Finding Optimal Apertures in *Kepler* Data

Jeffrey C. Smith\(^1,2\), Robert L. Morris\(^1,2\), Jon M. Jenkins\(^1\), Stephen T. Bryson\(^1\), Douglas A. Caldwell\(^1,2\), and Forrest R. Girouard\(^1,3\)

\(^1\) NASA Ames Research Center, Moffett Field, CA 94035, USA; jeffrey.smith@nasa.gov
\(^2\) SETI Institute, 189 Bernardo Avenue, Suite 100, Mountain View, CA 94043, USA
\(^3\) Logyx LLC, 425 N. Whisman Road, Mountain View, CA 94043, USA

Received 2016 March 2; accepted 2016 May 24; published 2016 October 20

Abstract

With the loss of two spacecraft reaction wheels precluding further data collection for the *Kepler* primary mission, even greater pressure is placed on the processing pipeline to eke out every last transit signal in the data. To that end, we have developed a new method to optimize the *Kepler* Simple Aperture Photometry (SAP) photometric apertures for both planet detection and minimization of systematic effects. The approach uses a per cadence modeling of the raw pixel data and then performs an aperture optimization based on signal-to-noise ratio and the *Kepler* Combined Differential Photometric Precision (CDPP), which is a measure of the noise over the duration of a reference transit signal. We have found the new apertures to be superior to the previous *Kepler* apertures. We can now also find a per cadence flux fraction in aperture and crowding metric. The new approach has also been proven to be robust at finding apertures in *K2* data that help mitigate the larger motion-induced systematics in the photometry. The method further allows us to identify errors in the *Kepler* and *K2* input catalogs.

Key words: methods: data analysis – techniques: image processing – techniques: photometric – planets and satellites: detection

Online material: color figures

1. An Overview of the *Kepler* Data Pipeline

*Kepler*’s primary science objective is to determine the frequency of Earth-size planets transiting their Sun-like host stars in the habitable zone. This daunting task demands an instrument that is capable of measuring the light output from each of over 150,000 stars simultaneously, with an unprecedented photometric precision of 30 parts per million (ppm) at 6.5-hour intervals for 12th magnitude stars. The *Kepler* Science Pipeline is tasked with processing the *Kepler* data and detecting the transiting planet signals.

The *Kepler* Science Pipeline is divided into several components in order to allow for efficient management and parallel processing of data (Jenkins et al. 2010a). Raw pixel data downlinked from the *Kepler* photometer are calibrated by the Calibration module (CAL) to produce calibrated target and background pixels (Quintana et al. 2010) and their associated uncertainties (Clarke et al. 2010). The calibrated pixels are then processed by the Photometric Analysis module (PA) to fit and remove cosmic rays and sky background and then re-find optimal apertures and extract simple aperture photometry from the background-corrected, calibrated target pixels (Twicken et al. 2010). This paper details improvements to the search for optimal apertures within the PA component.

The final step to produce light curves is performed in the Pre-search Data Conditioning module (PDC; Smith et al. 2012; Stumpe et al. 2012, 2014), where signatures in the light curves correlated with systematic error sources from the telescope and spacecraft, such as pointing drift, focus changes, and thermal transients, are removed. Additionally, PDC identifies and removes sudden pixel sensitivity dropsouts (SPSDs), which result in abrupt drops in measured flux with partial recovery of pixel sensitivity on timescales ranging from hours to days (Van Cleve & Caldwell 2009). PDC also identifies residual isolated outliers and fills data gaps (such as during intra-quarter downlinks) so that the data for each quarterly segment are contiguous when presented to later pipeline modules. In a final step, PDC adjusts the light curves to account for excess flux in the optimal apertures due to starfield crowding and the fraction of the target star flux in the aperture to make apparent transit depths uniform from quarter to quarter as the stars move from detector to detector with each roll maneuver. Output data products include raw and calibrated pixels, optimal apertures, raw and systematic error-corrected flux time series, and centroids and associated uncertainties for each target star. The data products are archived to the Data Management Center.
(DMC) and made available to the public through the Mikulski Archive at STScI (MAST\footnote{http://stdatu.stsci.edu/kepler/}; McCauliff et al. 2010).

Data are then passed to the Transiting Planet Search module (TPS; Jenkins et al. 2010b), where a wavelet-based adaptive matched filter is applied to identify transit-like features with durations in the range of 1.5 to 15 hours. Light curves with transit-like features whose combined signal-to-noise ratios (S/N) exceed 7.1\(\sigma\) for a specified trial period and epoch are verified to be of high quality by several vetoes as discussed in Seader et al. 2015. The transits that pass all vetoes are designated as threshold crossing events (TCEs) and subjected to further scrutiny by the Data Validation module (DV; Wu et al. 2010). The 7.1\(\sigma\) threshold was set so that there is no more than one expected false alarm for Earth-like planets for the entire campaign, assuming Gaussian statistics (Jenkins 2002). DV performs a suite of statistical tests to evaluate the confidence in the transit detection, to reject false positives by background eclipsing binaries, and to extract the physical parameters of each system (along with associated uncertainties and covariance matrices) for each planet candidate (Tenenbaum et al. 2010). After the planetary signatures are fitted, DV removes them from the light curves and invokes TPS again to search over the residual time series for additional transiting planets. This process repeats until no further TCEs are identified.

The Kepler mission uses forty-two 2200 \times 1024 pixel CCDs. Each CCD is composed of two output channels resulting in 84 “CCD channels,” which are also referred to as “module outputs.” On board storage and bandwidth constraints prevent the storage and downlink of all 96 million pixels per 30-minute cadence, so the Kepler spacecraft downlinks a specified collection of pixels for each target, called masks. These pixel masks are selected by considering the object brightnesses, local background, and the S/N on each pixel. A “halo” is then added to the masks, which is an extra ring of pixels around the signal pixels to include a margin of error (Bryson et al. 2010b). A final adjustment to the masks is made by adding an extra column to the data upstream side of the masks to provide for data to account for Local Detector Electronics (LDE) undershoot, which is a distortion of low-signal pixels immediately downstream of high-signal pixels (Van Cleve & Caldwell 2009). During the process of finding target masks, an optimal aperture is also found to define the pixels containing the core of the target image and to generate simple aperture photometry;\footnote{In simple aperture photometry, the brightness of a star in a given frame is measured by summing up only the pixel values in the core image that increase S/N.} this is the first optimal aperture finding method (Method \#1) and is discussed in Section 3. Past archived Kepler data releases (prior to Data Release 25) solely utilized this method for finding optimal apertures. As of Data Release 25, Method \#2 is now utilized and is discussed in Section 4.

