

# Photometric analysis in the *Kepler* Science Operations Center pipeline

Joseph D. Twicken<sup>\*a</sup>, Bruce D. Clarke<sup>a</sup>, Stephen T. Bryson<sup>b</sup>, Peter Tenenbaum<sup>a</sup>, Hayley Wu<sup>a</sup>,  
Jon M. Jenkins<sup>a</sup>, Forrest Girouard<sup>c</sup>, Todd C. Klaus<sup>c</sup>

<sup>a</sup>SETI Institute/NASA Ames Research Center, MS 244-30, Moffett Field, CA, USA 94035-1000;

<sup>b</sup>NASA Ames Research Center, MS 244-30, Moffett Field, CA, USA 94035-1000;

<sup>c</sup>Orbital Sciences Corporation/NASA Ames Research Center, MS 244-30, Moffett Field, CA,  
USA 94035-1000

## ABSTRACT

We describe the Photometric Analysis (PA) software component and its context in the *Kepler* Science Operations Center (SOC) Science Processing Pipeline. The primary tasks of this module are to compute the photometric flux and photocenters (centroids) for over 160,000 long cadence (~thirty minute) and 512 short cadence (~one minute) stellar targets from the calibrated pixels in their respective apertures. We discuss science algorithms for long and short cadence PA: cosmic ray cleaning; background estimation and removal; aperture photometry; and flux-weighted centroiding. We discuss the end-to-end propagation of uncertainties for the science algorithms. Finally, we present examples of photometric apertures, raw flux light curves, and centroid time series from *Kepler* flight data. PA light curves, centroid time series, and barycentric timestamp corrections are exported to the Multi-mission Archive at Space Telescope [Science Institute] (MAST) and are made available to the general public in accordance with the NASA/Kepler data release policy.

**Keywords:** Kepler, transit photometry, light curve, raw flux, photocenter, centroid.

## 1. INTRODUCTION

The *Kepler Mission* is designed to detect (habitable) Earth-size planets transiting Sun-like stars<sup>1</sup>. The spacecraft was launched on 6 March 2009 into an Earth-trailing heliocentric orbit with a period of 373 days. The pointing of the *Kepler* photometer is maintained to support imaging of the same star field continuously for the life of the mission (nominally 3.5 years for the primary mission). The field of view of the *Kepler* photometer is ~115 square degrees. Incident light is collected by 42 charge-coupled device (CCD) detectors (94.6 million total pixels) on the focal plane assembly. There are two readout channels (or module outputs) per CCD, for a total of 84 on the focal plane. Short exposures are integrated onboard to produce one image every 29.4 minutes for over 160,000 long cadence<sup>2</sup> (LC) targets, and one image every 0.98 minutes for 512 short cadence<sup>3</sup> (SC) targets. The spacecraft is rolled by 90 degrees on a quarterly basis so that the solar panels are continuously directed toward the Sun.

Science data acquired in-flight are processed in the *Kepler* Science Operations Center (SOC) Science Processing Pipeline<sup>4,5</sup>. Pixel values are calibrated for each cadence in the Calibration (CAL) software component<sup>6</sup>. Raw flux light curves are extracted and target photocenters (centroids) are computed in the Photometric Analysis (PA) component described in this paper. Systematic and other errors are corrected in the Presearch Data Conditioning (PDC) component<sup>7</sup>. Long cadence corrected flux light curves are then subjected to a search for transiting planets. The Transiting Planet Search (TPS) pipeline module<sup>8</sup> returns a Threshold Crossing Event (TCE) for each target and trial transit pulse that exceeds the specified detection threshold. A transiting planet model is fitted to the corrected flux light curves in the Data Validation (DV) pipeline module<sup>9,10</sup> for targets with TCEs. A transit signature obtained from the fitted parameters is subsequently removed from the corrected flux time series for each candidate planet, and a search is conducted for additional candidate planets. A suite of automated tests is performed when no additional candidate planets can be identified. The purpose of the automated tests is to facilitate identification of the true candidate planets from the large number of false positive transiting planet detections (astrophysical and otherwise).

The first primary task of PA is to extract raw flux light curves from calibrated pixels in the apertures of the respective long and short cadence targets. Prior to computation of the photometric flux for each target, so-called Argabrightening

\*joseph.twicken@nasa.gov; kepler.nasa.gov

events are (optionally) mitigated, cosmic rays are (optionally) removed, and a background estimate is subtracted from the pixels in the target apertures. The second primary task of PA is to estimate CCD coordinates of the aperture photocenter (centroid) for each target and cadence from background-removed pixels associated with the respective targets. Other tasks of PA are to compute barycentric timestamp corrections per target and cadence based on a spacecraft trajectory reconstruction, and to compute metrics (brightness and encircled energy) for monitoring instrument performance and, potentially, to support systematic error correction in the PDC component.

Raw flux light curves, centroid time series, and barycentric timestamp corrections are exported to the Multi-mission Archive at Space Telescope [Science Institute] (MAST) and are made available to the general public in accordance with the NASA/Kepler data release policy. Corrected flux light curves generated in PDC are also exported to the MAST. These account for systematic errors due to data and spacecraft anomalies, and also for excess flux in the target apertures due to background sources.

An overview of PA and the flow of data through the pipeline module are presented in Section 2. PA science algorithms are described in Section 3; cosmic ray cleaning is discussed in Section 3.1; background estimation and removal are discussed in Section 3.2; aperture photometry is discussed in Section 3.3; and flux-weighted centroiding is discussed in Section 3.4. A summary and conclusions are presented in Section 4.

## 2. PHOTOMETRIC ANALYSIS OVERVIEW AND DATA FLOW

The primary tasks of the PA pipeline module are to compute raw flux light curves and centroids of target stars from calibrated pixels in their respective apertures. The standard PA unit of work<sup>11,12</sup> for LC and SC science data processing is a single module output for a duration of one quarter (LC) or one month (SC).

