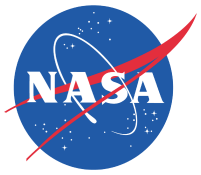

Planet Detection Metrics:

Automatic Detection of Background Objects Using the Centroid Robovetter

KSCI-19115-001

Fergal Mullally

May 22, 2017

NASA Ames Research Center
Moffett Field, CA 94035

Prepared by: Fergal Mullally Date: 5/22/17
Fergal Mullally, *Kepler* Science Office

Approved by: Natalie Batalha Date: 5/22/17
Natalie Batalha, *Kepler* Project Scientist

Approved by: Michael R. Haas Date: 5/22/17
Michael R. Haas, *Kepler* Science Office Director

Document Control

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

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

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

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

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

The correct citation for this document is: Mullally, F. 2017, *Planet Detection Metrics: Automatic Detection of Background Objects using the Centroid Robovetter* (KSCI-19115-001).

DOCUMENT CHANGE LOG

CHANGE DATE	PAGES AFFECTED	CHANGES/NOTES
May 22, 2017		Original release

Contents

1 Introduction **6**

2 Algorithm Inputs **7**

3 The Algorithm **8**
3.1 Identifying Valid Difference Images 9
3.2 Ensuring Sufficient Data for Reliable Analysis 11
3.3 Identifying Obvious Background Objects 11
3.4 Computing Valid Centroids 11
3.5 Estimating Centroid Offset Uncertainty 12
3.6 Identifying False Positives Based on Centroid Offsets 14

4 Algorithm Weaknesses and Mitigations **15**
4.1 Saturated Stars 15
4.2 Difference-Image Quality Metric 15
4.3 Bright, Nearby Variable Star 15
4.4 Bright, Far Away Star 15
4.5 Centroid Significance 16

5 Performance **16**
5.1 Agreement with Catalog of Burke et al. (2014) 16
5.2 Testing Against Simulated Transits 16
5.3 DR24 Minor Flag Names 18

6 Conclusions **20**

7 Acknowledgements **21**

Abstract

We present an automated method of identifying background eclipsing binaries masquerading as planet candidates in the *Kepler* planet candidate catalogs. We codify the manual vetting process for *Kepler* Objects of Interest (KOIs) described in Bryson et al. (2013) with a series of measurements and tests that can be performed algorithmically. We compare our automated results with a sample of manually vetted KOIs from the catalog of Burke et al. (2014) and find excellent agreement. We test the performance on a set of simulated transits and find our algorithm correctly identifies simulated false positives ≈50% of the time, and correctly identifies >99% of simulated planet candidates.

1 Introduction

In support of its primary goal of determining the frequency of Earth-size planets around sun-like stars, the *Kepler* mission produces regular catalogs of newly discovered planet candidates (Borucki et al., 2011a,b; Batalha et al., 2013; Burke et al., 2014; Rowe et al., 2015; Mullally et al., 2015; Coughlin et al., 2016). With each catalog, our techniques for identifying false positives improved, but were long dominated by a manual process involving many trained astronomers inspecting a series of metrics and searching for evidence that a given *Kepler* Object of Interest (KOI) was not a planet. This team is known as the Threshold Crossing Event Review Team, or TCERT[1]. A detailed description of the manual approach is given in Rowe et al. (2015), and some estimates of the repeatability of the decisions is given in Mullally et al. (2015).

The true reliability of these catalogs is still under active study (Thompson et al., 2017). Fressin et al. (2013) estimated the false positive rate for KOIs vetted as planet candidates to range from 10–20%, while Désert et al. (2015) found a rate of 8.8%. Mullally et al. (2015) warns that the reliability of the long period ($\gtrsim 200$ day) sample may be significantly worse.

One line of evidence that TCERT considers is whether the pixels that change in brightness during a transit are consistent with the hypothesis that the transit is occurring on the target star. If the sky locations of the target star and transit source are well resolved, this is a relatively easy measurement; for unresolved sources we rely on the measured photometric centroid shift during transit.

