



Kepler: A Search for Terrestrial Planets

Kepler Data Characteristics Handbook

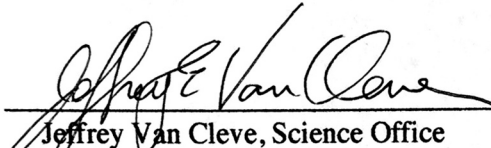
KSCI-19040-005

Data Analysis Working Group (DAWG)

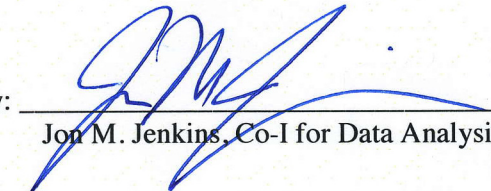
15 December 2016


NASA Ames Research Center

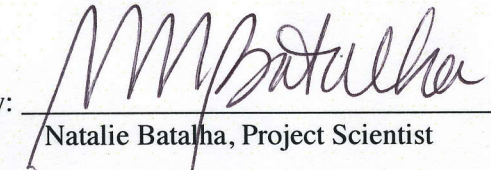
Moffett Field, CA. 94035

Prepared by:  Date 12/15/16
Jeffrey Van Cleve, Science Office

Approved by:  Date 12/15/2016
Doug Caldwell, Instrument Scientist

Approved by:  Date 12/15/2016
Jon M. Jenkins, Co-I for Data Analysis & DAWG Lead

Approved by:  Date 12/15/16
Michael R. Haas, Science Office Director

Approved by:  Date 12-15-2016
Natalie Batalha, Project Scientist

DOCUMENT CHANGE LOG

From KSCI-19040-004

Section	Change/Update
All	Citation by name and year not by number Modernize references where necessary Remove forward-looking statements
preface	Make less admonitory
1.1	Extend to end of mission with uniform formatting
1.2	Extend to end of mission and clarify meaning of FFI time stamp
1.3	Modernize major document list
1.4	Update
3.1	Replace with CDDP tracking plot from DRN 25; update scatter plot; update Table 4
4.6	post- <i>Kepler</i> mod 7 failure
4.7	Wheel 4 failure, rest
4.8	This problem has been corrected
5.11	Preamble about MAP for SC
5.11	Generate similar figure for Q5 data and assess stability of spurious signal over mission.
5.11	Check internal consistency of table, reference Baran, correct first line. Make units consistent (uHz)
5.12	Tidy format
5.17	Added new section for fixed bugs in SC pipeline processing
6	Remove forward statements about reprocessing Q0-Q14 for correct time
6.3	No longer needed
6.4	Tidy up. Remove cautions and forward-looking statements
9	PASP style, check and clean up references
11	SOC Pipeline history appendix removed in favor of recent KDRNs and KDPH

The Data Characteristics Handbook is the collective effort of the Data Analysis Working Group (DAWG), which is composed of Science Office (SO), Science Operations Center (SOC), and Guest Observer (GO) Office members as listed below:

Jon Jenkins, Chair

Doug Caldwell, Co-Chair

Barclay, Thomas

Bryson, Stephen T.

Burke, Christopher J.

Campbell, Jennifer

Catanzarite, Joseph

Christiansen, Jessie L.

Clarke, Bruce D.

Coughlin, Jeffrey L.

Girouard, Forrest

Haas, Michael R.

Klaus, Todd

Kolodziejczak, Jeffery (MSFC)

Li, Jie

McCauliff, Sean D.

Morris, Robert L.

Mullally, Fergal

Quintana, Elisa V.

Rowe, Jason

Sabale, Anima

Seader, Shawn

Smith, Jeffrey Claiborne

Still, Martin D.

Tenenbaum, Peter G.

Thompson, Susan E.

Twicken, Joe

Uddin, Akm Kamal

Van Cleve, Jeffrey E.

Zamudio, Khadeejah

When citing this document, please use the following reference:

J. E. Van Cleve, J. L. Christiansen, J. M. Jenkins, D. A. Caldwell, T. Barclay, S. T. Bryson, C. J.

Burke, J. Campbell, J. Catanzarite, B. D. Clarke, J. L. Coughlin, F. Girouard, M. R. Haas, T. C.

Klaus, J. J. Kolodziejczak, J. Li, S. D. McCauliff, R. L. Morris, F. Mullally, E. V. Quintana, J.

Rowe, A. Sabale, S. Seader, J. C. Smith, M. D. Still, P. G. Tenenbaum, S. E. Thompson, J. D.

Twicken, A. K. Uddin, and K. Zamudio, 2016, *Kepler Data Characteristics Handbook* (KSCI-19040-005).

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

1. Introduction	9
1.1 Dates, Cadence Numbers, and Units	10
1.2 Full Frame Images	13
1.3 Document Overview	14
2. Release Description	15
3. Evaluation of Performance	16
3.1 Overall	16
4. Historical Events	19
4.1 Kepler Mission Timeline To Date	19
4.2 Safe Mode	19
4.3 Loss of Fine Point	20
4.4 Attitude Tweaks Every Few Days	20
4.5 Variable FGS Guide Stars	21
4.6 Module 3 Failure	22
4.7 Reaction Wheel Failures	23
4.8 CH Cyg Corrupts the Smear Correction	23
5. Ongoing Phenomena	25
5.1 Image Motion	25
5.2 Focus Changes	26
5.3 Momentum Desaturation	29
5.4 Reaction Wheel Zero Crossings	29
5.5 Downlink Earth Point	31
5.6 Manually Excluded Cadences	31
5.7 Incomplete Apertures Produce Discontinuities at Quarter Boundaries	31
5.8 Argabrightening	32
5.9 Pixel Sensitivity Dropouts	33
5.10 Short Cadence Requantization Gaps	34
5.11 Spurious Frequencies in SC Data	35
5.12 Propagation of Uncertainties	38
5.13 Onboard Spacecraft Errors	39
5.14 Coronal Mass Ejections - Data Loss and Detector Changes	39

5.15	Attenuation of Long Period Signals	40
5.16	Anomaly Summary Figure	41
5.17	Fixed Bugs in SC Pipeline Processing	43
6.	Time and Time Stamps.....	44
6.1	Overview	44
6.2	Time Transformations, VTC to BKJD	44
6.3	Absolute Timing Accuracy with Kepler.....	46
7.	Ensemble Cotrending Basis Vectors	48
7.1	Introduction	48
7.2	Generation of CBVs	48
7.3	Using CBVs.....	49
7.4	Cautions	50
7.5	Obtaining CBVs	50
8.	Contents of Supplement	51
9.	References	52
10.	List of Acronyms and Abbreviations	53

Preface

The *Kepler* Data Processing Pipeline was developed for automated photometry of over 160,000 stars, and thus has always had to balance the Benthamite goal of doing the greatest good for the greatest number against the hazard of signal distortion for a particular target. In its final version, the pipeline used the multi-scale Bayesian Maximum A Priori (msMAP) method of systematic error removal, which produced significantly improved results for most long cadence targets compared to the pipeline version available soon after launch in 2009 (see papers by Smith *et al.* 2012 and Stumpe *et al.* 2012, 2014). Since even this final version of the pipeline may not be optimal for a particular target or science purpose, *Kepler* also provides target pixel files so that users can perform their own photometry. In addition, *Kepler* provides cotrending basis vectors (see §7.0) for all quarters so that users can better remove systematics from light curves generated by the pipeline or by themselves from target pixel files, while weighing the trade between systematic noise removal and signal fidelity given their target characteristics and science goals.

Investigators are strongly encouraged to study the *Kepler* Data Characteristics Handbook (KDCH) and Data Release Notes (DRNs) for all data sets that they intend to use; these documents in turn often refer to the *Kepler Instrument Handbook* (KIH) for basic information about the construction and operation of the *Kepler* flight system. We advise against publication of results based on *Kepler* light curves without such careful consideration and due diligence by the end user, and dialogue with the Science Office or Guest Observer Office where appropriate.

Users are encouraged to notice and document artifacts or issues with the processing and report them to the *Kepler* Science Office at keplergo@mail.arc.nasa.gov



Like Waldseemüller's 1507 map, the Maximum A Priori (MAP) method for PDC and other improvements to light curve correction implemented in SOC Pipeline 8.0 and later versions will show us more of the light-curve world than we knew before, while leaving some hazards to navigation around the edges. Far-ranging light curve mariners are given the tools to help navigate in such perilous circumstances by the Guest Observer Office at <https://keplerscience.arc.nasa.gov/software.html>

1. Introduction

The *Kepler* Data Characteristics Handbook (KDCH) provides a description of all phenomena identified in the *Kepler* data throughout the mission, and an explanation for how these characteristics are handled by the final version of the *Kepler* Data Processing Pipeline (SOC 9.3). The KDCH complements the *Kepler* Data Release Notes (KDRNs), which document phenomena and processing unique to a data release. The original motivation for this separation into static, explanatory text and a more “journalistic” set of figures and tables in the KDRN was for the user to become familiar with the Data Characteristics Handbook, then peruse the short Notes for a new quarter, referring back to the Handbook when necessary. With the completion of the *Kepler* mission and the final Data Release 25, both the KDCH and the DRN encompass the entire *Kepler* mission, so the distinction between them is in the level of exposition, not the extent of the time interval discussed.

In addition to the Data Characteristics Handbook, the following documents are important for users of *Kepler* data. All but items 5 and 6 are at <http://archive.stsci.edu/Kepler/documents.html>

1. *Kepler* Data Processing Handbook (KSCI-19081) or KDPH (Jenkins *et al.*, 2017). The KDPH describes how pixels downlinked from the spacecraft are converted by the *Kepler* Data Processing Pipeline (henceforth just “the pipeline”) into the data products delivered to the MAST archive. The KDPH is under revision for the final *Kepler* pipeline release (Jenkins *et al.*, 2017). Refer to Part II of the KDPH for the definitive reference on pipeline processes such as black level estimation (Dynamic Black Correction), pixel level calibration (CAL), light curve extraction (Photometric Analysis, PA), optimal aperture calculation (PA-COA), and systematic error correction (Presearch Data Conditioning, PDC).
2. *Kepler* Archive Manual (KDMC-10008) or KAM (Thompson *et al.*, 2016). The KAM describes the format and content of the data products, and how to search for them.
3. *Kepler* Instrument Handbook (KSCI-19033) or KIH (Van Cleve and Caldwell, 2016), which provides information about the design, performance, and operational constraints of the instrument, and an overview of the types of pixel data that are available. Users will need to be familiar with the material in Sections 2 and 4.2-4.5 of the KIH to fully benefit from this document and the accompanying Notes.
4. The *Kepler* Data Release Notes (KDRNs) document the specific instances of the phenomena described in the remainder of this document for each release of data.
5. The *Kepler* Mission Special Issue of *Astrophysical Journal Letters* (Volume 713, Number 2, 2010 April 20) contains papers providing background on mission definition (Koch, *et al.* 2010), target selection (Batalha, *et al.* 2010), science operations (Haas *et al.* 2010), the *Kepler* point spread function (Bryson, *et al.* 2010), instrument performance (Caldwell *et al.* 2010), and the data processing pipeline (Jenkins *et al.* 2010a). Two papers discuss the characteristics of the Long Cadence (Jenkins *et al.* 2010b) and Short Cadence data (Gilliland *et al.* 2010), respectively. Numerous additional papers also provide early science results in both planet detection and asteroseismology, placing the use of *Kepler* data in context.
6. Papers describing important post-launch improvements to the pipeline, including descriptions of the calibration of the pixels in CAL (Quintana *et al.* 2010), aperture photometry in PA (Twicken *et al.* 2010), light curve conditioning and systematic error removal in PDC (Smith *et al.* 2012; Stumpe *et al.* 2012 and 2014), and optimal aperture calculation in PA-COA (Smith *et al.* 2016).

The KDCH will from time to time compare the characteristics of the ultimate (DR 25) to the those of the penultimate (DR 24) release to illustrate the benefits of pipeline improvements, but is not meant to be a running historical narrative of all pipeline processing or errors found and fixed. For users interested in the provenance and characteristics of earlier releases used for already-published scientific papers, earlier versions of this document and the relevant KDRNs are available at MAST.

The *Kepler* mission was reconstituted as the *K2* mission after the failure of the 2nd of *Kepler*'s 4 reaction wheels in May 2013 (Howell *et al.*, 2014; Van Cleve *et al.*, 2016). The characteristics of the *K2* data, while remaining superior to the best ground-based photometry, are sufficiently different from those of *Kepler* data to warrant a separate document. No further revisions of the KDCH are planned, as both data collection and processing have concluded for *Kepler*.

1.1 Dates, Cadence Numbers, and Units

A set of coadded and stored pixels obtained at a specific time is referred to as a *cadence*, and the total amount of time over which the data in a cadence is coadded is the *cadence period*. The two cadence periods in use are Long Cadence (LC) and Short Cadence (SC). Each cadence consists of a series of frames that each include a 6.02 s exposure time and a 0.52-s readout time. For Long Cadence and Full Frame Images (FFI), 270 frames are coadded, for a total of 1765.5 s = 0.4904 h. For Short Cadence, 9 frames are coadded, for a total of 58.85 s. 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), and also increments even when no cadences are being collected. RCI is calculated from the first valid Cadence of a Quarter (LC) or Month (SC). For example, the first LC of Q1 has an RCI = 1 and CIN = 1105, while the last LC of Q1 has RCI = 1639 and CIN = 2743. A list of the cadences associated with each quarter of data is given below. Times have been rounded to the nearest second.

Table 1. Long Cadence Data

Q	First Cadence MJD midTime	Last Cadence MJD midTime	First Cadence UT midTime	Last Cadence UT midTime	Number CINs	Start CIN	End CIN
0	54953.03815	54962.74411	2-May-2009 00:54:56	11-May-2009 17:51:31	476	568	1043
1	54964.01099	54997.48122	13-May-2009 00:15:50	15-Jun-2009 11:32:58	1639	1105	2743
2	55002.01748	55090.96492	20-Jun-2009 00:25:10	16-Sep-2009 23:09:29	4354	2965	7318
3	55092.72221	55181.99660	18-Sep-2009 17:19:59	16-Dec-2009 23:55:06	4370	7404	11773
4	55184.87774	55274.70384	19-Dec-2009 21:03:56	19-Mar-2010 16:53:32	4397	11914	16310
5	55275.99115	55370.66003	20-Mar-2010 23:47:16	23-Jun-2010 15:50:27	4634	16373	21006
6	55371.94733	55461.79386	24-Jun-2010 22:44:09	22-Sep-2010 19:03:10	4398	21069	25466
7	55462.67251	55552.04909	23-Sep-2010 16:08:24	22-Dec-2010 01:10:42	4375	25509	29883
8	55567.86468	55634.84602	6-Jan-2011 20:45:09	14-Mar-2011 20:18:16	3279	30657	33935
9	55641.01696	55738.42395	21-Mar-2011 00:24:25	26-Jun-2011 10:10:30	4768	34237	39004
10	55739.34343	55832.76587	27-Jun-2011 08:14:33	28-Sep-2011 18:22:51	4573	39049	43621
11	55833.70579	55930.82669	29-Sep-2011 16:56:20	4-Jan-2012 19:50:26	4754	43667	48420
12	55931.90966	56014.52273	5-Jan-2012 21:49:55	28-Mar-2012 12:32:44	4044	48473	52516
13	56015.23787	56105.55441	29-Mar-2012 05:42:32	27-Jun-2012 13:18:21	4421	52551	56971
14	56106.63736	56203.81957	28-Jun-2012 15:17:48	3-Oct-2012 19:40:11	4757	57024	61780

Q	First Cadence MJD midTime	Last Cadence MJD midTime	First Cadence UT midTime	Last Cadence UT midTime	Number CINs	Start CIN	End CIN
15	56205.98550	56303.63768	5-Oct-2012 23:39:07	11-Jan-2013 15:18:16	4780	61886	66665
16	56304.59804	56390.46005	12-Jan-2013 14:21:11	8-Apr-2013 11:02:28	4203	66712	70914
17	56391.72690	56423.50115	9-Apr-2013 17:26:45	11-May-2013 12:01:40	1556	70976	72531

