Kepler Data Release 19 Notes
Q14

KSCI-19059-001
Data Analysis Working Group (DAWG)
Susan E. Thompson (Editor)

Prepared by: Susan E. Thompson Date: Feb. 4, 2013
Susan E. Thompson, Kepler Science Office, for the DAWG (page 4)

Approved by: J. Christiansen Date: 2/4/2013
Jessie L. Christiansen, Editor-in-Chief

Approved by: Date: 2/4/2013
Jon Jenkins, Co-I for Data Analysis & DAWG Lead

Approved by: Michael R. Haas Date: 2/4/2013
Michael R. Haas, Kepler Science Office Director

Document Control

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

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

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

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

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

The Data Characteristics Handbook and accompanying Data Release Notes are the collective effort of the Data Analysis Working Group (DAWG), composed of Science Office (SO), Science Opertions 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.
Catanzarite, Joseph
Christiansen, Jessie L.
Clarke, Bruce D.
Coughlin, Jeffery L.
Girouard, Forrest
Haas, Michael R.
Hall, Jennifer
Ibrahim, Khadeejah
Klaus, Todd
Kolodziejczak, Jeffery (MSFC)
Li, Jie
McCauliff, Sean D.
Morris, Rob
Mullally, Fergal
Quintana, Elisa V.
Rowe, Jason
Sabale, Anima
Seader, Shawn
Smith, Jeffrey Claiborne
Still, Martin D.
Tenenbaum, Peter G.
Thompson, Susan E.
Twicken, Joe
Uddin, Akm Kamal

The correct citation for this document is: S. E. Thompson, J. L. Christiansen, J. M. Jenkins, D. A. Caldwell, T. Barclay, S. T. Bryson, C. J. Burke, J. Catanzarite, B. D. Clarke, J. Coughlin, F. Girouard, M. R. Haas, J. R. Hall, K. Ibrahim, T. C. Klaus, J. J. Kolodziejczak, J. Li, S. D. McCauliff, R. Morris, F. Mullally, E. V. Quintana, J. Rowe, A. Sabale, S. Seader, J. C. Smith, M. D. Still, P. G. Tenenbaum, J. D. Twicken, and A. K. Uddin, 2013, Kepler Data Release 19 Notes (KSCI-19059-001).

Contents

1 Introduction

These Data Release Notes provide information specific to the release of Q14 data, processed with SOC Pipeline 8.3. These Notes contain the summary figures and tables for this quarter. The Kepler Data Characteristics Handbook (Christiansen et al., 2012b) discusses most of the known phenomena found in the Kepler data in more detail.

1.1 Dates and Cadence Numbers for Q14

Contents of Data Release 19–Cadence Data

Q.m		First Cadence MJD midTime	Last Cadence MJD midTime	First Cadence UT midTime	Last Cadence UT midTime	Num CINs	Start CIN	End CIN
14	LC	56106.6374	56203.8196	28-Jun-2012 15:17:47	03-Oct-2012 19:40:10	4757	57024	61780
14.1	SCM1	56106.6275	56137.5020	28-Jun-2012 15:03:34	29-Jul-2012 12:02:49	45330	1699180	1744509
14.2	SCM2	56138.6469	56168.8062	30-Jul-2012 15:31:34	29-Aug-2012 19:20:59	44280	1746190	1790469
14.3	SCM3	56169.6651	56203.8294	30-Aug-2012 15:57:48	03-Oct-2012 19:54:23	50160	1791730	1841889

Contents of Data Release 19–Full Frame Images

Q	Class	Filename	UT Start	UT End
Q14	FFI	KPLR2012211123923	2012-07-29 12:09:58	2012-07-29 12:39:23
Q14	FFI	KPLR2012242195726	2012-08-29 19:28:01	2012-08-29 19:57:26
Q14	FFI	KPLR2012277203051	2012-10-03 20:01:25	2012-10-03 20:30:51

1.2 The SOC Pipeline for Q14

Data Release 19 was processed with the SOC Pipeline 8.3. For details on how Kepler processes the data through the front-end of the pipeline (modules CAL, PA, PDC), please see the Data Processing Handbook (Fanelli et al., 2011). Notable changes and improvements to the pipeline in 8.3 include the following:

