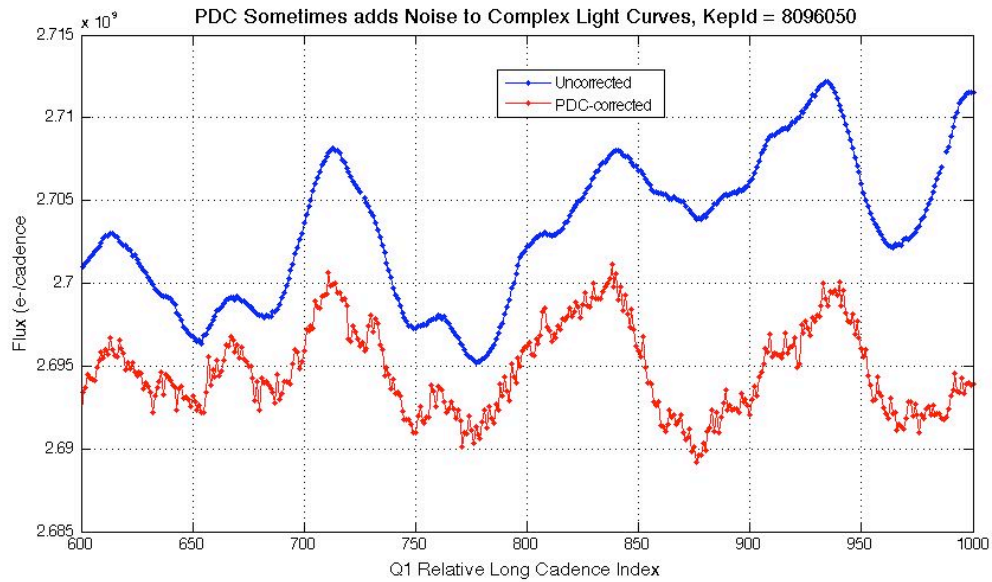


**Figure 7: Q3 example of PDC correction of a non-harmonically variable star light curve with discontinuities. The RMS of the corrected (red) curve is about 8x the noise calculated from the read and shot noise in the calibrated pixels, and is believed to be almost entirely due to intrinsic stellar variability.**

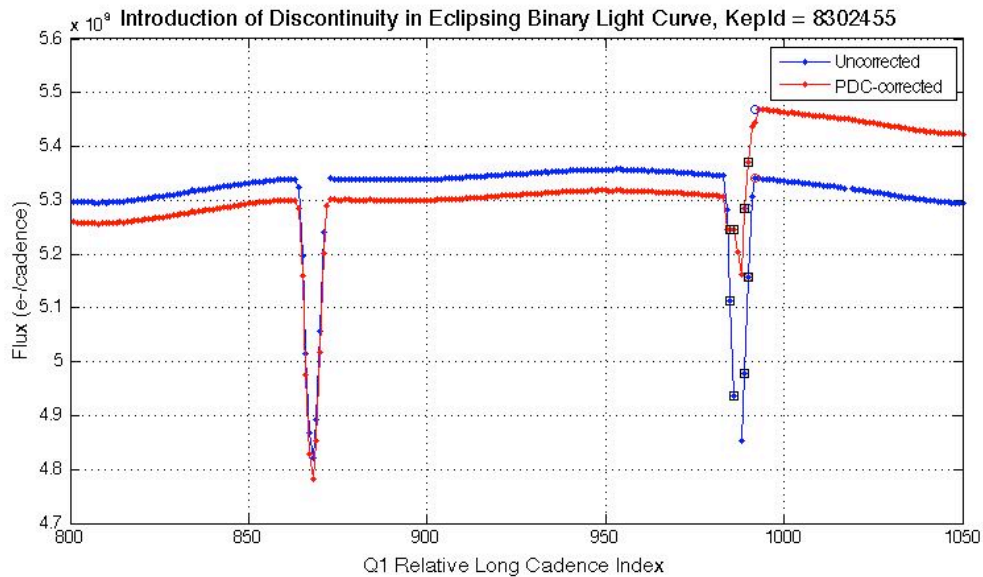


**Figure 8: Q1 example of an unidentified and hence uncorrected target-specific discontinuity.**





**Figure 9: Q1 example of PDC adding short-period noise to an intrinsically complex light curve.**



**Figure 10: Q1 Example of PDC misidentifying an eclipse as a discontinuity, and mistakenly introducing a discontinuity into the light curve.**

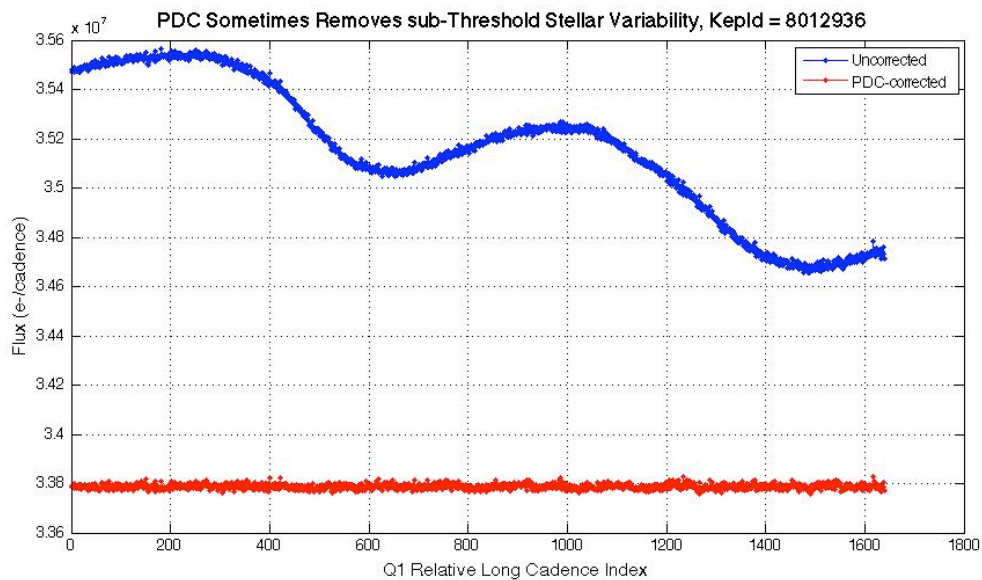
#### 4.4.3 Removal of Astrophysical Signatures

PDC can remove astrophysical signatures if they are:

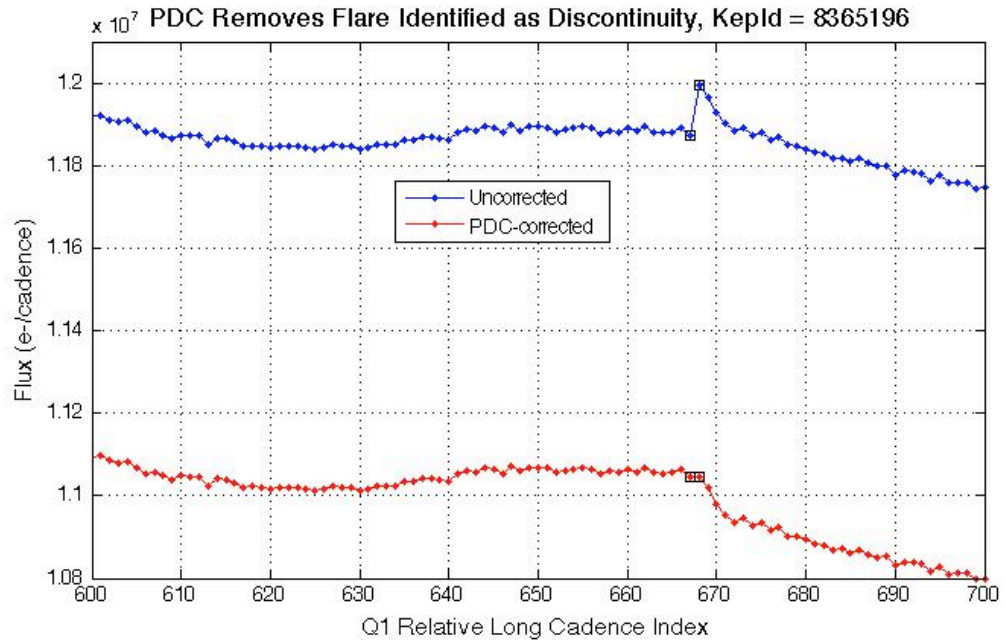
1. Harmonic, but have periods  $> 5d$  and fall below PDC's detection threshold for stellar variability. In Release 5, the center-peak threshold is 0.25%, which for otherwise quiet stars allows a harmonic with a peak-to-peak amplitude of 0.5% to go undetected.
2. Spikes a few Cadences wide (Figure 12), such as flares.

3. More or less linear ramps over the processing interval.
4. Harmonic signals above the threshold, but the harmonic fit does not produce a good fit, and current algorithm fails to recognize that cotrending has performed badly.
5. Non-harmonic signals for which current algorithm fails to recognize that cotrending has performed badly
6. Harmonic signals above the threshold for which the fit is good, but PDC incorrectly determines that target was cotrended well when treated as non-variable (Figure 11).

A thorough study of astrophysical signal distortion by PDC is planned, but has not been performed to date. Users may be helpful to this effort by reporting light curves, in which they suspect that a signal has been distorted or removed, to the Science Office at [kepler-scienceoffice@lists.nasa.gov](mailto:kepler-scienceoffice@lists.nasa.gov).



**Figure 11: Q1 example of PDC removal of harmonic stellar variability which falls below the harmonic variability detection threshold after initial cotrending. The amplitude of the harmonic is  $\pm 0.5\%$ , exceeding the threshold of  $\pm 0.25\%$ , before initial cotrending. The cotrending reduced the variability below the threshold, so the identification of the target as harmonic was considered mistaken as per Section 4.4.1 Step 8, hence the star was reprocessed as a quiet star, and a harmonic term was not identified and set aside for later restoration.**



**Figure 12: Q1 example of an astrophysical event, possibly a flare, identified by PDC and removed from the corrected light curve. Open black squares show astrophysical events identified as discontinuity anomalies and “corrected.”**

## 5. Lost or Degraded Data

In this Section, we discuss Cadences which are essentially lost to high-precision photometry due to planned or unplanned spacecraft events. Particularly important and unexpected phenomena are written up as Kepler Anomaly Reports (KARs) or SOC change requests (KSOCs). Puzzling features of the data products, which do not succumb to cursory examination by the DAWG, are written up as K-DAWG tickets to initiate and track more thorough investigations.

Some of the tables and figures in this Section are the same as in the Release 2 Notes, since the detection of the reported phenomena does not rely on changes to the Pipeline between Releases 2 and 5. There are fewer kinds of lost or degraded data to report in Q0 and Q1 compared to the Q2 and Q3 data, as the total duration of Q0 and Q1 is about half a full Quarter, thus allowing less opportunity for mischief.

### 5.1 Momentum Desaturation

Solar radiation torque causes angular momentum to build up in the reaction wheels, which then must be desaturated by thruster firings when the wheels spin up to their maximum operating RPM. Desats occur every 3 days. The spacecraft (S/C) is not designed to maintain Fine Point control during these events, and enters Coarse Point mode. The subsequent image motion is sufficient to spoil the photometric precision of data collected during desats, and a few minutes after desats during which the spacecraft restores Fine Point control. One LC and several SCs are affected for each desaturation.

The momentum dump Cadences have -Infs in the delivered light curve files, but finite values in the uncalibrated and calibrated pixels. The dump Cadences are listed in Table 3 so that users of time series will know which -Infs are due to desats, and users of pixel data will know which Cadences to exclude from their own analyses. Tables of the more numerous SCs afflicted by desats is included in the Supplement, though they duplicate some of the information in the SC data anomaly tables (Section 5.3.6).