Table 2. Short Cadence Data

Q.m	First Cadence MJD midTime	Last Cadence MJD midTime	First Cadence UT midTime	Last Cadence UT midTime	Number CINs	Start CIN	End CIN
0	54953.02828	54962.75399	2-May-2009 00:40:43	11-May-2009 18:05:45	14280	5500	19779
1	54964.00112	54997.49110	13-May-2009 00:01:37	15-Jun-2009 11:47:11	49170	21610	70779
2.1	55002.00760	55032.80035	20-Jun-2009 00:10:57	20-Jul-2009 19:12:30	45210	77410	122619
2.2	55032.82146	55062.79687	20-Jul-2009 19:42:54	19-Aug-2009 19:07:29	44010	122650	166659
2.3	55063.86010	55090.97480	20-Aug-2009 20:38:32	16-Sep-2009 23:23:42	39810	168220	208029
3.1	55092.71233	55123.05554	18-Sep-2009 17:05:46	19-Oct-2009 01:19:59	44550	210580	255129
3.2	55123.91444	55153.95114	19-Oct-2009 21:56:47	18-Nov-2009 22:49:39	44100	256390	300489
3.3	55156.01562	55182.00647	21-Nov-2009 00:22:30	17-Dec-2009 00:09:19	38160	303520	341679
4.1	55184.86786	55215.92625	19-Dec-2009 20:49:43	19-Jan-2010 22:13:48	45600	345880	391479
4.2	55216.80557	55245.73887	20-Jan-2010 19:20:02	18-Feb-2010 17:43:58	42480	392770	435249
4.3	55245.80085	55274.71371	18-Feb-2010 19:13:14	19-Mar-2010 17:07:45	42450	435340	477789
5.1	55275.98128	55307.50964	20-Mar-2010 23:33:02	21-Apr-2010 12:13:53	46290	479650	525939
5.2	55308.77720	55336.40275	22-Apr-2010 18:39:10	20-May-2010 09:39:58	40560	527800	568359
5.3	55337.09818	55370.66991	21-May-2010 02:21:23	23-Jun-2010 16:04:40	49290	569380	618669
6.1	55371.93745	55399.03173	24-Jun-2010 22:29:56	22-Jul-2010 00:45:41	39780	620530	660309
6.2	55399.87019	55430.78555	22-Jul-2010 20:53:04	22-Aug-2010 18:51:11	45390	661540	706929
6.3	55431.68529	55461.80374	23-Aug-2010 16:26:49	22-Sep-2010 19:17:23	44220	708250	752469
7.1	55462.66263	55492.78108	23-Sep-2010 15:54:11	23-Oct-2010 18:44:45	44220	753730	797949
7.2	55493.53780	55522.73674	24-Oct-2010 12:54:26	22-Nov-2010 17:40:55	42870	799060	841929
7.3	55523.61607	55552.05897	23-Nov-2010 14:47:09	22-Dec-2010 01:24:55	41760	843220	884979
8.1	55567.85481	55585.54963	6-Jan-2011 20:30:55	24-Jan-2011 13:11:28	25980	908170	934149
8.2	55585.61161	55614.70837	24-Jan-2011 14:40:43	22-Feb-2011 17:00:03	42720	934240	976959
8.2	55614.77035	55634.85590	22-Feb-2011 18:29:18	14-Mar-2011 20:32:30	29490	977050	1006539
9.1	55641.00708	55677.41909	21-Mar-2011 00:10:12	26-Apr-2011 10:03:29	53460	1015570	1069029
9.2	55678.11451	55706.61871	27-Apr-2011 02:44:54	25-May-2011 14:50:56	41850	1070050	1111899
9.3	55707.25283	55738.43383	26-May-2011 06:04:05	26-Jun-2011 10:24:43	45780	1112830	1158609
10.1	55739.33356	55769.45201	27-Jun-2011 08:00:19	27-Jul-2011 10:50:54	44220	1159930	1204149
10.2	55770.29047	55801.73710	28-Jul-2011 06:58:16	28-Aug-2011 17:41:26	46170	1205380	1251549
10.3	55802.57556	55832.77575	29-Aug-2011 13:48:49	28-Sep-2011 18:37:05	44340	1252780	1297119
11.1	55833.69591	55864.77474	29-Sep-2011 16:42:07	30-Oct-2011 18:35:38	45630	1298470	1344099
11.2	55865.53147	55895.73165	31-Oct-2011 12:45:19	30-Nov-2011 17:33:35	44340	1345210	1389549
11.3	55896.61098	55930.83656	1-Dec-2011 14:39:49	4-Jan-2012 20:04:39	50250	1390840	1441089
12.1	55931.89978	55958.40149	5-Jan-2012 21:35:41	1-Feb-2012 09:38:09	38910	1442650	1481559
12.2	55959.11735	55986.49770	2-Feb-2012 02:48:59	29-Feb-2012 11:56:41	40200	1482610	1522809
12.3	55987.39746	56014.53260	1-Mar-2012 09:32:20	28-Mar-2012 12:46:57	39840	1524130	1563969
13.1	56015.22799	56047.49198	29-Mar-2012 05:28:19	30-Apr-2012 11:48:27	47370	1564990	1612359

Q.m	First Cadence MJD midTime	Last Cadence MJD midTime	First Cadence UT midTime	Last Cadence UT midTime	Number CINs	Start CIN	End CIN
13.2	56048.18740	56077.42721	1-May-2012 04:29:51	30-May-2012 10:15:11	42930	1613380	1656309
13.3	56078.32697	56105.56428	31-May-2012 07:50:50	27-Jun-2012 13:32:34	39990	1657630	1697619
14.1	56106.62748	56123.50495	28-Jun-2012 15:03:34	15-Jul-2012 12:07:07	24780	1699180	1723959
14.2	56138.64693	56168.80625	30-Jul-2012 15:31:35	29-Aug-2012 19:21:00	44280	1746190	1790469
14.3	56169.66514	56203.82944	30-Aug-2012 15:57:48	3-Oct-2012 19:54:24	50160	1791730	1841889
15.1	56205.97562	56236.80925	5-Oct-2012 23:24:54	5-Nov-2012 19:25:19	45270	1845040	1890309
15.2	56237.77031	56267.88875	6-Nov-2012 18:29:15	6-Dec-2012 21:19:48	44220	1891720	1935939
15.3	56268.72721	56303.64756	7-Dec-2012 17:27:11	11-Jan-2013 15:32:29	51270	1937170	1988439
16.1	56304.58817	56309.81849	12-Jan-2013 14:06:58	17-Jan-2013 19:38:37	7680	1989820	1997499
16.2	56321.15981	56357.46965	29-Jan-2013 03:50:08	6-Mar-2013 11:16:18	53310	2014150	2067459
16.3	56358.61461	56390.46992	7-Mar-2013 14:45:03	8-Apr-2013 11:16:41	46770	2069140	2115909
17.1	56391.71703	56414.09114	9-Apr-2013 17:12:31	2-May-2013 02:11:15	32850	2117740	2150589
17.2	56419.30239	56423.51103	7-May-2013 07:15:26	11-May-2013 12:15:53	6180	2158240	2164419

Table 3. Full Frame Images

Q	Filename	UT Start	UT End
0	KPLR2009114174833	24-Apr-2009 17:19:08	24-Apr-2009 17:49:33
0	KPLR2009114204835	24-Apr-2009 20:19:10	24-Apr-2009 20:48:35
0	KPLR2009115002613	24-Apr-2009 23:56:48	25-Apr-2009 00:26:13
0	KPLR2009115053616	25-Apr-2009 05:06:51	25-Apr-2009 05:36:16
0	KPLR2009115080620	25-Apr-2009 07:36:55	25-Apr-2009 08:06:20
0	KPLR2009115131122	25-Apr-2009 12:41:56	25-Apr-2009 13:11:22
0	KPLR2009115173611	25-Apr-2009 17:06:46	25-Apr-2009 17:36:11
0	KPLR2009116035924	25-Apr-2009 03:29:58	26-Apr-2009 03:59:24
2	KPLR2009170043915	19-Jun-2009 04:09:49	19-Jun-2009 04:39:15
2	KPLR2009231194831	19-Aug-2009 19:19:05	19-Aug-2009 19:48:31
2	KPLR2009260000800	16-Sep-2009 23:38:34	17-Sep-2009 00:08:00
3	KPLR2009292020429	19-Oct-2009 01:35:04	19-Oct-2009 02:04:29
3	KPLR2009322233047	18-Nov-2009 23:01:21	18-Nov-2009 23:30:47
3	KPLR2009351005245	17-Dec-2009 00:23:19	17-Dec-2009 00:52:45
4	KPLR2010019225502	19-Jan-2010 22:25:37	19-Jan-2010 22:55:02
4	KPLR2010020005046	20-Jan-2010 00:21:21	20-Jan-2010 00:50:46
4	KPLR2010049182302	18-Feb-2010 17:53:37	18-Feb-2010 18:23:02
4	KPLR2010078174524	19-Mar-2010 17:15:58	19-Mar-2010 17:45:24
5	KPLR2010111125026	21-Apr-2010 12:21:01	21-Apr-2010 12:50:26
5	KPLR2010140101631	20-May-2010 09:47:06	20-Mar-2010 10:16:31
5	KPLR2010174164113	23-Jun-2010 16:11:48	23-Jun-2010 16:41:13
6	KPLR2010203012215	22-Jul-2010 00:52:49	22-Jul-2010 01:22:15
6	KPLR2010234192745	22-Aug-2010 18:58:19	22-Aug-2010 19:27:45
6	KPLR2010265195356	22-Sep-2010 19:24:31	22-Sep-2010 19:53:56
7	KPLR2010296192119	23-Oct-2010 18:51:53	23-Oct-2010 19:21:19
7	KPLR2010326181728	22-Nov-2010 17:48:03	22-Nov-2010 18:17:28
7	KPLR2010356020128	22-Dec-2010 01:32:03	22-Dec-2010 02:01:28
8	KPLR2011024134926	24-Jan-2011 13:20:01	24-Jan-2010 13:49:26

Q	Filename	UT Start	UT End
8	KPLR2011053174401	22-Feb-2011 17:14:35	22-Feb-2011 17:44:01
9	KPLR2011116104002	26-Apr-2011 10:10:37	26-Apr-2011 10:40:02
9	KPLR2011145152723	25-May-2011 14:57:58	25-Mar-2011 15:27:23
9	KPLR2011177110110	26-Jun-2011 10:31:44	26-Jun-2011 11:01:10
10	KPLR2011208112727	27-Jul-2011 10:58:01	27-Jul-2011 11:27:27
10	KPLR2011240181752	28-Aug-2011 17:48:27	28-Aug-2011 18:17:52
10	KPLR2011271191331	28-Sep-2011 18:44:06	28-Sep-2011 19:13:31
11	KPLR2011303191211	30-Oct-2011 18:42:45	30-Oct-2011 19:12:11
11	KPLR2011334181008	30-Nov-2011 17:40:43	30-Nov-3011 18:10:08
11	KPLR2012004204112	04-Jan-2012 20:11:47	04-Jan-2012 20:41:12
12	KPLR2012032101442	01-Feb-2012 09:45:16	01-Feb-2012 10:14:42
12	KPLR2012060123308	29-Feb-2012 12:03:42	29-Feb-2012 12:33:08
12	KPLR2012088132324	28-Mar-2012 12:53:58	28-Mar-2012 13:23:24
13	KPLR2012121122500	30-Apr-2012 11:55:30	30-Apr-2012 12:25:00
13	KPLR2012151105138	30-May-2012 10:22:12	30-May-2012 10:51:38
13	KPLR2012179140901	27-Jun-2012 13:39:35	27-Jun-2012 14:09:01
14	KPLR2012211123923	29-Jul-2012 12:09:58	29-Jul-2012 12:39:23
14	KPLR2012242195726	29-Aug-2012 19:28:01	29-Aug-2012 19:57:26
14	KPLR2012277203051	30-Oct-2012 20:01:25	30-Oct-2012 20:30:51
15	KPLR2012310200152	05-Nov-2012 19:32:27	05-Nov-2012 20:01:53
15	KPLR2012341215621	06-Dec-2012 21:26:56	06-Dec-2012 21:56:22
15	KPLR2013011160902	11-Jan-2013 15:39:37	11-Jan-2013 16:09:03
16	KPLR2013038133130	07-Feb-2013 13:02:05	07-Feb-2013 13:31:30
16	KPLR2013065115251	06-Mar-2013 11:23:26	06-Mar-2013 11:52:51
16	KPLR2013098115308	08-Apr-2013 11:23:43	08-Apr-2013 11:53:08

Figures and tables in this document and the KDRN present results in CIN, RCI, or Modified Julian Date (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 BJD, as described in detail in §6.2.4, is done on a target-by-target basis in the files available from MAST. The MJD of a cadence refers to the time at the midpoint of the cadence.

As of July 2011, details of which cadences are affected by most of the phenomena described in this document are available in the light curve and target pixel FITS files themselves, available from MAST. However, some processing details remain in the Data Release Notes, so users are encouraged to check these documents as well.

1.2 Full Frame Images

The FFIs contain a non-linear World Coordinate System (WCS) using the simple imaging polynomial (SIP) conventions of Shupe *et al.* (2005). The WCS solution is calculated from the motion polynomials based on the quarterly long cadence processing. Prior to DR 24, the pipeline did not calculate a motion polynomial directly from the FFI, but instead used the solution from the nearest LC. The WCS solution is typically accurate to 0.1 pixels (0.4 arcseconds) when compared to the 2MASS point source catalogue (Skrutskie *et al.* 2006). For additional details, see the Archive Manual.

The first 8 “golden” FFIs (Q0) and the first FFI of Q2 were not associated with a valid target table

and were not reprocessed in DR 24 or 25, and hence will have lower accuracy than typical of the other FFIs. The first FFI of Q2 is furthermore degraded since this FFI was taken at the very start of the quarter and we entered a safe mode shortly thereafter. When we subsequently returned to normal operation the pointing was offset from the pointing when the FFI was taken. We have used a linear solution for the WCS for this FFI and only expect the WCS to be accurate to ~ 1 pixel for CCDs near the edges of the focal plane.

The FFI time stamps are barycenter-corrected times corresponding to the middle of each output channel (§6).

1.3 Document Overview

In §2, we describe the generation and contents of a data release. Our evaluation of the current precision of the data is outlined in §3. A description of the historical events that have affected the *Kepler* data is provided in §4; similarly a description of the ongoing phenomena affecting the data is provided in §5. §6 outlines the generation and precision of the time stamps associated with *Kepler* data. §7 describes the ensemble Cotrending Basis Vectors for use in cotrending flux time series. §8 details the obsolescence of the Data Release Notes Supplement. A list of references is included in §9, and the acronyms used throughout this and other *Kepler* documents are defined in §10. See the KDPH for a description of the final pipeline.

2. Release Description

A *data set* refers to the data type and observation interval during which the data were collected. The observation interval for Long Cadence data is usually a *quarter*, indicated by Q[n], though Q0 and Q1 are 10 days and one month duration respectively, instead of the typical three-month quarter, and Q17 was truncated by the end of the mission. Short Cadence targets can be changed every month, so SC observation intervals are indicated by Q[n].M[m], where m = 1,2 or 3, indicating the Month within that Quarter. Released Science Operations Center (SOC) software has both a release label in the form of a version number (e.g., 8.2), and a revision number which precisely identifies the revision of the code corresponding to that label. For example, the code used to produce Data Release 25 has the release label “SOC Pipeline 9.3”.

A given data set was reprocessed as the software improved, and hence is the subject of multiple releases, of which DR 25 is the last. The combination of data set and data processing description defines a *data product*, and a set of data products simultaneously delivered to MAST 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.” Each data release is accompanied by a set of Data Release Notes, which document the phenomena occurring in and to that set of data.

Data products are made available to MAST users as FITS files, described in the KAM. The data are available both as light curves and as target pixel files; also available are the monthly full frame images (FFIs), and the ancillary, background and collateral (ABC) files. See the KAM for a description of these files. The light curve files include both uncorrected (SAP_FLUX) and systematic error-corrected (PDCSAP_FLUX) flux time series for simple aperture photometry. The FITS files contain the header keyword DATA_REL, which unambiguously associates a data release with the relevant Data Release Notes and the header keyword QUARTER, which identifies when the data were acquired.

Target pixel files contain the raw and calibrated pixels collected with the *Kepler* spacecraft. Similar to the light curve files, the target pixel files are FITS binary tables, organized by target. For the target pixel files, the FITS binary table contains a time series of images for the raw counts, the calibrated pixels, the background flux, and the removed cosmic rays. The intent of the pixel level data is to provide users enough information to perform their own photometry independent of the SOC pipeline. For details on how these files are formatted, please see the KAM.

3. Evaluation of Performance

3.1 Overall

The Combined Differential Photometric Precision (Jenkins *et al.*, 2010c; Christiansen *et al.*, 2012) of a photometric time series is the effective white noise seen by a transit pulse of a given duration, typically the duration of a transit or other phenomena 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 h CDPP of 20 ppm for a star with a planet exhibiting 84 ppm transits lasting 6.5 h leads to a single transit S/N of 4.2σ .

The CDPP performance has been discussed recently by Christiansen *et al.* (2012) and Gilliland *et al.* (2011, 2015). Nonetheless, the following cautions continue to apply:

1. Stellar variability and many instrumental effects are not in general broadband, colored noise processes.
2. Since giant stars are intrinsically variable at the level of *Kepler*'s precision, they are excluded from calculations of CDPP performance. A simple, but not foolproof, way to do this is to include only stars with high surface gravity ($\log g > 4$). A discussion of the stellar and instrument noise properties can be found in Gilliland *et al.* (2011, 2015).
3. Given the image artifacts discussed in detail in the KIH and Kolodziejczak *et al.* (2010), it is not generally possible to extrapolate noise as $1/\sqrt{\text{time}}$ for those channels afflicted by artifacts, which are flagged by the pipeline in DR 25 as described in DRN 25 §A.1.1.

The Transiting Planet Search (TPS) software module formally calculates CDPP on 6 h timescales and 13 other transit durations as a function of cadence for each target. The temporal median of the 6 h CDPP (TMCDPP) for each target is then divided by $\sqrt{13/12}$ to scale the results from 6 h to the Missions' 6.5-hour benchmark time scale, which is half the central transit duration of an Earth-size planet transiting a solar type star in a one-year orbit. The distribution of TMCDPP with *Kepler* magnitude is shown in **Figure 1**, and separated clearly into two branches, mostly corresponding to giants with $\log g < 4$ and dwarfs with $\log g > 4$.

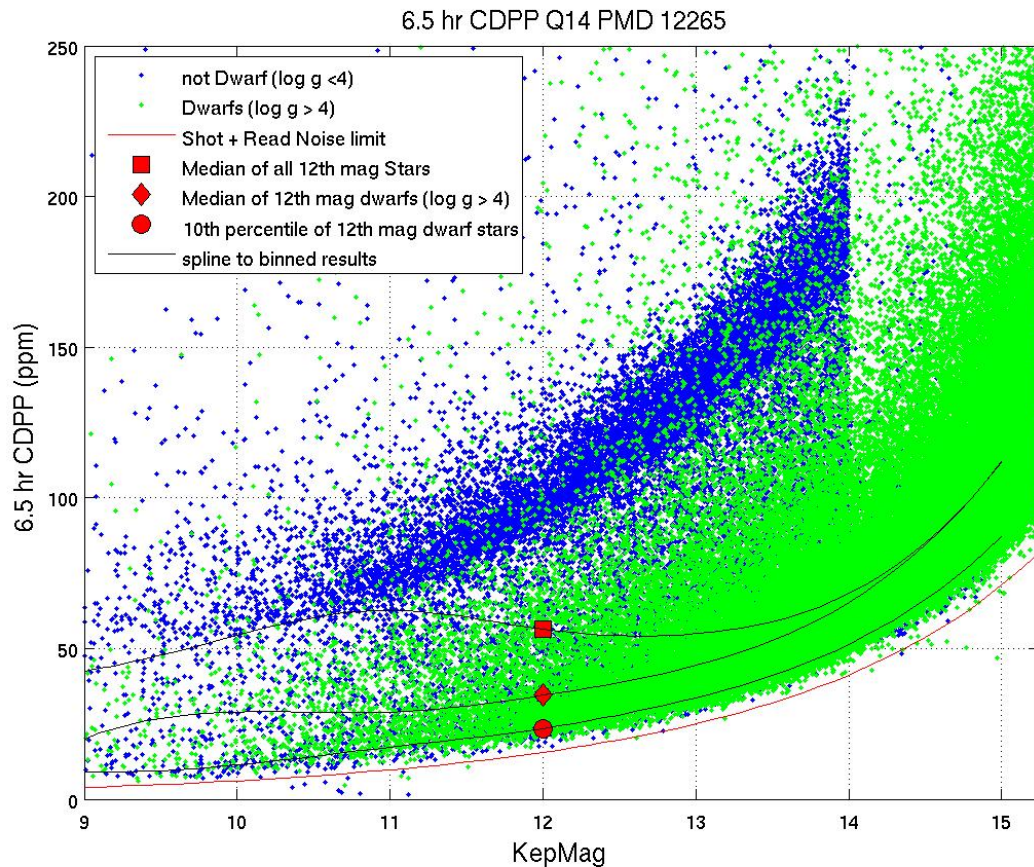


Figure 1: 6.5 h Temporal Median (TM) of the Q14 CDPP time series calculated by the SOC Pipeline 9.3 TPS module for stars between $K_p=9-15$. The 6 h TMCDDPs have been divided by $\sqrt{13/12} = 1.041$ to approximate 6.5 h TMCDDPs. Stars on the planetary target list with $\log g > 4$, which are likely to be dwarf stars, are shown as green + symbols; other stars are marked with blue + symbols. Black lines are spline fits by magnitude to the 10th percentile and median of dwarf stars, and to the median of all 12th mag stars.