PA science data includes calibrated background and target pixels. Due to the large volume of calibrated pixel data, PA is typically executed in multiple invocations<sup>12</sup>. For LC units of work, all of the background pixel data is processed in the first PA invocation. Background pixels are utilized to estimate the background level to be subtracted from each target pixel on each cadence. Target pixels are then processed in as many subsequent invocations as required. Calibrated pixel data for targets with PPA\_STELLAR labels are processed first in PA (although they may span multiple PA invocations). After all PPA\_STELLAR pixels have been processed, then calibrated pixels for targets without PPA\_STELLAR labels are processed. Targets with the PPA\_STELLAR label represent gold standards for motion polynomial fitting, centroid seeding, and computation of PA metrics. Approximately 200 of these are pre-selected per module output<sup>4</sup>. There are no background pixels in SC units of work, so the processing of SC science data is restricted to target pixels only.

Data flow in the PA pipeline module is shown in Figure 1. Data generally flows from left to right and top to bottom in the figure. Functional blocks in the diagram will be referenced with bold type where they are discussed in the text, and section numbers will be included for the algorithms that are described in detail later in this paper.

**Mitigate Argabrightening Events.** It has been observed in flight data that there are cadences for which the flux values of all background and target pixels are elevated<sup>4</sup>. These have been dubbed *Argabrightening* events after Vic Argabright, the Ball Aerospace engineer who first observed and reported them. There has been speculation about the source of the Argabrightening events, but the true source is unknown. If Argabrightening mitigation is enabled (by PDC module parameter<sup>11</sup>), then Argabrightening detection statistics are formulated for each cadence and a threshold is applied to identify the Argabrightening cadences. All background and target pixel time series are subsequently gapped for Argabrightening cadences in each of the PA invocations in the unit of work. Argabrightening statistics are formulated from the background pixels for LC units of work and from target pixels outside of the optimal apertures of the respective targets in the first invocation for SC units of work.

**Clean Cosmic Rays (3.1).** If cosmic ray cleaning is enabled, cosmic rays are identified and removed from the calibrated background (LC only) and target pixels. Cosmic ray metrics are computed separately from the background and target cosmic ray event lists. Background cosmic ray events and metrics are written to the pipeline Data Store after the background pixels have been processed in the first invocation of LC units of work. Target cosmic ray events and metrics are written to the Data Store after the last target invocation in all PA units of work.

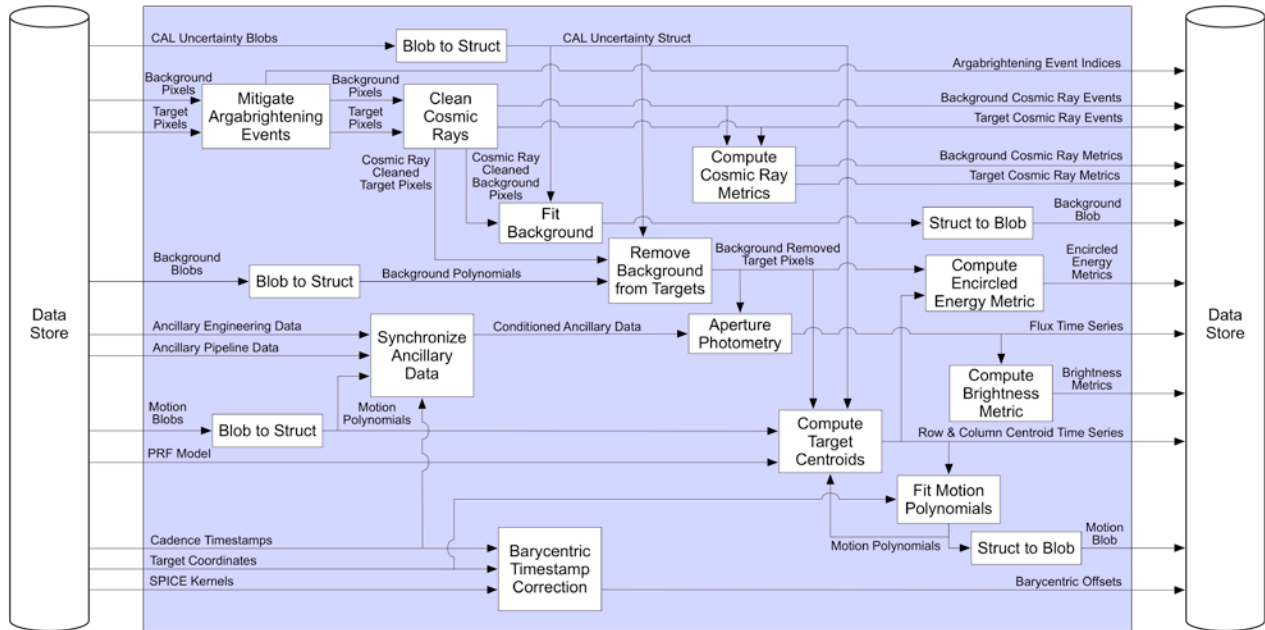


Figure 1: Data flow diagram for the Photometric Analysis (PA) pipeline module. Inputs are shown at the left and outputs are shown at the right. Inputs are obtained from the Data Store and outputs are written to the Data Store.

**Fit Background and Remove Background from Targets (3.2).** For LC units of work, a two-dimensional (background) polynomial is fitted to the cosmic ray-cleaned background pixels on a cadence by cadence basis. The background is removed from cosmic ray-cleaned target pixels by evaluating the background polynomial at the coordinates of the given target pixels and subtracting the resulting background estimates from the target pixel values on a cadence by cadence basis. For SC units of work, background polynomials must be normalized and interpolated at the short cadence timestamps before they can be evaluated to estimate the background.

**Aperture Photometry (3.3).** Flux time series are obtained by performing aperture photometry on the background-removed target pixels. In Simple Aperture Photometry (SAP), pixel values are summed for all pixels in the optimal aperture of a given target. SAP is the only photometry method currently supported by PA (SOC release 6.1).

**Compute Target Centroids (3.4).** Row and column centroid time series are computed from the background-removed target pixels on a cadence by cadence basis. Flux-weighted centroids are computed for all PA targets. Centroids based on fitting a predetermined Pixel Response Function<sup>13</sup> (PRF) to calibrated target pixel values in the respective target apertures may also be computed for PPA\_STELLAR or other targets. PRF-based centroiding, however, will not be discussed further in this paper.