**Table 3: Momentum dumps in Q0 and Q1 and the corresponding Long Cadences. CIN = Cadence Interval Number, RCI = Relative Cadence Index.**

Q0 LC			
CIN	RCI	Date(MJD)	
703	136	54955.79669	
848	281	54958.75956	
950	383	54960.84379	
Q1 LC			
CIN	RCI	Date(MJD)	
1246	142	54966.89213	
1392	288	54969.87544	
1538	434	54972.85874	
1684	580	54975.84205	
1830	726	54978.82535	
1976	872	54981.80866	
2122	1018	54984.79196	
2268	1164	54987.77527	
2414	1310	54990.75857	
2560	1456	54993.74188	
2706	1602	54996.72518	

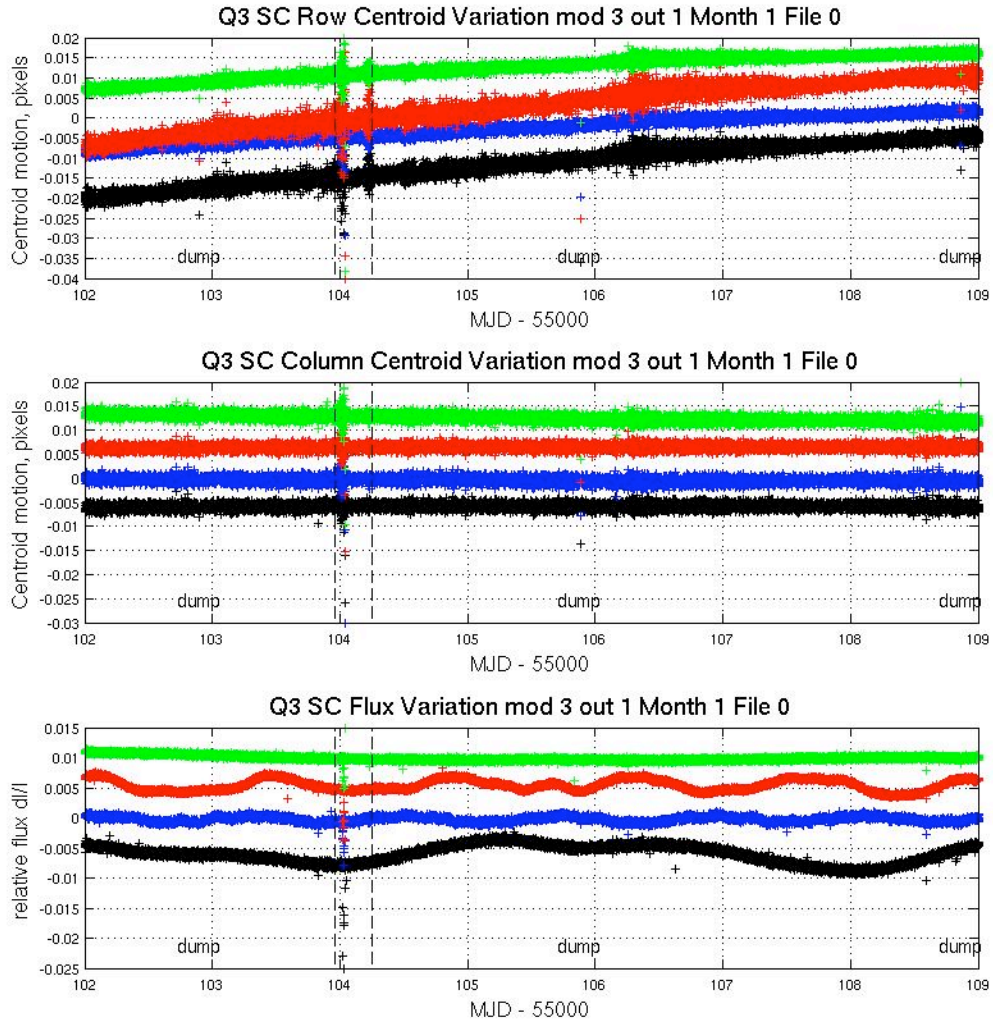
## 5.2 Reaction Wheel Zero Crossings

*The descriptive part of this Section is unchanged since Release 4, since the Q3 example shown is a particularly clear illustration of the problem. The Table, however, has been revised to show the zero crossing times for Q1.*

Another aspect of spacecraft momentum management is that some of the reaction wheels cross zero angular velocity from time to time. The affected wheel may rumble and degrade the pointing on timescales of a few minutes. The primary consequence is an increased noise in the Short Cadence centroids, and pixel and flux time series. The severity of the impact to the SC flux time series seems to vary from target to target, with all SC targets showing some impact on the centroid and pixel time series. In some cases, we observe negative spikes of order  $10^{-3}$  to  $10^{-2}$  in SC relative flux time series (Figure 13), and these Cadences must be excluded from further analysis. The impact on Long Cadence data is much less severe in both amplitude and prevalence. Zero crossings are not gapped in Release 5, and users will have to use Table 4 to identify possibly afflicted Cadences.

In Figure 13, the noise in centroids, and loss of flux, occurs on multiple stars during the zero crossing, so this noise is not the result of an uncorrected cosmic ray event or other local transient. Neither is it due to the momentum dumps (Section 5.1), at MJD – 55000 = 102.90, 105.88, and 108.86, for which one or two Cadences right after the dump may have bad pointing, but are not flagged as data gaps by the Pipeline. The zero crossings occur at distinctly different times than the momentum dumps.

Since the Pipeline does not flag zero crossings as anomalous data in Release 5, the zero crossing events in Q1 are shown in Table 4. Events were identified in reaction wheel telemetry, which is not sampled synchronously with Cadences. For each zero crossing event, the last Cadence ending before the event and the first Cadence beginning after the event were identified. Overlap between events is due to this rounding of Cadence numbers at times when the slowest wheel had nonzero speed for a time interval shorter than 2 Cadence periods.



**Figure 13: Example from Q3 of the effect of reaction wheel speed zero crossing on SC flux and centroids. The plots show row and column centroid motion, and the relative flux change, in the neighborhood of zero crossings. The data on several stars are overplotted in different colors in each panel of the Figure; vertical dashed black lines bracket the times during which at least one wheel had zero speed according to its telemetry. The curves are offset for clarity, and momentum dumps are labeled.**

**Table 4: Zero crossing events in Q1, defined as the time from first to last zero crossing in the event, rounded to the nearest Cadence. There were no zero crossings in Q0.**

Kepler Data Release 5 Notes Supplement

Zero Crossing Listing for Q1

Prepared by the DAWG

This file created: 24-May-2010 14:38

MJD 55340.60973

Column Definitions			Cadence Numbers			
LC	midTime	MJD	Long Cadence		Short Cadence	
Event#	Start	End	Start	End	Start	End
1	54985.221	54985.364	2143	2150	52768	52959
2	54985.405	54985.609	2152	2162	53052	53326
3	54994.110	54994.416	2578	2593	65839	66258
4	54995.152	54996.725	2629	2706	67353	69649

### 5.3 Data Anomalies

#### 5.3.1 Safe Mode

From time to time, the Kepler Spacecraft will go into Safe Mode, because of an unanticipated sensitivity to cosmic radiation, or unanticipated responses to command sequences. While each individual event is unexpected, it is not unusual for newly-commissioned spacecraft to experience them, until the in-orbit idiosyncrasies of the flight system are understood. Kepler's Flight Software has been modified to leave the LDE on during radiation-induced resets of the RAD750 processor, so that data degradation due to thermal transients after an LDE power cycle does not occur during this kind of Safe Modes.

There were no Safe Modes in Q0 or Q1.

#### 5.3.2 Loss of Fine Point

From time to time, the Kepler spacecraft will lose fine pointing control, rendering the Cadences collected useless for photometry of better than 1% precision. The cause of these Loss of Fine Point (LOFP) events is presently under investigation. While the LOFPs are treated as lost data by the Pipeline, users with sources for which ~1% photometry is scientifically interesting may wish to look at the pixel data corresponding to those Cadences. Recent Flight Software upgrades are expected to reduce the frequency and duration of LOFPs.

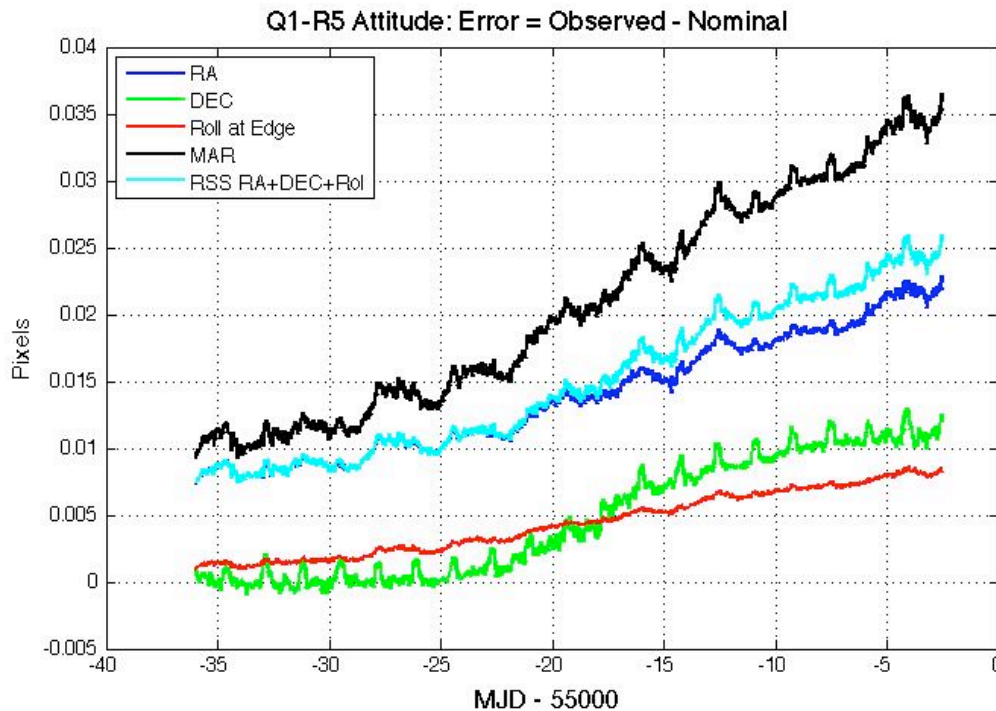
There were no LOFPs in Q0 or Q1.

#### 5.3.3 Pointing Drift and Attitude Tweaks

Daily reference pixels are used by the SOC/SO to measure S/C attitude. The SOC PDQ software uses centroids of 3-4 stars per module/output to determine the measured boresight attitude compared with the pointing model (which accounts for differential velocity aberration). The Photometer Attitude Determination (PAD) software performs a similar calculation to reconstruct the attitude using the Long Cadence science data when the data are processed after each downlink, and reprocessed on a Quarterly basis before delivery to MAST. The PAD attitude

errors (RA, Dec, roll) for Q1 are shown in Figure 14. The maximum attitude residual (MAR) is the largest distance between the expected and actual location of a star in its aperture, for a given Cadence. Since continued attitude drift would invalidate target aperture definitions and lead to large photometric errors, small attitude adjustments (“tweaks”) are performed if necessary to ensure that MAR is always  $< 100$  mpix. In Q2, tweaks were necessary, and they introduced discontinuities into the data which were not fully compensated by the initial version of the Pipeline. Parameter changes in the FGS centroiding algorithm, which were implemented at the start of Q3, have apparently greatly diminished the boresight drift.