Eclipsing binaries (EBs) can contaminate the planet sample at all planet radii. If the EB shares a similar line-of-sight to another star, flux dilution (i.e., the fact that many stars may contribute light to the aperture, but only one star dims during the transit) may reduce the measured transit depth to that expected from a much smaller body. Accurately vetting the catalog to identify such false positives is a key step in making an accurate estimate of the frequency of planets at all radii (e.g., Burke et al. (2015)).

[1]The KOIs are drawn from the set of Threshold Crossing Events (TCEs) identified by the *Kepler* pipeline (Jenkins, 2017).

The algorithm introduced here, dubbed the Centroid Robovetter, automates this process of testing for background eclipsing binaries (BGEBs)[2]. First used in Mullally et al. (2015) in a supervised fashion, a slightly improved version was applied autonomously for the Q1-Q17 DR24 catalog (Coughlin et al., 2016). Together with the Centroid Robovetter, ephemeris matching as described in Coughlin et al. (2014), and the metrics introduced by Thompson et al. (2015) and Mullally et al. (2016), Coughlin et al. (2016) uses a fully automated vetting pipeline for KOIs. An automated pipeline is faster, more objective, easier to test, and facilitates more accurate estimates of catalog completeness, an important ingredient in estimating occurrence rates. With a few additional modifications, this same Centroid Robovetter was used in the creation of the Q1-Q17 DR25 catalog (Thompson et al., 2017).

2 Algorithm Inputs

The Centroid Robovetter tries to measure statistically significant offsets between the target star and the source of the transit event. It relies on measuring centroid offsets based on fits to difference images for in- and out-of-transit cadences produced by the Data Validation module of the *Kepler* pipeline (DV, Wu et al., 2010). The technique, and the data products used, are described in detail in Bryson et al. (2013).

Images of the star during transit are created by summing the pixel images for in-transit cadences during a quarter (see Koch et al., 2010, and Haas et al., 2010, for an overview of spacecraft operations). Out-of-transit images are constructed in a similar manner by combining an equal number of cadences on either side of the transit. Difference images are created by subtracting the in-transit image from the out-of-transit (OOT) image. One difference image is created for every quarter in which one or more transits occur.[3]

DV then computes the shift in the photometric centroid during transit by fitting a model of the *Kepler* Pixel Response Function (PRF, Bryson et al., 2010b) to the (per-quarter) difference and OOT images. The mean shift and its significance is then computed, and the likelihood that the transit is due to a background object is calculated in a manner described in §6.3 of Bryson et al. (2013). DV also calculates the offset between the difference-image centroid and the position recorded in the *Kepler* Input Catalog (KIC, Brown et al., 2011).

There are a number of reasons why the centroid offset measured by DV should not be taken at face value:

- Our fit uncertainties are often dramatically underestimated (see § 3.5), so the significance of the shift should be measured from the scatter in multiple measurements. DV reports a significance even when the number of measurements is very low.

[2]Our definition of BGEB includes any transiting or eclipsing system that is not physically associated with the target, and includes signals better described as foreground events.

[3]To avoid corrupting the image, certain transits are excluded from the difference images (see Bryson et al. (2013)).

- For low signal-to-noise transits ($\lesssim 12$), the computed difference image is noise dominated, and the resulting centroid estimate is untrustworthy because it is often dominated by a single bright pixel. DV does not distinguish between high- and low-quality centroid estimates when computing the significance of a centroid offset.
- DV reports centroid values for saturated images, but these values are unreliable and should not be used.
- If the source of a transit signal is a background star that is incompletely captured in the mask (i.e., the set of pixels collected for a given target; see Bryson et al., 2010a), PRF fitting may fail to converge and miss an obvious false positive.
- In crowded fields, the OOT centroid may be systematically biased by the light from a nearby star. This leads to a large measured offset between the OOT and difference-image centroids, falsely suggesting the transit did not occur on the target star. While this problem can be mitigated by using the offset from the KIC position, which is less sensitive to influence from nearby stars, this KIC offset suffers from systematic errors that depend in detail on the quarters in which the centroids are measured. Accurately determining the significance of a KIC offset is challenging.