**Fit Motion Polynomials.** Two-dimensional (motion) polynomials are separately fitted to the collection of row and column centroids of the PPA\_STELLAR targets (if a sufficient number of such targets are available) as a function of target right ascension and declination. The motion polynomials are then evaluated at the celestial coordinates of the non-PPA\_STELLAR targets in order to seed the PRF-based centroid computation if PRF-based centroiding is enabled for such targets. The motion polynomials are written to the Data Store. In short cadence units of work, motion polynomials are obtained for centroid seeding by interpolation of long cadence motion polynomials provided to PA from the Data Store. The motion polynomials are computed cadence by cadence. In addition to providing seeds for centroids, motion polynomials are utilized elsewhere in the pipeline for focal plane geometry fitting<sup>14</sup>, systematic error correction<sup>7</sup>, attitude determination<sup>15</sup>, and computation of instrument performance metrics<sup>15</sup>.

**Compute Encircled Energy Metric and Compute Brightness Metric.** The encircled energy metric is computed for each cadence from the background-removed target pixels and centroids of the PPA\_STELLAR targets. This metric is a robust measure of the average radius required to capture a specified fraction of flux from the respective targets. The brightness metric is computed for each cadence from the raw flux light curves for the PPA\_STELLAR targets. This

metric is a robust measure of the average ratio of observed-to-estimated flux for the given targets. The raw flux light curves, row and column centroids, and encircled energy and brightness metrics are all written to the Data Store.

Uncertainties in calibrated background and target pixels are propagated by standard methods to uncertainties in the raw flux, centroids, metrics, and two-dimensional polynomials generated in PA. Uncertainties propagated in PA are also written to the Data Store. If rigorous Propagation of Uncertainties (POU) is enabled, uncertainties are propagated from covariance matrices for calibrated background and target pixels<sup>16</sup>. Otherwise, uncertainties in PA products are propagated only from uncertainties in calibrated background and target pixels under the assumption of statistical independence.

**Barycentric Timestamp Correction.** Barycentric timestamp corrections are computed for each target and cadence given the celestial coordinates of the target and a reconstructed spacecraft trajectory. The offset for each mid-cadence timestamp is computed such that the sum of timestamp plus offset yields the time that the flux would have been captured at the solar system barycenter. The barycentric offsets allow investigators to account for modulation of the timing of observed transits due to the heliocentric orbit of the photometer. Barycentric corrections produced in PA also compensate for small time slice offsets introduced in the multiplexed readout of the focal plane array.

### 3. PHOTOMETRIC ANALYSIS SCIENCE ALGORITHMS

PA science algorithms are discussed in this section with flow charts and illustrations where appropriate.

#### 3.1 Cosmic ray cleaning

Cosmic rays have been problematic for all space flight missions with CCD detectors. The expected cosmic ray flux rate<sup>2</sup> for the *Kepler* CCD array is  $5 \text{ cm}^{-2} \text{ sec}^{-1}$ . Each  $27 \text{ }\mu\text{m}$  pixel is therefore expected to receive a direct hit approximately three times per day. Energy may be deposited over a range of pixels for each hit, however, depending on the angle of incidence.

In the PA pipeline module, cosmic rays are cleaned (identified and subtracted) in both background and target pixels after Argabrightening events have been mitigated. Cosmic ray identification in the relatively low flux background pixels is more effective than in target pixels where deposited energy is often a small fraction of the total flux. Most cosmic rays are not expected to be detected, but the desire is to detect the cosmic rays with significant charge with a low false detection rate.

A flow chart describing the cosmic ray cleaning algorithm is shown in Figure 2. The cosmic ray identification process for PA background and target pixels is a general pixel time series outlier identification algorithm. To that end, both positive and negative outliers are identified. The algorithm therefore goes beyond identification of cosmic rays only, as these are restricted to deposits of positive charge. A parameter has been included in the module interface<sup>11,12</sup>, however, to define a threshold multiplier for negative events. This multiplier may be specified to be arbitrarily large; in that event, the algorithm is insensitive to negative flux outliers.

A two-dimensional array of all background or target pixel time series in a given PA invocation is first detrended along the time dimension with a low order polynomial. The detrended time series for the respective pixels are smoothed with a short median filter (relative to typical astrophysical events and stellar variations). The residuals (cosmic ray deltas) are then computed between the detrended time series and the time series obtained by median filtering each (detrended) pixel time series. A sliding Median Absolute Deviation (MAD) is computed on the residuals for each pixel time series to obtain a time-varying estimate of the residual scatter, and detection statistics are formulated by computing the ratio of the residuals to the MAD for each pixel and cadence. Outliers are identified by applying a threshold to the absolute value of the detection statistics.

Once the residual values that exceed the detection threshold have been identified, the number of outlier counts per pixel time series is determined. The pixels for which there are an excessive number of counts are identified and reprocessed as before with a longer duration median filter. The motivation for doing so is to prevent out-of-family cosmic ray identification based on the nature and variability of specific targets. Implicit is the assumption that cosmic rays or other outliers should be distributed uniformly among all background or target pixels.