Further information may be gleaned from examining the TMCDDP of subsets of the full target list, such as all targets with magnitude between 11.75 and 12.25 and $\log g > 4$, loosely referred to as “12th magnitude dwarfs”. Table 4 summarizes the median and percentile results for various target subsets in Q14. Note that the median CDPP for 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 as the stars get fainter. **Figure 2** shows the behavior of the TMCDDP for 12th magnitude dwarfs as a function of quarter and pipeline version.

A simple model of the noise floor can be calculated from the root-mean-square sum of shot noise and effective read noise - calculating this model over the benchmark 6.5-hour transit time gives the theoretical noise floor in the last column of Table 4.

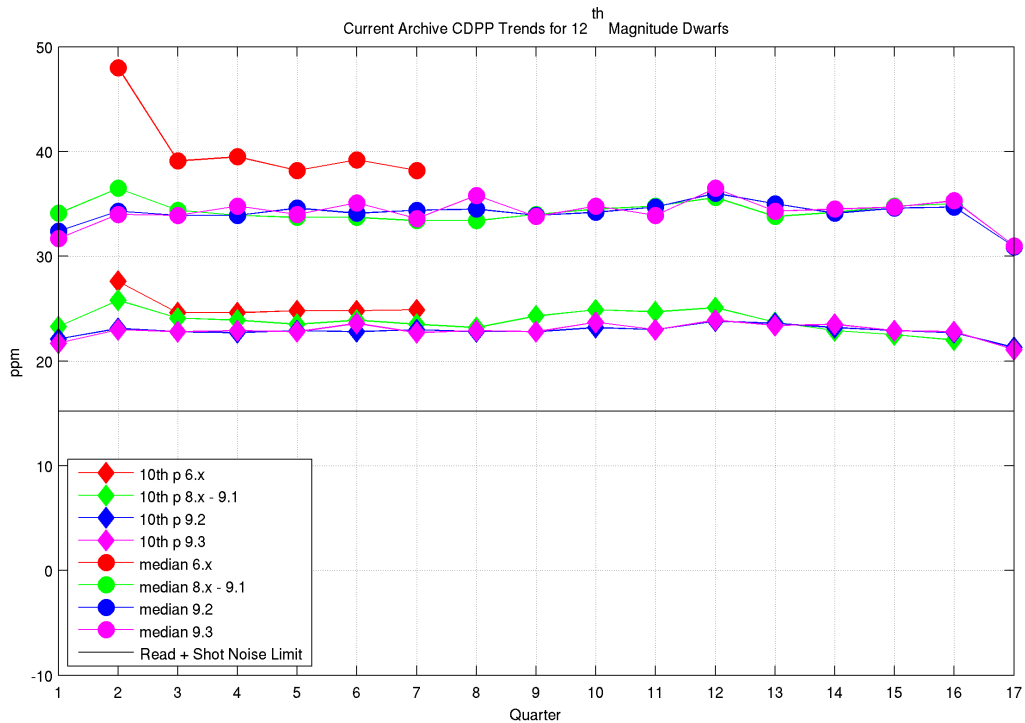


Figure 2: The 6.5 h TMCDDP for 12th magnitude dwarfs (log g > 4), as a function of quarter and processing version. We show both the 10th percentile value, i.e. the value below which 10% of the 12th magnitude dwarfs fall, and the median value for each quarter. The 6.x, 8.x and 9.x labels given in the legend refer to the version of the SOC pipeline used.

Table 4: Aggregate Statistics for the SOC 9.3 TMCDDP values for Q14. 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 CDDP, which does not include astrophysical noise.

Kp	ndwarfs	10th p	median	nstars	10th p	median	model
9	55	9.2	20.0	179	11.2	42.6	3.8
10	160	11.4	29.0	568	12.5	54.5	6.0
11	610	17.3	28.9	1712	19.1	62.7	9.5
12	2169	23.5	34.5	4385	25.0	56.4	15.2
13	6767	33.7	44.4	10476	35.1	54.9	24.4
14	13679	51.4	65.3	16582	52.2	68.2	40.1
15	27336	87.2	111.9	27466	87.0	111.9	68.8

4. Historical Events

This section describes the various avenues by which some *Kepler* data have been lost or degraded. More recent quarters of data suffer less from most of these effects since they have been mitigated where possible. For each quarter, a table (or, more recently, a figure) is produced for summarizing the data anomalies that occurred in that quarter and included in the relevant Data Release Notes. The types of anomalies that are included are described below.

4.1 Kepler Mission Timeline To Date

A graphical summary of some of the events detailed below is included as **Figure 3**.

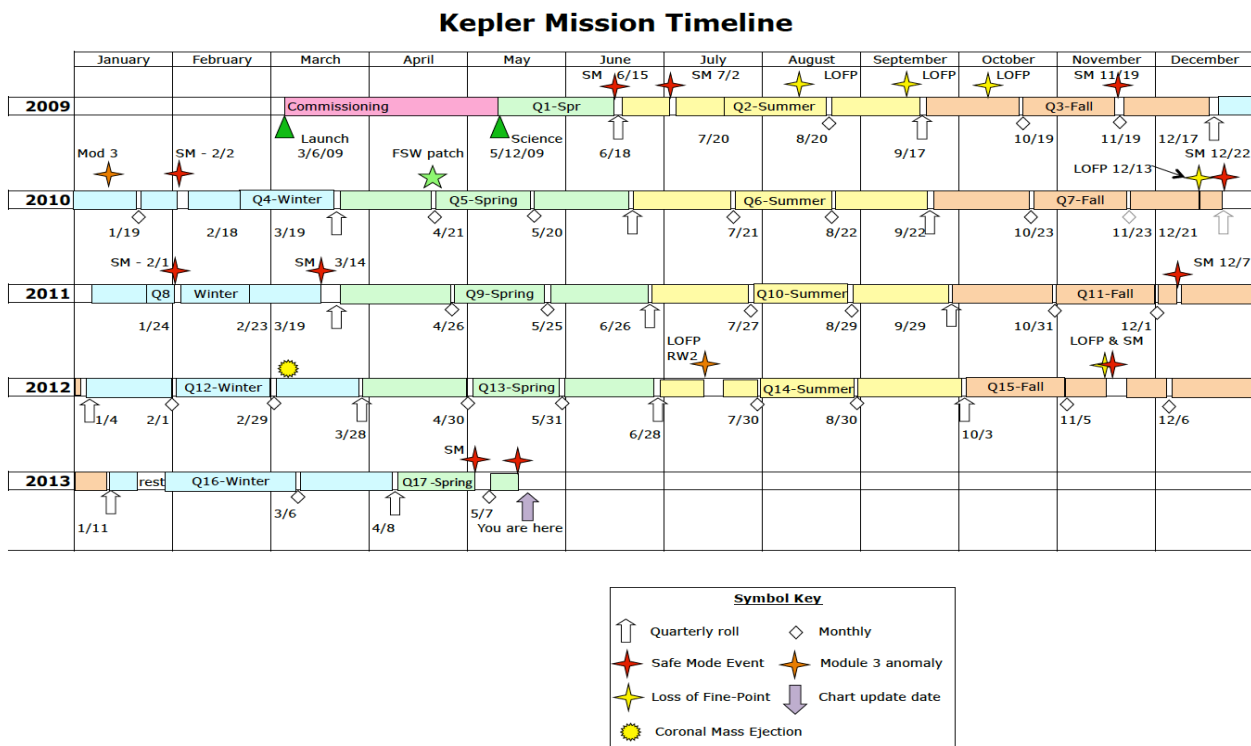


Figure 3: *Kepler* mission timeline showing historical events through the end of the mission.

4.2 Safe Mode

From time to time, the *Kepler* Spacecraft went into Safe Mode. While each individual event is unexpected, it is not unusual for a spacecraft to enter Safe Mode for a variety of reasons. Cadences lost during Safe Modes are identified in the FITS headers (KAM Table 2-3) and the Anomaly Summary figure in the DRN.

As indicated in **Figure 4**, there was a Safe Mode event during Q2 between MJD 55014.03 and 55016.19, corresponding to Long Cadences 3553-3659. The Local Detector Electronics (LDE) were turned off, but data previously collected remained in the solid state recorder for retrieval after Safe Mode recovery. Data collected after resumption of science observations show a photometric trend strongly correlated with a thermally- induced focus change related to the warming up of the back of the spacecraft, which is for the most part mitigated within the PDC module of the pipeline for the PDCSAP_FLUX time series. The *Kepler* Flight Software

has subsequently 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 is much less dramatic for this kind of Safe Mode.

4.3 Loss of Fine Point

From time to time, the *Kepler* spacecraft lost fine pointing control, degrading the photometric precision of those cadences. While the data obtained during LOFPs (Losses Of Fine Point) are treated as lost data by the pipeline, users with sources for which less precise photometry is scientifically interesting may wish to look at the pixel data corresponding to those cadences. Cadences affected by LOFPs are identified in the relevant quarter Anomaly Summary table or figure in the DRNs, and are also indicated in the FITS files at MAST with the quality flag column.

4.4 Attitude Tweaks Every Few Days

Reference pixels are used by the SOC Photometer Data Quality (PDQ) software module to measure spacecraft attitude during data collection. The Photometer Attitude Determination (PAD) software performs a similar calculation to reconstruct the attitude using the Long Cadence science data. **Figure 4** shows the PAD pointing solution for Quarter 2. 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. The RSS sum of RA, Dec, and roll errors is an upper bound on the rigid body component of MAR, and is also shown in the figure. Non-rigid body effects include change in telescope focus and resulting change in plate scale and location of stellar images on the Focal Plane Array (FPA).

Since continued attitude drift would invalidate target aperture definitions and lead to large photometric errors, small attitude adjustments (“tweaks”) were performed to ensure that the MAR remained less than 100 millipixels. In Q2, tweaks were necessary (shown in **Figure 4**), which introduced discontinuities into the data. Parameter changes in the FGS centroiding algorithm, which were implemented at the start of Q3, greatly diminished the boresight drift and eliminated the need for such attitude tweaks in subsequent quarters. The pipeline has largely corrected the discontinuities caused by these attitude tweaks (see KDPH for details).

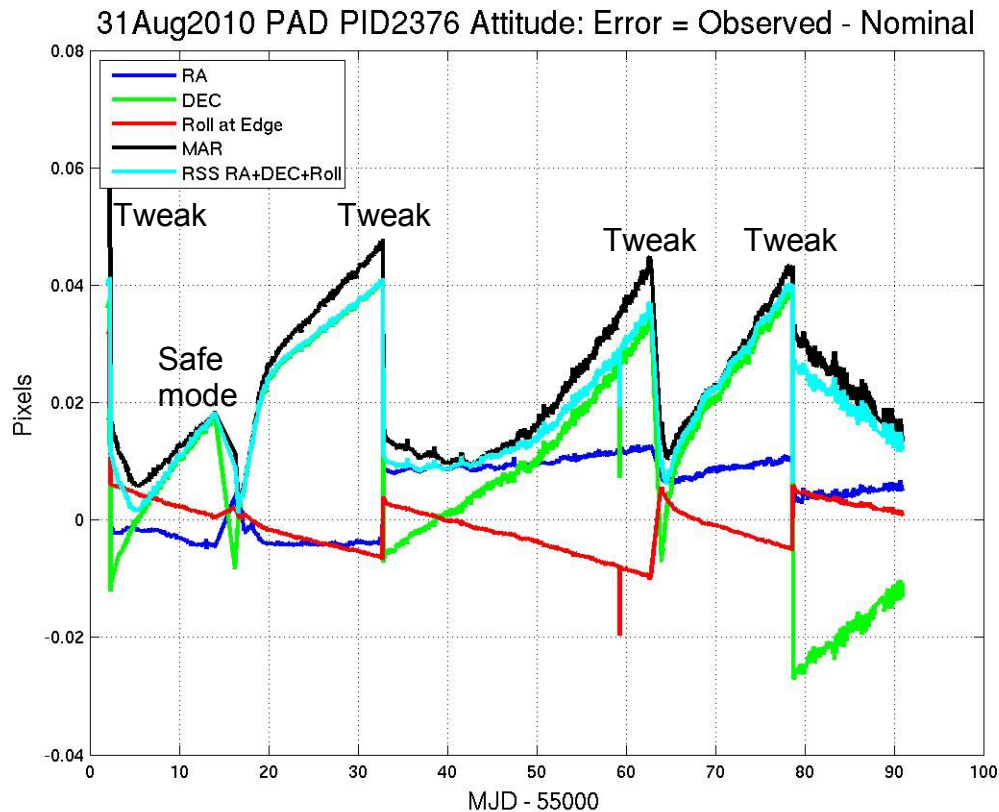


Figure 4: Attitude Offset in Quarter 2, calculated by PAD using Long Cadence data. The four large deviations near days 2, 32, 62 and 79 days (MJD-55000) are the result of attitude tweaks. The safe mode occurred around day 15 (MJD-55000). The roll is calculated for the edge of the focal plane.

4.5 Variable FGS Guide Stars

The first-moment centroiding algorithm used by the FGS (Fine Guidance System) did not originally subtract all of the instrumental bias from the FGS pixels. Thus, the calculated centroid of an FGS star depended on the star's flux when the star was not located at the center of the centroiding 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 ADCS (Attitude Determination and Control System), which attempts to keep the calculated attitude of the spacecraft constant, then moved the spacecraft 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. Users wishing to work with uncorrected flux time series or with the calibrated pixels need to be aware of possible FGS variability-induced signatures and not mistake them for features of their target flux time series.

The most egregious variable guide stars were replaced with quieter stars at the start of Quarter 2 (20 June 2009). One intrinsically variable star and one eclipsing binary continued to be used in the FGS, as shown in **Figure 5**. The effect of the intrinsically variable star can be seen as oscillations in the PAD attitude solution with the same period (2.9 d, see for example RA between MJD 55033 and 55055 in **Figure 4**).

The centroiding algorithm was updated to remove all of the instrumental electronic bias for Quarter

3 (19 September 2009) onwards, greatly diminishing the effect of stellar variability on calculated centroids. The sky background is not removed, but is expected to be negligible. Hence, FGS guide star variability is not a factor from Q3 onwards.

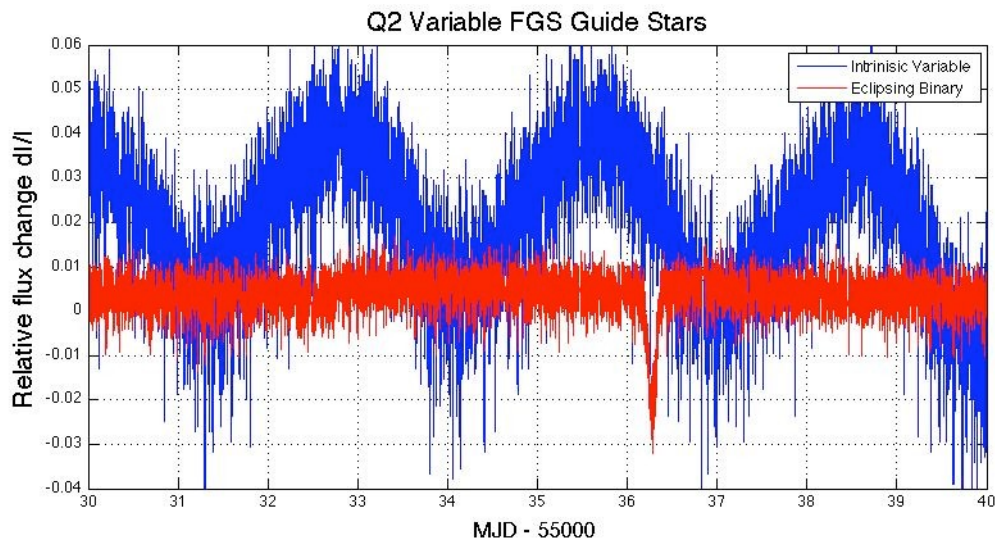


Figure 5: Quarter 2 flux time series of two variable FGS guide stars. One of the stars is an eclipsing binary with a period of 18.25 days, the other is an intrinsic variable with a period of 2.9 days. Only 10 days of data are shown here for illustration.

4.6 Module 3 Failure

All 4 outputs of Module 3 failed at 17:52 UTC Jan 9, 2010 during LC CIN 12935. Reference pixels showed loss of stars and black levels decreased by 75 to 100 DN per frame. FFIs show no evidence of photons or electrically injected signals. The start of line ringing and FGS crosstalk (see KIH, Sections 6.5 and 6.2 respectively) are still present after the anomaly, as shown in **Figure 6**.

The loss of the module led to consistent temperature drops within the LDE, telescope structure, Schmidt corrector, primary mirror, FPA modules, and acquisition/driver boards – which in turn affected photometry and centroids across the full focal plane.

After a review of probable causes, it was concluded that the probability of additional failures was remote, a conclusion supported by the continued operation of all the other modules for more than three years until module 7 failed after the end of the *Kepler* mission, but before the start of *K2*.

The impact of the loss of Module 3 on science observations is that 20% of the FOV suffered a one-quarter data outage every year as *Kepler* performed its quarterly rolls.