- The PDC (Presearch Data Conditioning) module of the pipeline has implemented multi-scale MAP, or msMAP. The algorithm, which will be fully described in a future paper, splits the data using a wavelet decomposition into three length-scale bands before fitting and removing systematic features. This reduces the chance that an astrophysical signal will be accidentally removed. The pipeline automatically compares the performance of msMAP to the regular MAP algorithm, and chooses the reduction with the better performance. Multiscale MAP is preferred for approximately 90% of targets.
- The way in which the CDPP time series (Christiansen et al., 2012a) is calculated by the Transiting Planet Search (TPS) module has been changed slightly to avoid spurious transit-like signatures at the ends of the data set. In order to generate the CDPP time series, the length of the data must be extended to a power of 2. Under SOC Pipeline 8.3 the extension is created by combining data from the beginning and end of the data string. Averaging these two results decreases noise in the extended portion of the data, which is not used to search for transits. Because of the finite window over which CDPP is calculated and due to the moving circular nature of the calculation, the CDPP values early and late in the data string will be reduced. For long multi-quarter searches the edge effects have little impact, but for the shorter quarterly CDPP metric calculation, which are reported in the FITS headers, edge-effects can result in reporting a slightly lower CDPP metric.

1.3 Kepler Mission Timeline to Date

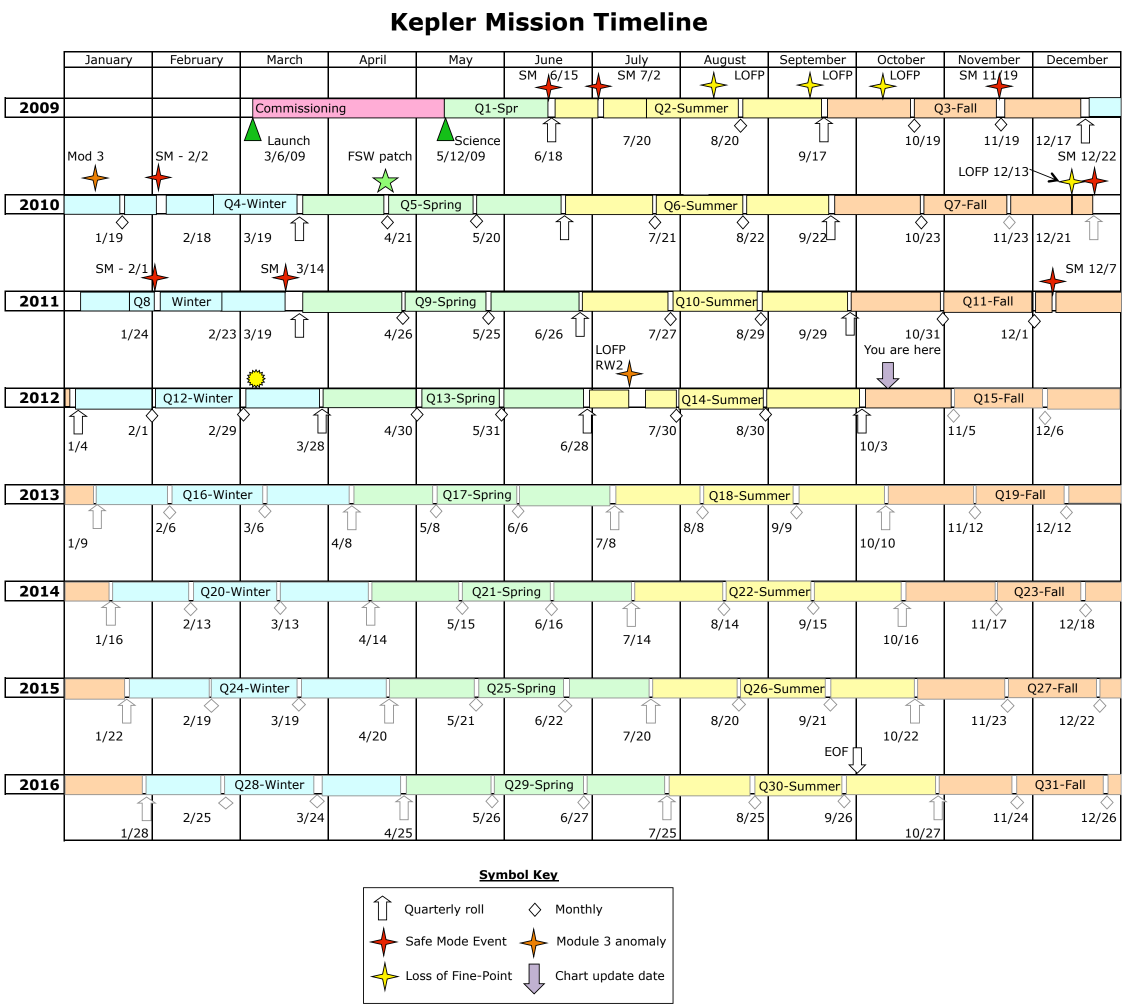

Figure 1: Kepler Mission Timeline as of the end of Q14. All future dates are tentative and subject to change.

2 Data Quality in Q14

2.1 Evaluation of CDPP

To understand the overall performance of the pipeline, we show the Temporal Median (TM) of the CDPP time series as calculated by the TPS pipeline for different versions of the SOC pipeline (Figure 2). We also provide the CDPP statistics for Q14, binned by magnitude, in Table 1.

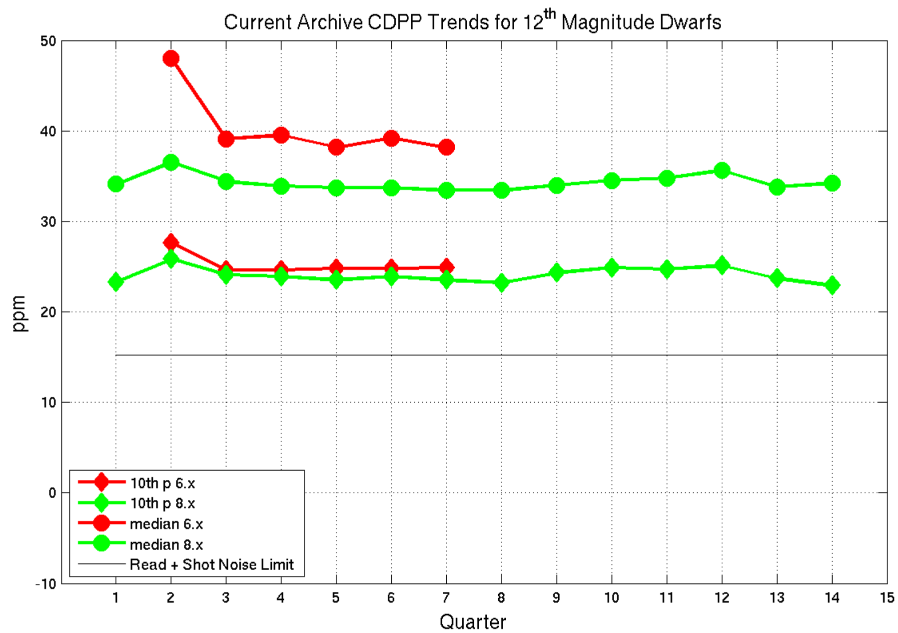

Figure 2: 6.5-h Temporal Median of the CDPP time series. The median (circles) and 10th percentile value (diamonds) for all dwarf stars between Kp=11.75–12.25 are given. The 6-h TMCDPPs have been divided by sqrt(13/12) = 1.041 to approximate 6.5-h TMCDPPs. A detailed discussion of the CDPP values is given in the Kepler Data Characteristics Handbook. The 6.x and 8.x labels given in the legend refer to the version of the SOC pipeline used. The reduction in CDPP for Q14 may be related to the new method used to calculated CDPP in Q14. The previous quarter's data points do not use this method.