Any automatic technique must account for these challenges to accurately identify false positives (FPs). To maximize the value of a catalog, it also must reliably identify corner cases where the identification may be suspect, so that additional oversight can be directed at the weakest identifications. In Mullally et al. (2015), such cases were vetted manually. In Coughlin et al. (2016) they are flagged for attention, but marked as planet candidates. This is consistent with the TCERT philosophy of "innocent until proven guilty" (Mullally et al., 2015), where strong evidence is required that a KOI is not a planet before marking it as a false positive. This maximizes the fraction of detectable planets in the final catalog, at the cost of incorrectly including some non-planets.

3 The Algorithm

The algorithm presented here is an implementation of the techniques suggested in Bryson et al. (2013). We mimic the manual steps detailed in that paper in an automatic fashion that speeds the process while removing human subjectivity. The algorithm proceeds in three main steps:

1. Identifying and rejecting low-quality difference images (§ 3.1),
2. Identifying sources clearly resolved from the target star (§ 3.3), and
3. Measuring statistically significant centroid motion during transit (§ 3.4–3.5).

We discuss the steps in detail below, and the MATLAB source code for the algorithm is available at `https://sourceforge.net/projects/keplercentroidrobovetter/`. Note that the code takes intermediate data products produced by the *Kepler* pipeline that have not been made public, and as such can not be run without considerable modification. The code is provided for documentation purposes only.

3.1 Identifying Valid Difference Images

We first check the images for saturation. Saturated pixels show near zero flux in the difference image and these images are not used. Stars are identified as saturated if they are listed in the KIC as being brighter than $K_{\rm p} = 11.5$. See § 5 of Bryson et al. (2013) for further discussion of the difference images of saturated stars.

Next, we check that the difference image is not noise dominated. We use a simple but effective method called connected-component labeling (Rosenfeld et al., 1966) to find contiguous clusters of three or more pixels brighter than some threshold, which we call labeled regions. The threshold was chosen (by trial and error) as the mean sky flux plus the root mean square (rms) scatter of a set of pixels. We compute both values by iteratively measuring the rms scatter and rejecting large values. This approach is fast and reliable, although it tends to be overly conservative, setting the threshold for a good difference image slightly higher than the

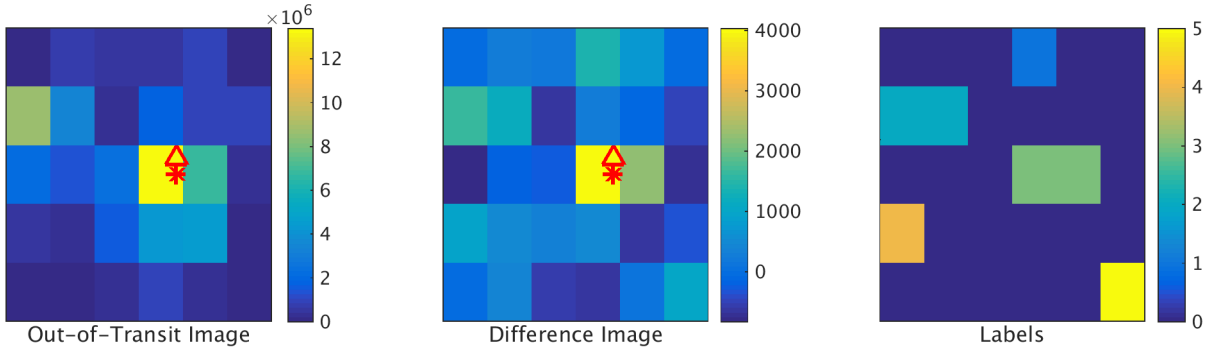

Figure 1: An example of a difference image misclassified as having a low SNR. The image at left shows the distribution of flux out-of-transit (OOT) for KIC 12647577 in Quarter 8, while the center image shows the difference image flux. The difference image looks qualitatively similar to the OOT image, so we expect the centroid measurement to be trustworthy. Indeed the KIC position (cross), OOT centroid (plus sign), and difference-image centroid (triangle) all have very similar positions (in fact the cross and plus overlap to produce an asterisk). The small centroid shift is consistent across the other quarters where this transit was observed (not shown). The image at right shows labeled regions, the groups of contiguous pixels above threshold. No group has greater than three pixels, so the image is incorrectly labeled as having a low SNR. We describe the construction of these images in §§ 2 & 3.1.