There were no tweaks in Q0 or Q1 since the drift did not grow to an unacceptable level.



**Figure 14: Attitude Error in Quarter 1, calculated by PAD using Long Cadence data. Pointing errors introduced by an intrinsically variable and an eclipsing binary FGS guide star can be seen in the RA and DEC, respectively, and are discussed in more detail in Section 6.2.**

### 5.3.4 Downlink Earth Point

Science data is downlinked once a month, and the spacecraft changes its attitude to point its fixed High Gain Antenna (HGA) at the Earth. Science data collection ceases, and the change in attitude induces a thermal transient in the Photometer. In Release 5, data collected after Earth Point are corrected in the same way as data after a Safe Mode.

There was one Earth Point in Release 5, which occurred between the end of Q0 and the start of Q1. This was a planned event, and had a modest impact on the data.

### 5.3.5 Manually Excluded Cadences

Occasionally, a Cadence is manually excluded, usually near a gap or discontinuity in the data which makes it difficult to exclude them automatically. It was not necessary to do this in Release 5.



**5.3.6 Anomaly Summary Table**

Table 5 shows a summary of the Anomalies for both LC and SC data. COARSE\_POINT in the SC table indicates normal attitude recovery after a momentum dump and not a Loss of Fine Point event (Section 5.3.2).

**Table 5: Anomaly Summary Table for Long and Short Cadences.**

LC CIN			Cadence midTime		Note
Start	End	Anomaly Type	MJD Start	MJD End	
790	790	ARGABRIGHTENING	54957.5642	54957.5846	Q0
914	915	ARGABRIGHTENING	54960.0980	54960.1388	Q0
972	972	ARGABRIGHTENING	54961.2831	54961.3035	Q0
1044	1104	EARTH_POINT			Q0-Q1 science break: data downlink.
2091	2282	ARGABRIGHTENING			Q1

Q	SC CIN		Anomaly Type	Cadence midTime		Note
	Start	End		MJD Start	MJD End	
Q0	12179	12180	ARGABRIGHTENING	54957.5771	54957.5785	
Q0	15898	15897	ARGABRIGHTENING	54960.1102	54960.1109	
Q0	17647	17651	ARGABRIGHTENING	54961.3015	54961.3049	
Q0	9564	9575	COARSE_POINT	54955.7960	54955.8042	
Q0	13913	13925	COARSE_POINT	54958.7582	54958.7671	
Q0	16969	16981	COARSE_POINT	54960.8397	54960.8486	
Q0-Q1	19780	21609	EARTH_POINT			Q0-Q1 science break: data downlink.
Q1	51205	51207	ARGABRIGHTENING			
Q1	56921	56922	ARGABRIGHTENING			
Q1	25850	25861	COARSE_POINT			
Q1	30229	30241	COARSE_POINT			
Q1	34610	34621	COARSE_POINT			
Q1	38990	39001	COARSE_POINT			
Q1	43370	43381	COARSE_POINT			
Q1	47750	47761	COARSE_POINT			
Q1	52130	52141	COARSE_POINT			
Q1	56510	56521	COARSE_POINT			
Q1	60890	60901	COARSE_POINT			
Q1	65270	65281	COARSE_POINT			
Q1	69650	69661	COARSE_POINT			

## 6. Systematic Errors

This Section discusses systematic errors arising in on-orbit operations, most of which will be removed from flux time series by PA or PDC (Section 4). While the Release 5 data is cotrended against image motion (as represented by the Cadence-to-Cadence coefficients of the motion polynomials calculated by PA), other telemetry items which may be used for cotrending the data in future releases are included in the Supplement so that users can at least qualitatively assess whether features in the time series look suspiciously like features in the telemetry items. This telemetry has been filtered and gapped as described in the file headers, but the user may need to resample the data to match the LC or SC sampling. In addition, PDC corrects systematic effects only in the flux time series, and this Section and Supplement files may be useful for users interested in centroids or pixel data.

Most of the events described in this Section are either reported by the spacecraft or detected in the Pipeline, then either corrected or marked as gaps. This Section reports events at lower thresholds than the Pipeline, which affect the light curves and therefore may be of interest to some users.

### 6.1 Argabrightening

*Argabrightening*, named after its discoverer, V. Argabright of BATC, is a presently unexplained diffuse illumination of the focal plane, lasting on the order of a few minutes. It is known to be light rather than an electronic offset since it appears in calibrated pixel data from which the electronic black level has been removed using the collateral data. It is not a result of gain change, or of targets moving in their apertures, since the phenomenon appears with the same amplitude in background pixels (in LC) or pixels outside the optimal aperture (in SC) as well as stellar target pixels. Many channels are affected simultaneously, and the amplitude of the event on each channel is many standard deviations above the trend, as shown in Figure 15. The method of detection is to calculate the median, for each Cadence and mod.out, of the calibrated background (LC) or out-of-optimal-aperture (SC) pixels, fit a parabola to the resulting time series, smooth the residual, and look for outliers in the difference between the residual and the smoothed residual. The Pipeline identifies Argabrightenings with this method, and subsequently treats those Cadences as gaps for all pixels in that channel. While the Pipeline processes each channel in isolation, all channels are marked as gaps if Argabrightenings are detected on more than half of the channels.

The Pipeline uses a rather high threshold of 100x the median absolute deviation (MAD) for LC and 60x for SC. While it appears that background subtraction has mostly removed this phenomenon from the delivered Long Cadence data, the residual effect has not been proven to be negligible in all cases, especially in Short Cadence data. There may also be significant Argabrightening events in both LC and SC, which do not exceed the thresholds. This Section gives a summary of events which exceed a 10x MAD threshold on at least 11 channels (Table 6 and Table 7), so that the user may consider whether some Cadences of interest might be afflicted by Argabrightening, but not identified as such by the Pipeline and gapped (i.e., -Inf in the light curve file). The Supplement contains these detection summaries as ASCII files.

The Supplement also contains the channel-by-channel background time series so users can identify low-level or few-channel Argabrightenings using their own criteria. These time series may also be useful for correcting SC data collected during Argabrightening events, since the Pipeline background correction interpolates LC background data to calculate the background for SC data. Users may notice some "chatter" in the background time series. A preliminary study shows that the problem is present in the calibrated background pixels, but not in the raw pixels, and is present in about 25% of the channels, with an amplitude up to 3% of the background. The reasons are still under investigation.

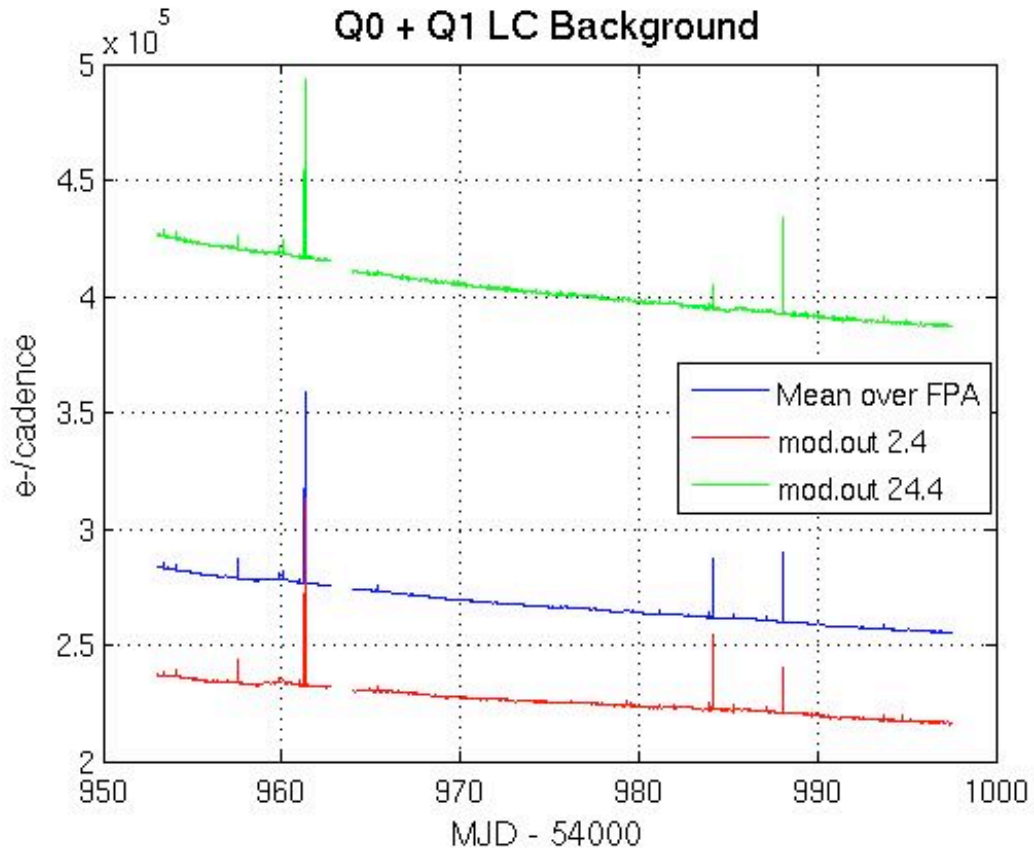


Figure 15: Background time series for Q0 + Q1 showing the average over all mod.outs, and the modules furthest from (2.4) and nearest to (24.4) the Galactic plane. The narrow spikes common to all 3 curves are Argabrightening events.

Table 6: Q0 and Q1 LC Argabrightening Events with amplitude greater than 10x the median absolute deviation (MAD), which occurred simultaneously on at least 11 of the 84 channels. The columns are (1) CIN = Cadence Interval Number for Argabrightening Cadences, (2) RCI = Relative Cadence Index for Argabrightening Cadences, (3) Date = Arg Cadence mid-Times, MJD, (4) Mean SNR over Channels of Arg Event, (5) N\_chan = Channels exceeding threshold in Arg Cadence, (6) N\_pipe = Channels exceeding default (Pipeline) threshold in ArgCadence. MAD is calculated on a channel-by-channel basis.