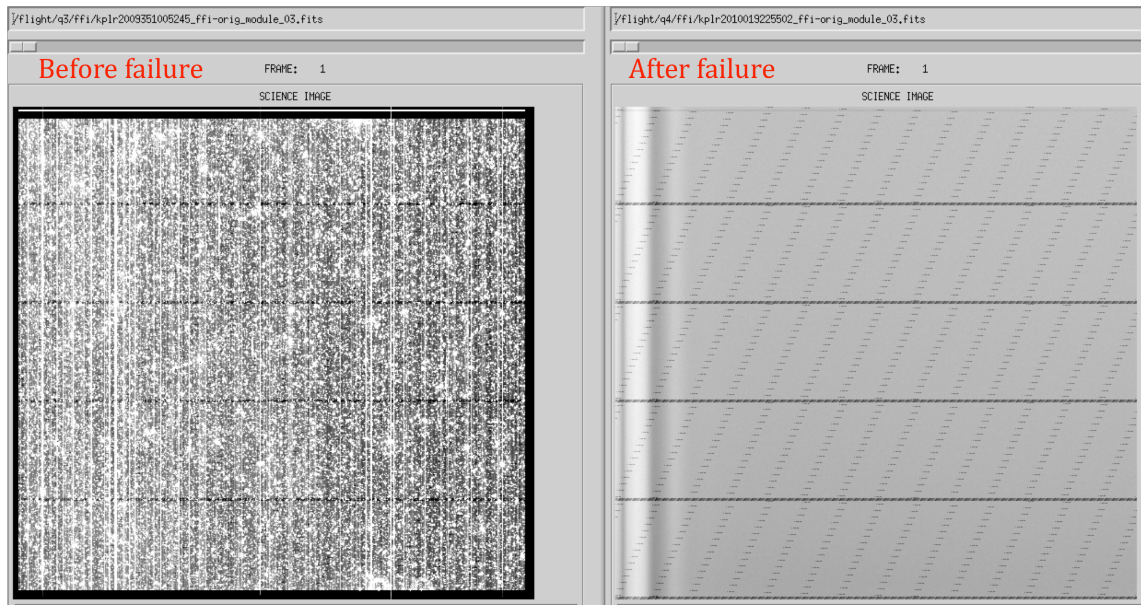


Figure 6: Permanent loss of Module 3. The left image is a normal uncalibrated FFI; the right is the image collected after the failure. Black-white image scaling is $1551 \text{ DN/cadence} = 5.7 \text{ DN/frame}$ for both images.

4.7 Reaction Wheel Failures

Kepler lost reaction wheel 2 due to excess friction on MJD 56122 (14 July 2012) and returned to science data collection using three reaction wheels on MJD 56128 (20 July 2012). The intervening six days of data have been excluded, as is normally the case for coarse point data. This change in attitude control occurred midway through Q14M1. However, *Kepler's* performance on three wheels appeared nominal, so the three- and four- wheel data were processed and exported as usual for Q14-Q17 (i.e., by month for short- cadence and by quarter for long-cadence). The wheels were rested in Q16 as per DRN 25 §16.3. Nonetheless, reaction wheel 4 failed on MJD 56423 (11 May 2013), ending the *Kepler* mission.

4.8 CH Cyg Corrupts the Smear Correction

The bright variable star CH Cyg (KIC 11913210) is unusual in that it is both variable and highly saturated when near maximum brightness. The star falls close enough to the masked smear region that saturated charge from the star spills into the masked smear. The largest values in these masked smear columns exceed the pixel data storage range on the spacecraft, causing the smear data values to wrap through zero. For some columns this wrapping can occur several times. Due to the highly-varying and incorrect smear values used, the calibrated pixel values and time series for targets on these columns are not valid. This problem has been mitigated by exclusion of the corrupted columns from the smear calculation. A more general discussion of corrupted smear and identifying afflicted columns using Archive data products is given in KAM §2.3.5.6.

The specific instances we have identified of CH Cyg impacting the smear collection include:

- For the duration of Q12, targets whose photometric apertures include pixels that lie in columns 632-638 on mod.out 2.1. See KIH Figure 24 for focal plane layout and mod.out labels.
- For the first few days of Q14, targets whose photometric apertures include pixels that lie in columns 630-634 on mod.out 24.1

5. Ongoing Phenomena

This section discusses systematic errors arising in nominal on-orbit operations, most of which are removed from the flux time series by PDC in the corrected flux time series product (PDCSAP_FLUX in the FITS files at MAST). As described in §7, PDC generates cotrending basis vectors (CBVs, see KAM §2.3.1.2 and §2.3.4) for investigators to use in their assessment and analysis of the flux time series if they choose to start from the “uncorrected” (SAP_FLUX) flux time series.

Most of the events described in this section are reported by the spacecraft or detected in the pipeline, then either corrected or marked as gaps (NaNs in the light curve and target pixel files).

5.1 Image Motion

The attitude determination and control system (ADCS) minimized the motion of the stellar images on the focal plane by monitoring the centroids of 40 stars on the fine guidance sensors (FGS) located in the corners of the focal plane. The ADCS compensated for the first order effects of velocity aberration, namely the apparently elliptical motion of the field of view about its inertial position, and the apparent rotation of the field of view due to differential velocity aberration (DVA). The ADCS could not compensate for higher order terms in (differential) velocity aberration across the FOV, which includes an apparent change in plate scale. The small change in location on the Kepler CCDs of each target in the field over a quarter is a combination of residual pointing errors (see for example §4.5), changes in focus, and these higher order DVA effects. Motion polynomials are calculated in the pipeline on a channel-by-channel and cadence-by-cadence basis to account for this local image motion. **Figure 7** shows a sample calculation of the motion of the center of mod.out 2.1 for Quarter 2. There is no requirement for smoothness in time of motion polynomials for cotrending and other purposes, so there is no fitting or smoothing across time. The image motion calculated from the motion polynomials for each target is available in the FITS files at MAST via the POS_CORR1 (column) and POS_CORR2 (row) arrays.

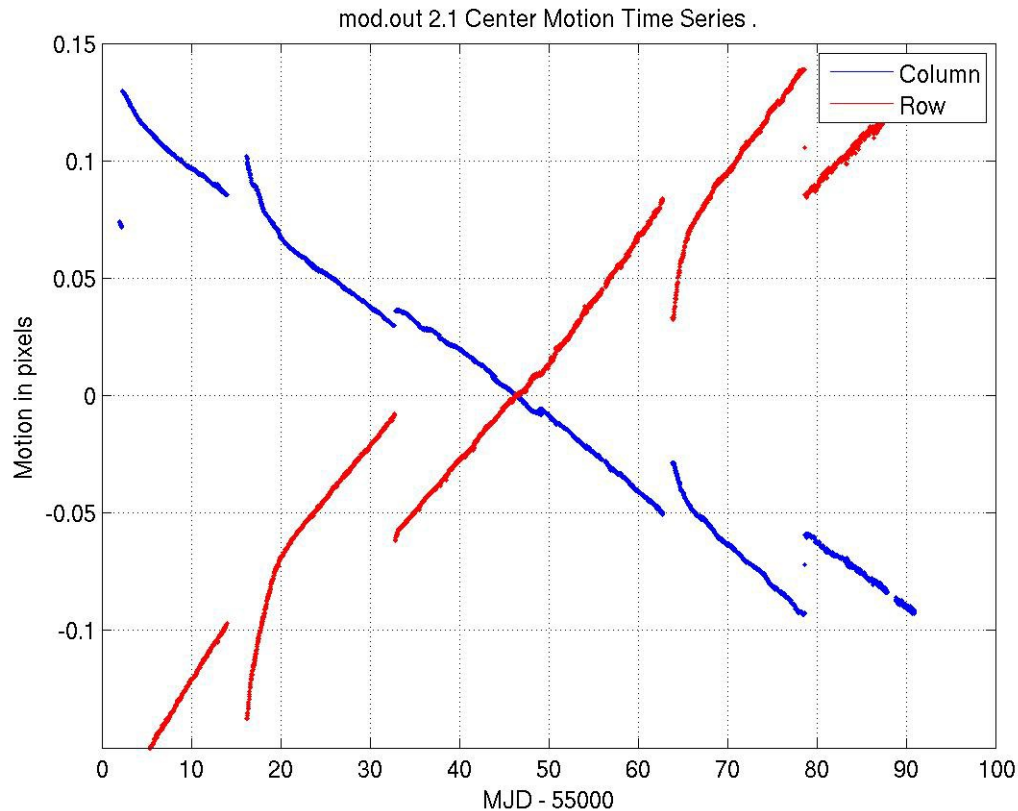


Figure 7: Mod.out 2.1 Center Motion Time Series for Q2, calculated from motion polynomials. The median row and column values have been subtracted. Since this mod.out is at the edge of the field, it shows large differential velocity aberration (DVA) with respect to the center of the field, as well as a higher sensitivity to focus jitter and drift. The four large discontinuities near days 2, 32, 62 and 79 (MJD-55000) are the result of attitude tweaks. A safe mode occurred around day 15 (MJD-55000).

5.2 Focus Changes

Examination of Q1 data (**Figure 8**) revealed that many of the science targets exhibit non-sinusoidal variations in their pixel time series with a period between 3 and 6 h. The behavior was less frequent at the beginning of Q1 and became progressively worse with time. Initially, this phenomenon was associated with reaction wheel momentum desaturation activities (§5.3), but became nearly continuous about 15 days into the observations. The problem persisted partway through Q3.

This focus change was observed in platescale metrics local to each channel defined by the motion of target star centroids relative to one another over time. This indicated a change in focus on timescales of 3 to 6 h that was initiated by the desaturation activities. Temperature sensors on the housings of reaction wheels 3 and 4 with the mnemonics TH1RW3T and TH1RW4T had the same time signature as image motion. These temperatures are shown in **Figure 9** and described in detail in the caption. The physical mechanism by which they coupled to focus was not well understood. Near the end of Q3, at MJD 55170 (5 December 2009), new Flight Software parameters were uploaded to substantially reduce the response threshold of the reaction wheel housing temperature controller, which in turn reduced the period of the heater cycle to below 20 minutes and also significantly reduced the amplitude of the temperature variations. Thermal cycling with a period shorter than the 3 day momentum management cycle is undetectable in TH1RW3T and TH1RW4T telemetry for Q8-Q13. The average temperatures of wheels 3 and 4 increased after the failure of

wheel 2 in Q14 (§4.7), reducing the need for heating at the end of a quarter. In Q4 and later, the dominant thermally-induced plate scale and focus changes over a quarter are the annual cycle shown in **Figure 10** and the 3-day signature of the momentum management cycle.

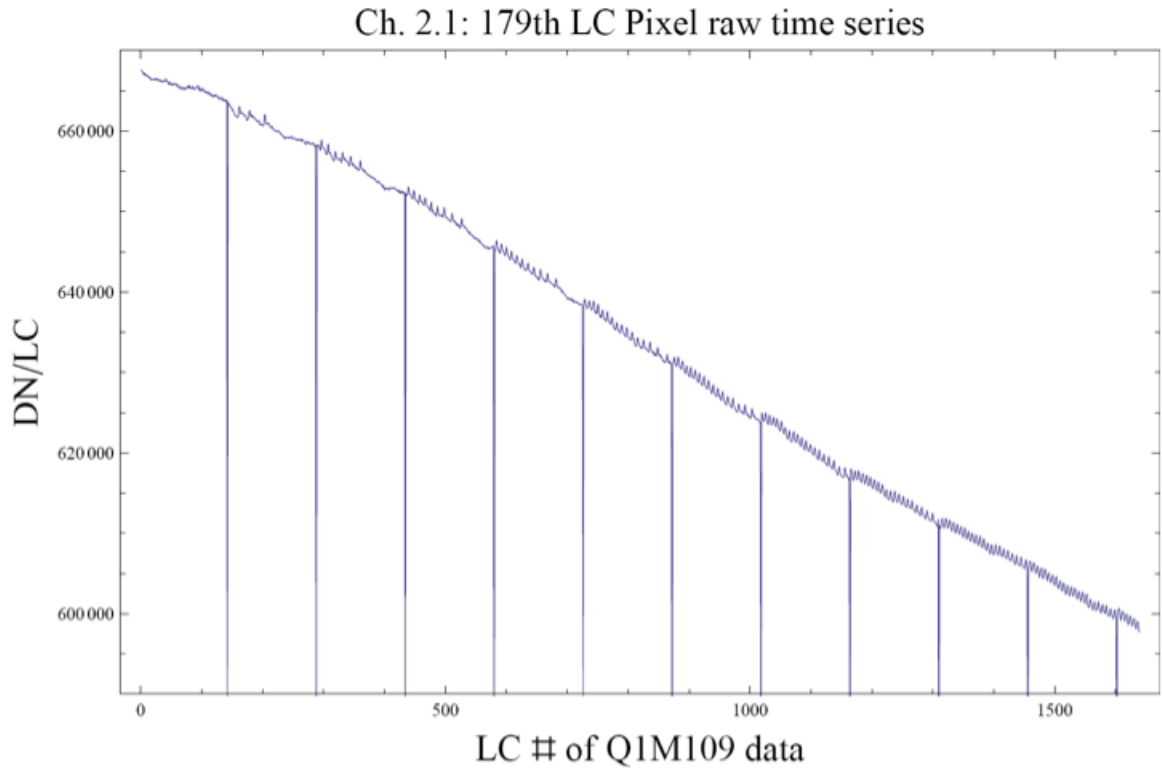


Figure 8: A good example of the 3- to 6-hr focus oscillation in a single raw pixel time series from Quarter 1. Similar signatures are seen in flux and plate scale. The large negative-going spikes are caused by reaction wheel desaturations (§5.3), which have not been removed from the flux time series in this plot. The abscissa is the Q1 relative cadence index, and the ordinate is Data Numbers (DN) per Long Cadence (LC).

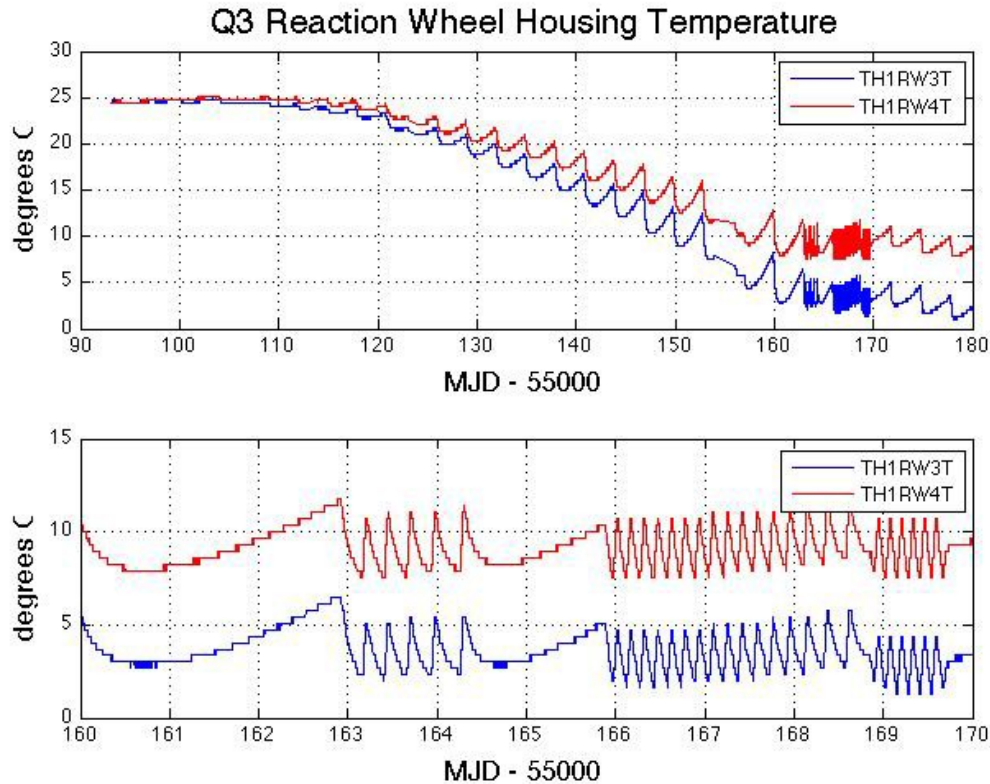


Figure 9: Reaction wheel housing temperatures during Q3. The upper panel shows that the temperature variation over most of the quarter is dominated by a slow seasonal drift and the 3-day period of wheel desaturations. The bottom panel shows that near the end of the quarter the reaction wheels have cooled sufficiently to engage the wheel housing heater, which then cycles on and off with a roughly 3-6 h period. Reducing the deadband on the temperature controller significantly mitigated the 3- to 6-hr variation after MJD 55170. The telemetry data in this Figure are smoothed with a 5-point median filter.

In the first few months of the mission, the data may show a secular variation of the focus driven by the outgassing of telescope components, in addition to the annual and momentum dump cycles driven by temperature changes in Flight System components discussed above. **Figure 10** indicates that the annual cycle dominates, with a good correlation between the focus, as measured by the Pixel Response Function (PRF) width, and several spacecraft temperature sensors.

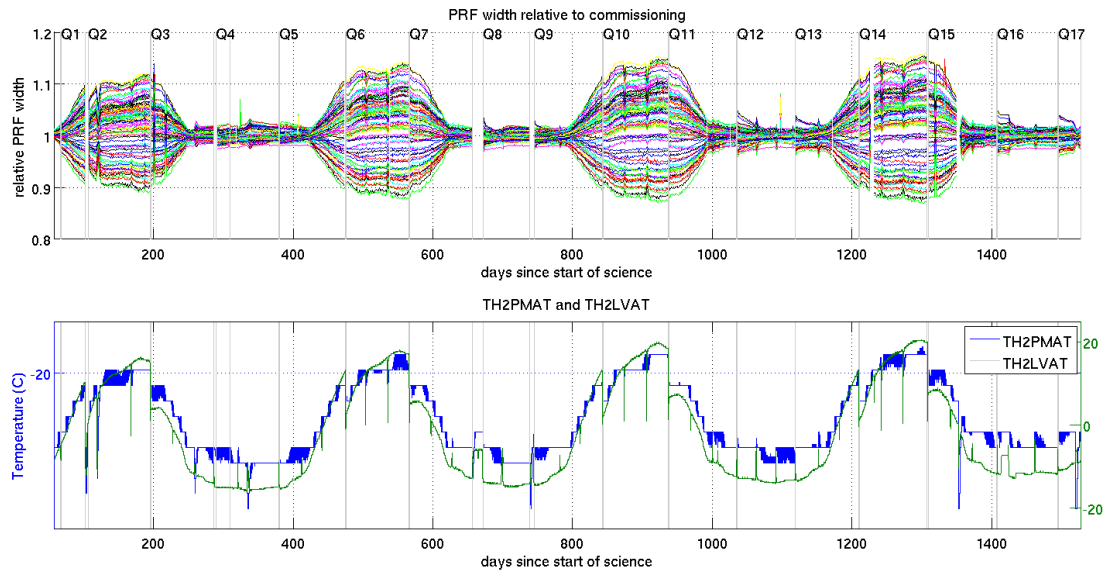


Figure 10: Correlation of variation in PRF width (top panel) with spacecraft temperature sensors TH2PMAT and TH2LVAT in the lower panel over the duration of the mission, demonstrating the seasonal nature of focus and PRF changes.