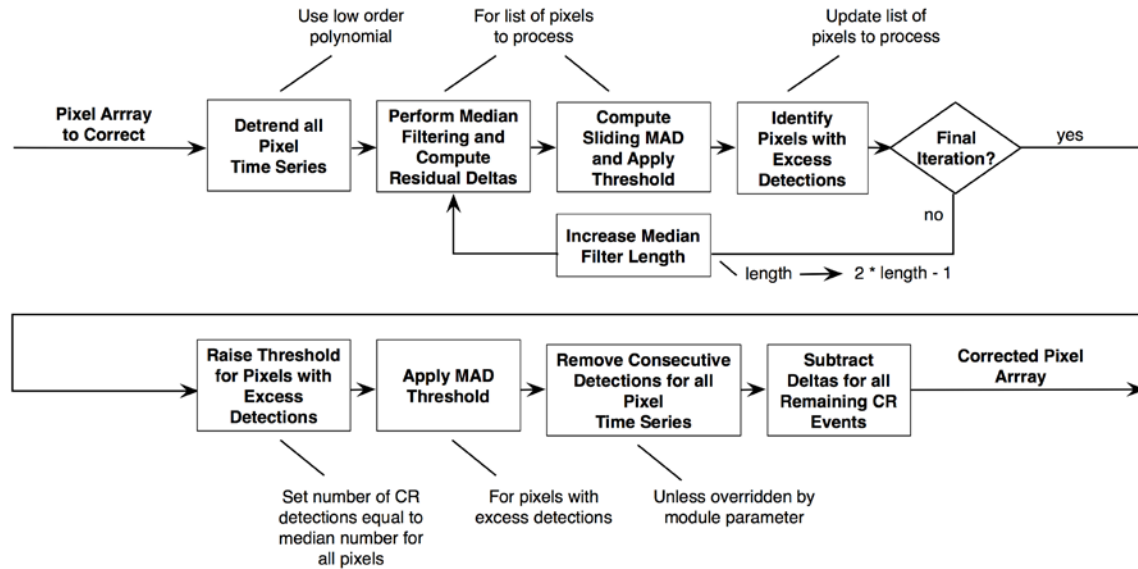


Figure 2: Flow chart for the cosmic ray cleaning algorithm. The loop in the upper path is executed twice in the current PA release (6.1).

After the process has repeated, the detection threshold is raised for those pixels that continue to have an excessive number of hits (in relation to the ensemble of background or target pixels in the given invocation). The number of cosmic rays identified for each of these pixels is set equal to the median number across the ensemble. Once outliers have been identified for all pixels and cadences, temporally consecutive events are invalidated as cosmic ray detections. Consecutive cosmic ray hits in a single pixel time series are extremely unlikely and almost certainly indicative of false detections. Finally, the residuals (cosmic ray deltas) after median filtering are subtracted from the original pixel array for the valid remaining cosmic ray events to obtain an array of cosmic ray-cleaned pixel values.

Cosmic ray detection statistics for a Q3 target<sup>17</sup> pixel are shown in Figure 3 along with the pixel time series before and after cosmic ray cleaning. Cosmic ray events above the specified 12 MAD detection threshold are circled in the original time series. The difference between the original and corrected pixel value for each cadence with an identified cosmic ray is equal to the residual (or delta) after median filtering. For each cosmic ray event, the mid-cadence timestamp is written to the Data Store along with the row and column coordinates of the pixel and the cosmic ray delta. Duplicate detections are first removed from the cosmic ray events lists. These may result if a pixel falls in multiple target apertures.

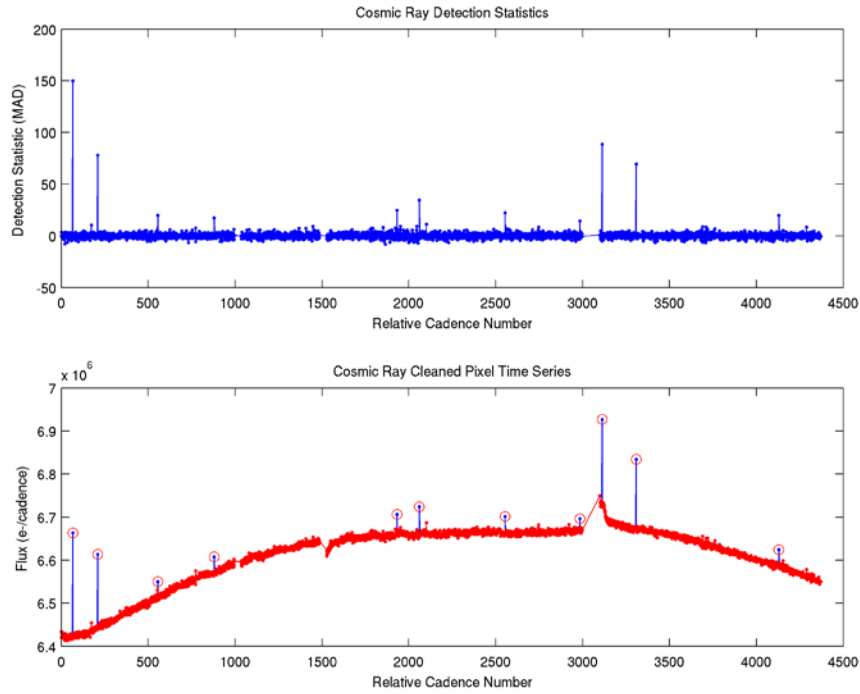


Figure 3: Cosmic ray detection statistics for a single Q3 pixel time series are shown in the upper panel. The target pixel time series before and after cleaning are shown in the lower panel. Cosmic ray events in the original pixel time series are circled. Flux transients following the monthly downlinks near cadences 1500 and 3000 are thermally induced.

### 3.2 Background estimation and removal

Background pixels are acquired at the long cadence rate in a grid pattern on each of the focal plane module outputs. 4,464 background pixels are defined on each channel. The background pixel time series represent spatial and temporal samples of the global background level and an attempt is made to prevent them from being corrupted by flux from neighboring stars or from saturated targets on the same columns. Nevertheless, there are background pixels in the flight data that exhibit high flux levels due to nearby or saturated targets. These pixels necessitate the development of a robust (against outliers) method for background estimation and subtraction.

Background estimation is performed in two steps. The process begins by fitting a two-dimensional (background) polynomial to the calibrated, cosmic ray-corrected background pixel values as a function of the CCD row and column coordinates for each cadence. This occurs on the first PA invocation in each long cadence unit of work. In all subsequent target invocations, the background is estimated by evaluating the background polynomial for each cadence at the CCD coordinates of the respective target pixels. This produces a spatially smooth estimate of the local background for each target without dedicating any background pixels to specific targets. In fact, the ratio of background pixels to targets is only on the order of 2:1 across the entire focal plane.