**Q0**

CIN	RCI	Date (MJD)	MeanSNR	N <sub>chan</sub>	N <sub>pipe</sub>
586	19	54953.40596	28.2	64	0
619	52	54954.08027	27.4	84	0
647	80	54954.65241	8.8	27	0
790	223	54957.57441	105.0	84	52
860	293	54959.00476	7.8	20	0
901	334	54959.84254	29.9	84	0

914	347	54960.10818	48.7	70	1
915	348	54960.12861	17.5	60	0
958	391	54961.00726	22.2	67	0
972	405	54961.29333	956.5	84	84
973	406	54961.31376	19.5	63	0
<b><u>Q1</u></b>					
1149	45	54964.91007	11.5	59	0
1170	66	54965.33918	32.2	84	0
1230	126	54966.56519	10.1	45	0
1271	167	54967.40297	7.9	26	0
1852	748	54979.27489	13.0	60	0
1943	839	54981.13435	17.0	74	0
2082	978	54983.97462	27.8	84	0
2091	987	54984.15852	351.6	84	81
2146	1042	54985.28237	23.4	76	0
2236	1132	54987.12139	18.1	75	0
2282	1178	54988.06134	397.2	84	84
2367	1263	54989.79819	5.0	11	0
2421	1317	54990.90161	5.5	12	0
2498	1394	54992.47499	3.3	12	0
2559	1455	54993.72144	28.0	84	0
2735	1631	54997.31776	6.3	27	0

**Table 7: Same analysis as Table 6, for Q0 and Q1 SC. Note consecutive detections of the largest events.**

<b><u>Q0</u></b>						
CIN	RCI	Date (MJD)	MeanSNR	N <sub>chan</sub>	N <sub>pipe</sub>	
6058	559	54953.40834	17.8	47	2	
6059	560	54953.40902	16.2	48	0	
7034	1535	54954.07311	35.1	80	5	
7870	2371	54954.64253	8.6	27	0	
7911	2412	54954.67046	8.2	27	0	
12179	6680	54957.57748	137.5	82	79	
13029	7530	54958.15643	6.3	12	0	
14282	8783	54959.00987	11.0	43	0	
15509	10010	54959.84561	41.8	82	15	
15898	10399	54960.11056	68.7	70	51	

15932	10433	54960.13372	21.1	60	2
17205	11706	54961.00078	7.4	17	0
17206	11707	54961.00147	13.3	49	0
17207	11708	54961.00215	6.6	25	0
17647	12148	54961.30184	974.5	82	82
17648	12149	54961.30252	284.4	82	81
17649	12150	54961.30320	35.0	77	8
17650	12151	54961.30388	12.4	40	1
17651	12152	54961.30456	5.2	18	0

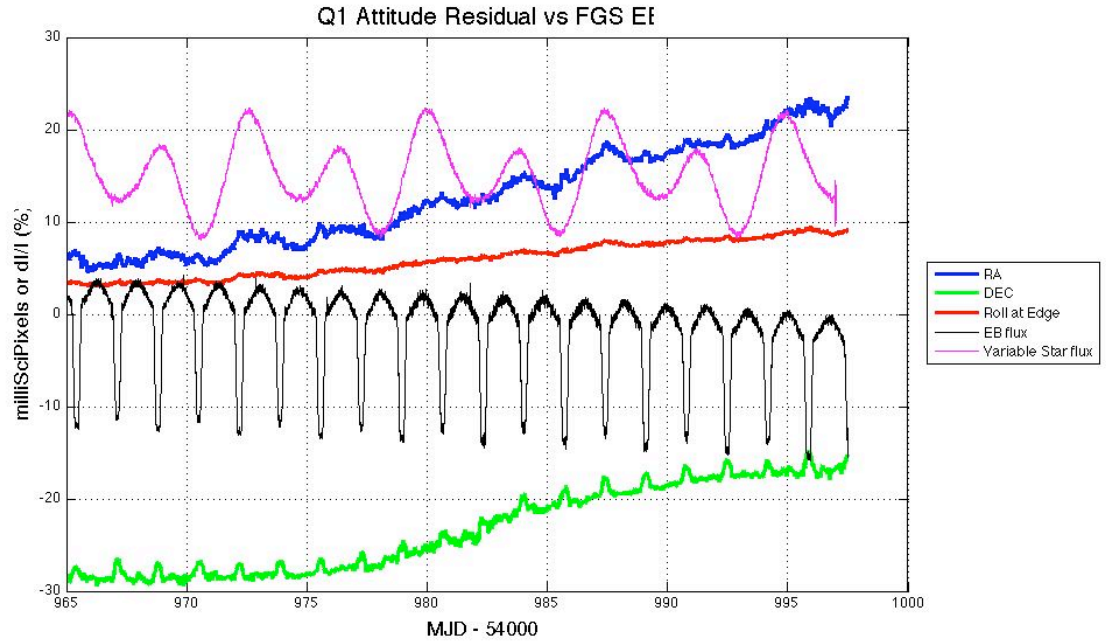
**Q1**

22952	1343	54964.91518	7.0	15	0
23586	1977	54965.34701	18.1	74	0
44038	22429	54979.27727	7.3	18	0
46769	25160	54981.13741	7.1	20	0
50932	29323	54983.97292	11.3	52	0
51205	29596	54984.15886	77.0	84	58
51206	29597	54984.15954	116.2	78	61
51207	29598	54984.16022	10.0	16	4
52852	31243	54985.28067	13.8	56	0
56921	35312	54988.05214	215.9	84	83
56922	35313	54988.05282	20.5	75	1
65246	43637	54993.72247	11.8	55	0

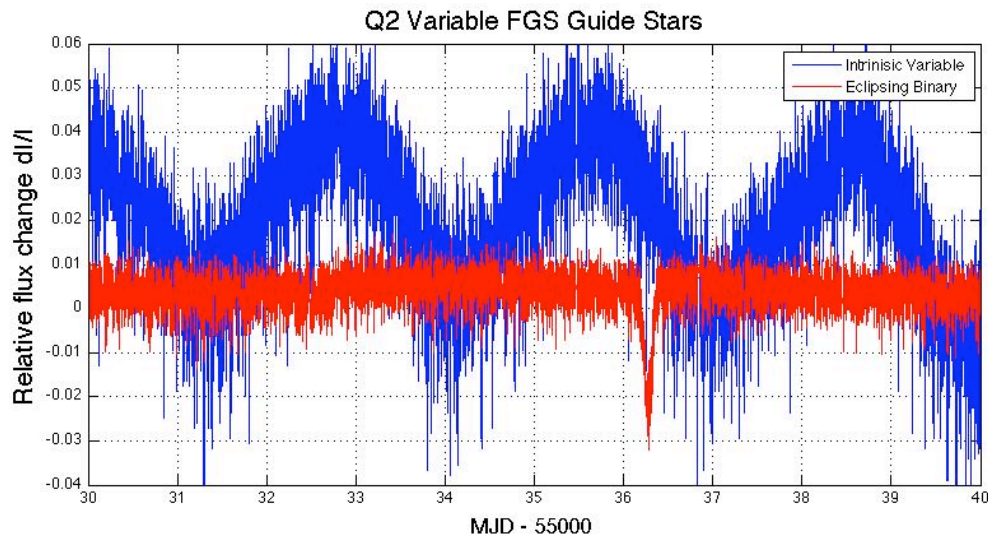
**6.2 Variable FGS Guide Stars**

The first-moment centroiding algorithm used by the FGS did not originally subtract all of the instrumental bias from the FGS pixels. Thus, the calculated centroid of an FGS star depended on the FGS star's flux when the star was not located at the center of the aperture. Variable stars then induced a variation in the attitude solution calculated from the centroids of 40 guide stars, 10 in each FGS module. The Attitude Determination and Control System (ADCS), which attempts to keep the calculated attitude of the S/C constant, then moved the S/C to respond to this varying input, with the result that the boresight of the telescope moved while the ADCS reported a constant attitude. Science target star centroids and pixel time series, and to a lesser extent aperture flux, then showed systematic errors proportional to the FGS star flux variation as shown in Figure 16. The Supplement provides users with the channel center motion time series if they wish to examine their uncorrected light curves for correlations with this motion.

At the start of Quarter 2 (6/20/2009), the most egregious variable stars were replaced with quieter stars. One intrinsic variable star and one eclipsing binary (EB) remain in the FGS; the Quarter 2 light curves for these stars are shown for historical interest in Figure 17. At the start of Quarter 3 (9/19/2009), the centroiding algorithm was updated to remove all of the instrumental background, greatly diminishing the effect of stellar variability on calculated centroids. The sky background is not removed, but is expected to be negligible. FGS guide star variability is not expected to be a factor starting with Q3.



**Figure 16: Quarter 1 attitude residual and the light curves of the two most variable FGS guide stars vs. time. The Figure shows that guide star variability leads to pointing errors with transit-like features which can be transferred to the light curves of science targets. One of the stars is an eclipsing binary (EB), while the other is an intrinsic variable. The EB flux is correlated with a pronounced transit-like signal in DEC, while the intrinsic variable star's flux is correlated with a sinusoidal component of the RA residual. These stars were removed from the FGS guide star list at the start of Q2.**



**Figure 17: Quarter 2 light curves of two variable FGS guide stars, which were present in Q0 and Q1 and remain on the FGS guide star list. One of the stars is an eclipsing binary (EB), while the other is an intrinsic variable. The period of the intrinsic variable is approximately 2.9 days, and the period of the EB is approximately 18.25 d. Only 10 days**

are shown, in order to resolve the intrinsic variability and the eclipse. The coupling of guide star variability to telescope pointing has been greatly diminished in Q3 and beyond.

### 6.3 Pixel Sensitivity Dropouts

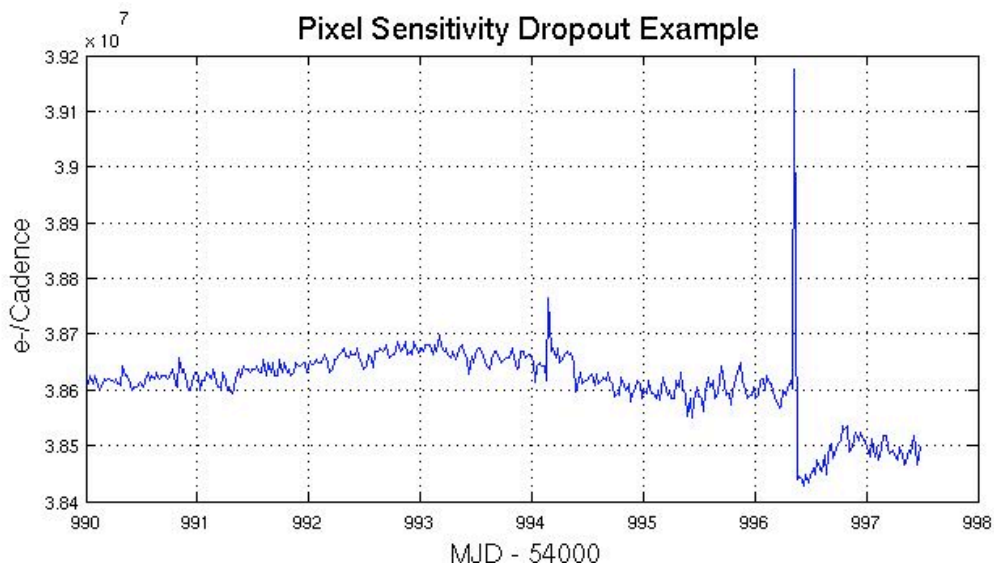
*This Section is unchanged from Release 4.*

Space-based focal planes respond to cosmic ray (CR) events in several ways:

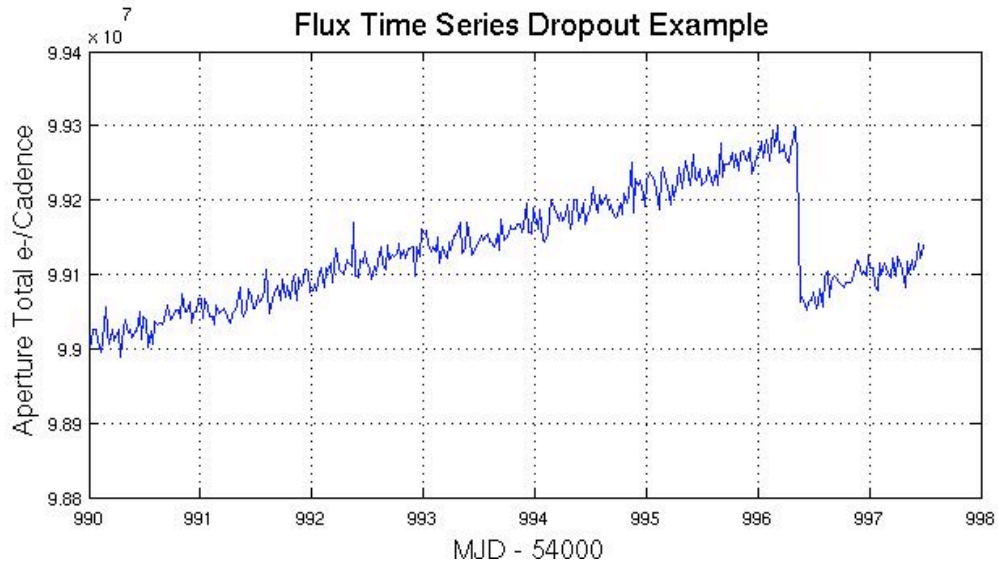
1. A transient response is induced by the charge deposited by the CR, and is cleared by the next reset (destructive readout) of the pixel.
2. Medium-term alteration of detector properties, which recover to their pre-event values after some time and resets without annealing.
3. Long-term alteration of detector properties, which are only restored by annealing the focal plane
4. Permanent damage

Typically, type 3 and 4 effects are caused by non-ionizing energy loss (NIEL), or “knock-on” damage, which can be caused by any baryonic particle.