2. Finding Optimal Apertures

Due to bandwidth considerations, data from all of the 84 million pixels in the Kepler focal plane cannot be stored for transmission for every cadence. Therefore specific objects (mostly stars) in the Kepler Field Of View (FOV) are identified as “targets”.\footnote{We will use the term “source” to reference any individual point source in the FOV, be it a star, Galaxy, or otherwise. The “targets” are almost all stars but a small number of “targets” are clusters, galaxies, or AGNs.} Groups of pixels around these objects are called masks and are extracted for storage and eventual transmission. Within each mask is a smaller collection of pixels called an “aperture” that is optimized for generating photometric light curves. For some targets, such as galaxies, the required apertures are explicitly specified and not optimized in the pipeline. For planetary transit targets the pixels in the aperture are selected to maximize the S/N and photometric precision for the target. The resulting collection of pixels is called an “optimal aperture.” For the Kepler Pipeline a single fixed optimal aperture is found for each data quarter of approximately 90 days. The maximum motion for Kepler data is 0.6 pixels over a 90-day quarter, with an average of 0.4 pixels, plus minor pointing drift. While using moving adaptive apertures is unavoidable for many ground-based photometric surveys, where the image motion can be several pixels and upward over a night of observations, it is not necessary for the pointing precision resembling that of Kepler.

Two principle methods are used for finding the optimal apertures. One is model-driven, and the other is data-driven. The first method (Method \#1) generates two pixel response function (PRF) model-derived synthetic full frame images (FFIs), one with and one without the target star (the PRF is discussed in Section 3). These FFIs are used to estimate the S/N of each pixel, allowing the selection of the pixel set that maximizes the S/N of the total flux of each target. The second method (Method \#2) uses a PRF model image fit to the acquired pixel data to find the target flux contribution to the scene. The method then uses the pure pixel scene data to estimate the noise. This second method has a further step where the aperture is optimized with Combined Differential Photometric Precision (CDPP, discussed in Section 4.2). The quality of the resultant aperture and light curve is compared between the two methods and the better of the two is chosen for each target. Method \#1 has been used for the Kepler mission since before launch and is how apertures are selected in flight. A brief description of this method is presented in Section 3 for reference. Method \#2 is novel to this paper.
3. Method #1: Using the Synthetic FFI and a Pure PRF Image Model

In this method the computation of optimal apertures is based on the generation of two synthetic FFIs for each CCD channel without the use of the acquired pixel data: one FFI with all stars and a second FFI with the target star removed (though still including the effects of the target star, such as smear). These images are then used to compare the signal from the target star with the noise from the target star, stellar background, and the instrument, thereby optimizing the aperture for S/N. By considering all the pixels in these two FFIs in a region around a target we can determine the pixel set whose sum maximizes the S/N for that target.

The undersampled Kepler point spread function (PSF), combined with the intra-pixel variability, causes the pixel response to be very sensitive to small motions of a point source’s centroid. These motions include spacecraft pointing jitter and differential velocity aberration and can be a significant source of signal noise and uncertainty. Therefore it is important to characterize each pixel’s response to image motion on a sub-pixel scale. We call the function describing the pixel’s response to a moving point source the PRF (Bryson et al. 2010a). The PRF accounts for the effects of the optical point spread function, the CCD detector responsivity function, spacecraft pointing jitter, and other systematic effects. We use the PRFs acquired during commissioning, the target catalog, and an estimate of the offsets ($\Delta x(t), \Delta y(t)$) over time to compute a synthetic FFI. This FFI is then used to estimate the set of pixels that provide an optimal signal for a target. The PRF model was determined during commissioning using long-cadence FFIs and a 15-minute cadence. The PRF was computed by partitioning each pixel into sub-pixel regions, each one of which describes the PRF as a two-dimensional polynomial in that sub-pixel region. For each sub-pixel region the polynomial was computed as a robust chi-squared fit to the star flux for all stars falling in that sub-pixel region using the commissioning data, as discussed in Bryson et al. (2010a). It accounts for components of the mean motion due to spacecraft pointing jitter and other effects that occur on a single long-cadence timescale. The pixel response function depends sensitively on the sub-pixel location of the point source’s centroid. We therefore approximate this dependence by characterizing the PRF on a $6 \times 6$ sub-pixel grid, so each pixel has 36 pixel response functions assigned to it, creating a super-resolution representation of the PRF on a 0.17 pixel grid. The sub-pixel grid accounts for intra-pixel variability. For each sub-pixel location the PRF is defined on a domain of $11 \times 11$ pixels centered on the point source centroid. The PRF model thus found is represented by interpolating a set of polynomial coefficients expressing the brightness of each pixel, given that pixel’s brightness as a function of offsets of the light falling on that pixel, by convolving the flux centered on each of that pixel’s sub-pixel position with the coefficients of the pixel response function for that sub-pixel position, and summing over the sub-pixel positions. Motion over longer timescales is explicitly accounted for by making the offsets ($\Delta x(t), \Delta y(t)$) time dependent.

The computation of a synthetic FFI uses several inputs (Bryson et al. 2010a):

- **Kepler Input catalog (KIC)** (Brown et al. 2011), providing J2000 right ascension (R.A.), declination (decl.), and magnitude in the Kepler bandpass of stellar targets in the Kepler field.
- **Pixel response function (PRF) model**, an observation-based super-resolution model of expected pixel values in response to point sources. The PRF model includes intra-pixel variability.
- **Focal Plane Geometry (FPG) and pointing model**, which includes measurements of the locations of the CCDs in the Kepler focal plane, models of the Kepler optics, and differential velocity aberration (DVA).
- **Saturation model**, which includes information about the well-depth of each output channel.
- **Zodiacal light model**, represented as a mesh of magnitude values on the sky.
- **Read noise model**, observed values of read noise for each output channel.
- **Charge Transfer Efficiency (CTE) model**, which describes how much flux is lost with each charge transfer during readout.

In essence, for each source in the KIC that falls on a channel, the source’s pixel position on the channel is computed from the KIC, using the FPG and pointing model and a copy of the appropriate PRF scaled by the brightness of that source and smeared by DVA motion. This is all added to the FFI in the appropriate pixel location. Due to the large number of sources in the KIC and the resolution required to estimate the DVA motion, the direct computation of the FFI would be impractically slow. To overcome this difficulty, both the PRF and the brightness of each pixel in the FFI are represented by two-dimensional polynomials, and most of the computations are performed in terms of the polynomial coefficients. The contribution from each individual source is included in the model by adding the coefficients of the PRF’s polynomial representation to the coefficients of the FFI pixel’s polynomial representation. Once this process is complete the FFI is generated by evaluating the FFI’s polynomial for each pixel. The synthetic FFI is completed by adding simulated spillover of saturated pixels, effects of charge transfer, smear due to shutterless operation, and zodiacal light.

The above computation of a synthetic FFI uses all known sources that fall on a channel. A subset of these sources are identified as target stars and the same procedure is used to generate synthetic images of each target’s contribution in
isolation. The isolated target images are then subtracted from the all-inclusive FFI to produce an FFI with the target flux removed. When removing the flux from a target star to compute the second FFI, care is taken to retain the smear signal from all other sources. A noise model is then used to estimate the noise in each pixel. This noise model includes the shot noise of the target, background signal, smear, and zodiacal light, as well as read and quantization noise. The S/N for each pixel is computed using

$$\text{SNR}_{\text{pixel}} = \frac{f_{\text{target}}}{\sqrt{f_{\text{target}} + f_{\text{back}} + \nu_{\text{read}}^2 + \nu_{\text{quant}}^2}},$$  \hspace{1cm} (1)$$

where $\nu_{\text{read}}$ is the channel’s read noise based on the read noise model and $\nu_{\text{quant}}$ is the quantization noise, given by

$$\nu_{\text{quant}} = \frac{n_c}{12} \left( \frac{w}{2n_b - 1} \right)^2,$$  \hspace{1cm} (2)$$

where $n_c$ is the number of cadences in a co-added observation ($n_c = 270$ for Kepler Long Cadence), $w$ is the well-depth, and $n_b$ is the number of bits in the analog-to-digital converter (14). The other two terms in (1), $f_{\text{target}}$ and $f_{\text{back}}$, are the target and background flux values from the FFI, and in the denominator, account for the Poisson shot noise (which scales as the square root of the flux).

