Global Erratum for *Kepler* Q0-Q17 & *K2* C0-C5 Short-Cadence Data

KSCI-19080-002

23 March 2016

NASA Ames Research Center
Moffett Field, CA 94035

Prepared by: Douglas A. Caldwell Date 3/23/2016
Douglas Caldwell, Instrument Scientist

Prepared by: Jeffrey E. Van Cleve Date 3/23/16
Jeffrey E. Van Cleve, Support Scientist

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

Document Control

Ownership

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

Control Level

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

Physical Location

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

Distribution Requests

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

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

Michael.R.Haas@nasa.gov

DOCUMENT CHANGE LOG

REVISION	CHANGE DATE	PAGES AFFECTED	CHANGES/NOTES
001	February 4, 2016	All	Original release
002	March 17, 2016	7, 8, 9, 11, 12-15	Fixed minor issues in Sections 2-4, added change log and Section 5
002	March 22, 2016	11-15	Disambiguate “bad” and “good,” including changing .csv file names; reorder Figs 4-7

Table of Contents

1. Summary

An accounting error has scrambled much of the short-cadence collateral smear data used to correct for the effects of *Kepler*'s shutterless readout. This error has been present since launch and affects approximately half of all short-cadence targets observed by *Kepler* and *K2* to date. The resulting calibration errors are present in both the short-cadence target pixel files and the short-cadence light curves for *Kepler* Data Releases 1-24 and *K2* Data Releases 1-7. This error does not affect long-cadence data.

Since it will take some time to correct this error and reprocess all *Kepler* and *K2* data, a list of affected targets is provided. Even though the affected targets are readily identified, the science impact for any particular target may be difficult to assess. Since the smear signal is often small compared to the target signal, the effect is negligible for many targets. However, the smear signal is scene-dependent, so time-varying signals can be introduced into any target by the other stars falling on the same CCD column. Some tips on how to assess the severity of the calibration error are provided below.

2. Severity and Scope

A Guest Observer identified a puzzling difference between the short- and long-cadence calibrated pixel data for a specific *K2* target and contacted the *Kepler* Science Center in November, 2015. The problem was traced to an accounting error for the short-cadence collateral smear data. Under some conditions, these data are passed incorrectly between the spacecraft and the ground, causing smear values to be assigned to the wrong columns within a target's aperture. In these cases, the pipeline's pixel-level calibration routine applies an erroneous smear correction that impacts *Kepler* Data Releases 1-24 and *K2* Data Releases 1-7.

Figure 1 shows the short- and long-cadence target pixel images for the *K2* target that triggered this investigation.

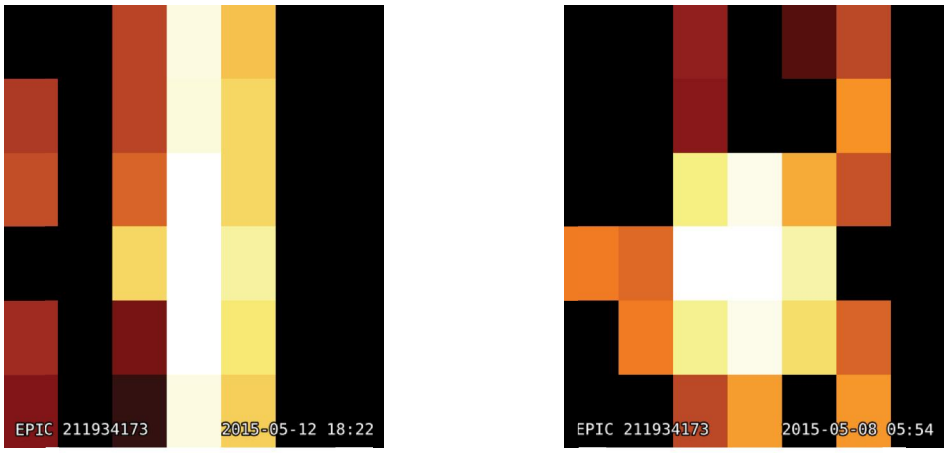


Figure 1: The discovery target (EPIC 211934172, C5) is improperly corrected for smear in the short-cadence image on the left, but properly corrected in the long-cadence image on the right. The bad smear correction results in striping in the short-cadence image, so that unlike the long-cadence image, it no longer appears star-like.