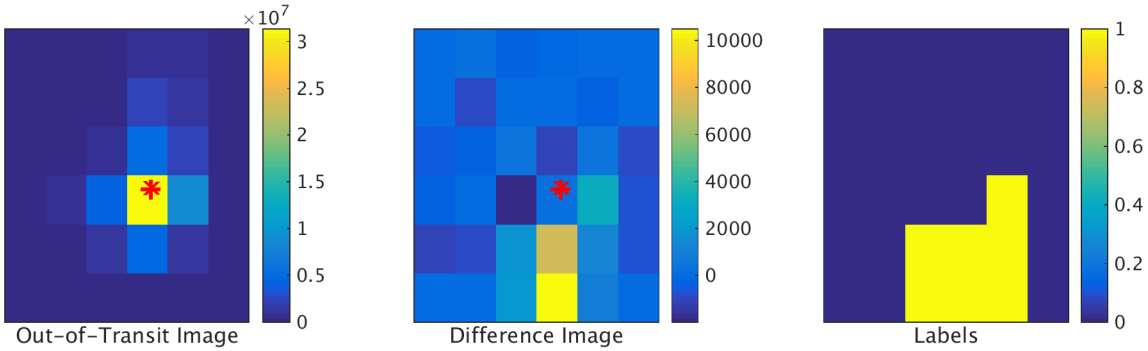

Figure 2: The background source of the transit for KIC 11100670 is incompletely captured by the mask in Quarter 5, and PRF fitting fails to converge in the difference image (evidenced by the lack of a triangle symbol). We identify this object as a false positive because the location of the brightest pixel at the edge of the mask in the difference image (center panel) is inconsistent with the KIC location of the star. The panels and symbols are defined as in Figure 1.

ideal (see Figures 1 and 2).

When we find a star in the difference image, we still reject the difference image if it fails any of the following three tests:

1. Bleed trails from saturated stars and column effects (Coughlin et al., 2014) are identified as labeled regions that are not sufficiently round (i.e., defined as a set of contiguous pixels covering twice as many columns as rows (or vice versa)). This test identifies systematics, but is insensitive to asymmetries in the PRF.

2. Difference images may contain deeply negative-valued pixels caused by imperfect detrending during creation. These negative values bias the centroid measurements, and such images should be ignored. We reject difference images where the flux from the most negative pixel *adjacent to the brightest pixel* is below a threshold,

$$F_{\rm min} < -0.5 F_{\rm max}$$

   where $F_{\rm max}$ is the flux in the brightest pixel.

3. In rare cases, the difference image may be inverted because the star is brighter during the transit. It is beyond the scope of the algorithm to determine whether such events are weak transits around variable stars, or a case where pulsations are incorrectly identified as a transit. If any quarter's difference image shows an inverted PRF, the KOI is flagged for attention (see § 5.3).

3.2 Ensuring Sufficient Data for Reliable Analysis

Not all quarters of data necessarily produce a difference image because some quarters may contain no transits, or some transits may be lost because they overlap with other spacecraft events, such as data downlinks, safe modes, or transits from another object in the system (Bryson et al., 2013). If a KOI has fewer than three good difference images, we conclude there is insufficient data to rely on a centroid measurement. Following our "innocent until proven guilty" philosophy, we treat such KOIs as planet candidates, and set flags to indicate why they passed (see § 5.3).

Our difference-image quality metric is too conservative, and often incorrectly flags quarters as having a low SNR. If the number of good images falls below threshold because of SNR issues, we mark the KOI as a planet candidate and set a warning flag.