Given a collection of pixels, the S/N of the collection is given by Equation (1) summed over the pixels in the collection. Optimal pixel selection begins by including the pixel with the highest S/N. The next pixel to be added is the pixel that results in the greatest increase in S/N of the collection. Initially the collection S/N will increase as pixels are added. After the bright pixels in the target have been added, dim pixels dominated by noise cause the S/N to decrease. The pixel collection with the highest S/N defines the optimal aperture.

This computation of the optimal aperture is based on a method originally implemented in the Kepler End-to-End Model (ETEM), described in Jenkins (2004) with modifications described in (Bryson 2008) and (Bryson et al. 2010b). It is a model-driven approach and does not use the actual CCD pixel data in the computation. It relies on an accurate PRF, pointing knowledge, target catalog, and representation of noise sources. Herein lies its main deficiency: errors in the PRF model or target catalog position and magnitude directly result in aperture errors. Incomplete accounting for background objects by the catalogs used, focus errors, stellar variability, and saturation also impair the method. While this method is fast and reliably selects an aperture, the above errors lead to compromised photometry on some targets. Using collected pixel data to identify and correct these errors is the motivation for the method described in the next section. Sections 8 and 9 discuss how the photometry has improved with the new method.

4. Method #2: Using a PRF-Based Image Model and the Pixel Scene

The driving factor to improve upon the older aperture finding routine is to allow the data to speak for itself when finding both the background noise and the target signal. The Kepler catalog is quite complete to 17th magnitude, but dimmer background objects can still contribute to the noise and yet not be modeled in Method #1. There can also be an error of up to $\pm 0.5$ brightness in magnitude and 4 arcseconds in position for known catalog objects, which further contributes to errors in the found optimal aperture (Brown et al. 2011).

Method #2 fits an image model to the scene and can thereby update the catalog targets’ position and brightness. It then calculates the signal flux for each pixel in the mask based on this model, which is then used as the numerator in the S/N ratio. Note that although the image model exploits the existing PRF models, this is not PRF photometry per se; we are not generating light curves from the PRF model but merely using the modeled image for the numerator in the S/N calculation.

The optimal aperture Method #2 algorithm includes a multi-step process. For each cadence,

1. The PRF model is fit to the pixel scene using the catalog as initial values and then finds the contribution to the pixel flux from each catalog object plus the residual background. A description of the image modeling is given in Section 4.1 below.
2. The target center pixel is found based on the PRF model fit and this pixel is labeled as the first pixel.
3. The pixel adding order is found by calculating the S/N for each progressively larger aperture size and always choosing the next pixel that increases the S/N the most. This will order the pixels by their contribution to the S/N.
4. The peak in the S/N curve is found using Equation (3) below; this is the S/N-maximized optimal aperture for this cadence.

This part of Method #2 is very similar to Method #1 except for the use of real data. So, instead of using Equation (1), here we compute the S/N on a per cadence basis using the equation

$$\text{SNR}_k = \frac{\sum f_i}{\sqrt{k(\nu_{\text{read}}^2 + \nu_{\text{quant}}^2) + \sum i y_i}},$$  \hspace{1cm} (3)$$

where $f$ is the target image fit to the PRF model (as given in Equation (4) below), $y$ is the actual pixel data, including all background (this is the shot noise term), and the other two terms are the non-shot noise (read noise & quantization noise) as computed in Section 3, and $i$ is over each pixel in the aperture of size $k$. Note that the calibrated pixel data as processed in the Kepler CAL module is used (i.e. corrected for bias, flat-field, etc.). When ordering the pixels there is no
requirement for symmetric apertures and there can be holes in the aperture; however, the apertures must be contiguous in the 4-connected sense, i.e., each pixel in the aperture has a solid edge in contact with at least one other pixel in the aperture.

We now have an optimal aperture for each cadence. The Kepler mission pipeline creates photometric light curves using simple aperture photometry, which means a single aperture is used for each quarter. There are different ways we can construct the single optimal aperture and we find two apertures directly from this S/N-optimized, per cadence aperture. The first is a 95% union of each cadence aperture. That is to say, a pixel is included if greater than 5% of the cadences include that pixel in its optimal aperture. The other is a 50% union, or median aperture over all cadences. For targets with small centroid motion the median aperture is optimal since it finds the “core” pixel array containing the signal. The median is also better than a mean at not being skewed by transient blips in the centroid position. The median aperture therefore robustly removes outlier cadences. When there is large motion, however, the “core” aperture found with the median will clip too much of the flux far from the center of motion but the 95% union aperture is better at accounting for the “smear” of the flux over the larger pixel array.

By ordering the pixels as described above, the number of possible apertures in the N-pixel mask is reduced from $2^N - 1$ (all possible subsets of pixels minus the empty set) to N. With this ordering we can therefore afford to compute the much more costly estimate of Combined Differential Photometric Precision (CDPP; Jenkins et al. 2010b) for each of the N candidate apertures to find the aperture with the lowest CDPP. This CDPP-optimized aperture is discussed in Section 4.2.

4.1. Modeling Target Masks

The optimal aperture selection is determined by separating the flux contribution of the target star from that of other sources so that we can disentangle the numerator, or information, from the denominator, or noise, in the S/N. An accurate model of image formation enables this decomposition of background-subtracted pixel flux measurements into contributions from each point source in the instrument’s FOV. Our model consists of a scene description comprising the set of known sources $s = s_1, s_2, ..., s_N$ with celestial coordinates $(\alpha_n, \delta_n)$ and magnitudes, a model of background flux, a model of image motion on the focal plane, and a model of the Kepler PRF.