Once the background is estimated for each target pixel and cadence, the background is removed by subtracting the estimated value from the calibrated, cosmic ray-corrected target pixel value. Uncertainties in the background-removed target pixels are propagated from uncertainties in the background and target pixels by standard methods. The covariance matrix for the background polynomial is determined from uncertainties in the background pixels for each long cadence. The background polynomial covariance matrix is then propagated through the background estimation and subtraction processes. Uncertainties may be propagated in an approximate (but computationally inexpensive) fashion, or in a rigorous treatment (that utilizes, as a starting point, full covariance matrices<sup>16</sup> for the background pixels and target pixels in each stellar aperture) depending on the logical value of a PA module parameter.

Background removal for short cadence targets is complicated by the fact that there are no short cadence background pixels. In short cadence units of work, the long cadence background polynomials are provided as input to PA on the first target invocation. The background polynomials are then normalized and interpolated in time at the midpoints of the short cadence intervals. The background estimation and removal process proceeds with the interpolated background polynomials as before. It should be noted that changes in the background that occur on time scales shorter than the long cadence rate cannot be captured in the short cadence background estimation process.

The order for the two-dimensional polynomial that is fit to the background pixels for each long cadence is determined by default in the pipeline with the Akaike Information Criterion<sup>18</sup> (AIC). This criterion includes a penalty that increases with fit order and seeks to optimize the trade-off between fit order and goodness of fit. In PA, the order that is used for the background fit for all cadences is determined by the order (up to a specified maximum) that minimizes average AIC over all cadences in the unit of work. For a background fit of order  $K$ , the number of background polynomial coefficients for each cadence is  $(K+1) * (K+2) / 2$ .

For the set of calibrated, cosmic ray-corrected background pixels  $\mathbf{B}$ , let  $p_B$  and  $\sigma_B$  designate the pixel values and uncertainties for any given cadence such that  $p_{B,n}$  and  $\sigma_{B,n}$  represent the value and uncertainty of the  $n^{\text{th}}$  background pixel. Furthermore, let the (one-based) CCD row and column coordinates of the background pixels be designated by  $i_B$  and  $j_B$  respectively. The background polynomial  $x_{BP}$  is then determined for order  $K_B$  by minimizing the weighted  $\chi^2$  defined in equations (1) and (2) for the given cadence:

$$\chi_B^2 = \sum_{n \in B} \left[ \left( p_{B,n} - \sum_{k=0}^{K_B} \sum_{l=0}^k x_{BP,m} i_{B,n}^{k-l} j_{B,n}^l \right) / \sigma_{B,n} \right]^2 \quad (1)$$

where the background polynomial coefficient index is given by

$$m = k(k+1)/2 + l \quad (2)$$

For the set of pixels  $\mathbf{P}$  in a specified target aperture, the background level may then be estimated for the given cadence from the background polynomial  $x_{BP}$  and subtracted from the calibrated, cosmic ray-corrected target pixels  $p_T$  to obtain the background-removed target pixel values  $p_S$  as follows:

$$p_S = p_T - \hat{p}_B \quad (3)$$

where the CCD row and column coordinates of the target pixels are designated  $i_T$  and  $j_T$  respectively, and the background estimate for the  $n^{\text{th}}$  target pixel in  $\mathbf{P}$  is given by

$$\hat{p}_{B,n} = \sum_{k=0}^{K_B} \sum_{l=0}^k x_{BP,m} i_{T,n}^{k-l} j_{T,n}^l \quad (4)$$

The background polynomial coefficient index  $m$  is defined as in equation (2).

The weighted least squares problem defined in equation (1) is solved in an iterative, robust manner in order to deemphasize outlier background pixel values that would otherwise perturb the two-dimensional fit. The problem may still be posed for each cadence as one of  $Ax = b$ . The least squares solution is the familiar  $x = Tb$ , where  $T = (A'A)^{-1}A'$ . If the rigorous propagation of uncertainties is enabled in PA, the covariance matrix  $C_{BP}$  for the background polynomial  $x_{BP}$  is determined from the covariance matrix  $C_B$  of the background pixels<sup>16</sup> and the background polynomial transformation matrix  $T_{BP}$  for the given cadence by

$$C_{BP} = T_{BP} C_B T_{BP}' \quad (5)$$

Uncertainties in the background polynomial coefficients are then propagated to the background-removed pixels in the target aperture for the given cadence by

$$C_S = C_T + \hat{A}_T C_{BP} \hat{A}_T' \quad (6)$$

where  $C_T$  is the covariance matrix for the cosmic ray-corrected target pixels<sup>16</sup>, and the linear transformation to evaluate the background polynomial in the target aperture (4) is given by

$$p_B^\wedge = \hat{A}_T x_{BP} \quad (7)$$

### 3.3 Aperture photometry

It is not possible to save and later downlink all of the pixel values acquired on the focal plane array. The storage and bandwidth required to support that are excessive. Rather, an aperture is defined for each target that specifies the pixels necessary to support pixel-level calibrations, light curve extraction, and computation of the target photocenter. Only pixels in the specified target apertures (plus the background and collateral<sup>6</sup> pixels for black level and smear corrections in CAL) are written to the Solid State Recorder aboard the spacecraft and downlinked for pipeline processing. Great care is taken in generation of the aperture definitions<sup>19</sup> to ensure that all of the pixels required to support the *Kepler* science mission are captured. Pipeline software may be enhanced over time, but the science pixels may only be acquired once.

Within each target aperture, the subset of pixels required for photometric extraction of light curves is referred to as the *optimal aperture*. The size of the optimal aperture for any given target depends on a number of factors, including target magnitude, PRF, noise level, local crowding and Differential Velocity Aberration (DVA). The apertures and associated optimal apertures are redefined for each observing season as the targets move from one CCD to another with each quarterly roll of the photometer. Images illustrating the mean flux level for pixels (after background subtraction) in two sample target apertures are shown in Figure 4. The pixels in the optimal aperture for each target are marked with circles. The *Kepler* magnitude ( $Kp$ ) is 13 for the (PPA\_STELLAR) target on the left, and 10 for the saturated target on the right. Charge bleeds along the column where the flux concentration is highest for the saturated target. The optimal aperture is stretched vertically to permit flux to be captured in one saturated column or another as the target moves due to DVA.