Table 1: Aggregate statistics for the TMCDPPs by magnitude. Column Definitions: (1) Kepler Magnitude at the center of the 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 the bin. (3) 10th percentile TMCDPP for dwarfs in the bin. (4) Median TMCDPP for dwarfs in the bin. (5) Number of all stars in the bin. (6) 10th percentile TMCDPP of all stars in the bin. (7) Median TMCDPP for all stars in the bin. (8) Simplified noise model CDPP.

Kp mag	No. dwarfs	10th prctile	Median	No. stars	10th prctile	Median	Noise model
9.0	56	9.6	20.2	179	11.3	43.1	3.8
10.0	160	12.0	28.4	568	13.3	54.5	6.0
11.0	611	17.3	28.8	1712	19.1	61.9	9.5
12.0	2169	22.9	34.2	4385	24.4	55.1	15.2
13.0	6768	32.8	44.1	10476	34.2	54.3	24.4
14.0	13868	50.3	65.3	16582	51.1	68.0	40.1
15.0	27462	87.4	114.2	27466	87.4	114.2	68.8
16.0	14324	158.2	203.9	14324	158.2	203.9	127.8

2.2 Summary of Data Anomalies

Certain cadences are flagged to indicate a possible reduction of quality. See the QUALITY and SAP_QUALITY columns of the target pixel and light curve files, respectively. Cadences with data anomalies that affect the entire focal plane are shown in Figure 3. The meaning of the flags are explained in the Data Characteristics Handbook (Christiansen et al., 2012b) and Archive Manual (Thompson & Fraquelli, 2012).

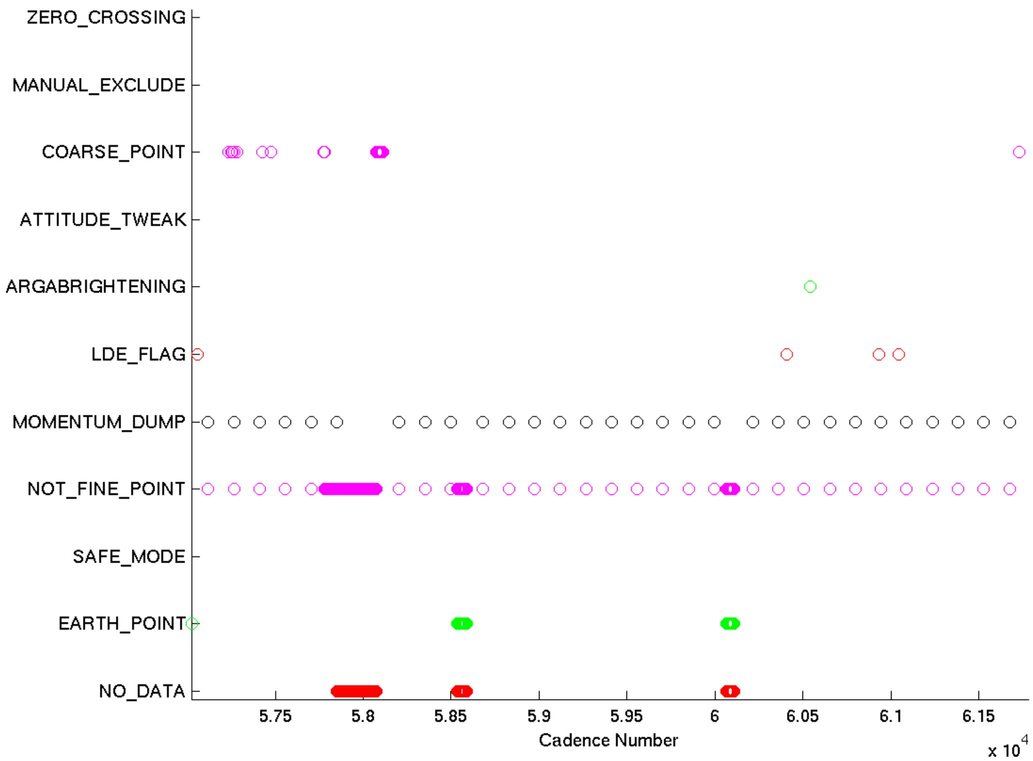

Figure 3: An overview of the location of the data anomalies flagged in Q14. "No_Data " is not an anomaly flag and simply indicates those cadences with no data collected (e.g. during Earth-point or Safe Mode events).