We model the calibrated mean electron flux $f$ at each pixel $p$ during cadence $c$ as a linear combination of PRF values for each source plus a background model,

$$f(p, c) = f_{\text{bkgd}}(p, c) + \beta(c) + \sum_{n=1}^{N} F_n(c) \overline{PRF}(p, x_n(c), y_n(c)).$$  \hspace{1cm} (4)

The total flux $F_n$ due to source $n$ is distributed among pixels according to the normalized PRF, referred to here as $\overline{PRF}$, which is obtained by evaluating the PRF over an 11-by-11 pixel grid centered at $x$, $y$ and dividing pixel $p$ by the sum of evaluated pixels. For more details on evaluating the PRF the reader is referred to Bryson et al. (2010a). The real-valued CCD coordinates $x_n(c)$ and $y_n(c)$ are estimates of the mean photocenter of source $s_n$ during cadence $c$. Centroid estimates are obtained from a polynomial model of image motion, which maps celestial coordinates to CCD coordinates,

$$x_n = P_x(\alpha_n, \delta_n) \tag{5}$$

$$y_n = P_y(\alpha_n, \delta_n). \tag{6}$$

Motion polynomials $P_x$ and $P_y$ for each channel and cadence are robustly fitted to images of a spatially distributed set of $\approx$200 bright, unsaturated, and uncrowded sources (Twicken et al. 2010). Background flux estimates $f_{\text{bkgd}}$ are obtained from similarly-constructed 2D polynomial models fitted to the grid of background pixels collected from each channel. The constant term $\beta$ is included to account for any local bias in the background flux estimates for the set of pixels being modeled.

An independent model is fitted to each set of pixels comprising a target mask. The scene model for a given mask is constructed from entries in the KIC (Brown et al. 2011), UKIRT (Lawrence et al. 2007), and Ecliptic Plane Input Catalog (EPIC)\footnote{EPIC is hosted at MAST (http://archive.stsci.edu/k2/epic/search.php).} and is initialized by ranking sources within 5 pixels from the edges of the target’s mask by their maximum expected S/N contribution to any pixel in the mask. Sources whose maximum S/N values fall below a threshold are discarded and up to 20 sources from the remaining set are admitted to the scene model in order of rank. Since certain regions of the FOV may be densely populated with dim stars, limiting the scene model in this way prevents wasting time and computing power on sources that do not make significant flux contributions.

We fit model parameters $F_n(c)$, $\beta_n(c)$, $\alpha_n$, and $\delta_n$ to the observed data by minimizing the weighted sum of chi-square residuals over all pixels and cadences in the model. Fine-tuning the source positions $\alpha_n$, $\delta_n$ in the fit compensates for both catalog errors and local biases in the motion polynomials. Position fitting is enabled only for sources with centroids inside the target mask that contribute sufficient flux, defined by a threshold parameter. We only allow for position variations if the star is sufficiently bright and its expected centroid lies within the mask. If the predicted centroid falls outside the mask, then the data would only capture the “skirt” of the PRF and its position would be too poorly constrained for a fit to be reliable. The minimization is constrained such that source flux is non-negative and perturbations $\Delta_n$ to catalog positions $\alpha_n$, $\delta_n$ are smaller than 1.5 pixels. The purpose of the fitting of $\Delta_n$ is...
to identify and correct errors in the catalog. But we have found
in some cases the fitter will drift far off in R.A. and decl. and so
the limit of 1.5 pixels (3.98 arcsec) is to inhibit these cases. A
drift of greater than 4 arcsec is considered a poor fit since
catalog errors are not expected to be any larger (Brown et al. 2011). Note that the 1.5 pixel limit is only to the catalog
right ascension and declination (αn, δn). The motion polyno-
mials can still allow for greater centroid motion than 1.5
pixels across the CCD, which is not uncommon for K2 data.
Also note that, while fitting F_n and β is a linear problem, fitting
source positions α_n and δ_n is inherently nonlinear. We address
the two components of the fit separately in an iterative scheme.
We use a Levenberg-Marquardt algorithm to optimize the
perturbation of source positions, while the non-negative least
squares method in (Kim et al. 2013) is used to fit the flux
model of Equation (4) for a given perturbation,

\[
\begin{align*}
\text{minimize} & \quad (w(\Delta) + 1) \sum_{p=1}^{n} \sum_{c=1}^{c} \chi^2(p, c; F(c), \beta(c), \Delta) \\
\text{subject to} & \quad F_n(c) \geq 0, \Delta_n \leq \Delta_{\text{max}}
\end{align*}
\]

(7)

and where

\[
\chi^2(p, c) = \frac{(f_{\text{observed}}(p, c) - f_{\text{model}}(p, c))^2}{\sigma(p, c)}.
\]

Measurement uncertainties σ(p, c) are obtained from the
Calibration (CAL) pipeline module along with calibrated pixel
flux f_{\text{observed}}. The heuristic weighting function w(\Delta) in
Equation (7) prevents source positions from converging on
the same peak in the data by increasing as sources move toward
one another from their catalog positions. Treating the sources
as charged particles of like signs, we compute the potential
energy of the source configuration defined by the catalog, as
well as that of the perturbed configuration (catalog + Δ). If the
energy of the perturbed configuration is greater than the catalog
configuration, then w(\Delta) is given the value of the difference
(perturbed—catalog). Otherwise it is set to zero. By basing the
weight on the energy increase, we penalize solutions in which
multiple stars tend toward the same position without penalizing
solutions that agree with the catalog. An illustration of the
model fitting process is given in Figure 1.

The fit yields an initial light curve estimate for the target star
as well as estimates of errors in its catalog position and
magnitude. The light curve estimate constitutes the numerator
of the S/N estimate in Equation (3) that is used for establishing
the optimal aperture and pixel adding order. F_n(c), α_n, and δ_n
can be used to identify errors in the Kepler KIC, UKIRT, and
EPIC Catalogs.

4.2. Optimizing the Aperture for Photometric Precision

An S/N is a very simple and fast calculation. One can easily
calculate it for every pixel combination and find a pixel adding
order in a tractable amount of time, even for large apertures.
However, if planet transit finding is the primary science goal, as it is
for the Kepler mission, a better metric to optimize the aperture
is CDPP (Jenkins et al. 2010b). Simply speaking, CDPP is an
estimate of how well a transit-like signal can be detected in a
stellar light curve. A 6-hour CDPP of 100 ppm means that a 100
ppm 6-hour transit signal has a detection statistic of 1σ. CDPP is a
significantly more sophisticated calculation than S/N, so we
cannot afford to calculate it for a large number of candidate
apertures. Fortunately, in Section 4 we have already found a pixel
adding order per cadence. But a further step is first needed since
CDPP is calculated on a flux time series, not on individual
cadences; therefore, a single “averaged” pixel adding order over
the entire quarter is needed. This single “average” pixel adding order
gives us a limited set of apertures where we can compute
CDPP. We calculate the “average” pixel adding order by taking the
“mode”, or most frequently occurring values, of the pixel
adding order per cadence, ensuring that every pixel is only
included once in the averaged pixel adding order. Once we have
this order, we begin with the center pixel (first pixel in the pixel
adding order) and generate a simple aperture light curve. CDPP is
then calculated for each sequentially larger aperture using the
found averaged pixel adding order.

The calculation of CDPP is explained in detail in Jenkins et al.
(2010b). Calculating CDPP for all 14 pulse durations used in the
planet transit search would be prohibitively expensive in the
pipeline. The performance benchmark for Kepler is a 6 hr CDPP
on a 12th magnitude dwarf star. We therefore restrict ourselves
to computing a 6-hour transit duration CDPP. Also considered is
that the total flux in the apertures change as we increase the
aperture size, which will directly change the absolute CDPP
values. We therefore median normalize the light curves to place
each aperture on equal footing. The CDPP thus computed is a
time series giving the detection statistic for each cadence. We
take a robust root-mean square value of this time series to obtain
a single CDPP value as a function of aperture size.