3.3 Identifying Obvious Background Objects

Having identified good quality difference images, we next search for obvious background objects. When an eclipse happens on a nearby star that is incompletely captured by the mask, the PRF fit may fail (see Figure 2), so we need to search for such events directly. Our approach is again simple, but effective. We map the KIC position of the star onto the pixel grid, and measure its distance from the center of the brightest pixel in the mask. If this separation is greater than 1.5 pixels, the image indicates that the transit is on a background object. If two-thirds of the images indicate a background object, we mark the KOI as a false positive. If fewer than two-thirds, but > 3 quarters indicate a background object, we flag the KOI as requiring further attention. Similarly, if there are fewer than four good difference images and one indicates a background object, the KOI is flagged for attention. These threshold values were chosen by experiment to best reproduce the results of TCERT and avoid falsely incriminating KOIs.

3.4 Computing Valid Centroids

If a KOI passes the obvious background object test, we then check the evidence from the centroid fits. If the transit happens on the target star, the PRF centroid should not move during transit, and the OOT and difference-image centroids should agree. If they disagree, that is evidence that the transit is occurring on a background object. Again, we insist on at least three measurements before making a decision to fail a KOI.

DV measures the centroid offset per quarter in two ways: by comparing the difference image and out-of-transit centroids (the OOT offset), and comparing the difference-image centroid to the KIC position of the star (the KIC offset). The OOT offset has smaller biases in uncrowded fields, but the KIC offset performs better when more than one star contributes significant flux to the mask. In these cases, the OOT centroid can be "pulled" away from the true position by the contaminating flux. This can lead to the OOT centroid lying far from the photocenter of

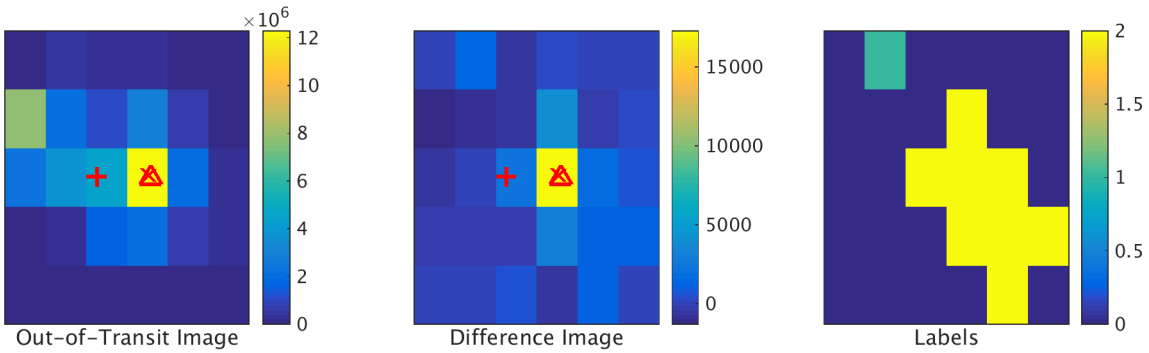

Figure 3: Two stars occupy the mask for KIC 9958962. The OOT centroid (plus sign) is biased by the second star and the best fit solution lies in the wings of the target. In cases like this we recognize that the OOT centroid is inconsistent with the KIC and fall back on the KIC offset. These images are based on the transits of the second KOI (period of 90 days) in Quarter 3. The panels and symbols are defined as in Figure 1.

the star, while the difference-image centroid is consistent with the transit being on the target, giving the false impression of a large shift during the transit (see Figure 3 for an example).

To determine which centroid to use, we check that the OOT centroid is consistent with the KIC position. If it is not, we fall back on using the KIC position. To determine what level of disagreement is significant, we looked at the distribution of offsets between the OOT centroid and the KIC position for 16,000 measurements across 1,300 stars in Burke et al. (2014). We show the two one-dimensional distributions in Figure 4. The distributions are well modeled in the core by a Gaussian distribution with a standard deviation of 40 millipixels, but there is a long tail with significantly larger values. If the offset for a given KOI is > 0.5 pixels in any one quarter, or the median offset is > 80 millipixels (i.e., twice the standard deviation of the fit), we fall back on the KIC offset, otherwise we trust the OOT offset.

We measure the mean offset and its statistical significance using an unweighted robust least squares fit to the column and row offsets for the good quarters. We use the robustfit[4] function in MATLAB, with the default bi-square weight function and the default tuning constant of 4.685. This is a departure from Bryson et al. (2013), who recommend weighting the fit by the formal centroid uncertainty. We find weighted fits can be unduly biased by outliers with low formal uncertainty.