Clarifications on select flags in Figure 3 are listed here:

- ARGABRIGHTENING refers to cadences where the multiple-channel Argabrightening flag (flag 0x07, decimal value 64) was set. The single channel Argabrightening flag (0x0D, decimal value 4096) is not represented on this plot.
- 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 flag 0x03 (decimal value 4) in the FITS files.
- LDE_FLAG refers to flags set by the Spacecraft when a 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.

3 Notable Features of the Q14 Data

In this section we discuss features of the data that occurred during collection or processing that are either new to Q14, significantly different than previous quarters, or not discussed in the Data Characteristics Handbook (Christiansen et al., 2012b). A more complete listing of events that are known to affect the data are discussed in the Data Characteristics Handbook.

3.1 Reaction Wheel Failure

Kepler lost reaction wheel 2 due to excess friction on 2012-07-14 (MJD 56122) and returned to science data collection using three reaction wheels on 2012-07-20 (MJD 56128). The intervening six days of data has 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 appears nominal, so the three- and four-wheel data have been processed and exported as usual (i.e., by month for short-cadence and by quarter for long-cadence).

3.2 Missing Short Cadence Flags

A small number of short cadences were not marked as COARSE_POINT during the reaction wheel failure. These cadences are not suitable for science, and should be removed before analysing short cadence data. The affected short cadences range from 1721709 to 1731878, or MJD 56121.975 to 56128.902.

3.3 Coronal Mass Ejection

The spacecraft was impacted by a small Coronal Mass Ejection on 2012-06-25, 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.

3.4 Error in Barycentric Times

The barycenteric times currently reported in the TIME columns and the headers of all Kepler data products have an error. The times are reported in the UTC (Coordinated Universal Time) system, not in TDB (Barycentric Dynamical Time) as the headers of the files state. As a result the time stamps are incorrect by approximately one minute: the sum of the number of leap seconds and the offset between Atomic Time (TAI) and Terrestrial Time (TT). Except for the addition of one leap second in Q14, the reported times are internally consistent and this error is only apparent when comparing Kepler times to other observations with timing accuracies better than a couple of minutes. However, all Kepler results that report an absolute barycentric time (e.g., the epoch of a planet transit) have this error, including those reported in published papers and in the Kepler archives (i.e., NExScI and MAST).

The following data file types are affected: ffi_cal, ffi_uncert, lpd-targ, spd-targ, llc, slc, and bkg. The Kepler times can be corrected to the TDB system by adding 66.184 seconds to the reported barycentric times for all cadence numbers less than or equal to 57139 in LC (1702663 in SC). For times after this cadence, add 67.184 seconds. This cadence was taken during the first month of Q14 at the time of the most recent leap second, UTC 2012-06-30 23:59:60. This simple additive correction does not account for the relativistic correction between the UTC and TDB systems, which is of order 1.6 ms and significantly less than the 50 ms precision of the Kepler clock.

The Kepler project will correct this error during a future reprocessing effort in mid 2013. Meanwhile, Kepler will continue reporting all epochs and periods in the current time system until the data files at MAST can be uniformly corrected. When the Kepler times have been corrected, users will be alerted via data release notes, the MAST web site, and statements in published papers.

3.5 A Few Corrupted Targets on Module 24.1

Targets whose photometric apertures include pixels that lie in columns 630-634 on mod.out 24.1 (channel 81, skygroup 1) may have an incorrect smear correction for the first few days of Q14, corrupting the data on these targets. From the beginning of the quarter until the loss of reaction wheel 2 on 2012-07-14 (MJD 56122), the bright variable star CH Cyg (KIC 11913210) was very bright, causing these columns to be affected by charge bleeding into the masked smear region on mod.out 24.1. These affected columns were not excluded during calibration, causing targets that fall near the columns occupied by CH Cyg to have an incorrect smear correction. The high 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 occurs several times early in the quarter. Due to the highly-varying and incorrect smear values used, calibrated pixel values, and all calculated time series for targets on these columns are not valid. This problem will be corrected in a future processing of the data.