CDPP is dangerous to use by itself to find the optimal
aperture. If, for example, a background object is brighter than
the target, then the CDPP of an aperture centered on the
background object will very likely be less than one centered on
the target and the aperture will grow to engulf the brighter
background object. The S/N, on the other hand, is very robust
against background contamination, the numerator in the S/N
being purely from the modeled target star. Any contamination
will only decrease the S/N value. A combination of the two
metrics can therefore simultaneously exclude background
objects and optimize for planet detection. We combine the
two and find k to minimize

\[
\text{CDPP}_k = (\text{SNR}_k - \text{SNR}) / \sigma_{\text{CDPP}},
\]

(9)

where CDPP_k is a smoothed CDPP curve for an aperture
composed of k pixels. The smoothing is necessary since the
CDPP measurement has more scatter than the S/N curve,
which is typically very smooth. $\sigma_{\text{S/N}}$ and $\sigma_{\text{CDPP}}$ are the variance in the CDPP and S/N curves, the ratio of these scales the S/N curve to the same range as the CDPP curve so that they are given equal weight in the optimization. SNR is the median S/N value over aperture sizes $k$. The minimum of the above curve is chosen and this is the CDPP-optimized aperture.

5. **Selecting the Best Aperture using CDPP and Logistic Regression**

We have now found four apertures:

1. Pure PRF image model-derived (Method #1);
2. Real pixel data-derived S/N-optimized Median;
3. Real pixel data-derived S/N-optimized Union; and
4. Real pixel data-derived CDPP-optimized.

Each of these apertures can yield better performance for a subset of *Kepler* targets. Since transit signal detection is the primary goal of the *Kepler* mission, we principally select the optimal aperture based on CDPP. An estimate of the uncertainty in the CDPP measurement is used to bias the selection to just apertures 2-4. This is because we believe an aperture selected using flight data will be superior in most cases to a purely model-based aperture. So if we are within the uncertainty in the measurement, we will pick the Method #2 aperture. However, in certain cases, the data-derived aperture can fall into a local minimum during the optimization procedure and not identify the true optimal aperture, hence the model-based aperture (Method #1) is still considered, which has more consistent and robust behavior (by virtue of it being based on a synthetic model).

For about 0.5% of targets method #2 has been shown to be problematic in spite of it having a lower CDPP. CDPP is agnostic to whether or not the proper target is chosen. For example, if one of the 4 apertures incorrectly centers on a background object that is brighter than the target then the CDPP metric could select that aperture. Method #1, on the other hand, is more robust to these cases, but at the expense of overall poorer apertures. So for the 0.5% of targets we identify, reverting to Method #1 is a move to revert to the more conservative aperture. A logistic regressive heuristic model was developed to help identify these corner cases. We identified several key metrics that can be used in aggregate to identify the bad corner cases, including

1. Net change in flux in the aperture between Method #1 and Method #2;
2. Fractional change in aperture between Method #1 and Method #2;
3. Fractional change in flux between Method #1 and Method #2;
4. Ratio of total mask used in the Method #2 aperture;
5. And etc.

Any one of these predictors is not necessarily sufficient to identify poor performance. Logistic regression (see James et al. 2014) was therefore used to identify the appropriate combination of predictors. To find the proper regressive model, a training set is needed to “train” the model. Each of the metrics was calculated for all targets on *Kepler* CCD module channel output 2.1 for quarter 13, 1358 targets in total. The targets were ranked by response to each of the metrics. Then for each target in order of rank, the diagnostic information generated by the aperture finding algorithms (an example of which is shown in Figures 3 and 4) was examined to identify poor aperture selection in Method #2. Targets continued to be tallied down to lower predictor response until all remaining Method #2 apertures appeared to be good. This resulted in a training set of 100 targets, with 10 having incorrectly chosen one of the

![Figure 1. Single model of an actual *Kepler* target mask. The image model is constructed from a background model, a normalized PRF model evaluated for each source, and estimates $F_s$ of the total electron flux contributed by each source. The celestial coordinates $(\alpha_s, \delta_s)$ of each source are mapped to CCD coordinates $(x_n, y_n)$ by a polynomial model of image motion. Note that in this example the centroid of source $s_1$ lies slightly outside the mask, but it still contributes significant flux to pixels inside the mask. Also note that the color spaces of the component images have been stretched to aid visualization. (A color version of this figure is available in the online journal.)](image-url)

---

Publications of the Astronomical Society of the Pacific, 128:124501 (16pp), 2016 December

Smith et al.
Method #2 apertures. Multinomial logistic regression was then used to model a response $p$ on the predictors:

$$p(X) = \frac{\exp(\beta_0 + \beta_1 X_1 + \cdots + \beta_N X_N)}{1 + \exp(\beta_0 + \beta_1 X_1 + \cdots + \beta_N X_N)},$$

(10)

where $X = (X_1, \ldots, X_N)$ are the $N$ predictors and $\beta = (\beta_0, \ldots, \beta_N)$ are the coefficients plus intercept $\beta_0$. The coefficients were fit using maximum likelihood and the $p$-values for the identified coefficients were then used to rank the suitability of each of the predictors. After a couple iterations of dimensional reduction we found the three predictors, 1, 3, and 4 from above, are most effective at identifying poor aperture selection. The other predictors, such as the fractional change in aperture, were found to be poor predictors and we zeroed their coefficients, $\beta_2$, etc. This zeroing resulted in better separation in the prediction value $p(X)$ between good and bad apertures.\footnote{This observation emphasizes the common adage that more variables in a model does not necessarily produce a better model!}

Several other predictors were also tried but not listed above as none were found to enhance the separability of good and bad apertures.

The prediction $p(X)$ is a real function spanning $[0, 1]$, so we must determine the prediction discrimination threshold for poor apertures, $\tau$. The threshold value that minimizes false negatives was chosen. The logic was that whenever an aperture found by Method #2 is deemed poor, we error conservatively by falling back on the more robust Method #1 aperture. Once our predictive model was trained we used two different CCD channels, 7.3 and 13.2, to test the model (3934 targets in total). These two channels have different image characteristics than the training channel, 2.1, so they are good channels with which to test the robustness of the method. The three predictors and coefficients were evaluated in Equation (10) to return a prediction value, and when $p(X) > \tau$, the target aperture was flagged as poorly chosen and we reverted to Method #1. The table of confusion for the test data is given in Table 1. We see that the total number of truly poor apertures is $\frac{25}{3934} \approx 0.5\%$ and we correctly identify 22 of those, leaving only 3, or about 0.05\%, not identified. The 18 false positives are not considered

Figure 2. Flow chart showing the steps in finding optimal apertures in Kepler data. (A color version of this figure is available in the online journal.)
problematic because, as explained above, we choose to error conservatively. This logistic regressive method can therefore correctly identify 90% of the small number of poor apertures, reducing the total number of poor apertures down to only 0.05%, which translates to only about 1 target per Kepler CCD channel, or \( \sim 75 \) total out of the \( \sim 150,000 \) targets processed by Kepler.