3.5 Estimating Centroid Offset Uncertainty

Our offsets are measured by fitting the PRF model to the flux distribution across the pixels. The model, described in Bryson et al. (2010b), is based on commissioning data (Bryson et al.,

[4] http://www.mathworks.com/help/stats/robustfit.html

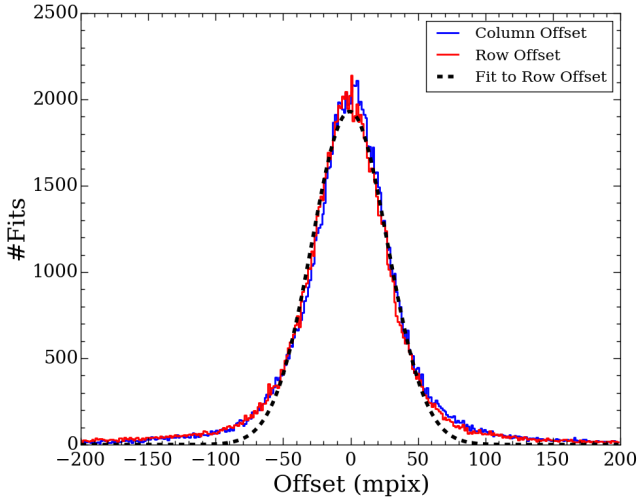

Figure 4: Distribution of offsets between the PRF fit to the OOT image and KIC position. The red solid line shows the difference in CCD row, the blue line shows the difference in CCD column. The black dashed line is a Gaussian fit to the row offsets. The core of the distribution is well modeled by a Gaussian, but there are more stars in the wings than the model predicts. The large offsets are due to OOT centroids being biased in crowded fields.

2017). The actual PRF changes with the temperature profile across the focal plane, itself a function of orbital phase and spacecraft orientation (see Figure 10 of Van Cleve et al., 2016). Because the actual PRF differs from the model, our formal position uncertainty underestimates our scatter by up to an order of magnitude.

To address this issue we compute three estimates of uncertainty and take the largest. Following Bryson et al. (2013), we estimate our uncertainty from both the rms scatter in the individual offsets, $\Delta c_{\rm rms}$, and from a bootstrap analysis. We compute the bootstrap uncertainty, $\Delta c_{\rm bs}$, from the rms of many distributions which are sampled with replacement from the set of measured centroids for a KOI. We find that the Bryson et al. (2013) approach of computing Q^2 distributions (where Q is the number of quarterly centroid measurements) does not produce repeatable results. Instead we sample every permutation of drawing Q samples from Q values, up to a limit of 50,000. We estimate the uncertainty on the uncertainty of the bootstrap centroid, $\Delta^2 c_{\rm bs}$, by dividing $\Delta c_{\rm bs}$ by the square root of the number of trials.

Our third measure of uncertainty, $\Delta c_{\rm formal}$, is the formal error on the average of the offsets, and is given by the hypotenuse of the individual formal uncertainties from the PRF fit. This guards against the rare case where the quarterly measurements randomly scatter closer to each other than the formal uncertainty suggests, biasing the previous two measurements into overestimating the significance of the offset.

We then choose our final value for centroid uncertainty as:

$$(\Delta c)^2 = \max(\Delta c_{\rm rms}, \Delta c_{\rm bs}, \Delta c_{\rm formal})^2 + (\Delta c_{\rm sys})^2, \quad (1)$$

where $\Delta c_{\rm sys}$ is a systematic uncertainty term set to 0.066" for OOT centroids and 0.160" for KIC centroids. The former number is taken from Bryson et al. (2013), and the latter corresponds to the $1\,\sigma$ width in Figure 4.

If only three or four quarterly offsets are available, $\Delta^2 c_{\rm bs}$ can be quite large. We refuse to fail a KOI if changing the value of Δc by $\pm 3\Delta^2 c_{\rm bs}$ would change the disposition from pass to fail (or vice versa). A flag is set to indicate when this occurs. This process is complex, but necessary to avoid failing valid planet candidates.