For users of the corrected (PDCSAP_FLUX) flux time series, the focus changes are well represented by the cotrending basis vectors (KAM §2.3.4).

5.3 Momentum Desaturation

Solar radiation torque causes angular momentum to build up in the reaction wheels, which were then desaturated by thruster firings to keep wheel speeds within operating limits. Desats occur every ~ 3 days (146 Long Cadence intervals). The spacecraft was not designed to maintain Fine Point control during these events, and entered Coarse Point mode. The subsequent image motion was sufficient to spoil the photometric precision of data collected during desats, and a few minutes after desats during which the spacecraft restored Fine Point control. One LC and several SCs are affected for each desaturation.

The desat cadences have NaNs in the delivered light curve files. Prior to Data Release 14, a table of the desat cadences was provided in the Data Release Notes each quarter, so that users of time series would know which NaNs were due to desats. Since the release of the version 2.0 FITS files at MAST, desat cadences are indicated in the quality flag column, and users are now directed to those files to locate these cadences; see KAM Table 2.3.

5.4 Reaction Wheel Zero Crossings

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. All SC targets show some impact, although the severity varies from target to target. In some cases, we observe negative spikes of $\sim 10^{-3}$ to 10^{-2} in SC relative flux time series (see, for example, **Figure 11**), and these cadences should be excluded from further analysis. These cadences are marked with bit 5 of the QUALITY flag in the FITS header (KAM Table 2.3).

The impact on Long Cadence data is much less severe in both amplitude and prevalence. Since the

failure of reaction wheel 2 in Q14M1 (see §4.7), the momentum management strategy has minimized the time each reaction wheel spends at low speeds, significantly decreasing the number of cadences affected by zero crossings from Q14M2 onwards. In **Figure 11**, 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 CR event or local transient. The zero crossings occur at different times than the momentum dumping wheel desaturations (§5.3).

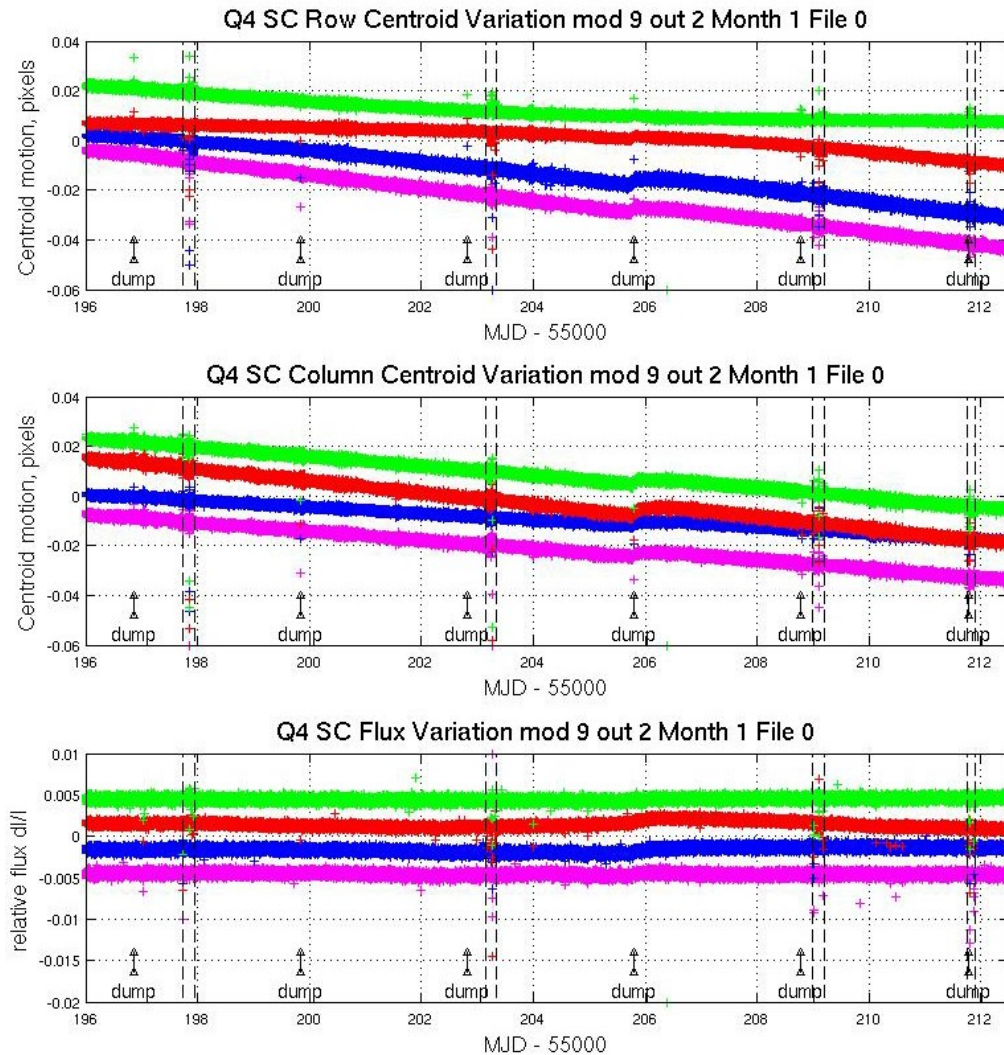


Figure 11: Example from Q4 of the effect of reaction wheel speed zero crossings on SC flux and centroids. The plots show row and column centroid motion, and the relative flux change, in the neighborhood of zero crossings. The data on several stars are overplotted in different colors in each panel of 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. The kink in the data at MJD = 55205.72 is the failure of module 3.

5.5 *Downlink Earth Point*

Science data were downlinked once a month, and the spacecraft changed its attitude to point its fixed High Gain Antenna (HGA) at the Earth. Science data collection stopped during downlink, and the change in attitude induced a thermal transient in the Photometer. Data collected after Earth Point are corrected by PDC in the same way as data after a Safe Mode. Cadences affected by Earth Point are identified in the Anomaly Summary table or figure in the relevant Data Release Notes. These cadences are also indicated in the FITS files at MAST in the quality flag column with bit 4; see KAM Table 2.3.

5.6 *Manually Excluded Cadences*

Occasionally, a cadence was manually excluded before pipeline processing, usually near a gap or discontinuity in the data that would make it difficult to identify and exclude automatically. Users are encouraged to view these cadences with skepticism. These cadences are also identified in the Anomaly Summary table or figure, and in the quality flag column of the FITS files at MAST; see KAM Table 2.3.

5.7 *Incomplete Apertures Produce Discontinuities at Quarter Boundaries*

Since a target moved to a different CCD with each quarterly roll, it sampled a different pixel response function (PRF) and was assigned a different optimal aperture. This changed both the crowding metric and the flux fraction, so some mismatch in flux across quarter boundaries is expected. However, some targets show unusually large flux changes or flux slope discontinuities across quarters. In these cases, quarterly changes in relative feature depths have also been observed. In each case, the problem has been with sources that saturate three or more pixels (*Kepler* magnitude 11 or brighter) and have optimal apertures (Bryson *et al.*, 2010) that omit pixels with bleeding charge in one or more quarters. This problem can be substantially mitigated by including the missing pixels and redoing the photometry using the target pixel files. However, if charge falls outside the full target aperture (which includes a halo of pixels around the estimated optimal subset), then information is irretrievably lost.

The most important reasons for an incorrect optimal aperture are:

- Variability of sources, when that variability exceeds a few percent, since the optimal aperture was designed for a fixed *Kepler* magnitude.
- Inability of the focal plane nonlinearity model to predict in detail the length and position of the charge bleed pixels in a column containing a saturated source.

The saturation model used for Q0-Q9 data collection accommodated 25%/75% to 75%/25% asymmetries in the length of the up and down bleed by collecting extra pixels along the saturating columns, but larger asymmetries resulted in incomplete capture of the bleeding charge.

As the mission progressed, visual inspection revealed those stars with poorly captured saturation. The *Kepler* magnitudes of these stars have been adjusted so that they are assigned larger apertures for subsequent quarters. A new empirical saturation map (KDPH chapter on Target Aperture Definition, TAD) was also implemented for target definition starting in Q10, so only data for targets with uncaptured bleed should continue to exhibit this problem.

For stars which are neither variable above the few percent level, nor saturating, a data-driven approach to aperture definition in the PA-COA pipeline module (KDPH and Smith *et al.* 2016) has supplanted the model-driven approach of the early mission, largely obviating the need for users to re-do the photometry (PA) for most targets. Small flux and slope discontinuities are removed at quarter boundaries in the planet-finding part of the pipeline (see KDPH) when joining multiple quarters of data.

5.8 Argabrightening

Argabrightening, named after its discoverer, V. Argabright of BATC, is a diffuse illumination of the focal plane, lasting on the order of a few minutes, possibly due to impact-generated debris (Witteborn *et al.*, 2011). It is known to be illumination 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. Spatial variation within a mod.out is significant for some events (**Figure 12**). While low-spatial-frequency changes in background are removed by subtraction of Pipeline-generated background polynomials for Long Cadence data, users are cautioned about Argabrightening cadences because of the possible fine spatial structure, possible errors in the nonlinearity model, and the absence of direct background measurements for Short Cadence, for which interpolation of Long Cadence background values is required.

The method of detecting Argabrightenings is:

1. Calculate the median, for each cadence and mod.out, of the calibrated background (LC) or out-of-optimal-aperture (SC) pixels,
2. Detrend the data by fitting a parabola to the resulting time series and subtract the fit.
3. High-pass filter the detrended data by median filtering with a 25 cadence wide filter, and subtract that median-filtered curve from the detrended data to form the residual background flux time series.
4. Calculate the Median Absolute Deviation (MAD) of the residual. The Argabrightening statistic S_{Arg} is then the ratio of the residual to the MAD.
5. Find values of S_{Arg} which exceed T_{MAD} , the single-channel threshold, and subsequently treat those cadences as gaps for all pixels in that channel. These detections are indicated by the single-channel Argabrightening flag (bit 13) in the data QUALITY column in the archive files (KAM Table 2-3).
6. A multi-channel event is recorded on a given cadence if at least 10 single-channel events are recorded. In these cases the Argabrightening event is deemed to have interfered with the Fine Guidance Sensors and caused a brief pointing excursion. These pointing excursions have the potential to degrade photometric quality for all targets, not just those on the channel directly affected by the Argabrightening. Multi-channel events are indicted by bit 7 in the data QUALITY column in the archive files (KAM Table 2-3). If this flag is set for a given long cadence, it is also set for all corresponding short cadences.

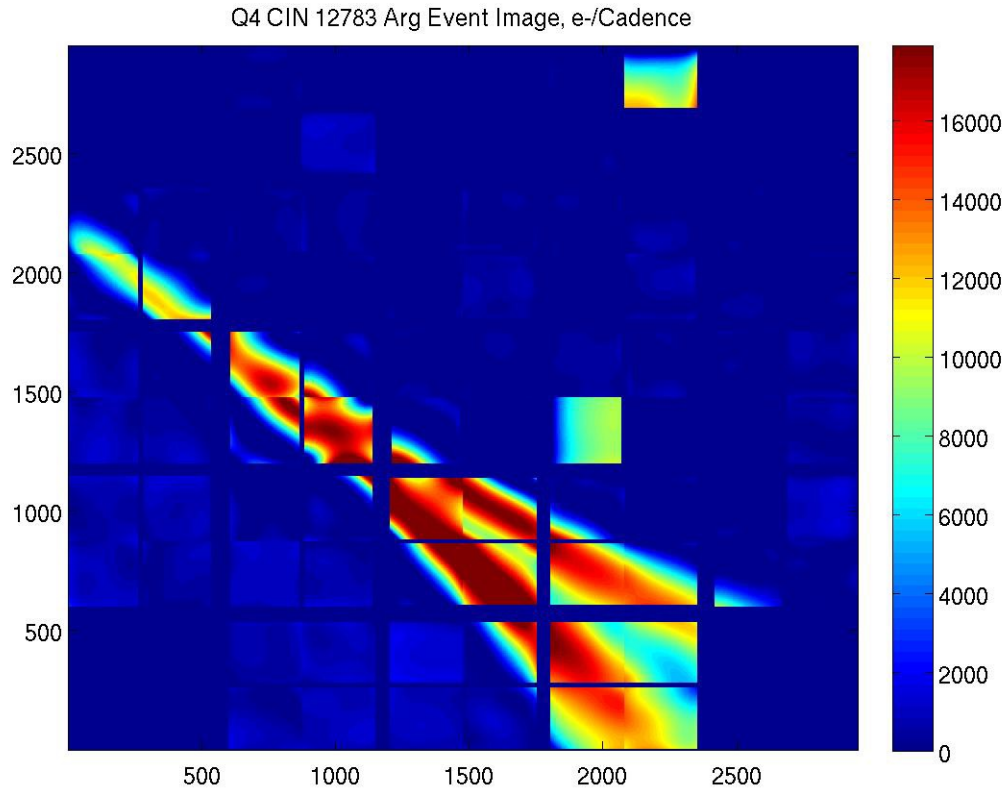


Figure 12: Image of Argabrightening event in Q4, CIN 12783. The image was formed by calculating 4th - order background polynomials for each mod.out, then subtracting the quiescent background values from the background during the event. Note structure on sub-mod.out scales. Square features in mod.outs 4.3 and 14.2 are believed to be uncorrected “chatter” in the calibrated background pixels and not real illumination features.

5.9 Pixel Sensitivity Dropouts

The CCD detectors on board the spacecraft 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 near or at their pre-event values after some time and reset without annealing, called Sudden Pixel Sensitivity Dropouts (SPSDs).
3. Long-term alteration of detector properties, which would only be restored by annealing the focal plane (which was never exercised).
4. Permanent damage.

Typically, annealable and permanent damage are caused by non-ionizing energy loss, or “knock-on” damage, which can be caused by any baryonic particle.

Transient charge deposits are removed by the pipeline’s CR detection algorithm (Twicken *et al.* 2010; Morris *et al.* 2016). The PDC module of the pipeline corrects SPSDs in the flux time series, as shown in **Figure 13**. The SPSD algorithm is described briefly in Stumpe *et al.* 2012 and in more detail by Morris *et al.* 2016. Users examining pixel data and uncorrected (SAP_FLUX) flux time series will still need to remain aware of them. Cadences where these effects have been identified are tagged in the FITS files (see KAM §2.3.1.1). Annealable and permanent damage effects do not appear common and are not mitigated in the pipeline.

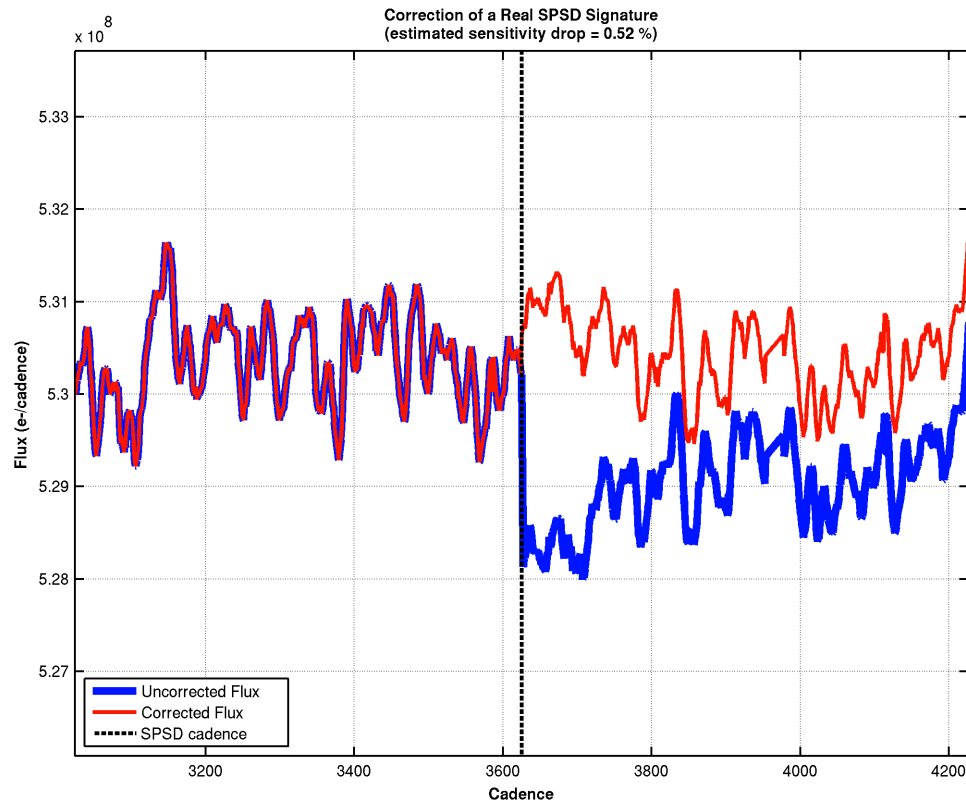


Figure 13: The effect of a sudden pixel sensitivity dropout (SPSD). The black dashed line shows the cadence where the SPSD occurs. The blue curve follows the PA data, while the red line shows the PDC correction of the SPSD.

5.10 Short Cadence Requantization Gaps

Short Cadence pixels at mean intensities $>20,000$ e- show banding as shown in **Figure 14**, with quantized values of number of electrons preferred. This is the result of the onboard requantization (KIH §7.4; Jenkins and Dunnuck, 2011; Jenkins, Caldwell, and Gilliland, 2004), and is considered benign since in the overall extraction the flux time series is near the Poisson limit. These requantization gaps are expected, and a necessary cost associated with achieving the required compression rates on board the *Kepler* spacecraft. However, the phenomenon is described here so that users will not suspect an undiagnosed problem.