6. Per Cadence Flux Fraction and Crowding Metric

No aperture will collect all the flux from any one object; the point spread function insists that some flux will spill over any finite aperture. We therefore calculate a flux fraction in aperture, or flux fraction for short, to quantify the fraction of target flux in the photometric aperture. Likewise, background objects will inevitably spill into the aperture to some degree. The crowding metric gives the fraction of flux in the aperture that is due to the target star. Up until the improvements made to the Kepler pipeline discussed in this paper, the mission was limited to a single flux fraction and crowding metric value per quarter. Since a target image model can be computed for every cadence, a flux fraction and crowding metric can now be computed per cadence. These per cadence values can be used to further reduce systematic trends in the data. However, we have not yet incorporated this information into the systematic error removal in PDC.

7. Saturated Pixels

Very bright targets result in CCD pixel saturation. For saturated targets the image modeling in Section 4.1 does not function properly because the PRF model does not account for saturation. We therefore always choose Method #1 for saturated targets and use the purely synthetic image model. Saturated charge is spilled along columns, with the fraction of charge spilled up and down determined by the saturation model mentioned in Section 3. The pixel values are set to the saturation values in the synthetic image created in Method #1 and the aperture is computed. Ground tests indicate that the well-depth and saturation spill direction vary across the focal plane (Van Cleve & Caldwell 2009), including variation within a channel. Therefore a buffer is added to the optimal aperture for saturated targets. This buffer is applied equally up and down the column and is added to the target’s optimal aperture, but the pixel values themselves are not changed. If the buffer size is smaller than the optimal aperture in either direction, then

![Figure 3. Mask scene, fitted PRF image, and four found apertures for Kepler ID 7846328. The color bar is in units of electrons per second. (A color version of this figure is available in the online journal.)](image)
the buffer size is ignored in that direction. In principle, there is no reason the saturation model cannot be used in the image modeling in Section 4.1 and future improvements could incorporate this model.

8. Application to Kepler Data

In Figure 2 we see a general flow chart of all the steps in aperture selection discussed in this paper. The purple box shows Method #1 and the aquamarine box shows Method #2, which loops over all cadences. We find four different apertures in total, force contiguity of the apertures, and finally select the best. A per cadence flux fraction and crowding metric is then computed for the chosen aperture. In practice, for Kepler data, we don’t actually loop over all cadences but just every tenth cadence. We found no degradation in performance and yet achieved a nearly factor of ten decrease in processing time, allowing the pipeline to continue to finish in a timely manner. The image motion in Kepler data is a small fraction of a pixel per cadence, so every tenth cadence is sufficient to capture the full motion.

As an example, Figure 3 shows the four found apertures for Kepler ID 7846328. The pixel image colors give the background-removed pixel data in this target’s mask. The found target centers on each cadence are plotted as red dots (the centroid motion is not large for this example) and the median center for each background object is a black dot. In this example we see two dim background objects with black dots and one bright background object with a center just off the mask. The four found apertures are labeled as crosses, x’s, circles, and diamonds, as referenced in the legend. The contour lines show isophotes of the target object’s image model, which is used as the numerator in Equation (3). The pixel data (with

| True Positive | False Negative |
|---------------|----------------|
| 22            | 3              |
| False Positive| True Negative  |
| 18            | 3891           |

Figure 4. The aperture fitting diagnostic figures for Kepler ID 7846328. (A color version of this figure is available in the online journal.)

Table 1

Table of Confusion for Identifying Poor Aperture Selection

| True Positive | False Negative |
|---------------|----------------|
| 22            | 3              |
| False Positive| True Negative  |
| 18            | 3891           |

Publications of the Astronomical Society of the Pacific, 128:124501 (16pp), 2016 December Smith et al.
background not removed) is used in the denominator of the S/N. The aperture selection process is illustrated for this same target in Figure 4. The upper right plot shows the S/N versus aperture size using the average pixel adding order with a maximum clearly present. Note that the actual S/N-optimized aperture is found for every cadence but here we just plot the average S/N over all cadences. The error bars represent the standard deviation in the S/N curves over all cadences. This target has little motion or stellar variability, so the S/N curve is very similar for all cadences. The lower right plot shows calculated CDPP versus the number of pixels in the aperture in the average pixel adding order as blue dots and the smoothed CDPP curve as a magenta line. The cyan curve is the combined CDPP and S/N curve, similar to Equation (9). Combining the CDPP with S/N allows for a single minimum to be more pronounced. The magenta curve, on the other hand, has two potential local minima near 0 and 32 pixels. Neither would be a proper aperture for this target. The dip in the magenta CDPP curve beginning at pixel index 30 is due to the aperture engulfing the bright background object as shown by the pixels labeled “30-32” in Figure 3. If the pixel mask for this target was even larger and the CDPP curve extended out to cover the entire bright background object, then the minimum in the CDPP curve would include both objects, as occurs in many other target examples. The upper left plot shows the photometric light curve for each of the four apertures. In this case the Method #1 and median S/N-optimized apertures contain the same pixels (as also seen in Figure 3), so the blue curve lies directly under the red. The legend contains the calculated rms CDPP for each of these four light curves and the CDPP-optimized light curve (cyan curve) is the clear winner at 60 ppm versus 91 ppm for the Method #1 aperture. We can also see that the cyan curve has dramatically reduced instrumental systematics versus the other curves, so it is the clear superior aperture and light curve. Systematic features in the data comparable in length and size to transits will obscure transit signals and therefore increase CDPP. It is therefore reasonable that the CDPP-optimized light curve will also have minimized systematics. The three-day reaction wheel heater cycle visible in the black and red curves is a good example of a systematic signal that can obscure a transit signal.

One may suspect that the CDPP-optimized aperture will always win the CDPP test. If we were optimizing the aperture purely on CDPP then this would be the case. But as discussed above, finding a global CDPP minimum would require $2^N - 1$ evaluations of CDPP for a pixel mask of size $N$. This is an insurmountable computational endeavor for an automated pipeline. We therefore must reduce the number of CDPP computations, which we do by finding the average pixel adding order. Doing so risks not finding the true optimized adding order and therefore the CDPP-optimized aperture can drift off in aperture space and not find the best aperture. We also have to take care not to engulf background objects, further complicating the CDPP optimization. The heuristic logistic regressive model will also revert to the Method #1 aperture in some cases. Considering these caveats, the actual aperture used on all targets is a combination of the four found apertures. For quarter 13, over 166,791 targets, the breakdown for aperture selection is

- Pure PRF image model-derived = 7.6%;
- Real pixel data-derived S/N-optimized Median = 21.3%;
- Real pixel data-derived S/N-optimized Union = 22.5%; and
- Real pixel data-derived CDPP-optimized = 48.6%.