Since the data cannot be reprocessed immediately, the project has identified all affected targets (see Section 4). To assess the impact on affected targets, users should inspect the calibrated target pixels (see Section 2.3.2 of the *Kepler* Archive Manual). An improperly corrected smear signal may show up as an anomalously bright, or dark, column in the calibrated target pixel image (*cf* Figure 1). Comparison of the short-cadence pixel image with the long-cadence pixel image (which is not affected by this problem) can be used to estimate the magnitude of any contaminating signal.

3. Detailed Description

The short-cadence collateral pixel accounting problem results from a misinterpretation as to how the *Kepler* spacecraft processes and stores short-cadence data. For each observed target, the spacecraft stores the target pixel data values at either the short-cadence period (~1 minute), or the long-cadence period (~30 minutes). The spacecraft also stores calibration, or "collateral," pixels at the two cadence periods. At long-cadence, all 1070 rows of the trailing black collateral signal and all 1100 columns of the masked and virtual smear collateral signals are saved for each CCD detector channel (see sections 2.6.3 and 4.5 of the *Kepler* Instrument Handbook for a descriptions of these signals).

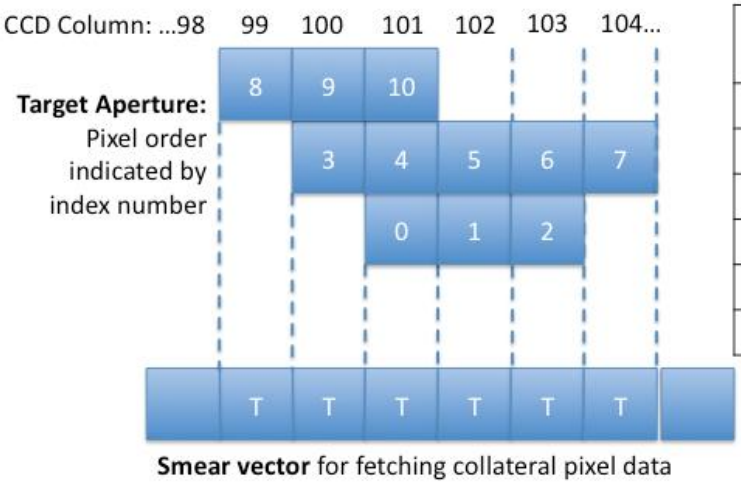
For short-cadence, only the collateral pixels associated with the individual target apertures are stored in order to minimize on-board storage requirements. That is, only trailing black rows overlapping the target aperture rows and smear columns overlapping the target aperture columns are stored (see section 2.6.3.2 and Figure 7 of the *Kepler* Instrument Handbook). Furthermore, the row/column addresses of these collateral pixels are not saved, so their order within the downlinked data file is dependent on the order in which the target pixels are processed on-board the spacecraft. This mapping order is maintained on the ground in a "Pixel Mapping Reference File" (PMRF), which specifies the target id, data type, and row or column offset, for each downlinked data file. There are separate PMRFs for the long-cadence target pixel data, long-cadence collateral pixel data, short-cadence target pixel data, and short-cadence collateral pixel data. Within each PMRF, the order of the data is determined by the order in which the spacecraft processes the pixels.

The processing order is determined by the order of the pixel offsets within the aperture definition associated with the target. In what follows, the "target aperture" refers to the full set of pixels stored by the spacecraft for a given target. For each target, the spacecraft looks up the reference row and column, and then steps through the relative row and column offsets defined in the target aperture table to decide which pixels to retrieve. When all of the offsets for the target aperture have been processed, the spacecraft processes the next target. Short-cadence processing involves an added step: as each offset is processed, the spacecraft flags the row in a trailing black array and the column in a smear array. When it has completed the processing for all offsets in the target aperture, it retrieves the flagged trailing black pixels and the flagged masked and virtual smear pixels *in numerical order* and stores them for later downlink.