Type 1 effects are removed by the Pipeline’s CR detection algorithm. At this point in the mission, type 3 effects do not appear to be common enough to warrant the disruption of the observing schedule that would be caused by annealing, and both type 3 and type 4 effects will eventually be mitigated by updating the bad pixel map used for calibration. Type 2 effects are not corrected by the Pipeline at the pixel level (Figure 18). In Release 5, the Pipeline corrects the aperture flux discontinuities (Figure 19) resulting from these pixel discontinuities (Section 4.4), though users examining pixel data and uncorrected light curves need to remain aware of them.



**Figure 18: Pixel time series from Q1 (Release 2) showing discontinuity after large CR event. CRs have not been removed by the Pipeline at this stage of processing. Target: KeplerID = 7960363, KeplerMag = 13.3. Dropouts are not corrected on a pixel-by-pixel basis in Release 5.**



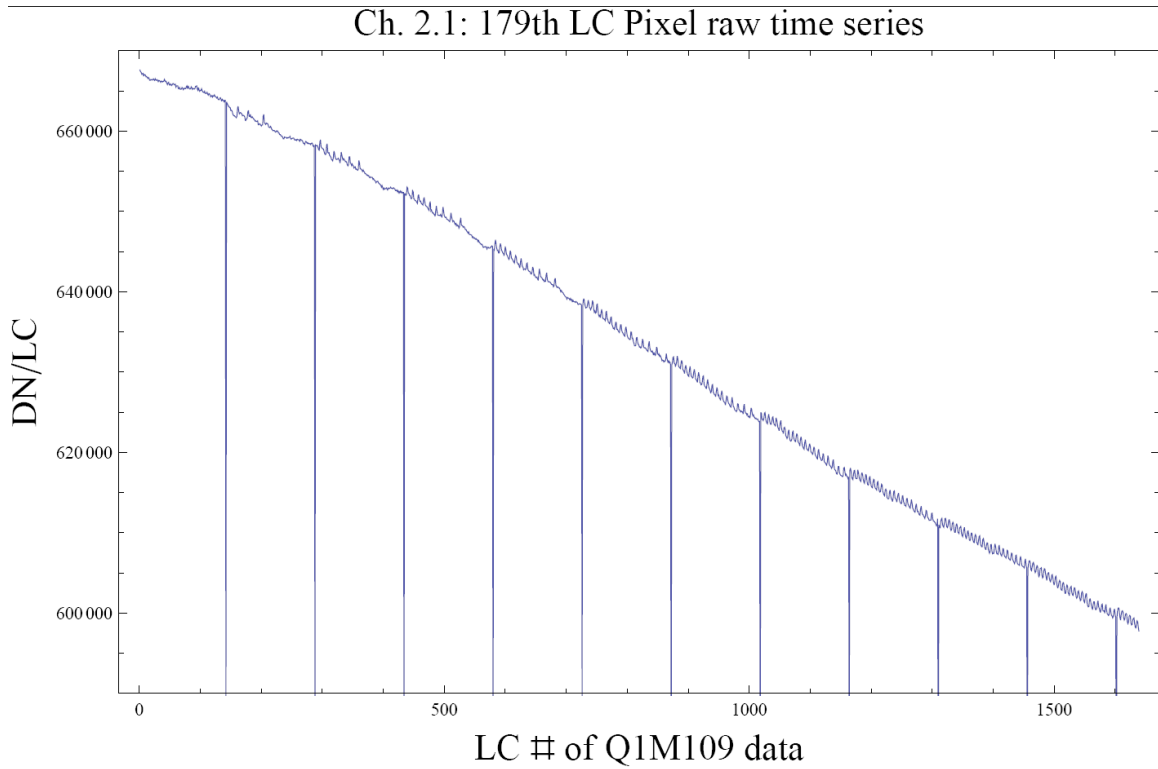
**Figure 19: Same event as for the previous Figure as seen in the uncorrected Simple Aperture Photometry (SAP) light curve produced by PA. CR hits have been removed by PA. This figure is presented for historical interest, as PDC identifies most of these discontinuities and removes before producing the corrected light curves (Section 4.4).**

#### **6.4 Focus Drift and Jitter**

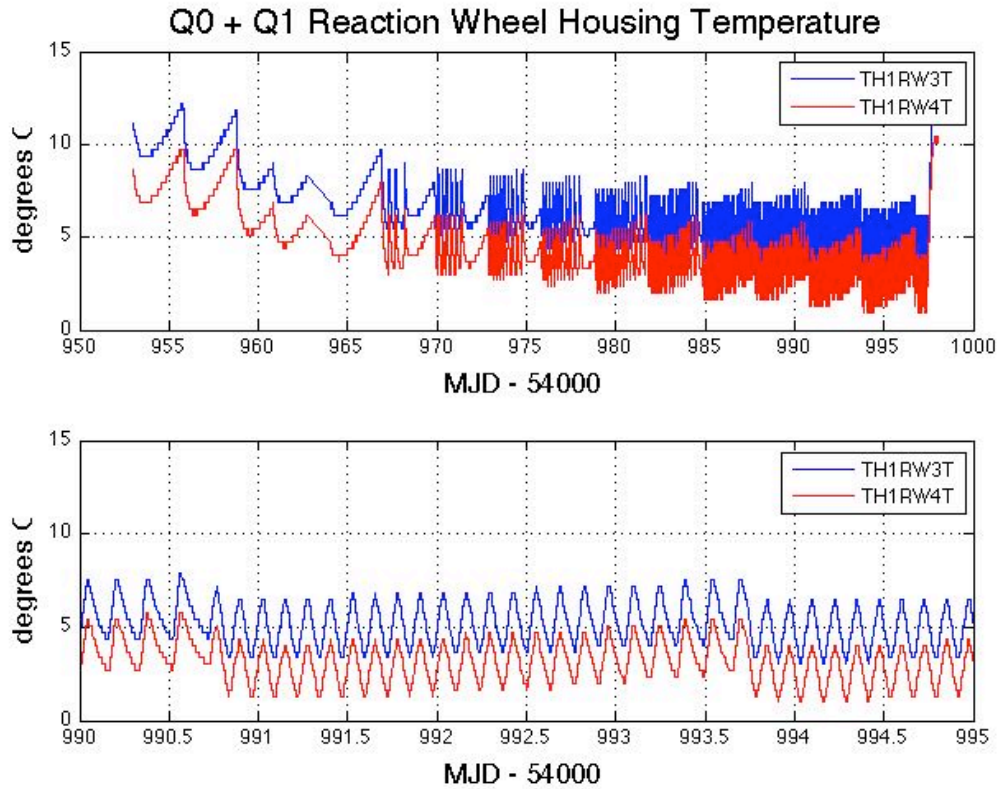
Examination of Q1 data (Figure 20) revealed that many of the science targets exhibit non-sinusoidal variations in their pixel time series with a period between 3 and 6 hours. The behavior was less frequent at the beginning of Q1 and becomes progressively worse with time. Initially, this phenomenon was associated with desaturation activities, but became nearly continuous about 15 days into the observations.

This jitter is observed in platescale metrics local to each mod.out defined by the motion of target star centroids relative to one another over time. This indicates that we are seeing a change in focus at timescales of 3 to 6 hours and that the behavior appears to be initiated by the desat activities. Reaction wheel temperature sensors with the mnemonics TH1RW3T and TH1RW4T have the same time signature, but the physical mechanism by which they couple to focus is still under discussion. At the beginning of a Quarter, the reaction wheel heaters do not cycle on and off, and the temperature changes have the same 3 day interval as the planned desaturations (Figure 21). Later in the Quarter, the heaters cycle with a 3 to 6 hr period. Near the end of Q3, at MJD = 55170, new Flight Software parameters were uploaded to substantially reduce the deadband on the reaction wheel housing temperature controller, and subsequent to that date the 3 to 6 hr cycle in both the temperature telemetry and the focus metric were eliminated.



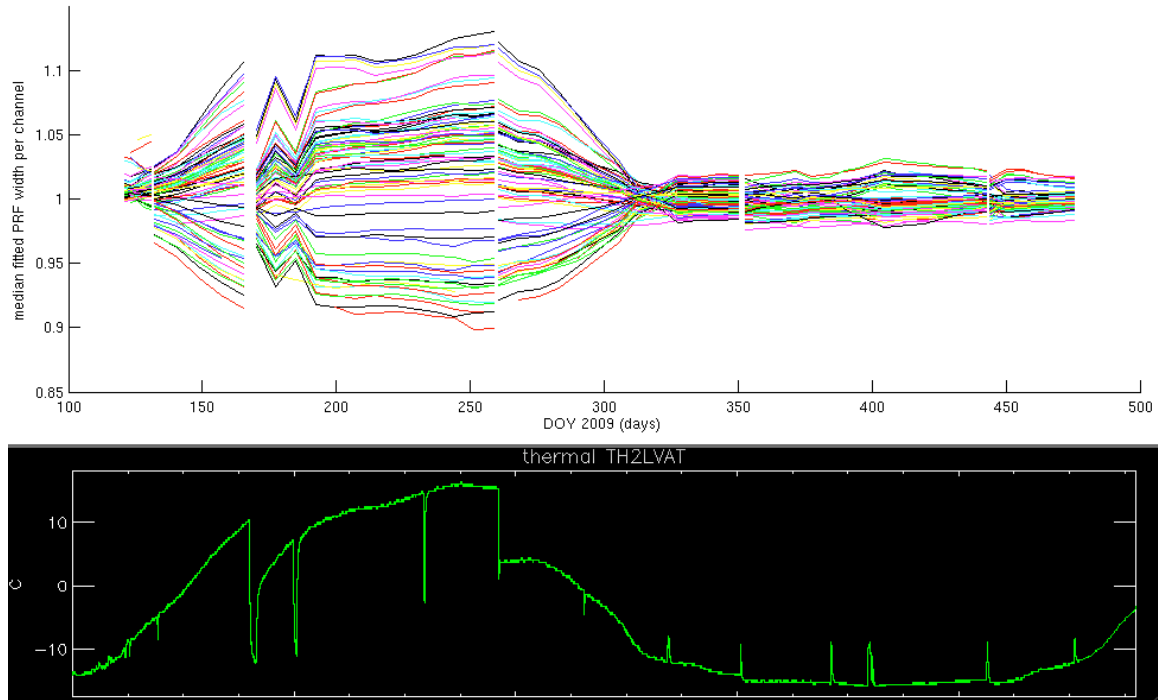


**Figure 20:** A good example of the 3 to 6 hr Focus Oscillation in a single raw pixel time series from Quarter 1. Similar signatures are seen in flux and plate scale. The large negative-going spikes are caused by desaturations (Section 5.1), which have not been removed from this time series in this plot. The abscissa is the Q1 relative Cadence index, and the ordinate is Data Numbers (DN) per Long Cadence (LC).