Of the 7.6% that use Method #1, 0.9% are due to the logistical regressive heuristic model identifying when the Method #2 aperture is poor despite it having the lowest CDPP. Note also that 21.3% chose the S/N-optimized median aperture over the CDPP-optimized aperture. This is again due to the CDPP-optimized aperture not being optimized solely on CDPP, and also because many times the S/N- and CDPP-optimized apertures have very similar CDPP values. The fractional change in aperture between Method #1 and Method #2 does have an annular shape across the FOV, where targets further away from the ring of best focus require more correction. This is in agreement with expectations. We rely on a static PRF model that was derived from data acquired in a particular state of focus during commissioning in May 2009; our image models are therefore generally better in regions that more closely match that state. A graphical representation of the selected aperture over the FOV is shown in Figure 5. We see the dominance of the CDPP-optimized aperture selection and an even distribution over the FOV for all aperture types but with a slight preference for Method #2 apertures near the center of the FOV. There are no “Haloed Saturated” apertures; this aperture type is reserved for K2 data, as discussed in Section 9.

The full *Kepler* data set of all 18 Quarters of data has been reprocessed using the newer aperture selection method discussed in this paper. Figure 6 shows a histogram of the overall improvement in CDPP for simple aperture photometry light curves using the newer method. The old method used solely Method #1. The revised apertures reduce CDPP for the vast majority of targets and quarters, with 77% showing an improvement, 13% showing some degradation, and 10% showing no change. The majority of those showing no change are due to the algorithm choosing Method #1. In 14% of cases CDPP is reduced by 10% or more. The next component in the *Kepler* processing is PDC (Stumpe et al. 2012), so light curves generated with simple aperture photometry in PA are not the final light curves generated by the *Kepler* mission. In PDC we remove systematic trends in the data, so the CDPP values generated in PA are much larger than those seen by the transiting planet search (TPS) algorithm. But with the improvements in this paper, PDC is now beginning with lower noise light curves than before, so the resultant PDC light curves.
have even lower CDPP, which directly translates into better sensitivity to transit detection in TPS. We have shown that the decrease in CDPP does indeed propagate through PDC.

9. Application to K2 Data

In the K2 mission (Howell et al. 2014) periodic thruster firings are used to compensate for the loss of two of the four reaction wheels. These firings occur up to once every 6 hours, or every 12th long cadence, and the resulting barrel axis roll motion can be well over 4 arcseconds for the outer CCD channels. The roll drift and thruster firing repeats producing a characteristic “sawtooth” pattern in uncorrected light curves. Method #1 as implemented is unable to account for the K2 roll motion, so the Method #1 apertures are typically of low quality for the higher motion targets. Method #2, in contrast, can be directly applied to the K2 mission data and there are very few changes to the algorithm. To properly account for the large, fast motion we must find the aperture on every cadence. There are no changes in how the optimal aperture is selected but due to this large motion the union aperture is most often found to have the lowest CDPP, as shown in the following aperture selection breakdown for Campaign 4 with over 17,278 targets:

- Pure PRF image model-derived = 34.4%;
- Real pixel data-derived S/N-optimized Median = 4.3%;
- Real pixel data-derived S/N-optimized Union = 53.3%; and
- Real pixel data-derived CDPP-optimized = 8.0%.

Of the 34.4% that chose Method #1, 19.4% are saturated or custom targets where Method #1 is always chosen. This leaves only 15.0% where the Method #1 aperture was considered.

Figure 5. Chosen aperture for Kepler Quarter 13 over the FOV. The aperture selection is evenly distributed over the FOV, with a strong preference for the CDPP-optimized aperture.

(A color version of this figure is available in the online journal.)

Figure 6. Histogram of fractional CDPP improvement when comparing light curves produced from the new aperture method with those produced from the original Method #1. Positive values indicate improved photometric precision. This histogram is over all Kepler targets and each target is counted separately for each quarter. That is, the sum of the histogram is (number of targets) x (number of quarters).

(A color version of this figure is available in the online journal.)
better than the ones selected by Method #2. A graphical representation of the selected apertures over the FOV is shown in Figure 7. In the figure, we see an annular dependence where targets further from the barrel axis (i.e. where the roll motion is larger) strongly prefer the union aperture. Near the center of the FOV, where the roll motion is smaller, the Median and CDPP apertures are more preferentially chosen. For saturated targets and due to the large roll motion, Method #1 will not always

Figure 7. Chosen aperture for K2 Campaign 4 over the FOV. The annular dependence is due to the large roll motion about the instrument barrel axis. (A color version of this figure is available in the online journal.)

Figure 8. Mask scene, fitted PRF image, and four found apertures for EPIC ID 204115036. (A color version of this figure is available in the online journal.)
collect all relevant pixels. We therefore add a two pixel halo around each Method #1-derived saturated target aperture, these targets are labeled as “Haloed Saturated” in the figure and are evenly distributed. Note that the Method #1 apertures are clustered. The vast majority of these are custom targets where the Method #1 aperture is always chosen and custom targets tend to be clustered in regions of interest to specific Guest Observer Office funded projects.9

To illustrate the performance on K2 data, Figures 8 and 9 give the pixel scene, apertures, and diagnostic curves for EPIC ID 204115036, processed during Campaign 2 on module output 2.3. These two figures are directly analogous to the example figures given above for Kepler data. Here, we clearly see the much larger motion (red centroid dots) in the K2 mission. The Method #1 aperture in Figure 8 does not properly account for the large roll motion and therefore finds an aperture (the plus symbols) that is slightly off of the target centroid. The Method #2 median and CDPP-optimized apertures (the x’s and circles, respectively) do find the proper center but do not fully account for the motion. The union aperture, however, is both properly centered and accounts for the full motion of the target centroid. Three bright background objects are very near to the target. Care must be taken that these objects do not contaminate the target’s optimal aperture. In Figure 9 we see the diagnostic curves. Note that the “average” S/N curve no longer peaks at the S/N-optimized median aperture size (i.e., the red circle in the upper right plot is not on the blue curve peak). This is indicative of high centroid motion and that the S/N-optimized median aperture is changing over time and therefore the “averaged” S/N-optimized curve will not account for the motion or find a good optimal aperture. This can also be seen in the large error bars, which represent the standard deviation in the S/N curves over all cadences. The spread is large compared to the sharp peak in the curve. We also see in the lower right plot that the CDPP curve (magenta curve) has minima as we engulf the background objects, whereas the combined CDPP and S/N curve does not (cyan curve). But neither the median nor the CDPP-optimized aperture is the best aperture. The upper left plot shows that the union aperture has, by far, the

9 See http://keplerscience.arc.nasa.gov for details on the Guest Observer program.
lowest CDPP and is the best choice. The large image motion relative to the optimal aperture results in the uncorrected light curves, flux fraction, and crowding metric time series showing a strong “sawtooth” behavior. We could increase the aperture size until no sawtooth is present but doing so would cause the S/N to decrease and the broad spectrum noise to increase to unacceptable levels; we would simply be trading a distinct noise source for broad spectrum noise. The distinct sawtooth signal can be removed in PDC, whereas broad spectrum noise is very difficult to remove. We therefore are best served by minimizing the noise at the expense of a stronger sawtooth signal.

The above discussion invites the natural conclusion that adaptive apertures that adjust on every cadence might be optimal for K2 data. The authors do not disagree with this observation. Unfortunately, the K2 mission data processing pipeline is an adaptation of the Kepler processing pipeline, which was committed to a single fixed optimal aperture per quarter. The time and resources were not available to modify the pipeline to allow for adaptive apertures. Given this constraint we have found a smaller aperture to be generally better.