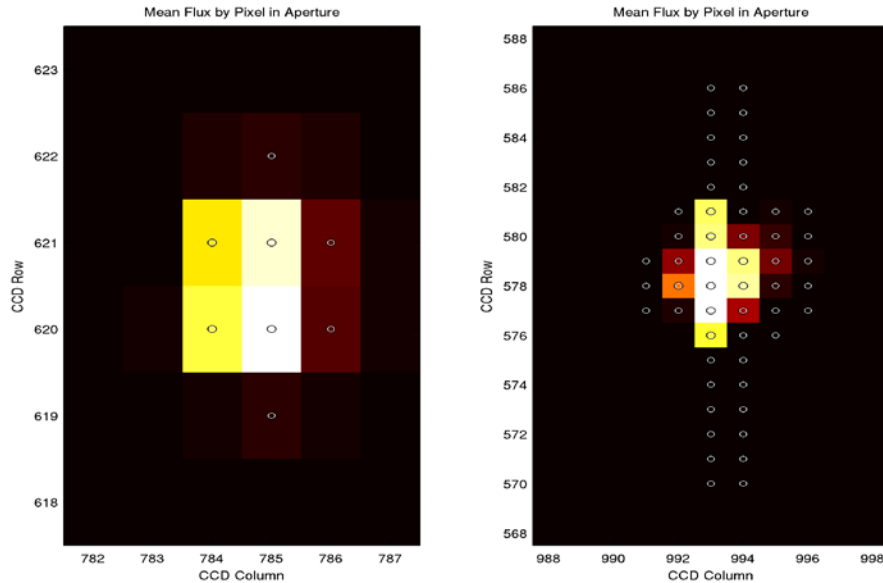


Figure 4: Mean flux levels for pixels in two target apertures. The PPA\_STELLAR target on the left is 13<sup>th</sup> magnitude. The saturated target on the right is 10<sup>th</sup> magnitude. Pixels in the optimal aperture for each target are indicated with circles. The color map range for the saturated target is nearly ten times larger than that of the PPA\_STELLAR target.

The optimal aperture does not necessarily contain all of the stellar flux for a given target. The *flux fraction* in the aperture refers to the ratio of target flux contained in the optimal aperture to the total flux of the target. Furthermore, not all flux in the optimal aperture is due to the primary target. The *crowding metric* refers to the fraction of flux in the optimal



aperture that is due to the target. The flux fractions and crowding metrics are computed for each target table when the apertures are defined<sup>19</sup> and are provided to the pipeline modules that require them. Excess flux due to crowding in the optimal apertures is removed<sup>7</sup> when the light curves are corrected in PDC.

Light curves are computed in the current release (6.1) of PA by a method referred to as Simple Aperture Photometry (SAP). It is possible that additional photometric methods will be supported in future releases. Once the calibrated target pixels have had cosmic rays corrected and background removed, the raw flux is obtained per target and cadence by the unweighted summation of pixels in the associated optimal aperture. For the set of target pixels  $\mathbf{O}$  in the optimal aperture of a specified target, the raw flux  $f_R$  is computed by SAP from the background removed pixels  $p_S$  in the target aperture for a given cadence by

$$f_R = \sum_{n \in \mathbf{O}} p_{S,n} = t_{SAP} p_S \quad (8)$$

where the  $n^{\text{th}}$  element of the row vector  $t_{SAP}$  is equal to one if the  $n^{\text{th}}$  pixel is in the optimal aperture  $\mathbf{O}$ , or zero otherwise.

It should be noted that if a data gap exists for any pixel and cadence in the optimal aperture of a given target, then a data gap is set for the raw flux value (and associated uncertainty) for that target and cadence. Discontinuities cannot be artificially introduced into the light curves by extracting the raw flux from only a subset of the pixels in the optimal aperture. As this is written, it has not been observed in-flight that data availability is pixel dependent. Rather, it has always been true that either all or none of the pixels are valid for a given cadence.

Some representative light curves for long cadence targets are shown in Figure 5. Raw flux is plotted versus cadence number for all cadences in the unit of work (Q3 for module output 7.3). A quiet target is shown in the upper panel, a variable target in the middle panel, and an eclipsing binary is shown at the bottom. Multiple cadence data gaps are due to loss of fine point and two monthly downlinks. Single cadence gaps due to momentum desaturations and mitigated Argabrightening events are also present (but not visible) in the light curves.

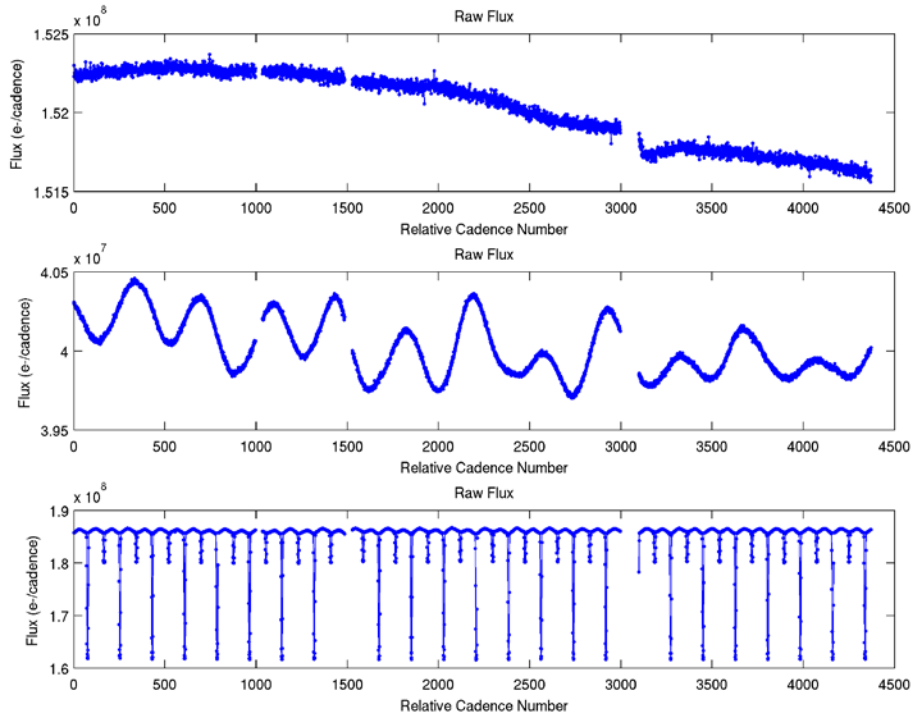


Figure 5: Raw flux light curves for representative targets computed by SAP in Q3 for module output 7.3. A quiet PPA\_STELLAR target is shown in the upper panel, a variable target in the middle panel, and an eclipsing binary in the lower panel.