**Figure 21: Reaction wheel housing temperatures during Q0 and Q1. The upper panel shows that temperature variation in Q0 is dominated by a slow seasonal drift and the 3 day period of wheel desaturations. Shortly after the start of Q1, however, the reaction wheels have cooled sufficiently to engage the wheel housing heater, which then cycles on and off with a period of 3 to 6 hr (bottom panel). The telemetry data in this Figure is not plotted for times when the spacecraft is not in Fine Point, and is smoothed with a 5 point median filter.**

The DAWG investigated whether there is a secular variation of the focus and PRF width driven by the outgassing of telescope components, in addition to the seasonal and momentum dump cycles driven by temperature changes in Flight System components discussed above. Preliminary results indicate that the seasonal cycle dominates, with a good correlation between PRF width variation and the temperature of the Launch Vehicle Adapter (TH2LVAT), as shown in Figure 22. The pattern is expected to repeat, now that a full year of science data collection has passed.



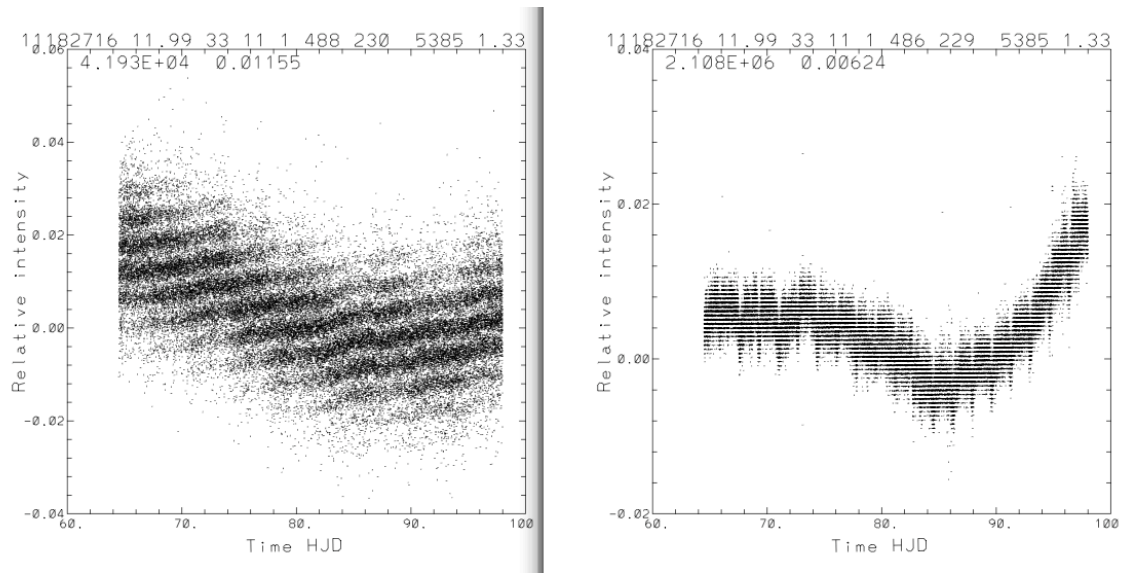
**Figure 22: Correlation of variation in PRF width with Launch Vehicle Adapter temperature TH2LVAT, demonstrating the seasonal nature of focus and PRF changes.**

For users of the PDC output, the focus changes are mostly captured by the motion polynomial coefficients used for cotrending. For users doing their own cotrending, the mod.out center motion time series provided in the Supplement will represent much of the image motion resulting from focus changes, for all targets on the corresponding mod.out. However, they do not represent local plate scale changes, which may contribute systematic errors to the light curves of individual targets on that mod.out. Thus the reaction wheel and Launch Vehicle Adapter temperature sensor telemetry for Q0 and Q1 are also provided in the Supplement.

## 6.5 Requantization Gaps

*This Section is unchanged since Release 4.*

Pixels at mean intensities  $>20,000$  e<sup>-</sup> show banding as shown in Figure 23, with quantized values of number of electrons preferred. This is the result of the onboard requantization (KIH Section 7.4), and is considered benign since in the overall extraction the light curve is near the Poisson limit. These requantization gaps are expected, and a necessary cost associated with achieving the required compression rates on board Kepler. However, it is pointed out here so that users will not suspect a problem.

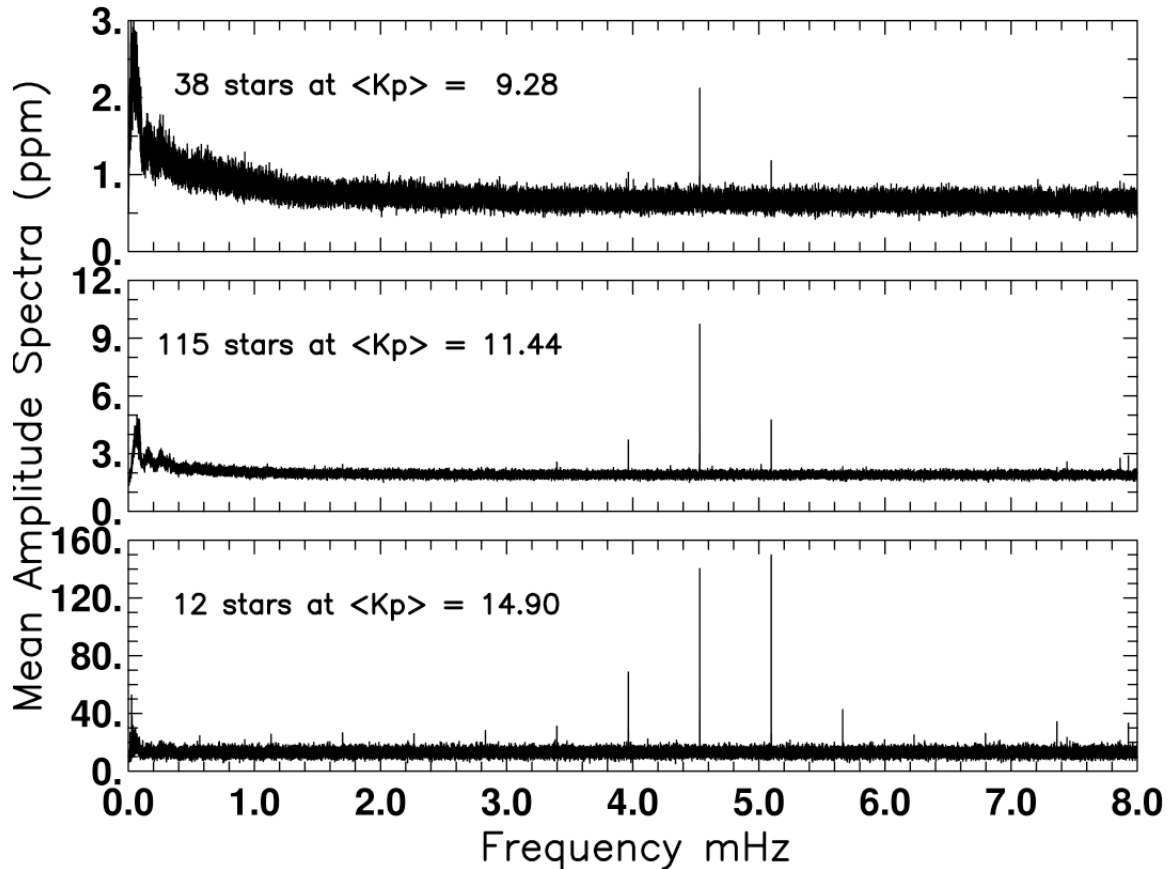


**Figure 23: Requantization gap example in Q1 SC pixel time series. The ‘band gaps’ scale with mean intensity (42,000 e- left, 2.1e6 right). See KIH Section 7.4 for a discussion of quantization and the (insignificant) information loss it entails.**

### 6.6 Spurious Frequencies in SC Data with Spacing of $1/LC$

Spurious frequencies are seen in SC flux time series, and pixel data of all types – including trailing black collateral pixels. The frequencies have an exact spacing of  $1/LC$  interval, as shown in Figure 24. As the SC data are analyzed in the frequency domain in order to measure the size and age of bright planetary host stars, the contamination of the data by these spurious frequencies will complicate these asteroseismology analyses, but will not compromise the core Kepler science. The physical cause of this problem is still under discussion, though the problem might be remedied with a simple comb notch filter in future releases even if no ancillary data can be found that exhibits these features.

This feature was first reported in Q1 data (Ref. 8). It has now been identified in pre-launch ground test data as well as Q3 flight data, and is therefore considered a normal feature of the as-built electronics. It is not an artifact introduced by the Pipeline, since it appears in raw trailing black collateral data.



**Figure 24: Mean amplitude spectra over samples of quiet stars from Q1, spanning more than a factor of 100 in brightness, showing spurious frequencies. The 1/LC-Cadence artifacts at the fundamental of 0.566391 mHz and all harmonics are visible for the faint star set in the bottom panel. Even at 9th magnitude in the upper panel this artifact remains a dominant spectral feature from the 7th and 8th harmonics. From Gilliland et al. (Ref. 8).**

### 6.7 Known Erroneous FITS header keywords

The automatically generated FFI WCS FITS header keyword values are known to be incorrect. The MAST staff have updated the 8 “Golden” FFIs from Q0 with WCS headers which are accurate to the extent that image distortion across a mod.out can be neglected (2-3 pixels error at the margins of a mod.out), as described by the file WCSREADME.txt on the MAST FFI ftp site: <http://archive.stsci.edu/pub/kepler/ffi>. For the other FFIs, users interested in sources for which light curves are unavailable, but for which KeplerIDs exist, may use the procedure for locating sources in FFIs provided by the Guest Observer office at: <http://keplergo.arc.nasa.gov/DataAnalysisInspectionFFI.shtml>. The pixel coordinate to sky coordinate tables in the KIH Supplement may be of some use for locating the corners and edges of mod.outs as projected on the sky; images of the mod.outs projected on the sky are shown at <http://archive.stsci.edu/kepler/images.html>.

The WCS keywords in the light curve and target pixel file headers are also incorrect. However, the MODULE and OUTPUT keyword values are correct, and the row and column values are in the files, so it is an easy matter to find the location of an observed target in an FFI to see the surrounding area.

## 7. Data Delivered – Format

### 7.1 FFI

The FFIs are one FITS file per image, with 84 extensions, one for each module/output. See the KIH to map the extension table number = channel number onto module and output.

### 7.2 Light Curves

*Users are strongly encouraged to read this section as there are significant differences in format between Release 5 and Release 2 of the same data set.*

Light curves have file names like kplr<kepler\_id>-<stop\_time>, with a suffix of either llc (Long Cadence) or slc (Short Cadence), and a file name extension of fits.

A light curve is time series data, that is, a series of data points in time. Each data point corresponds to a measurement from a Cadence. For each data point, the flux value from simple aperture photometry (SAP) is given, along with the associated uncertainty. Only SAP light curves are available at this time. The centroid position for the target and time of the data point are also included.

The light curves are packaged as FITS binary table files. The fields of the binary table, all of which are scalar, are briefly described below and are listed in Table 8. There are 19 fields comprising 88 bytes per Cadence; however, fields 12-19 are not populated at this time. The FITS table header listed in the Appendix of the MAST manual is superseded by Table 8. The new keywords `DATA_REL` and `QUARTER` discussed in Section 2 are in the binary table header. The module and output are identified in the binary table extension header keywords `MODULE` and `OUTPUT`.