The bug was introduced in this last step. The PMRF ordering for these collateral pixels assumed that they were stored in the order given in the target aperture table, but the spacecraft stores them in numerical order within a target definition. The processing pipeline reads in the data and associates each value with the row/column defined in the PMRF. For targets whose rows or columns are not in strict numerical

order within the target aperture table, this effectively scrambles the short-cadence collateral data. Since the spacecraft always returns all rows/columns of collateral data at long-cadence, there is no scrambling of the long-cadence data. In practice, the target apertures have always been specified in increasing row order, so there has been no scrambling of the trailing black collateral data for either *Kepler* or *K2*.

Figure 2 is an example of a target aperture definition that results in scrambled short-cadence collateral smear data. In this sample target aperture, the pixels are processed in index order (0-10). The unique columns are sequentially identified in index order, not numerical order, in the PMRF (*PMRF Column* in the table). However, the spacecraft returns the collateral data in numerical order (*Data Column*), so the data will not be in the order specified by the PMRF.



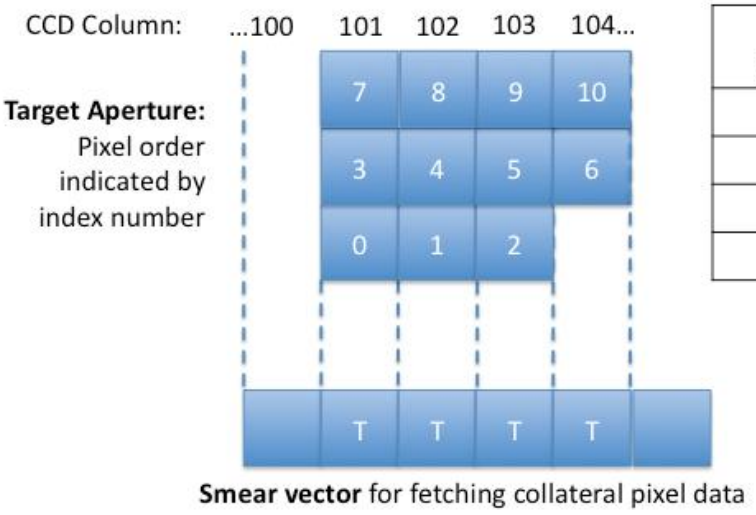
PMRF Column	DATA Column
101	99
102	100
103	101
100	102
104	103
99	104

Figure 2: A sample short-cadence target aperture is shown along with the smear vector used on the spacecraft to determine what collateral data columns to store. Columns in the PMRF are ordered according to the target aperture index order. The spacecraft processes SC target pixels according to target aperture order, but then retrieves and stores collateral data in numerical order based on flags in the smear vector.

Figure 3 is an example of a target aperture that has no scrambled collateral data. In this sample target aperture, the unique rows and the unique columns both occur in numerical order as the pixels are processed in index order (0-10). For apertures where this occurs, the PMRF mapping matches the spacecraft's data storage order.

Finally, target apertures like the "good" one shown in Figure 3, where the PMRF mapping is correct, can still suffer from scrambled smear values if they share columns with a "bad" target aperture like the one shown in Figure 2, where the PMRF mapping is incorrect. This occurs when the "good" aperture is processed before the "bad" aperture because the data processing pipeline only saves the last value for

collateral pixels that are downloaded more than once. The table of affected targets (see Section 4) includes those whose smear values were scrambled due to overlap with another short-cadence target.



PMRF Column	DATA Column
101	101
102	102
103	103
104	104

Figure 3: Sample target aperture where the PMRF correctly maps the short-cadence collateral data stored by the spacecraft. In this case, both the rows and columns appear in increasing order in the target aperture definition, matching the spacecraft storage order.

4. Identification and Recovery

For *Kepler*, there are 13,964 affected target-months out of a total of ~25,000 target-months. A complete listing of all *Kepler* short-cadence targets with scrambled smear is provided in the file *kepler_scrambled_short_cadence_collateral_target_list.csv*. This file includes target-months that span the entire *Kepler* mission (Q0-Q17M2). Each entry in the file contains the *Kepler* ID, the affected quarter and month, and the corresponding target table number. The entries are sorted by *Kepler* ID, so users can quickly identify targets with scrambled smear. Those targets impacted in multiple observing months are listed multiple times. Absent short-cadence targets are guaranteed to be free of this defect. No long-cadence targets are affected.