If the rigorous propagation of uncertainties is enabled, the variance  $v_F$  for the raw flux value computed by SAP in equation (8) is determined from the covariance matrix  $C_S$  for all pixels in the target aperture on the given cadence by

$$v_F = t_{SAP} C_S t_{SAP}^T \quad (9)$$

### 3.4 Centroiding

There is substantial motion of the target positions on the focal plane in the flight science data. The dominant source of long term target motion is DVA, which causes the targets to trace small but significant elliptical paths across the respective CCD detectors over the period of the heliocentric orbit of the photometer. The maximum motion due to DVA is 0.6 pixels per observing quarter<sup>2</sup>. There is also target movement due to pointing jitter, pointing drift, focus changes (that appear to result from temperature changes in the photometer), and commanded attitude adjustments to compensate for pointing drift. It should be noted that the photocenters of variable targets also move in crowded apertures.

Flux levels vary with target motion at even the sub-pixel level. Systematic effects that result in target motion therefore produce correlated signatures in the associated light curves. Precise computation of the target locations on each cadence is critical for correcting systematic errors later in the pipeline. It is also required for precise a posteriori definition of optimal target apertures, for precision reconstruction of spacecraft attitude, and for monitoring instrument performance. The photocenter of each target in its aperture is referred to as the target *centroid*. It should also be noted that analysis of centroid motion<sup>9</sup> is a critical tool for distinguishing between legitimate transits due to orbiting planets and apparent transits due to background eclipsing binaries.

Two centroiding methods are employed in PA. Flux-weighted centroids are computed for every target and cadence and are currently exported to the MAST. These centroids are computationally inexpensive and essentially determine the photometric center of mass in the target aperture. PRF-based centroids may additionally be computed for some or all targets in the pipeline. This method will not be described here. Flux-weighted centroids are computed in an aperture that includes the optimal aperture plus a single halo ring for each target. PRF-based centroids are computed in an aperture that includes all of the available pixels for each target. In both cases, uncertainties are propagated to the computed centroid row and column coordinates based on the Jacobian for each centroid computation.

Every centroid computed in PA is validated against the bounding box of the associated centroid aperture, and each is gapped in the event that it does not fall within the bounding box for any reason. Gaps are also set for PRF-based centroids that cannot be successfully computed for any target and cadence due to failure of the iterative fitting algorithm.

Let the subset of pixels in the aperture of a given target that are utilized for computation of the flux-weighted centroids on each cadence be denoted by  $W$ . If the row and column centroid coordinates are denoted by  $r$  and  $c$  respectively, then the flux-weighted centroid is computed for any given cadence by

$$r = \frac{\sum_{n \in W} i_{T,n} P_{S,n}}{\sum_{n \in W} P_{S,n}}, \quad c = \frac{\sum_{n \in W} j_{T,n} P_{S,n}}{\sum_{n \in W} P_{S,n}} \quad (10)$$

As in equation (4), the CCD row and column coordinates of the pixels in the target aperture are designated  $i_T$  and  $j_T$  respectively.

Flux-weighted centroid row and column time series for a long cadence PPA\_STELLAR target on module output 7.3 in Q3 are shown in Figure 6. The centroid row time series is shown in the upper panel and the centroid column time series in the lower panel. The dominant factor driving long term centroid drift is DVA. Transients in the respective centroid time series at the beginning of the quarter and following the monthly downlinks to Earth (near cadences 1500 and 3000) are due to focus changes which result from thermal changes in the photometer. It is interesting to note that the transients are more pronounced in the centroids than they are in the light curve for this target shown in the upper panel of Figure 5.

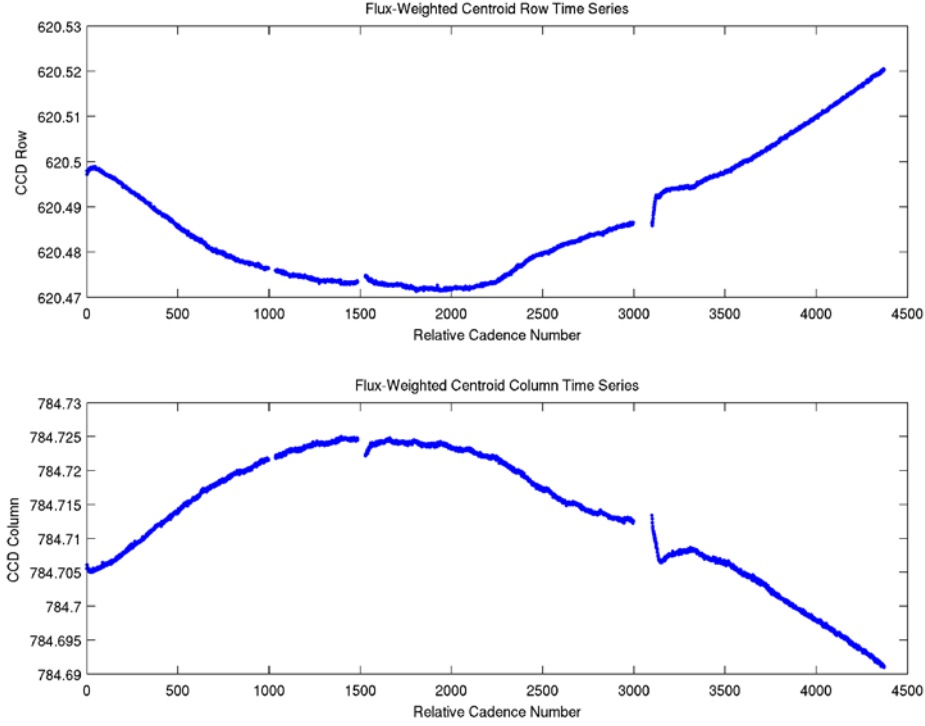


Figure 6: Flux-weighted centroid row and column time series for the PPA\_STELLAR target also shown in the left panel of Figure 4 and the upper panel of Figure 5. The data gaps are due to a loss of fine point anomaly and monthly data downlinks. Transients are due to focus changes that result from thermal changes in the photometer.