In Section 4.1 we mentioned that the fitted image model parameters can be used to identify errors in the Kepler catalog. Two examples of this were found in K2 Campaign 4 processing. There is a group of targets whose measured flux is more than twice that expected from their EPIC magnitudes. Figure 10 shows that these targets fall into spatial groups that are aligned with R.A. and decl. (shown as blue markers), rather than focal plane coordinates and strongly indicates that the cause of this anomaly is catalog errors. The source of this error is unknown and is not correlated with any particular Kepler target type. The other error identified in the figure is the scatter of red markers indicating targets whose brightness is overestimated. These targets are strongly correlated with “JHK” and “J” stars that were discovered to be K/M dwarfs that are not well represented in the data used to create the conversion from JHK/J to Kepler magnitude (Howell et al. 2010). Method #2 attempts to correct for these errors and find a true optimal aperture, but the PRF modeling can only go so far, so identifying these catalog errors is of benefit to the mission. Star catalogs generated for future campaigns will try to account for and fix these errors.
10. Conclusion

The above described optimal aperture finding method has been shown to provide superior photometry for Kepler data as summarized in Figure 6. We can now also be shown to provide superior photometry for the Kepler and K2 input catalogs. This has already proved useful in refinement of the K2 campaign field catalogs, improving photometry, and increasing the great wealth of knowledge obtainable with the mission.

The new approach has also been proven robust at finding apertures in K2 data and helping to mitigate the larger motion-induced systematics in the photometry. The older Method #1 does not perform well with the larger motion, so the new method is absolutely critical for extracting high-quality photometry with K2. Dynamic and moving apertures could potentially provide even better photometry but the mission did not have the resources to implement this further improvement.

The next processing component in the Kepler pipeline is PDC, which has been modified for use with K2 data (Van Cleve et al. 2016). It has been shown to remove up to 99% of the “sawtooth” systematic pattern in the data. Nevertheless, the PDC component of the K2 pipeline is probably the area where the most improvement in photometry could be achieved.

As with any working method, improvements could still be made. One improvement relates to how we find a single aperture from the per cadence optimal apertures in Section 4, Equation (3). As of now we find two apertures: (1) a 50% union, or median aperture and (2) a 95% union. Instead, we could allow the union percentile to be a parameter that we optimize versus S/N or CDPP. This would not require us to first find an average pixel adding order, which introduces an added approximation and sometimes an awkward pixel order. Another potential improvement is to identify and fit for uncataloged background objects. As of now, we use the KIC, UKIRT, and EPIC catalog R.A., decl., and magnitude with the motion polynomials (Equation (5)) of each object to model the scene and image motion. We do know that dim, uncataloged background objects do exist and can contaminate the scene fitting. A method to auto-find these objects would aid in more complete scene modeling. A final improvement would be a better characterization of the intra-pixel response variations. The PRF model obtained during commissioning has been proven to perform well for Kepler. However, we do know that it does not capture all variations in the flux in the presence of motion. Commissioning and PRF modeling are critical to any CCD-based observations and the limits of the PRF model are probably a limiting factor in the performance of our method.

Funding for the Kepler and K2 Missions is provided by NASA’s Science Mission Directorate. The authors acknowledge the efforts of the Kepler Mission team for obtaining the calibrated pixel and light curve data used in this publication. These data products were generated by the Kepler Mission science pipeline through the efforts of the Kepler Science Operations Center and Science Office. The Kepler Mission is lead by the project office at the NASA Ames Research Center. Ball Aerospace built the Kepler photometer and spacecraft, which is operated by the mission operations center at LASP. These data products are archived at the Mikulski Archive for Space Telescopes/NASA Exoplanet Science Institute. We thank the hundreds of people whose efforts made Kepler’s grand voyage of discovery possible. We especially want to thank the Kepler Science Operation Center and Science Office staff who design, build, and operate the Kepler Science Pipeline for putting their hearts into this endeavor.

Facility: Kepler.

References

Brown, T. M., Latham, D. W., Everett, M. E., & Esquerdo, G. A. 2011, AJ, 142, 112

Bryson, S. T. 2008, Target Optimal Aperture Selection, Tech. Rep. KADN-26108, NASA KPO@Ames Design Note

Bryson, S. T., Tenenbaum, P., Jenkins, J. M., et al. 2010a, ApJL, 713, 97

Bryson, S. T., Jenkins, J. M., Klaus, T. C., et al. 2010b, Proc. SPIE, 7740, 77401D

Clarke, B. D., Allen, C., Bryson, S. T., et al. 2010, Proc. SPIE, 7740, 774020

Howell, S. B., Rowe, J. F., Bryson, S. T., et al. 2012, ApJ, 746, 123

Howell, S. B., Sobeck, C., Haas, M., et al. 2014, PASP, 126, 398

James, G., Witten, D., Hastie, T., & Tibshirani, R. 2014, An Introduction to Statistical Learning, Springer Texts in Statistics (Springer)

Jenkins, J. M. 2002, ApJ, 564, 495

Jenkins, J. M. 2004, Kepler: A Search for Terrestrial Planets algorithm Theoretical Basis Documents, Tech. Rep. KSC-21008, NASA Kepler Design Document

Jenkins, J. M., Caldwell, D. A., Chandrasekaran, H., et al. 2010a, ApJL, 713, L87

Jenkins, J. M., Chandrasekaran, H., McCauliff, S. D., et al. 2010b, in SPIE Conference Series, 7740

Kim, D., Sra, S., & Dhillon, I. S. 2013, Optimization Methods and Software, 28, 1012

Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, MNRAS, 379, 1599

McCauliff, S., Cote, M. T., Girouard, F. R., et al. 2010, in SPIE Conference Series, 7740, 77400M

Quintana, E. V., Jenkins, J. M., Clarke, B. D., et al. 2010, in SPIE Conference Series, 7740

Seader, S., Jenkins, J. M., Tenenbaum, P., et al. 2015, ApJS, 217

Smith, J. C., Stumpe, C., Cleve, J. E. V., et al. 2012, PASP, 124, 1000

Stumpe, M. C., Smith, J. C., Catanzarite, J. H., et al. 2014, PASP, 126, 100

Stumpe, M. C., Smith, J. C., Cleve, J. E. V., et al. 2012, PASP, 124, 985

Tenenbaum, P., Bryson, S. T., Chandrasekaran, H., et al. 2010, in SPIE Conference Series, 7740

Twicken, J. D., Clarke, B. D., Bryson, S. T., et al. 2010, in SPIE Conference Series, 7740

Van Cleve, J. E., & Caldwell, D. A. 2009, Kepler Instrument Handbook, Tech. Rep. KSCI-19033, NASA Ames Research Center Kepler Mission

Van Cleve, J. E., Howell, S. B., Smith, J. C., et al. 2016, PASP, 128, 965

Wu, H., Twicken, J. D., Tenenbaum, P., et al. 2010, in SPIE Conference Series, Vol. 7740, 42W