3.6 Identifying False Positives Based on Centroid Offsets

We define the offset significance as offset divided by uncertainty. We use the cuts in offset and significance suggested in Bryson et al. (2013) (and shown as red lines in Figure 5) to decide whether a KOI is marked as a planet candidate, a false positive, or as a possible false positive. In the Q1-Q16 catalog (Mullally et al., 2015), the possible false positives were subjected to human scrutiny. In the Q1-Q17 DR24 catalog they were marked as candidates, based on our tests with simulated events (see § 5.2).

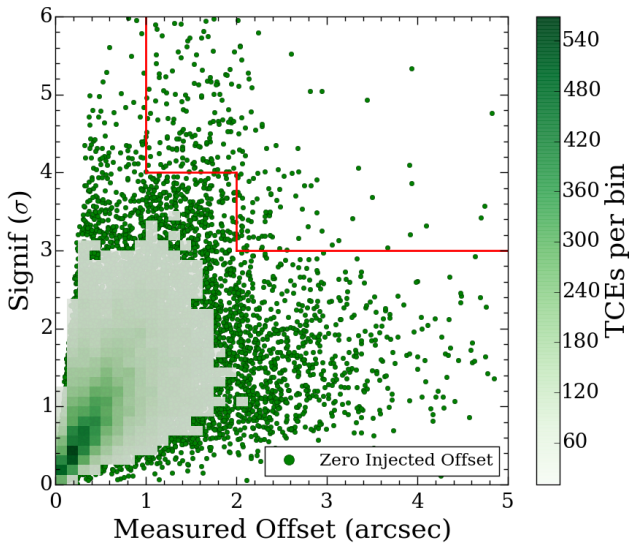

Figure 5: Distribution of measured offset and significance for transits injected on the target star (i.e., with no offset). Overly dense regions of this plot are shown as a shaded histogram for clarity. KOIs that land in the upper right region (as indicated by the red lines) are incorrectly marked as false positives using the limits set forth by Bryson et al. (2013).

4 Algorithm Weaknesses and Mitigations

4.1 Saturated Stars

The focus changes across the *Kepler* field of view, and the saturation magnitude is brighter in regions of poor focus. An obvious improvement in our algorithm would be to search for evidence of saturation in the pixel time series instead of applying a simple magnitude cut. Also, clearly resolved sources can often be identified visually in even heavily saturated cases, and a future version of the algorithm could extract useful information from the brightest stars.

4.2 Difference-Image Quality Metric

The weakest part of our approach is our difference-image quality metric. As shown in Figure 1, our metric sometimes marks difference images as having a low SNR even when the image quality is still good enough to produce trustworthy centroids. In Q1-Q17 DR24, 8% of KOIs were marked as having three or more difference images, but fewer than three that passed the SNR test. Some FPs were probably missed because we did not manually examine these cases.

4.3 Bright, Nearby Variable Star

If a bright, variable star is within the mask of the target star, the change in flux from the target during transit can be less than the change in flux from the variable star over the same interval. This variability can be mistaken for an eclipse signal from a background star. Because the timescale for variability is typically different than the transit period, the background variable is not usually bright in difference images from all quarters. We catch many such background variables by insisting that many quarterly difference images show a resolved source. This catches most examples, but some errors inevitably slip through. If there are only a small number of difference images available, it is not possible to distinguish between true background objects and variable stars based on difference images alone. Such cases are flagged as requiring additional attention.

4.4 Bright, Far Away Star

If the source of a false positive is a bright star that is many pixels away from the mask of collected pixels, then our connected component labeling will fail because the wings of the bright star's PRF contribute similar flux to every pixel in the mask and no contiguous group of pixels is brighter than the mean. Such FPs could in principle be detected in the difference images with an automated approach. Instead, these KOIs are found by the period-epoch matching algorithm Coughlin et al. (2014) or the ghostbuster metric (Thompson et al., 2017).