If the rigorous propagation of uncertainties treatment is enabled by the pipeline operator, the variances  $v_R$  and  $v_C$  for the respective row and column centroids on the given cadence are computed by

$$v_R = t_{FWR} C_S t_{FWR}', \quad v_C = t_{FWC} C_S t_{FWC}' \quad (11)$$

The covariance matrix  $C_S$  of the background-removed pixels in the target aperture was defined in equation (6). The Jacobian transformations  $t_{FWR}$  and  $t_{FWC}$  are row vectors where the  $n^{\text{th}}$  elements of the respective transformations are given by

$$t_{FWR,n} = \frac{i_{T,n} - r}{\sum_{v \in W} P_{S,v}}, \quad t_{FWC,n} = \frac{j_{T,n} - c}{\sum_{v \in W} P_{S,v}} \quad (12)$$

if the  $n^{\text{th}}$  pixel is in the flux-weighted centroid aperture  $W$ , or zero otherwise.

#### 4. SUMMARY AND CONCLUSIONS

The primary tasks of the Photometric Analysis module of the *Kepler* SOC Science Processing Pipeline are to compute the photometric flux and photocenters (centroids) for all long and short cadence targets from the calibrated pixels in their respective apertures. We first presented an overview of the PA module. We have then shown how cosmic rays are cleaned from the calibrated background and target pixel data. We have discussed the process for estimation and removal of background levels from the pixels in the apertures of the respective science targets. We have described the extraction of raw flux light curves by SAP from the background-removed pixels in the optimal apertures of the science targets. We have shown how flux-weighted centroids are computed per target and cadence. We have also described how

uncertainties are propagated to the various PA products from covariance matrices for the calibrated background pixels and the calibrated pixels in the respective target apertures. Raw flux light curves, centroid time series, and barycentric timestamp corrections produced by PA are exported to the MAST and made available to the general public in accordance with the NASA/Kepler data release policy.

## ACKNOWLEDGEMENTS

The authors would like to thank Greg Orzech, Bill Wohler and Michael Fanelli for reviewing this paper.

Funding for the *Kepler Mission* has been provided by the NASA Science Mission Directorate.

## REFERENCES

- [1] Koch, D.G., et al., “*Kepler Mission* Design, Realized Photometric Performance, and Early Science,” *Astrophysical Journal Letters*, 713 (2), L79-L86 (2010).
- [2] Jenkins, J.M., et al., “Initial Characteristics of *Kepler* Long Cadence Data for Detecting Transiting Planets,” *Astrophysical Journal Letters*, 713 (2), L120-L125 (2010).
- [3] Gilliland, R.L., et al., “Initial Characteristics of *Kepler* Short Cadence Data,” *Astrophysical Journal Letters*, 713 (2), L160-L163 (2010).
- [4] Jenkins, J.M., et al., “Overview of the *Kepler* Science Processing Pipeline,” *Astrophysical Journal Letters*, 713 (2), L87-L91 (2010).
- [5] Middour, C., et al., “*Kepler* Science Operations Center architecture,” *Proc. SPIE 7740*, in press (2010).
- [6] Quintana, E.V., et al., “Pixel-level calibration in the *Kepler* Science Operations Center pipeline,” *Proc. SPIE 7740*, in press (2010).
- [7] Twicken, J.D., et al., “Presearch data conditioning in the *Kepler* Science Operations Center pipeline,” *Proc. SPIE 7740*, in press (2010).
- [8] Jenkins, J.M., et al., “Transiting planet search in the *Kepler* pipeline,” *Proc. SPIE 7740*, in press (2010).
- [9] Wu, H., et al., “Data validation in the *Kepler* Science Operations Center pipeline,” *Proc. SPIE 7740*, in press (2010).
- [10] Tenenbaum, P., et al., “An algorithm for fitting of planet models to *Kepler* light curves,” *Proc. SPIE 7740*, in press (2010).
- [11] Klaus, T.C., et al., “The *Kepler* Science Operations Center pipeline framework,” *Proc. SPIE 7740*, in press (2010).
- [12] Klaus, T.C., et al., “The *Kepler* Science Operations Center pipeline framework extensions,” *Proc. SPIE 7740*, in press (2010).
- [13] Bryson, S.T., et al., “The *Kepler* Pixel Response Function,” *Astrophysical Journal Letters*, 713 (2), L97-L102 (2010).
- [14] Tenenbaum, P. and Jenkins, J.M., “Focal plane geometry characterization of the *Kepler* mission,” *Proc. SPIE 7740*, in press (2010).
- [15] Li, J., et al., “Photometer performance assessment in *Kepler* science data processing,” *Proc. SPIE 7740*, in press (2010).
- [16] Clarke, B.D., et al., “A framework for propagation of uncertainties in the *Kepler* data analysis pipeline,” *Proc. SPIE 7740*, in press (2010).
- [17] Haas, M.R., et al., “*Kepler* Science Operations,” *Astrophysical Journal Letters*, 713 (2), L115-L119 (2010).
- [18] Akaike, H., “A New Look at the Statistical Model Identification,” *IEEE Transactions on Automatic Control*, AC-19 (6), 716-723 (1974).
- [19] Bryson, S.T., et al., “Selecting pixels for *Kepler* downlink,” *Proc. SPIE 7740*, in press (2010).