By the return to science collection on 2012-07-20 (MJD 56128) CH Cyg had dimmed sufficiently that it no longer compromised the smear correction and remained so for the rest of the quarter.

3.6 PDC attenuates long period signals

Users interested in events with durations comparable to the length of a data set (3 months for long cadence, 1 month for short cadence) are advised to not use PDC. All PDC algorithms have difficulty distinguishing between astrophysical and instrumental signals on these timescales. 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 into real data and measuring the attenuating effect of PDC processing. We show the results in Figure 4 below. Signals with periods less $\sim$ 3 days are preserved, while periods greater than 20 days are almost entirely removed. Note that transits are not affected by this feature. Even if the interval between transits is long, from the perspective of PDC their duration is always short.

3.7 Treatment of PDC Short Cadence Data

The PDC module uses a modified version of the PDC-MAP (Presearch Data Conditioning -*Maximum A Posteriori*) algorithm to correct the SC data (Smith et al., 2012; Stumpe et al., 2012). Since PDC is applied to SC on a per-month basis instead of per-quarter (as is done for LC), each month may have a different mean flux level. PDC attempts to preserve the median flux of a data set. Since most Kepler targets have long term systematic trends, the median flux in any month is likely different than the median flux for the quarter. This can cause a step function in the short cadence PDC time series across monthly boundaries; see Figure 5. The offsets can be removed by simply dividing each month of data by the median flux in that month.

3.8 Obsolete PDC Keywords

PDC uses a new algorithm to remove systematic signals from the data, and some of the header keywords are now obsolete. These keywords will be replaced in a future version of the FITS files. In the meantime, the following keywords relating to PDC should be ignored: PDCVAR, the measure of the target variability, and PDCPRWT, the PDC prior weight. When the data is reprocessed and re-exorted, the keywords will be updated to reflect the new PDC algorithm.

3.9 Missing CBV Data for mod.out 6.1

The Q14 CBVs for mod.out 6.1 (channel 13) are incomplete. The basis vectors 9–16 were not reported as part of the data processing and the CBV FITS files contain an array of zeros for these vectors. Users should not use vectors 9–16 to detrend data on mod.out 6.1. Vectors 1 through 8, those most useful for detrending data, are accurately reported in the CBV FITS files. Also, the MJD times reported for mod.out 6.1 are only accurate to 10 seconds. The times reported for the other mod.outs are accurate and may be used instead.

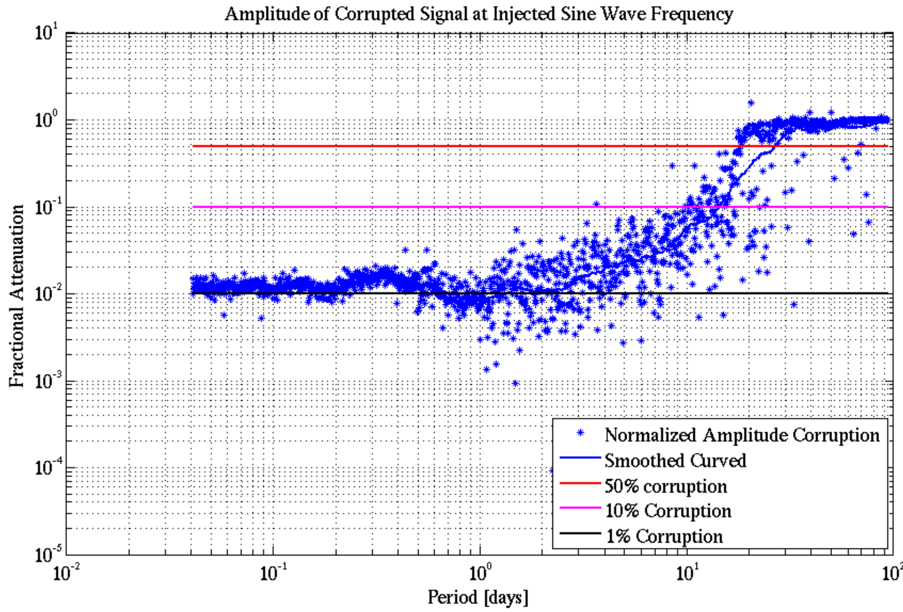

Figure 4: Signal attenuation in msMAP. Sine waves of different periods were injected into long cadence lightcurves before being processed with PDC. Their amplitudes before and after processing were compared. Each blue symbol represents the fractional change in amplitude of a given injected sine wave. Short period signals are attenuated by ∼1%, while long period signals were entirely removed.