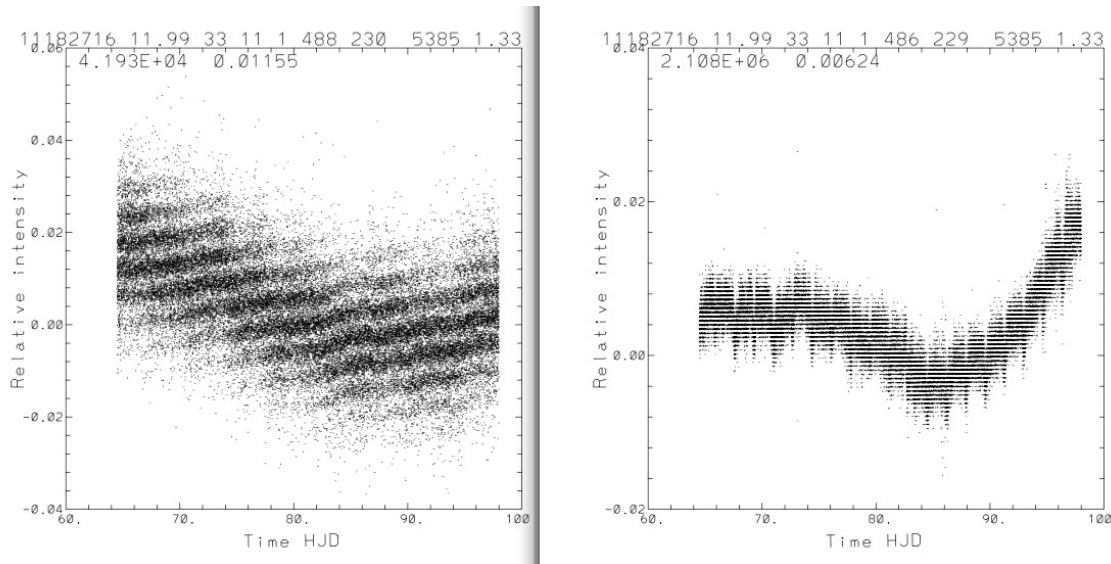


Figure 14: Requantization gap example in Q1 SC pixel time series. The ‘band gaps’ scale with mean intensity (42,000 e- left, 2.1e6 right). See KIH §7.4 for a discussion of quantization and the (insignificant) information loss it entails.

5.11 Spurious Frequencies in SC Data

The light curve systematic error correction scheme for SC uses the basis vectors and fit coefficients found by Bayesian MAP (Stumpe *et al.*, 2012; Smith *et al.*, 2012) for the corresponding LCs, since there are an insufficient number of SC light curves in an output channel to form a reasonably complete basis and implement MAP directly. This method (“quickMAP”) is described in more detail in the KDPH PDC chapter. Since the 30:1 spline interpolation of the LC basis vectors to the SC sampling times cannot well-represent signals above the LC Nyquist frequency (283.2 μHz), spurious signals at these higher frequencies will not be effectively removed by PDC, and mitigation consists of identifying common mode signals and either removing or ignoring them.

5.11.1 Integer Multiples of Inverse LC Period

Spurious frequencies are seen in SC flux time series, and pixel data of all types, including trailing black collateral pixels. The frequencies are exact multiples of the LC sampling rate (566.4 μHz), as shown in **Figure 15**. 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 asteroseismic analyses. The physical cause of this problem is not understood but is presumably related to power and thermal cycles in the LDE as events repeat themselves every LC period.

This feature was first reported in Q1 data (Gilliland *et al.*, 2010). 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, and is also therefore an additive rather than a multiplicative systematic error. These spurious signals were stable in frequency and amplitude throughout the mission.

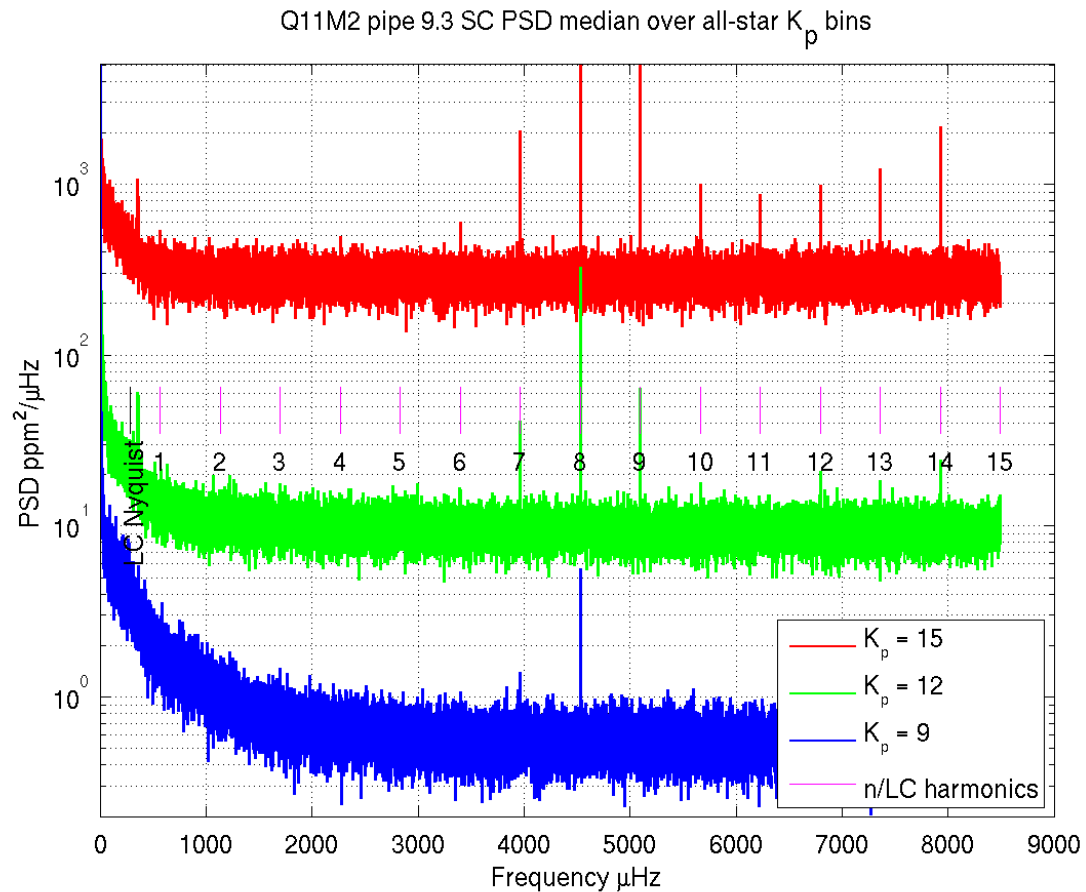


Figure 15: Spurious frequencies are observed in the power spectral density for all stars from Q11. The figure shows the median over stars in each of three magnitude bins, spanning more than a factor of 100 in brightness. The 6th – 14th multiples of the LC sampling frequency (566.4 μ Hz) are clearly visible for the 15th mag star. Even at 9th magnitude, the 8th multiple of the 1/LC artifact remains the dominant spectral feature (Figure inspired by Q1 results of Gilliland *et al.*, 2010).

5.11.2 Other Frequencies

Further analysis of SC data in Q1 and subsequent quarters showed several stars in which the SC data has peak power at 7865 μ Hz (\sim 127.16 seconds). This is not a harmonic of the 1/LC noise discussed in the previous section. Across the Q2M1 safe mode event, the phase shifted for both the 1/LC harmonics and for the 7865 μ Hz feature. Since stellar signals tend to stay at the same phase, the phase shift across Q2M1 is evidence that the n/LC and 7865 μ Hz features are instrumental. Peaks have also been reported at 7024, 7444, 7865, and 8286 μ Hz – consistent with a splitting of 421 μ Hz = 2375.3 s, or 39.59 minutes.

In Q0-Q2, multiple groups reported the issues around 80-95 μ Hz which correspond to about 3.2 h. The non-sinusoidal nature of these spurious signals leads to evenly spaced peaks, not unlike stellar oscillations. This is the same period as the temperature variation of the reaction wheel housing temperature (§5.2). Users are encouraged to examine the CBVs (§7) to strengthen the case that detected spectral features are astrophysical and not instrumental.

A list of spurious short cadence frequencies was supplied by Andrzej Baran (see also Baran, 2013) based on the Q0-Q14 data. The updated list of spurious frequencies noted in the SC data is given in Table 5.

There is some evidence of increased broadband noise for the brightest stars in the 1-3 mHz frequency band towards the end of the mission, perhaps due to increased rumble in reaction wheel 4 in the months before its mission-ending failure.

Table 5: List of Possible Spurious Frequencies in SC data above the LC Nyquist frequency (283 μ Hz). Users are advised to check detections against this list, and do their own tests for non-astrophysical common mode signals in SC data. Labels: RWTH = Reaction wheel housing temperature controller thermal cycling (believed not to be a problem from Q3 onward). U = unknown. Narrow lines are defined as $n/\Delta n > 50$, broad lines as $n/\Delta n < 50$.

SC spurious frequency summary							
freq	freq	period	period	period	period		
uHz	d ⁻¹	S	min	hr	d	Label	width
86.8	7.50	11520.00	192.000	3.2000	0.13333	RWTH	Broad
196.5	16.98	5088.34	84.806	1.4134	0.05889		Broad[6]
242.5	20.95	4124.11	68.735	1.1456	0.04773		V.broad [5]
290.0	25.06	3448.28	57.471	0.9579	0.03991	U1	Broad
340.0	29.38	2941.18	49.020	0.8170	0.03404	U2	Broad
344.9	29.80	2899.33	48.322	0.8054	0.03356		V.broad [4]
360.0	31.10	2777.78	46.296	0.7716	0.03215	U3	Narrow
362.8	31.35	2755.98	45.933	0.7656	0.03190		V.broad [3]
365.8	31.61	2733.31	45.555	0.7593	0.03164		Broad [2]
370.4	32.00	2700.00	45.000	0.7500	0.03125	U4	Narrow
421.0	36.37	2375.30	39.588	0.6598	0.02749	split U5-U8	Narrow
566.4	48.94	1765.56	29.426	0.4904	0.02043	1/LC	Narrow
1132.8	97.87	882.78	14.713	0.2452	0.01022	2/LC	Narrow
1680.1	145.15	595.25	9.921	0.1653	0.00689		[10]
1699.2	146.81	588.52	9.809	0.1635	0.00681	3/LC	Narrow
1803.4	155.81	554.52	9.242	0.1540	0.00642		[1]
1825.2	157.50	547.88	9.131	0.1522	0.00634		[9]
2092.6	180.80	477.88	7.964	0.1327	0.00553		[8]
2265.6	195.74	441.39	7.357	0.1226	0.00511	4/LC	Narrow
2832.0	244.68	353.11	5.885	0.0981	0.00409	5/LC	Narrow
3398.3	293.62	294.26	4.904	0.0817	0.00341	6/LC	Narrow
3964.7	342.55	252.22	4.204	0.0701	0.00292	7/LC	Narrow
4428.4	382.61	225.82	3.763	0.0627	0.00261		
4451.4	384.60	224.65	3.744	0.0624	0.00260		[11]
4531.1	391.49	220.70	3.678	0.0613	0.00255	8/LC	Narrow
4995.8	431.64	200.17	3.336	0.0556	0.00232		[1]
5016.2	433.40	199.35	3.323	0.0554	0.00231		[11]

5097.5	440.43	196.17	3.270	0.0545	0.00227	9/LC	Narrow
5564.4	480.76	179.72	2.995	0.0499	0.00208		Broad [7]
5584.5	482.40	179.07	2.984	0.0497	0.00207		[11]
5663.9	489.36	176.56	2.943	0.0490	0.00204	10/LC	Narrow
6230.3	538.30	160.51	2.675	0.0446	0.00186	11/LC	Narrow
6796.7	587.23	147.13	2.452	0.0409	0.00170	12/LC	Narrow
7024.0	606.87	142.37	2.373	0.0395	0.00165	U5	Narrow [12]
7363.1	636.17	135.81	2.264	0.0377	0.00157	13/LC	narrow
7444.0	643.16	134.34	2.239	0.0373	0.00155	U6	Narrow [13]
7865.0	679.54	127.15	2.119	0.0353	0.00147	U7	Narrow [14]
7929.5	685.11	126.11	2.102	0.0350	0.00146	14/LC	Narrow
8286.0	715.91	120.69	2.011	0.0335	0.00140	U8	Narrow [15]
8495.9	734.04	117.70	1.962	0.0327	0.00136	15/LC	narrow

[1] Noted in Q7.

[2] 31.52-31.70 d⁻¹, and harmonic at 63.38 c/d.

[3] 31.17-31.53 d⁻¹, noted in Q7.

[4] 29.46-30.14 d⁻¹.

[5] Noted in Q7.

[6] Broad harmonics at 33.95 d⁻¹ and 50.92 d⁻¹.

[7] 480.82 d⁻¹ in Q6.3, 480.69 d⁻¹ in Q7.1, 480.59 d⁻¹ in Q7.2, 480.49 d⁻¹ in Q7.3.

[8] Noted at 180.45 d⁻¹ in Q6.3.

[9] Also noted at 157.97 d⁻¹.

[10] Also noted at 145.36 d⁻¹.

[11] 384.6 d⁻¹, 433.4 d⁻¹ and 482.4 d⁻¹ have the same separation as the 1/LC artifacts but are shifted from the LC comb by about 6.98 d⁻¹.

[12] Evolved to 609.29 d⁻¹ by Q7.1.

[13] Evolved to 645.14 d⁻¹ by Q7.3.

[14] Evolved to 680.73 d⁻¹ by Q7.3.

[15] Evolved to 716.34 d⁻¹ by Q7.3.

[12-15] U5-U8 create a comb with separation of 36.4 d⁻¹ evolving to 35.7 d⁻¹ by Q7.3.

5.12 Propagation of Uncertainties

Since DR 8, the propagation of uncertainties (POU) has been turned on as standard procedure for all data. Between DR 8 and DR 22, uncertainties were calculated from a fully propagated covariance matrix every 24th Long Cadence; the variances were then linearly interpolated for the remaining cadences, in order to save time. Users of these flux uncertainties had noted that for targets that have large variations on timescales less than 12 h, the uncertainties can vary unphysically due to interpolation between largely divergent values. An example is shown in **Figure 16**. DR 23 (Q17) was the first attempt at fully sampled POU, but contained a bug remedied in later releases. For DR 24 and DR 25, the LC error estimates are fully sampled and propagated for the entire mission data set. The "fully sampled POU" is a shot noise estimate (minimal POU) plus a bias term derived from the difference between POU calculated using full covariance on a set of decimated cadences and minimal POU on those same cadences.

For bright sources, users have the option of largely ignoring the systematic noise and reconstructing a set of uncertainties using: $\sigma^2 = (\text{shot noise})^2 + (\text{read noise})^2$. The number of pixels used in the optimal aperture can be obtained from either the light curve or target pixel files. The read noise for each channel is given in the headers of those files with the keyword READNOIS.

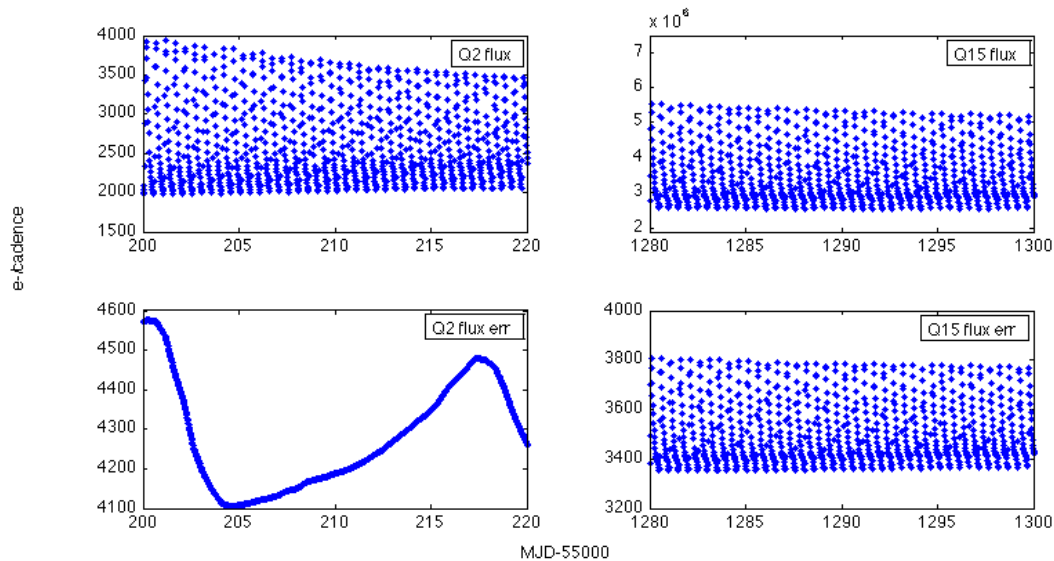


Figure 16: Upper panels: Flux time series for V450 Lyr. Lower panels: The corresponding flux uncertainties in the light curve and target pixel files. The large variations seen in Q2 are due to the decimation in the sampling of the error calculation and subsequent interpolation. In Q15, a new uncertainty calculation has been implemented that better estimates the uncertainty for each cadence.

5.13 Onboard Spacecraft Errors

Cadences can also be affected by errors reported by the spacecraft. These cadences are identified with the QUALITY flag in the FITS headers from Data Release 20 onwards (LDE flag in **Figure 18**, Detector Anomaly flag bit 15 in KAM Table 2.3). We identify these cadences in the Anomaly Summary table or figure in the relevant Release Notes so that users can identify the cause of missing cadences. These errors are primarily related to memory and synchronization issues.

5.14 Coronal Mass Ejections - Data Loss and Detector Changes

5.14.1 Quarter 12

After the increase in solar activity at the start of 2012, the spacecraft was impacted by several coronal mass ejections. Data quality during Q12 was affected by three coronal mass ejections (CMEs). Cadences impacted by the CME were marked with the Manual Exclude flag (bit 9, decimal 256) in the SAP QUALITY column of the light-curve files (and the QUALITY column in the target pixel files). Users are strongly discouraged from using the data collected during these events.

During CMEs, the flux of charged particles from the Sun impacting the spacecraft can increase by many orders of magnitude, causing an increase in measured dark current, as well as the "cosmic ray" count. The fine guidance sensors are also impacted, so the pointing of the spacecraft can deviate from the nominal value by many millipixels.

Users should note that the focal plane underwent some long-term changes after the CMEs in Q12. These effects are particularly noticeable after the third, and most powerful, CME in that quarter. In particular:

- The dark current rose slightly. This increase does not materially affect data quality.