The following data values are given for each data point in a light curve:

- barycentric time and time correction for the midpoint of the Cadence
- for the simple aperture photometry (pixel sum) of optimal aperture pixels
  - first-moment centroid position of the target and uncertainty
  - uncorrected flux value and uncertainty. Gap Cadences are set to -Inf
  - corrected flux value and uncertainty. Gap Cadences are set to -Inf

**Table 8: Available light curve data table fields, modified after the MAST manual KDMC-10008 (August 30, 2009): SAP replaces OAP, and data in columns 12-19 is not available and are filled with -Inf. Time units are the same in Releases 3-5.**

Column Number	Field Name	Data Type	Bytes	Description	Units
1	barytime	1D	8	barycentric time BJD – 2400000. See Section 7.4 for detailed discussion.	days
2	timcorr	1E	4	barycentric time correction. See Section 7.4 for detailed discussion	seconds
3	Cadence_number	1J	4	Cadence number (CIN)	N/A
4	ap_cent_row	1D	8	row pixel location	pixels
5	ap_cent_r_err	1E	4	error in row pixel location	pixels
6	ap_cent_col	1D	8	column pixel location	pixels
7	ap_cent_c_err	1E	4	error in column pixel location	pixels
8	ap_raw_flux	1E	4	SAP uncorrected flux	e- / Cadence
9	ap_raw_err	1E	4	SAP uncorrected flux error	e- / Cadence
10	ap_corr_flux	1E	4	SAP corrected un-filled flux	e- / Cadence

Column Number	Field Name	Data Type	Bytes	Description	Units
11	ap_corr_err	1E	4	SAP corrected un-filled flux error	e- / Cadence

#### Data Types:

1D – double precision floating point.

1E – single precision floating point. Note that, although all SOC calculations and internal data representation is double-precision, the SAP fluxes and errors are reported as single-precision floats, which will give roundoff errors of approximately 0.11 ppm (*Numerical Recipes* Chapter 20 & confirmed by numerical experiments on MAST and internal SOC data).

1J – 32 bit integer

See Section 7.4 for a discussion of time and time stamps.

If you are an IDL user, the `tbget` program in the `astrolib` library extracts the data. If you are an IRAF user, `tprint` can be used to dump an ascii table of selected row and column values.

### 7.3 Pixels

*Target Pixel Data Files are not currently available as of 6/15/2010, but will be in the near future.*

Target **pixel** data files contain all the pixels for a target from all Cadences, while target **Cadence** files contain pixels from all targets for a single Cadence. Up to 5 different files are produced for each target. These consist of the Long and Short Cadence pixel data for the target, the collateral pixels for the Long and Short Cadences, and the background pixels.

Both original pixel values and calibrated flux values are in the pixel data files. The original pixel value is the integer value as recorded on the spacecraft. The calibrated pixel value is that provided by the SOC, and is equal to the output of CAL with cosmic rays removed. The calibrated pixels have **not** had background subtracted. Unfortunately, since the background is not subtracted, and users are not presently provided with the list of pixels in the optimal aperture, there is no simple way to construct the uncorrected (PA output) light curve from the ‘calibrated’ pixels. Users should be aware that the format and content of the target pixel files is the subject of vigorous discussion (Section 7.5), in the hopes of remedying this situation.

The target pixel data files are archived as a dataset. A request for the data will return all extensions that were archived with the dataset.

Pixel data table fields are described in the Kepler Archive Manual (KDMC-10008).

### 7.4 Time and Time Stamps

The primary time stamps available for each Cadence in both LC and SC time series are intended to provide proper BJD times corrected to the solar system barycenter, at the flux-weighted mid-point of the Cadences, and are uniquely determined for each star individually.

*Users are urged to read this Section as a close reading may help them avoid attempting to do follow-up observations at the wrong time.*

#### 7.4.1 Overview

The precision and accuracy of the time assigned to a Cadence are limited by the intrinsic precision and accuracy of the hardware and the promptness and reproducibility of the flight software time-stamping process. The Flight System requirement, including both hardware and software contributions, is that the absolute time of the start and end of each Cadence is known to  $\pm 50$  ms. This requirement was developed so that knowledge of astrophysical event times would

be limited by the characteristics of the event, rather than the characteristics of the flight system, even for high SNR events.

Several factors must be accounted for before approaching the 50 ms limit:

1. Relate readout time of a pixel to Vehicle Time Code (VTC) recorded for that pixel and Cadence in the SSR. The VTC stamp of a Cadence is created within 4 ms after the last pixel of the last frame of the last time slice of that Cadence is read out from the LDE.
2. VTC to UTC of end of Cadence, using Time Correlation Coefficients. Done by DMC, with precision and accuracy to be documented.
3. Convert UTC to Barycentric JD. This is done in PA (Section 4.3) on a target-by-target basis. The amplitude of the barycentric correction is approximately  $(a_K/c)\cos\beta$ , where  $a_K \sim 1.02$  AU is the semi-major axis of Kepler's approximately circular ( $e_K < 0.04$ ) orbit around the Sun,  $c$  the speed of light, and  $\beta$  is the ecliptic latitude of the target. In the case of the center of the Kepler FOV, with  $\beta = 65$  degrees, the amplitude of the UTC to barycentric correction is approximately +/- 211 s. BJD is later than UTC when Kepler is on the half of its orbit closest to Cygnus (roughly May 1 – Nov 1) and earlier than UTC on the other half of the orbit. This correction is done on a target-by-target basis to support Kepler's 50 ms timing accuracy requirement.
4. Subtract readout time slice offsets (See KIH Section 5.1). This is done in PA (Section 4.3) in Release 4. The magnitude of the time slice offset is  $t_{rts} = 0.25 + 0.62(5 - n_{slice})$  s, where  $n_{slice}$  is the time slice index (1-5) as described in the KIH. Note that this will in general be different from Quarter to Quarter for the same star, as the star will be on different mod.outs, so the relative timing of events across Quarter boundaries must take this into account.

## 7.4.2 Time Stamp Definitions for Release 5

Cadence files:

JD = Julian Date

MJD = Modified Julian Date

MJD = JD - 2400000.5

1.  $STARTTIME(i) = \text{MJD of start of } i^{\text{th}} \text{ Cadence}$
2.  $END\_TIME = \text{MJD of end of } i^{\text{th}} \text{ Cadence}$
3.  $MID\_TIME(i) = \text{MJD of middle of Cadence} = (STARTTIME(i) + END\_TIME(i))/2$
4.  $JD(i) = MID\_TIME(i) + 2400000.5$

Releases 4 and 5 light curves, with barycentric and time slice corrections:

1.  $timcorr(i) = dtB(i) - t_{rts}$ , where  $dtB(i)$  = barycentric correction generated by PA, a function of Cadence  $MID\_TIME(i)$  and target position, and  $t_{rts}$  is the readout time slice offset described in Section 7.4.1. Units: seconds.
2.  $BJD(i) = \text{barycentric Julian Date} = timcorr(i)/86400 + JD(i)$ . Units: days
3.  $barytime(i) = \text{Barycentric Reduced Julian Date} = BJD(i) - 2400000 = timcorr(i)/86400 + MID\_TIME(i) + 0.5$ . Units: days
4.  $LC\_START = \text{MJD of beginning of first Cadence (uncorrected)}$ . Units: days
5.  $LC\_END = \text{MJD of end of last Cadence (uncorrected)}$ . Units: days



Or, as is summarized in the FITS table header:

COMMENT barytime(i)- timcorr(i)/86400 - 0.5 = utc mjd(i) for cadence\_number(i)

Where utc mjd(i) for cadence\_number(i) is the same as MID\_TIME(i)

### The difference between Release 3 and Releases 4 and 5

In Release 3,  $t_{\text{rts}} = 0$  for all targets, while in Release 4  $t_{\text{rts}}$  is calculated as described in Section 7.4.1. That is the only difference.

### The vexing matter of the 0.5 days

Users should note that barytime follows the same conventions as Julian Date, and astronomers in general; that is, the day begins at noon. MJD, on the other hand, follows the convention of the civil world: that the day begins at midnight. If  $\text{timcorr} = 0$ , then  $\text{MJD} = \text{barytime} - 0.5 \text{ d}$  and  $\text{barytime} = \text{MJD} + 0.5 \text{ d}$ .

### **7.4.3 Caveats and Uncertainties**

Factors which users should consider before basing scientific conclusions on time stamps are:

1. The precise phasing of an individual pixel with respect to the Cadence time stamp (not understood to better than  $\pm 0.5 \text{ s}$ ) at this time.
2. General and special relativistic effects in the calculation of the barycentric correction. For example, time dilation at Kepler with respect to a clock at rest with respect to the solar system barycenter, but outside the Sun's gravity well, is  $7.5 \times 10^{-9} = 0.23 \text{ s/yr}$  – so these effects cannot be dismissed out of hand at this level, and must be shown to be negligible at the level of Kepler's time accuracy requirement of 50 ms or corrected for.
3. The existing corrections have yet to be verified with flight data.
4. Light travel time and relativistic corrections to the user's target, if the target is a component of a binary system.

The advice of the DAWG is not to consider as scientifically significant relative timing variations less than the read time (0.5 s) or absolute timing accuracy better than one frame time (6.5 s) until such time as the stability and accuracy of time stamps can be documented to near the theoretical limit.

### **7.5 Future Formats Under Discussion**

The Science Office recognizes that the MAST products are deficient in several ways, and are discussing the following improvements:

1. Provision of an aperture extension, which will tell users which pixels were used to calculate uncorrected flux time series and centroids.
2. Data quality flags, encoding much of the information on lost or degraded data and systematic errors provided in these Notes in a way that will spare users the drudgery of fusing the data in the Supplement with the light curve data.
3. A local WCS coordinate system derived from linearized motion polynomials with Cadence-to-Cadence corrections from the mid-time of the data set. These corrections are a target-specific image motion time series for users to use in their own systematic error correction and are thus an improvement on the mod.out center motion time series provided in the Supplement.
4. Packaging target pixel files as columns of images rather than as pixel lists.