References

Christiansen, J. L., Jenkins, J. M., Caldwell, D. A., Burke, C. J., Tenenbaum, P., Seader, S., Thompson, S. E., Barclay, T. S., Clarke, B. D., Li, J., Smith, J. C., Stumpe, M. C., Twicken, J. D., & Van Cleve, J. (2012a). The Derivation, Properties, and Value of Kepler"s Combined Differential Photometric Precision. *PASP*, *124*, 1279–1287.

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

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

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

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

Thompson, S. E., & Fraquelli, D. (2012). Kepler Archive Manual. *KDMC-10008-004*.

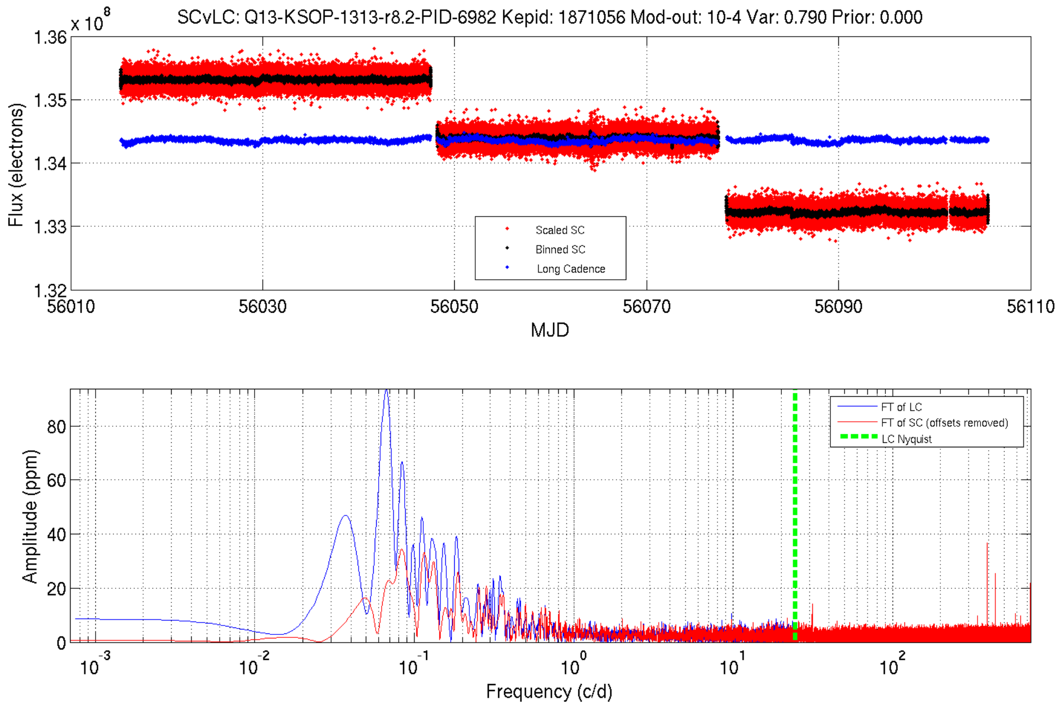

Figure 5: **Top Panel:** Long cadence (blue) and short cadence (red) PDC time series for KIC 1871056 showing different mean fluxes across different months. The short cadence is scaled to show the same average flux per cadence as the long cadence data. The black line shows the short cadence binned to the long cadence exposure time. The sudden jumps in mean flux for different months is an artifact of processing. **Bottom Panel:** Fourier transform of long cadence (blue) and short cadence (red) data. The constant offsets have been removed from the short cadence data. The vertical dashed line shows the LC Nyquist frequency.