- A small number of pixels show a pronounced drop in sensitivity after the largest CME. When a target star falls on one or more of these pixels, the mean measured flux will be lower after the CME than before. In the majority of such cases, there is no requisite pre-SPSD baseline available for the SPSD correction algorithm to use for identifying CME-driven SPSDs. Hence, PDC fails to correct most of them and often introduces additional low-frequency artifacts into the light curve.

5.14.2 Quarter 14

The spacecraft was also impacted by a small Coronal Mass Ejection on MJD 56103 (25 June 2012) in the first month of Q14 data. The effects can be seen for an approximately 16-hour period, from long cadences 57519 to 57551, in the collateral data and in the background flux time series. Data quality was not degraded to the point of flagging or exclusion.

5.15 Attenuation of Long Period Signals

From Release 20 onwards, users should note that PDC attenuates long period signals in the data. Users interested in events with durations longer than 10 days (i.e., with most of their spectral power $< 1 \mu\text{Hz}$) should not use these PDC light curves, but rather should do their own photometry using the calibrated pixels or the uncorrected light curves and the Cotrending Basis Vectors (CBVs, §7).

All PDC algorithms have difficulty distinguishing between astrophysical and instrumental signals on these timescales. In its pursuit of good noise performance on transit time scales, msMAP assumes all long period signals are systematic and removes them. We quantified the effect on an astrophysical signal by injecting sinusoidal signals of different periods and fixed amplitude equal to 0.1% times the median flux level into real data. We then measured the attenuating effect of the PDC processing. The results are shown in **Figure 17** and Gilliland *et al.* 2015. Signals with periods less than ~ 3 days are very well preserved, while signals with periods greater than 20 days are almost entirely removed. Larger input signals show relatively better preservation at long periods, while smaller amplitudes show more damping (Gilliland *et al.* 2015). Note that transits, or any sharp, periodic feature, are not affected by this corruption. Though the interval between transits is long, their durations are always short compared to PDC's bandpass.

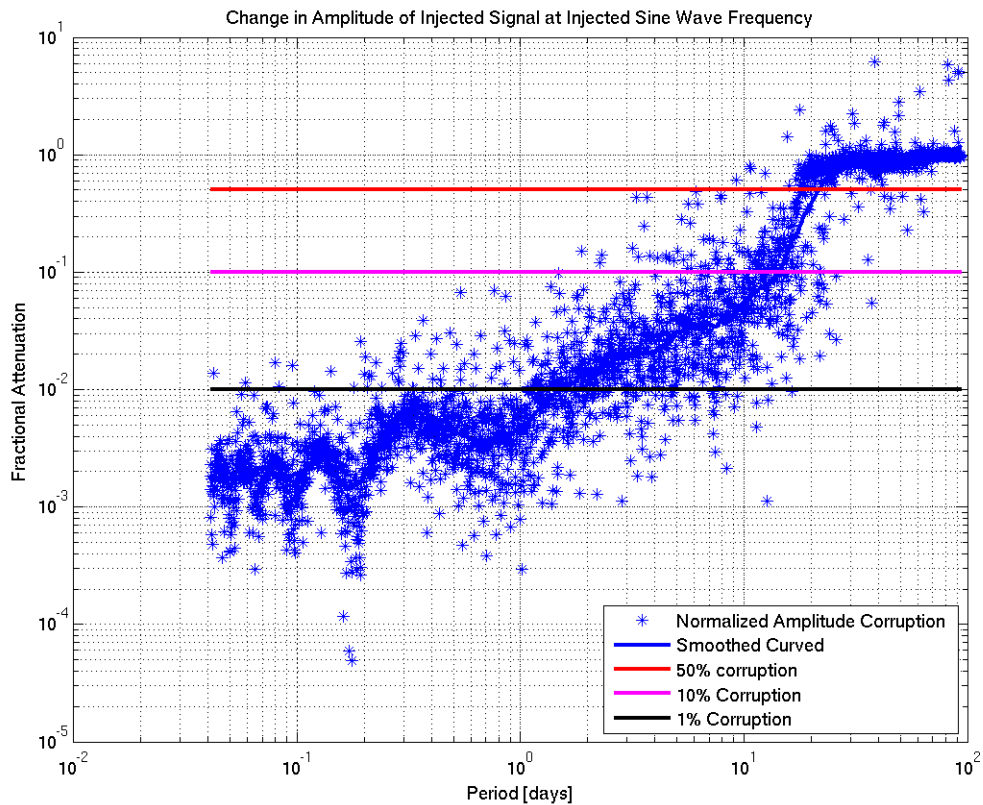


Figure 17: Signal attenuation as a function of period in PDC using msMAP. Sine waves of different periods and 0.1% relative amplitude were injected into long cadence lightcurves before being processed with PDC. Their amplitudes before and after processing were compared. Each blue symbol represents the fractional change in amplitude of a given injected sine wave. Short period signals were attenuated by less than 1%, while long period signals were entirely removed.

5.16 Anomaly Summary Figure

The anomalies that are identified for purposes of pipeline processing are graphically summarized in the relevant Release Notes for each quarter. **Figure 18** shows an example from Q14 (Data Release Notes 19). The afflicted cadences are marked in detail in the FITS file quality flags (KAM Table 2-3).

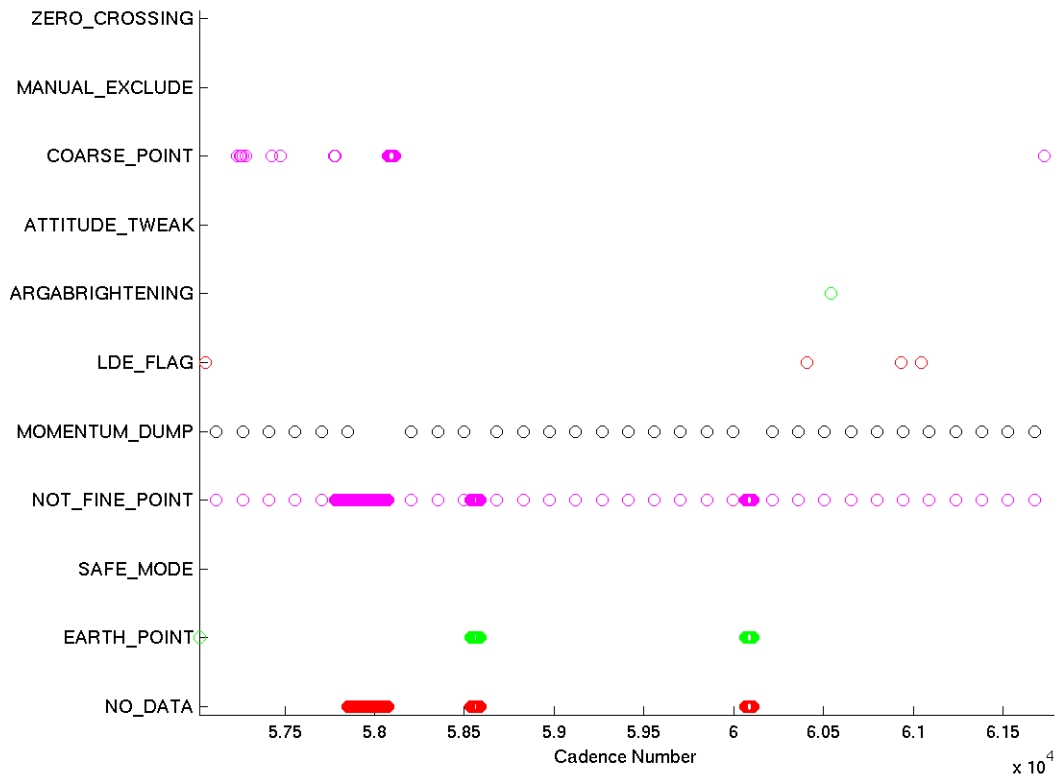


Figure 18: An overview of the location of the data anomalies flagged in Q14. "No_Data" is not an anomaly flag and simply indicates those cadences with no data collected (e.g., during Earth-point or Safe Mode events).

Clarifications on select flags in **Figure 18** are listed here:

- ARGABRIGHTENING refers to cadences where the multiple-channel Argabrightening flag (bit 7, decimal value 64) was set. The single channel Argabrightening flag (bit 13, decimal value 4096) is not represented on this plot. See §5.8 for more discussion of Argabrightenings.
- COARSE POINT refers to cadences where the pointing of the telescope drifted by more than 0.5 millipixels from the nominal value. NOT FINE POINT refers to cadences where the telescope's fine guidance sensor reported that the telescope was not in fine point mode. These flags are combined as bit 3 in the QUALITY field of the archive FITS files (KAM Table 2-3).
- LDE FLAG refers to flags set by the Spacecraft when an error was detected in the Local Detector Electronics (LDE) or the on-board memory. The pipeline does not process these cadences and only raw pixels are available. See §5.13 for more details. This is bit 15 in the QUALITY field of the archive FITS files (KAM Table 2-3).

5.17 Fixed Bugs in SC Pipeline Processing

These historical notes for major bugs in the SC pipeline, now fixed in DR 25, are an aid to evaluating already-published *Kepler* papers which used data processed with earlier pipelines.

In April 2015 it was noticed that the DR 24 short-cadence (SC) data, which was processed with the SOC 9.2 pipeline, is incorrect and virtually unusable for projects requiring high photometric precision. The problem was traced to an error in the pipeline module that calibrates the pixels, specifically in how it handles the black model calculated by the Dynablack algorithm (Kolodziejczak *et al.*, 2010; KDPH). While this is a major problem for SC data, there is no impact on the long-cadence data for either DR 24 or DR 25. More detail can be found in the Erratum to DRN 24 (KSCI-19064-002a-DRN24err.pdf) at MAST.

In March 2016 it was discovered that *Kepler* and *K2* short-cadence collateral smear data had been scrambled for many targets throughout the entire *Kepler* mission, due to a misunderstanding of the pixel readout order in the unique, target-specific way that short cadence collateral data are selected and read out (See KIH §2.6.3.2). The problem and its impact are described in detail in project document KSCI-19080-002, available at MAST. The bottom line for *Kepler*: only a handful of targets are likely to be affected in a significant way. Comparison of the DR 25 data with earlier releases can quantify the impact on individual targets and extracted scientifically important parameters (Salabert *et al.*, 2016).

6. Time and Time Stamps

The primary time stamps available for each cadence in both LC and SC time series are intended to provide times corrected to the solar system barycenter, and are uniquely determined by the mid-cadence time stamp and the position-dependent correction for each star.

6.1 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 shall be known to ± 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 steps are required to approach the 50 ms requirement:

1. Relate readout time of a pixel to the 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 (KIH §5.1) of that cadence is read out from the LDE.
2. Convert VTC to UTC for each cadence, using Time Correlation Coefficients at the DMC.
3. Convert UTC to Barycentric JD. This is done in PA 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 is the speed of light, and β 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.
4. Subtract readout time slice offsets (see KIH Section 5.1). This is done in PA for each channel. The magnitude of the time slice offset is $t_{\text{rts}} = 0.25 + 0.62(5 - n_{\text{slice}})$ s, where n_{slice} is the time slice index (1-5) as described in the KIH and recorded in the archive file FITS headers. Note that this will in general be different from Quarter to Quarter for the same star, as the spacecraft rolls and stars move cyclically between different mod.outs. The relative timing of events across Quarter boundaries must then take this into account.

6.2 Time Transformations, VTC to BKJD

6.2.1 Vehicle Time Code

The VTC clock is undisciplined and drifts with respect to Coordinated Universal Time (UTC). To maintain cadence boundaries to within ± 2 s of their ideal UTC times, the LDE was resynchronized after safemodes and during most quarterly roll science breaks. **Figure 19** shows the measured drift in LDE time stamp with respect to UTC. As a consequence of this drift, the cadence mid-times are not evenly spaced in UTC, and the cadence durations also change, although the variation in duration is $< 10^{-6}$ of a LC duration and therefore the photometry cannot be affected at more than that level (**Figure 20**).

When the data is downloaded to Earth, the Mission Operations Center converts VTC to UTC, correcting for leap seconds and any drift in the spacecraft clock, as measured from telemetry.

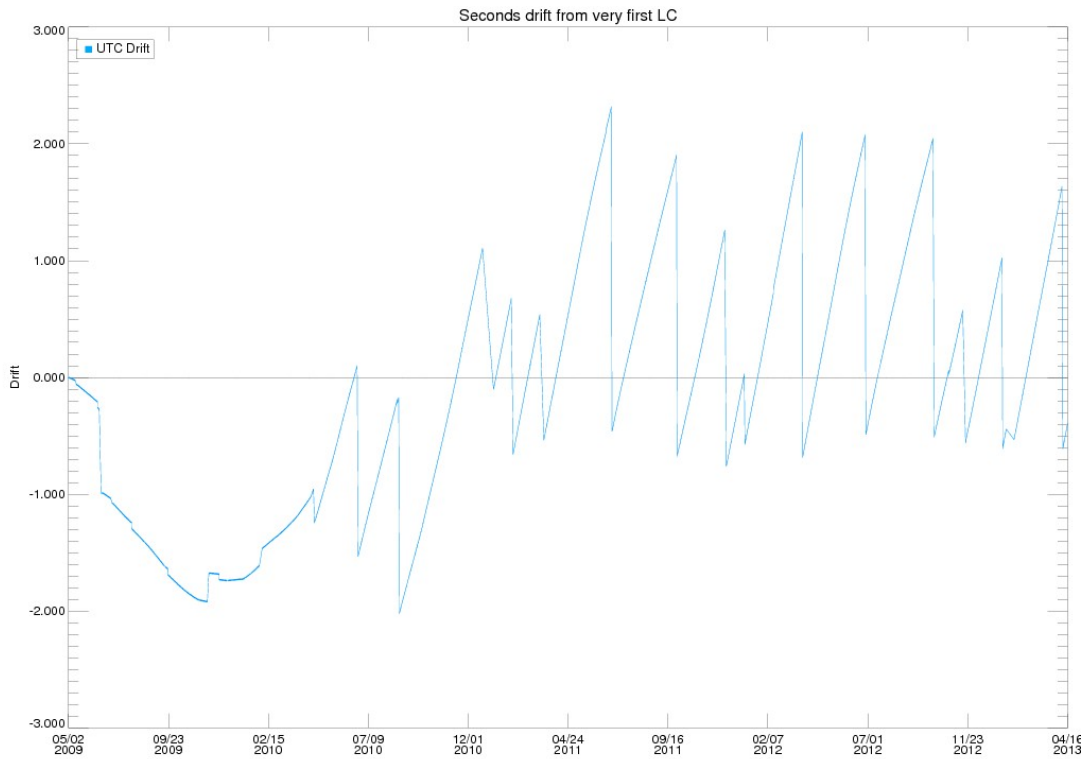


Figure 19: The drift in LDE time stamps, given in seconds, is plotted with respect to UTC. Since the sudden rate change in the drift near MJD 55293 (7 April 2010) the time stamp (derived from VTC) has run consistently slow relative to UTC. To compensate, more frequent and larger LDE synchs have been performed, causing the saw-tooth discontinuities. The times given in the FITS headers are corrected for this drift.



Figure 20: The change in the LC duration, given in seconds, with time, due to the drift in the onboard spacecraft clock. The two-week running average is plotted in blue, and the cumulative average with time is plotted in green. Each point on the green line is the median of all blue points from earlier epochs. The overall change is at the sub- millisecond level and therefore should not be of concern to the user.

6.2.2 Barycentric *Kepler* Julian Date

The contemporary value of BJD (~ 2.5 million days) is too large to be stored with millisecond precision in an eight byte, double precision, floating point number. To compensate, *Kepler* reports the value of BJD-2454833.0. This time system is referred to as Barycentric *Kepler* Julian Date (BKJD). The offset is equal to the value of JD at midday on 2009-01-01. BKJD has the added advantage that it is only used for corrected dates, so it is more difficult to confuse BKJD dates with uncorrected JD or MJD.

6.3 Absolute Timing Accuracy with *Kepler*

To constrain the absolute timing accuracy of observations reported by *Kepler*, we can compare timing measurements obtained with the spacecraft to simultaneous, independent measurements of the same astronomical phenomenon. The pulsating sdB star, KIC 10139564 (Baran *et al.*, 2012) was observed in short cadence with *Kepler*, and also observed simultaneously from two ground based observatories, Moletai Astronomical Observatory and Terskol Observatory. These observatories were part of the Whole Earth Telescope (Provencal *et al.*, 2012) collaboration's DAMP04 campaign. The Whole Earth Telescope has long experience tying observations from different telescopes onto a uniform time system (Nather *et al.*, 1990).

For each month of short cadence *Kepler* data, we fit a sine curve with period 182.71975 seconds to the data and measured the best fit phase for that month. We performed the same measurement on data from each of the ground based observatories. We compare the measured phase for each dataset

to that expected if the phase was constant (the O-C value). We show the results in **Figure 21**. The small filled circles show the variation in phase for *Kepler* data, while the open squares show the ground-based results. The volume of ground-based data is much smaller than that obtained by *Kepler* and uncertainties are correspondingly larger. If the period of pulsation is changing at a constant rate, we would expect a parabolic increase in the value of O-C (see Kepler *et al.* 1991 for a derivation). We show the best-fit parabola to the *Kepler* data as a red line. This change in period might be due to the cooling of the star, or a change in its internal structure. Of more interest is the difference between the phase recorded by *Kepler* and the ground-based observations. Using the corrected *Kepler* timestamps (see the Note at the start of this section), in the worst case the disagreement is significant at 1.5σ . From this measurement we conclude that the absolute times are correct to within 5 seconds with 84% confidence and within 7 seconds to 97.5% confidence.