## 8. References

1. "Initial Assessment Of The Kepler Photometric Precision," W.J. Borucki, NASA Ames Research Center, J. Jenkins, SETI Institute, and the Kepler Science Team (May 30, 2009)
2. "Kepler's Optical Phase Curve of the Exoplanet HAT-P-7," W. J. Borucki *et al.*, *Science* Vol 325 7 August 2009 p. 709
3. "Pixel Level Calibration in the Kepler Science Operations Center Pipeline," E. V. Quintana *et al.*, SPIE Astronomical Instrumentation conference, June 2010.
4. "Photometric Analysis in the Kepler Science Operations Center Pipeline," J. D. Twicken *et al.*, SPIE Astronomical Instrumentation conference, June 2010.
5. "Presearch Data Conditioning in the Kepler Science Operations Center Pipeline," J. D. Twicken *et al.*, SPIE Astronomical Instrumentation conference, June 2010.
6. Dave Monet, private communication.
7. "Initial Characteristics of Kepler Long Cadence Data for Detecting Transiting Planets," J. M. Jenkins *et al.*, *ApJ Letters* **713**, L120-L125 (2010)
8. "Initial Characteristics of Kepler Short Cadence Data," R. L. Gilliland *et al.*, *ApJ Letters* **713**, L160-163 (2010)
9. "Overview of the Kepler Science Processing Pipeline," Jon M. Jenkins *et al.*, *ApJ Letters* **713**, L87-L91 (2010)
10. "Discovery and Rossiter-McLaughlin Effect of Exoplanet Kepler-8b," J. M. Jenkins *et al.*, submitted to *ApJ* <http://arxiv.org/abs/1001.0416>
11. "Kepler Mission Design, Realized Photometric Performance, and Early Science," D. Koch *et al.*, *ApJ Letters* **713**, L79-L86 (2010)
12. "Selection, Prioritization, and Characteristics of Kepler Target Stars," N. Batalha *et al.*, *ApJ Letters* **713**, L109-L114 (2010)
13. "Kepler Science Operations," M. Haas *et al.*, *ApJ Letters* **713**, L115-L119 (2010)
14. "The Kepler Pixel Response Function," S. Bryson *et al.*, *ApJ Letters* **713**, L97-L102 (2010)
15. "Instrument Performance in Kepler's First Months," D. Caldwell *et al.*, *ApJ Letters* **713**, L92-L96 (2010)

## 9. List of Acronyms and Abbreviations

ACS	Advanced Camera for Surveys
ADC	Analog to Digital Converter
ADCS	Attitude Determination and Control Subsystem
AED	Ancillary Engineering Data
ARP	Artifact Removal Pixel
BATC	Ball Aerospace & Technologies Corp.
BG	BackGround pixel of interest
BOL	Beginning Of Life
BPF	Band Pass Filter
CAL	Pixel Calibration module
CCD	Charge Coupled Device
CDPP	Combined Differential Photometric Precision
CDS	Correlated Double Sampling
CR	Cosmic Ray
CSCI	Computer Software Configuration Item
CTE	Charge Transfer Efficiency
CTI	Charge Transfer Inefficiency
DAA	Detector Array Assembly
DAP	Data Analysis Program
DAWG	Data Analysis Working Group
DCA	Detector Chip Assembly
DCE	Dust Cover Ejection
DIA	Differential Image Analysis
DMC	Data Management Center
DNL	Differential Non-Linearity of A/D converter
DSN	Deep Space Network
DV	Data Validation module
DVA	Differential Velocity Aberration
ECA	Electronic Component Assembly
EE	Encircled Energy
EOL	End of Life
ETEM	End-To-End Model of Kepler
FFI	Full Field Image
FFL	Field Flatteners Lens
FGS	Fine guidance sensor
FOP	Follow-up Observation Program
FOV	Field of View
FPA	Focal Plane Assembly
FPAА	Focal Plane Array Assembly

FSW	Flight Software
GCR	Galactic Cosmic Ray
GO	Guest Observer
GUI	Graphical User Interface
HGA	high-gain antenna
HST	Hubble Space Telescope
HZ	Habitable Zone
I&T	Integration and Test
INL	Integral Non-Linearity of A/D converter
IRNU	Intra-pixel Response Nonuniformity
KACR	Kepler Activity Change Request (for additional data during Commissioning)
KAR	Kepler Anomaly Report
KCB	Kepler Control Box
KDAH	Kepler Data Analysis Handbook
KIC	Kepler Input Catalog
KSOP	Kepler Science OPERations
KTD	Kepler Tech Demo (simulated star field light source)
LC	Long Cadence
LCC	Long Cadence Collateral
LDE	Local Detector Electronics
LGA	low-gain antenna
LOS	Line of Sight
LPS	LDE Power Supply
LUT	look-up table
LV	Launch Vehicle
MAD	Median Absolute Deviation
MAST	Multi-mission Archive at STSci
MJD	Modified Julian Date = JD - 2400000.5
MOC	Mission Operation Center
MORC	Module, Output, Row, Column
NVM	Non-Volatile Memory
OFAD	Optical Field Angle Distortion
PA	Photometric Analysis module
PAD	Photometer Attitude Determination (Pipeline S/W)
PDC	Pre-Search Data Conditioning module
PID	Pipeline instance Identifier (unique number assigned to each run of the Pipeline)
PM	Primary Mirror
PMA	Primary Mirror Assembly
POI	Pixels of Interest

PPA	Photometer Performance Assessment (Pipeline S/W)
ppm	parts per million
PRF	Pixel Response Function
PRNU	Pixel Response Non-Uniformity
PSD	power spectral density
PSF	Point Spread Function
PSP	Participating Scientist Program
PWA	Printed Wiring Assembly
QE	Quantum Efficiency
RC	Reverse Clock
S/C	Spacecraft
S/W	Software
SAO	Smithsonian Astrophysical Observatory
SC	Short Cadence
SCo	Schmidt Corrector
SDA	Science Data Accumulator
SNR	Signal-to-Noise Ratio
SO	Science Office
SOC	Science Operations Center
SOL	Start-of-Line
SSR	Solid State Recorder
SSTVT	Single-String Transit Verification Test
STScI	Space Telescope Science Institute
SVD	Singular Value Decomposition
TAD	Target and Aperture Definition module
TDT	Target Definition Table
TPS	Transiting Planet Search module
TVAC	Thermal Vacuum testing

## 10. Contents of Supplement

The Supplement is available as a full package (DataReleaseNotes\_05\_SupplementFull.tar) and a short package suitable for emailing (DataReleaseNotes\_05\_SupplementSmall.tar). The small package does not contain

Q1-SC-MAST-R5\_background.txt

Q1-SC-MAST-R5\_background.mat

Q0-SC-MAST-R5\_background.txt

Q0-SC-MAST-R5\_background.mat

Q0Q1\_TH1RW34T\_MJD\_gap.txt

Q0Q1\_TH12LVAT\_MJD\_gap.txt

### 10.1 Pipeline Instance Detail Reports

These files list the Pipeline version and parameters used to process the data, so that the Pipeline results in this Release can be reconstructed precisely at some future time. Multiple files for the same data set are needed if the Pipeline needs to be re-run from a particular step, as was the case for Q0 LC data. The file names are:

Q0\_LC\_r6.1\_ksop435\_CAL\_complete\_Pipeline\_Instance\_Detail\_Report\_100512.txt

Q0\_LC\_r6.1\_ksop435\_PA\_complete\_Pipeline\_Instance\_Detail\_Report\_100512.txt

Q0\_LC\_r6.1\_ksop435\_PDC\_to\_PPA\_Pipeline\_Instance\_Detail\_Report\_100512.txt

Q0\_SC\_r6.1\_ksop435\_Pipeline\_Instance\_Detail\_Report\_100512.txt

Q1\_LC\_r6.1\_ksop435\_CAL\_to\_PPA\_as-run\_Pipeline\_Instance\_Detail\_Report\_100421.txt

Q1\_SC\_r6.1\_ksop435\_as-run\_Pipeline\_Instance\_Detail\_Report\_100421.txt

### 10.2 Thermal and Image Motion Data for Systematic Error Correction

These files are provided so that users can perform their own systematic error correction, if they conclude that the methods used by PDC are not suitable for their targets and scientific goals. It is important to remember that inclusion of additional time series to the cotrending basis set may not improve the results if the cotrending time series are noisy, poorly sampled, or nearly degenerate. The thermal AED will, in general, have to be resampled to match the Cadence times, and on physical grounds it may be more effective to cotrend against bandpass-filtered AED as separate basis vectors. See the SPIE PDC paper (Ref. 5) for a brief discussion of synchronizing ancillary data to mid-Cadence timestamps, and the use of synchronized AED as a cotrending basis set.

#### 10.2.1 Mod.out Central Motion

On rare occasions (<2% of the points), users may notice some “chatter” in the motion time series, which results from a known problem with the motion polynomial fitting algorithm and not actual jumps in telescope attitude or CCD position. A more robust, iterative algorithm has been identified will be implemented in future Pipeline software to remedy this problem. Users will also clearly see DVA and the signatures of the variable FGS guides stars (Section 6.2) and the reaction wheel heaters (Section 6.4) in the motion time series.

Files: Q1-MAST-R5\_central\_column\_motion.txt and Q1-MAST-R5\_central\_row\_motion.txt -- Channel central column (row) motion from motion polynomials for all channels, sampled at the Long Cadence period.

Column Descriptions

1. Cadence Interval Number
  2. Relative Cadence Index
  3. Gap Indicator. 1 = Momentum Dump or Loss of Fine Point
  4. Cadence mid-Times, MJD
- 5-88. Mod.out center column (row) for each channel. Units: pixels. mod.outs are shown

Q1-MAST-R5\_central\_motion.mat – MATLAB file containing both row and column motion; this will spare MATLAB users the drudgery of parsing the text files.

### 10.2.2 Average LDE board Temperature

Q0Q1\_LDE\_averageBoardTemp.txt – average of the ten LDE board temperatures.

Column descriptions:

1. MJD - 54900, units: d, sampling 6.92E-04 d = 59.75 s
2. Average temperature, units: C

### 10.2.3 Reaction Wheel Housing Temperature

Q0Q1\_TH1RW34T\_MJD\_gap.txt – Reaction wheel housing temperature. Data are gapped for desats and between Q0 and Q1 and median-filtered with a box width = 5 samples.

Column definitions:

1. MJD – 54000, units: d, sampling (unfiltered) = 58.0 s
2. TH1RW3T – units: C
3. TH1RW4T – units: C

### 10.2.4 Launch Vehicle Adapter Temperature

Q0Q1\_TH12LVAT\_MJD\_gap.txt -- Launch Vehicle Adapter Temperature. Data are gapped for desats and between Q0 and Q1 and median-filtered with a box width = 5 samples.

Column definitions:

1. MJD – 54900, units: d, sampling (unfiltered) = 58.0 s
2. TH1LVAT – units: C
3. TH2LVAT – units: C

## 10.3 Background Time Series

Q0-LC-MAST-R5\_background.txt

Q0-SC-MAST-R5\_background.txt

Q1-LC-MAST-R5\_background.txt

Q1-SC-MAST-R5\_background.txt

Column definitions

1. Cadence Interval Number
2. Relative Cadence Index for Argabrightening Cadences
3. Gap Indicator. 1 = No Data, Momentum Dump, or Loss of Fine Point
4. Cadence mid-Times, MJD

5. Mean background current averaged over FPA, e-/Cadence. All zeros = no SC targets

6-89. Mod.out background in e-/Cadence for each channel. mod.outs are shown

Corresponding MATLAB files are provided to spare MATLAB users the drudgery of parsing the text files.

## **10.4 Flight System Events**

### Argabrightening Detections

ArgAgg\_Q0\_LC\_PID278\_MADT010\_MCT10\_Summary.txt

ArgAgg\_Q0\_SC\_PID338\_MADT010\_MCT10\_Summary.txt

ArgAgg\_Q1\_LC\_PID359\_MADT010\_MCT10\_Summary.txt

ArgAgg\_Q1\_SC\_PID379\_MADT010\_MCT10\_Summary.txt

Column Definitions:

1. Cadence Interval Number for Argabrightening Cadences
2. Relative Cadence Index for Argabrightening Cadences
3. Arg Cadence mid-Times, MJD
4. Mean SNR over Channels of Arg Event
5. Channels exceeding threshold in Arg Cadence
6. Channels exceeding default threshold in ArgCadence

### Out of Fine Point Cadence Lists

Q0\_LC\_isNotFinePoint.txt

Q0\_SC\_isNotFinePoint.txt

Q1\_LC\_isNotFinePoint.txt

Q1\_SC\_isNotFinePoint.txt

## **10.5 Calibration File READMEs**

The calibration file names are not listed in the headers of the light curves and target pixel files. The calibration file names listed in the FITS headers of Cadence files and FFIs are not, in general, correct. The calibration files actually used in Release 5 are:

kplr2008072318\_gain.readme.txt

kplr2008102416\_read-noise.readme.txt

kplr2008102809\_undershoot.readme.txt

kplr2009060215\_linearity.readme.txt

kplr2009060615-mmo\_2d-black.readme.txt

kplr2009062300\_lsflat.readme.txt

kplr2009062414-MMO\_ssflat.readme.txt