The situation is similar for *K2*. There are 301 affected short-cadence targets (out of a total of 665) listed in *K2_scrambled_short_cadence_collateral_target_list.csv*. This file contains the EPIC ID, campaign, and target table number for every affected target observation during C0-C5 (*i.e.,* the problem was identified and fixed before the release of the C6 short-cadence data in *K2* Data Release 8).

This problem existed before launch and is present in *Kepler* Data Releases 1-24 and *K2* Data Releases 1-7. Since the error was discovered before the final *Kepler* processing, this problem will be fixed for all short-cadence data collected by *Kepler* in Data Release 25, which is scheduled for release in summer, 2016. For *K2*, the updated short-cadence target pixel files will be released in summer, 2016. The fix for C0-C2 will be delayed until fall, 2016, to align with the planned processing and release of long-cadence photometry for these early campaigns. Fortunately, all required smear data is available to completely recover from this accounting error.

5. Impact on Photometry

We investigated the impact of scrambled smear on pipeline SC light curves, so *Kepler* users can understand the validity of published results based on archival light curves, and *K2* users can estimate the prevalence and magnitude of erroneous signals in their own photometry. We found that *Kepler* users need not be concerned about errors from scrambled smear in archival light curves unless the optimal aperture contains smear from a bright variable star. *K2* users, on the other hand, need to be aware that as many as one in four targets identified as scrambled smear (§4) may show significant differences when the smear data is properly ordered. This is not unexpected, since the much greater image motion in *K2* induces much greater temporal variability in collateral smear data even if the stars contributing to smear are themselves nonvariable. As a result, *K2* users should carefully examine their calibrated SC pixel data versus the corresponding LC pixel data for scrambled smear targets if they wish to proceed before *K2* C0-C5 reprocessing. Our method and conclusions are discussed in more detail below.

Test Data and Method

The *Kepler* pipeline's Photometric Analysis (PA) module generated simple aperture photometry (SAP) SC light curves for 10 channels of Q5M1 using properly-ordered smear data, which we compared to archival DR24 (SOC 9.2) results. We used the unreleased *K2* SC PA to generate SAP light curves for all channels of C6, with one run containing some scrambled smear and a second run using properly-ordered smear. While *K2* is not delivering SC light curves to the archive, we expect that light curves generated by users from calibrated pixels will show qualitatively similar features.

To calculate the significance of the smear scrambling on SAP, we first matched up the light curves of the same KIC or EPIC number, and normalized them by dividing by the median and subtracting 1. We then calculated Δ, the robust 1-99% standard deviation of the difference of these normalized light curves, and the Nyquist (or cadence-to-cadence difference) noise N_{Ny}. N_{Ny} is 1.48 times the median absolute deviation (MAD) of the first difference of the properly-ordered smear light curve; for a zero mean unit-variance Gaussian noise time series, N_{Ny} is sqrt(2). The significance of the change is then defined as $S = \Delta/N_{Ny}$. Histograms of S were generated, and example light curves were plotted to validate the use of S as an impact metric. We found that $S << 1$ differences were invisible to the eye, while $S > 1$ light curve pairs were obviously different. $S = 1$ is a good cutoff for significant impact, as shown in Figure 4.

We also found that properly-ordered smear target light curves changed slightly on channels in which a scrambled smear target was reprocessed with the properly-ordered smear, because of an interpolation fit involving the ordered set of all short-cadence collateral data for a channel; in all these cases, $S << 1$. For properly-ordered

smear targets on channels without scrambled smear targets, the light curves were identical in the two pipeline runs, as expected.