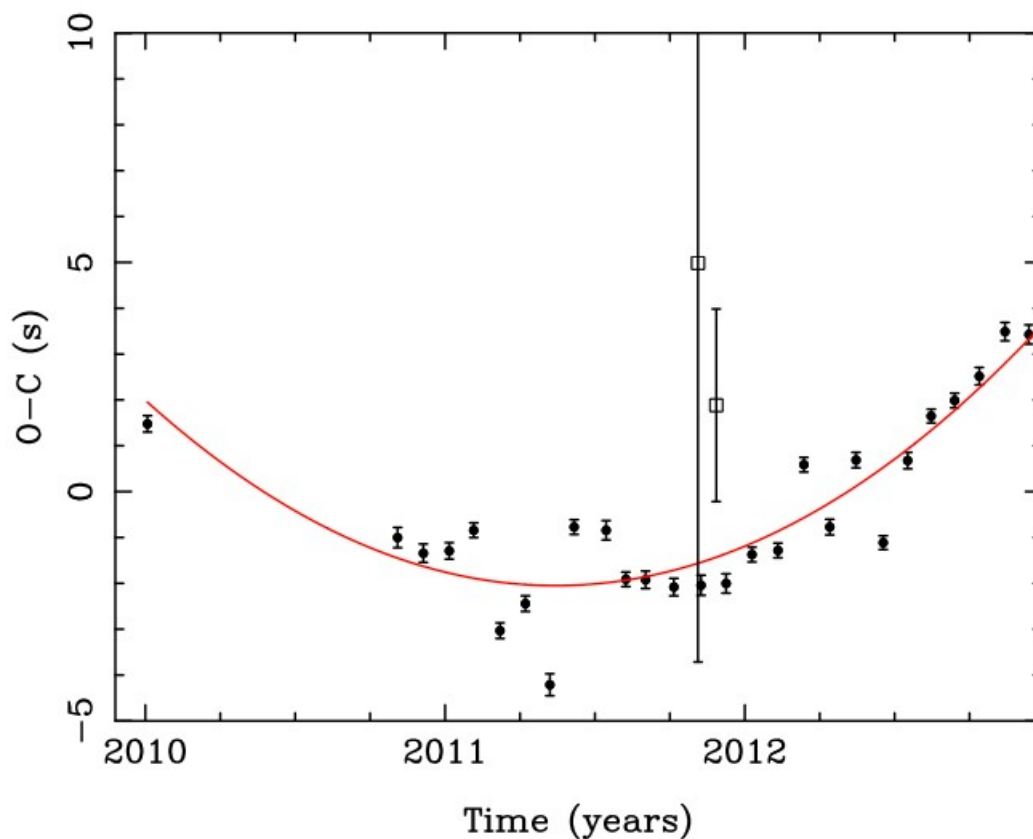


Figure 21: Comparison of the times of events recorded by *Kepler* (filled circles) and ground based observatories (open squares), along with a parabolic fit to the *Kepler* data (Baran *et al.*, 2012). The ground-based and *Kepler* times agree to within 7 seconds with greater than 97% confidence, establishing the absolute times of *Kepler* with that accuracy. The internal scatter of the *Kepler* data around the fit is ~ 100 ms $1-\sigma$.

7. Ensemble Cotrending Basis Vectors

7.1 Introduction

We infer the presence of systematic errors in the *Kepler* flux time series from the correlations observed between them, since we do not expect the stars themselves to have correlated signals. These correlations can be represented as linear combinations of orthonormal functions, called cotrending basis vectors (CBVs), which in some sense represent most of the correlated features in a reference ensemble of flux time series for a given quarter and output channel (mod.out). Light curve correction can be performed by finding the ‘best’ fit of these CBVs to a light curve, and removing the fit. The CBVs of a reference ensemble of highly correlated stars are provided so that users can perform their own systematic error removal, using their own tools or those supplied by the GO Office.

From SOC Pipeline 8.3 onwards, both the ‘standard’ CBVs calculated in the earlier versions (Stumpe *et al.*, 2012; Smith *et al.*, 2012) and new multi-scale CBVs (Stumpe *et al.* 2014) are computed by the pipeline, which then determines the most appropriate correction for each target. We continue to provide the standard CBVs for users to correct their own systematics, but the multi-scale CBVs will not be an archived data product.

Systematic error correction of the SAP_FLUX flux time series using the CBVs may be preferable or necessary to using the delivered PDCSAP_FLUX flux time series, since:

1. Users of target pixel files who generate their own flux time series will need to do their own systematic error removal, though it is up to the user to understand whether a set of CBVs which well-represent systematic errors in uncorrected pipeline flux time series is also a good representation of their systematic errors given differences in aperture size and their method of extracting a flux time series from the pixels,
2. Users can decide how many CBVs to use to fit systematic trends, and
3. Users can use the CBVs with non-least squares fitting methods such as MAP (Jenkins *et al.*, 2012) or the lasso (Tibshirani, 1996).

7.2 Generation of CBVs

The method for CBV generation is detailed in Smith *et al.* (2012). Briefly, the method:

1. Normalizes each flux time series with the mean and standard deviation of the time series,
2. Calculates the correlations between normalized flux time series and selects the 50% most correlated stars as the reference ensemble,
3. Estimates the intrinsic variability of each star in the reference ensemble by removing a 3rd order polynomial fit, and calculating the standard deviation of the detrended light curve normalized by flux uncertainty (roughly, read and shot noise). Stars with a normalized detrended standard deviation 30% greater than the median value on a particular mod.out are rejected from the reference ensemble,
4. Performs a Singular Value Decomposition (SVD) of the median-removed, median normalized flux time series of these most correlated stars. The CBVs are in this case the SVD principal components, though a more general nomenclature is used to allow users to contemplate non-SVD approaches. Users are provided with the leading 16 components. Users are also provided with the SVD principal values for each mod.out to inform their decision about how many components to use in their fits.

7.3 Using CBVs

To use the CBVs for least-squares fitting, subtract the mean uncorrected flux from the uncorrected flux time series of interest and divide by the mean. Since the basis is orthonormal, the linear least-squares fit coefficient of the n th CBV is simply the inner product of the median-removed, median-normalized uncorrected flux time series with the n th CBV. Subtract the fit to get the corrected (median-removed, median-normalized) flux time series.

Convenient CBV tools, including robust fitting and time window exclusion, are provided by the Guest Observer Office as part of PyKE and are available from <http://Keplerscience.arc.nasa.gov/ContributedSoftwareKepcotrend.shtml>. The site provides instructions on how to install the software and specific instructions on how to use the tools to fit the CBVs. Users should note that, unlike the SOC Pipeline, these tools do not include scalar amplitude corrections for the fraction of target flux captured in the optimal aperture, or the fraction of the total flux in that aperture which is from the target star. These quantities (the flux fraction in aperture and the crowding metric, respectively) are available for the optimal apertures computed by the SOC as keywords in the FITS files at MAST (see KAM §2.3.1.2).

Figure 22 shows that 8 or fewer components generally capture most of the systematic error for all mod.outs, though 16 are provided if users wish to make their own decisions. **Figure 23** shows the strongest 8 CBVs for Q5, in decreasing strength from left to right and top to bottom.

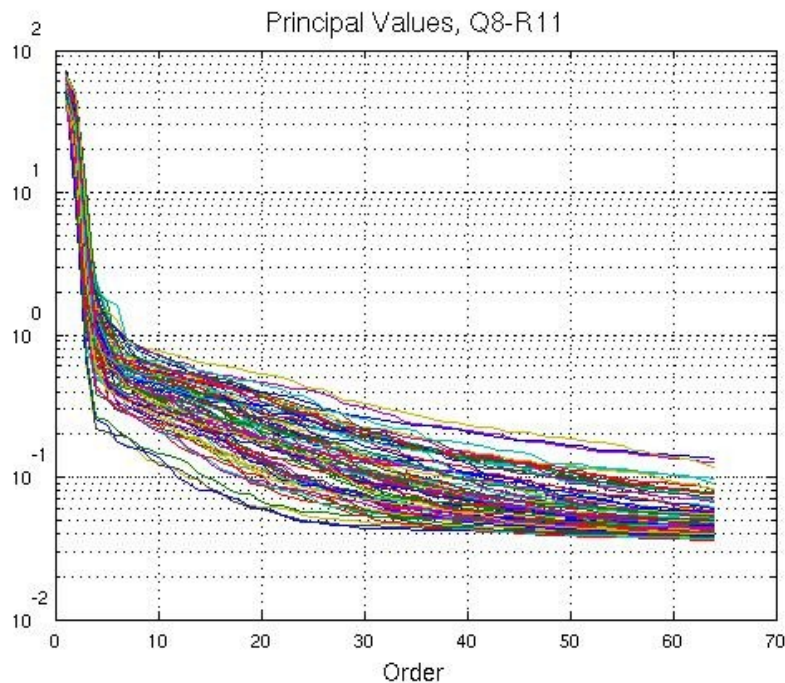


Figure 22: Principal values of SVD-extracted cotrending basis vectors for Quarter 8, Release 11, showing that most of the systematic error can be accounted for by the first 8 or fewer components. Users are provided with the first 16 components for each mod.out, as well as the principal values that are plotted in this Figure.

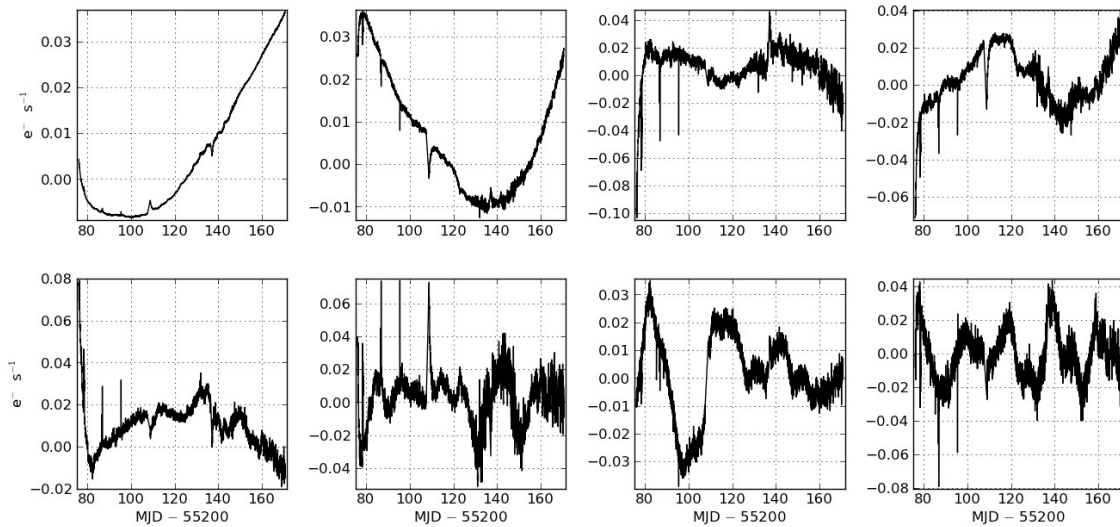


Figure 23: These panels show the strongest eight basis vectors from analysis of Q5, in decreasing order of importance from left to right and top to bottom.

7.4 Cautions

1. Channels with strong Moiré pattern drift (see the Instrument Handbook, §6.7) have less stable weak (>5th order) CBVs, in the sense that random constant perturbations of the reference ensemble give results which differ by significantly more than a sign and a constant. While the stability of components to the 10th order has been verified for good channels, the pipeline default for all channels is to use only the first 8.
2. While most variable stars have been removed from the reference ensemble, it is still possible that some weaker CBVs have been influenced by noisy stars or by variable stars whose signal slips through the detrending filter. Users should exclude bad CBVs from their fits, and in general be cautious about reporting results with the same period and phase as one of the basis vectors used in the fit.
3. No vetting of CBVs 9-16 has been done, as the pipeline does not generate the robust fit coefficients for these CBVs since they are not used in MAP. These CBVs should be used with extreme caution.

7.5 Obtaining CBVs

The cotrending basis vectors can be downloaded in FITS files from MAST separate from the data download (KAM §2.3.4). There is one file per quarter containing 84 extensions, one for each channel. Each extension contains 16 basis vectors along with the cadence and MJD of the observations. The cadences found in the basis vector file match the number of cadences in the light curve file for that quarter. To ensure that the data was processed through the same version of the pipeline as the stars used to create the basis vectors, the keyword `DATA_REL` in the data's light curve file should match that found in the basis vector file. A new basis vector file was provided each time the data were reprocessed.

8. Contents of Supplement

There is no longer a Supplement to the KDCH, since the cadences affected by the various phenomena described in Sections 4 and 5 are now identified within the files themselves, typically by the quality flag in the FITS headers. See Table 2.3 in the *Kepler* Archive Manual for more details.

9. References

Publications

- Batalha, N. M., *et al.* 2010, ApJL, 713, L109
 Baran, A. S., Reed, M. D., Stello, D., *et al.* 2012, MNRAS, 424, 2686
 Baran, A. S. 2013, Acta Astronautica, 63
 Bryson, S. T., *et al.* 2010, ApJL, 713, L97
 Bryson, S. T., Jenkins, J. M., Klaus, T. C. *et al.* 2010, Proc. SPIE, Vol. 7740, 77401D
 Caldwell, D. A., *et al.* 2010, ApJL, 713, L92
 Christiansen, J. L., Jenkins, J. M., Caldwell, D. A., *et al.* 2012, PASP, 124, 1279-1287
 Gilliland, R. G., Jenkins, J. M., Borucki, W. J., *et al.* 2010, ApJ, 713, L160
 Gilliland, R. L., Chaplin, W. J., Dunham, E. W., *et al.* 2011, ApJS, 197, 6
 Gilliland, R. L., Chaplin, W. J., Jenkins, J. M., Ramsey, L. W., and Smith, J. C. 2015, AJ, 150, 133
 Haas, M. R., *et al.* 2010, ApJL, 713, L115
 Howell, S. B., Sobeck, C., Haas, M. R., *et al.* 2014, PASP, 126, 398
 Jenkins, J. M., *et al.* 2010a, ApJL, 713, L87
 Jenkins, J. M., *et al.* 2010b, ApJL, 713, L120
 Jenkins, J. M., Chandrasekaran, H., McCauliff, S. D., *et al.* 2010c, Proc SPIE, Vol. 7742, 77400D
 Jenkins, J. M., Dunnuck, J. 2011, Proc. SPIE Vol. 8146, 814602
 Jenkins, J. M., Smith, J. C., Tenenbaum, P. T., Twicken, J. D., and Van Cleve, J. E. 2012, "Planet Detection: The *Kepler* Mission" in Advances in Machine Learning and Data Mining for Astronomy, p. 355
 Kepler, S. O., Winget, D. E., Nather, R. E., *et al.* 1991, ApJL, 378, L45
 Koch, D. G., *et al.* 2010, ApJL, 713, L79
 Kolodziejczak, J. J., Caldwell, D. A., Van Cleve, J. E., *et al.* 2010, Proc. SPIE, Vol. 7742, 77421G
 Morris, R. L., Twicken, J. T., Smith, J. C., *et al.* 2016, in prep for PASP
 Nather, R. E., Winget, D. E., Clemens, J. C., Hansen, C. J., Hine, B. P., 1990, ApJ, 361, 309
 Provencal J. L., Montgomery M. H., A. Kanaan, A., *et al.* 2012, Ap J, 751, 91
 Quintana, E. V., Jenkins, J. M., Clarke, B. D., *et al.* 2010, Proc. SPIE, Vol. 7740, 77401X
 Salabert, D., García, R. A., Mathur, S., and Ballot, J. 2016, <https://arxiv.org/pdf/1611.05184v1.pdf>
 Shupe, D., Moshir, M., Makovoz, D., and Narron, R. 2005, ASP Conf. Ser., 347, 491
 Skrutskie, M. F., Cutri, R. M., Stiening, R., *et al.* 2006, AJ, 131, 1163
 Smith, J. C., Stumpe, M. C., Van Cleve, J. E., *et al.* 2012, PASP, 124, 1000
 Smith, J. C., Morris, R. L., Jenkins, J. M., Bryson, S. T., Caldwell, D. A., Girouard, F. R. 2016, PASP, 128:124501
 Stumpe, M. C., Smith, J. C., Van Cleve, J. E., *et al.* 2012, PASP, 124, 985
 Stumpe, M. C., Smith, J. C., Catanzarite, J. H., *et al.* 2014, PASP, 126, 100
 Tibshirani, R., 1996. Journal of the Royal Statistical Society B58(1), 267
 Twicken, J. D., Clarke, B. D., Bryson, S. T., *et al.* 2010, Proc. SPIE Vol. 7740, 774023
 Van Cleve, J. E., Howell, S. B., Smith, J. C., *et al.* 2016, PASP, 128:075002
 Witteborn, F., Van Cleve, J. E., Borucki, W., and Argabright, V. 2011, Proc. SPIE, Vol. 8151, 815117

Kepler project documents available at the MAST *Kepler* web site:

<http://archive.stsci.edu/Kepler/documents.html>

Jenkins, J. M., Caldwell, D. A., and Gilliland, R. L. 2004, KSOC-21008 "Algorithm Theoretical Basis Document" (ATBD)

Fanelli, M. N., *et al.* 2011, KSCI-19081-001 "*Kepler* Data Processing Handbook" (KDPH)

Jenkins, J. M., *et al.* 2017, KSCI-19081-002 "*Kepler* Data Processing Handbook" (KDPH)

Thompson, S. E., Fraquelli, D., Van Cleve, J. E., and Caldwell, D. A. 2016, KDMC-10008-006 "*Kepler* Archive Manual" (KAM)

Van Cleve, J. E., and Caldwell, D. A. 2016, KSCI-19033-002, "*Kepler* Instrument Handbook"

10. List of Acronyms and Abbreviations

ADCS	Attitude Determination and Control Subsystem
AED	Ancillary Engineering Data
BATC	Ball Aerospace & Technologies Corp.
BJD	Barycentric Julian Date
BKJD	Barycentric <i>Kepler</i> Julian Date
CAL	Calibration (pipeline software)
CBV	Cotrending Basis Vector
CCD	Charge Coupled Device
CDPP	Combined Differential Photometric Precision
COA	Calculate Optimal Aperture (pipeline module)
CR	Cosmic Ray
DAWG	Data Analysis Working Group
DV	Data Validation (pipeline software)
DVA	Differential Velocity Aberration
FFI	Full Field Image
FGS	Fine Guidance Sensor
FOV	Field of View
FPA	Focal Plane Assembly
GO	Guest Observer
HGA	High-Gain Antenna
JD	Julian Date
LC	Long Cadence
LDE	Local Detector Electronics
MAD	Median Absolute Deviation
MAST	Mikulski Archive for Space Telescopes
MJD	Modified Julian Date = JD - 2400000.5
PA	Photometric Analysis (pipeline software)
PAD	Photometer Attitude Determination (pipeline software)
PDC	Pre-search Data Conditioning (pipeline software)
PPA	Photometer Performance Assessment (pipeline software)
ppm	parts per million
PRF	Pixel Response Function
SC	Short Cadence
SNR	Signal-to-Noise Ratio
SO	Science Office
SOC	Science Operations Center
STScI	Space Telescope Science Institute
SVD	Singular Value Decomposition
TDB	Barycentric Dynamical Time
TPS	Transiting Planet Search (pipeline software)
UTC	Universal Time Coordinated
VTC	Vehicle Time Code