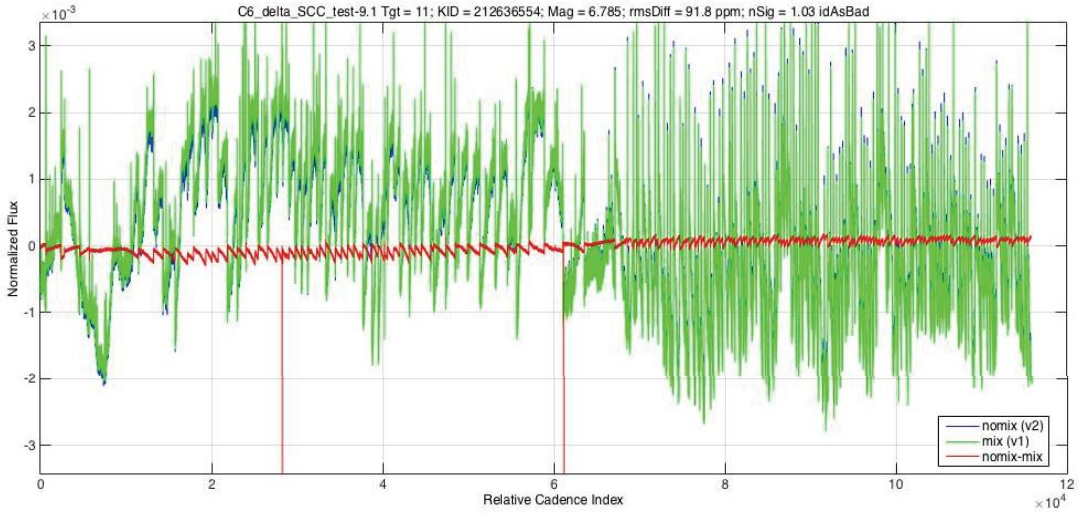


Figure 4: Marginally significant (S ~ 1) light curve difference in K2 C6 data. While the difference is detectable, it is much smaller than the signal amplitude. Green = scrambled smear light curve, blue = properly-ordered smear light curve, red = difference.

Impact on *Kepler* Photometry

For scrambled smear targets, $S < 0.3$ for all targets examined. Figure 5 compares the light curves of the largest S found. However, given the small number of channels examined, we may have missed some cases in which the optimal aperture contains scrambled smear from a variable star resulting in larger values of the light curve difference significance, S.

For properly-ordered smear targets, differences are completely negligible, with $S < 0.025$ in all cases.

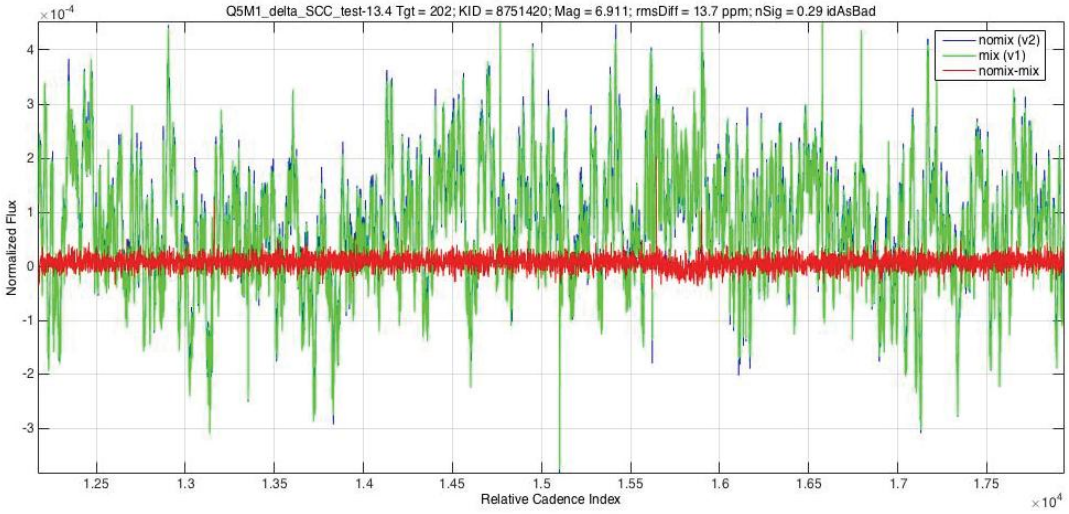


Figure 5: Most significant light curve difference (S = 0.3) found in *Kepler* Q5M1 data for a scrambled smear target. Green = scrambled smear SAP light curve, blue = reprocessed properly-ordered smear SAP light curve, and red = difference.

Impact on *K2* Photometry

For scrambled smear targets in *K2* C6, *S* can be as large as 8, as shown in Figure 6**.** The difference is quite obvious to the eye (Figure 7), so valid data for this target requires the properly-ordered smear.

For properly-ordered smear targets, differences are negligible as was found for *Kepler*. $S < 0.015$ with the exception of an RR Lyrae star with factor of 2 brightness variation, for which $S = 0.06$. Even in this case the light curve difference was invisible on the scale of the stellar signal.

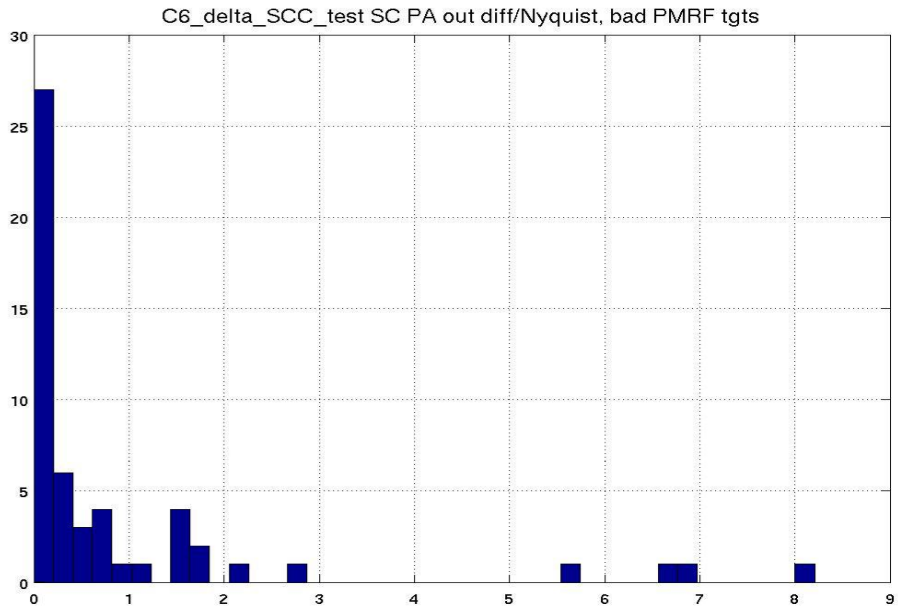


Figure 6: Histogram of significance of the difference (S) between light curves of the same target in two different pipeline runs, one run with scrambled smear and the other run with properly-ordered smear. For *K2* C6, 13 of 54 scrambled smear targets had S > 1, out of a total of 84 targets.

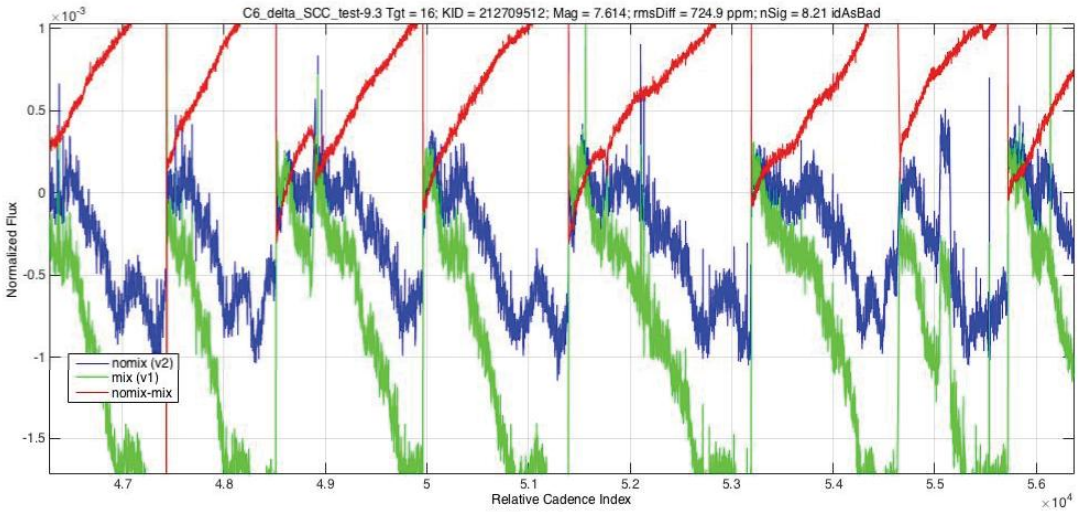


Figure 7: Most significant difference between light curves of the same target in two different pipeline runs for *K2* C6, one with scrambled smear and the other with properly-ordered smear. Green = scrambled smear light curve, blue = properly-ordered smear light curve, and red = difference.