4.5 Centroid Significance

In crowded fields, we identify badly measured OOT centroids by comparing with the KIC position of the star. If the OOT centroid is significantly offset from the KIC position, we assume the field is crowded and rely on the KIC position. Our estimate of significance is based on the average for a large number of stars (see Figure 4) and may not be appropriate for individual cases. It is possible, although unlikely, that a disposition is wrong because the difference-image centroid is incorrectly compared to the OOT centroid in a moderately crowded star field.

As a final precaution, we note that the KIC position of a star can also be incorrect in crowded fields (or for stars with high proper motions), leading to a small fraction of objects incorrectly having large and significant KIC offsets. To guard against this problem, we flag, but do not fail an object if it shows a statistically significant offset from the KIC position. Approximately a dozen M stars suffer from this problem because their high proper motions result in a measurable difference between their catalog and observed positions.

5 Performance

5.1 Agreement with Catalog of Burke et al. (2014)

Published catalogs of manually vetted KOIs provide a labeled dataset against which the algorithm can be trained and tested. We divided the KOIs in the Q1-Q8 catalog of Burke et al. (2014) into training and test sets. The training set was used to help develop the algorithm, then the test set was used to measure performance. We show the results in Table 1. The algorithm correctly predicts the TCERT disposition over 98% of the time. Of the 18 cases where the algorithm incorrectly labeled the KOI (where correct is defined as agreeing with the TCERT designation), we judge three cases to be errors on the part of TCERT, and two cases were ephemeris matches (see § 4.4).

5.2 Testing Against Simulated Transits

Christiansen et al. (2016) tested the performance of the *Kepler* pipeline used by Coughlin et al. (2016) by injecting transits at the pixel level and measuring the rate at which those injections were recovered as a function of period, injected depth, etc. Their method is similar to Christiansen et al. (2015), but was run on all available data, instead of just one year's worth. Some 42% of the recovered events were injected at a location slightly offset from the target's catalog position (up to 4" or one *Kepler* pixel). These offset injections allow us to test the performance of our algorithm in a controlled fashion.

Table 1. Agreement with TCERT for Q1-Q8

	Robovetter PC	Needs Scrutiny	Robovetter FP
TCERT PC	900	69	4
TCERT FP	14	3	47

Note. — TCERT PC refers to the objects labeled as planet candidates in Burke et al. (2014) and TCERT FP refers to those labeled as false positives. Rows indicate the previously published classification, and columns are the results of the Centroid Robovetter.

We looked first at the non-offset injections. We would expect most of these transits to pass our tests. In Figure 5, we see that the bulk of our measured offsets are less than 1" and less than $2\,\sigma$ significant. If the distribution were Gaussian, and our uncertainties well measured, we would expect 96 injections to be measured as greater than $3\,\sigma$ due to noise alone. Instead we find 1064 exceed this threshold, indicating our error calculations are unduly optimistic. That said, our choice of threshold results in less than 1% of non-offset injected KOIs being marked as clearly failing due to a significant offset, indicating that the vast majority of bona-fide, on-target events are passed by the algorithm.

In Figure 6 we plot a two-dimensional histogram of the fraction of off-source injected transits that were correctly marked as false positives as a function of the injected MES (i.e., multiple-event statistic; see Jenkins (2017)) and offset distance. We expect that large-offset, large-MES injections will frequently be marked FP, but small offset, small MES injections will incorrectly pass because we can't detect the offset with a sufficient SNR. In Bayesian terms, this figure can be interpreted as the likelihood of detection of a false positive for a given MES and offset, the prior is the astrophysical probability of there being a background source, and the posterior is the probability of detecting a background source in *Kepler* data.

Our uncertainty model includes a systematic term that means we rarely fail anything with an offset less than 1". For transits injected with MES as high as 20 and offsets >1.5", we typically only detect $\approx 50\%$ of the injections as false positives. We find the offsets are typically measured correctly to within the uncertainties, but the significance is often too low to claim an unambiguous detection of a false positive.

Christiansen et al. (2016) injected transits with a distribution of parameters intended to best measure the recoverability of low SNR events. They injected few high SNR (> 20) transits, so the right-hand side of the plot is strongly affected by Poisson noise due to the lack of injected

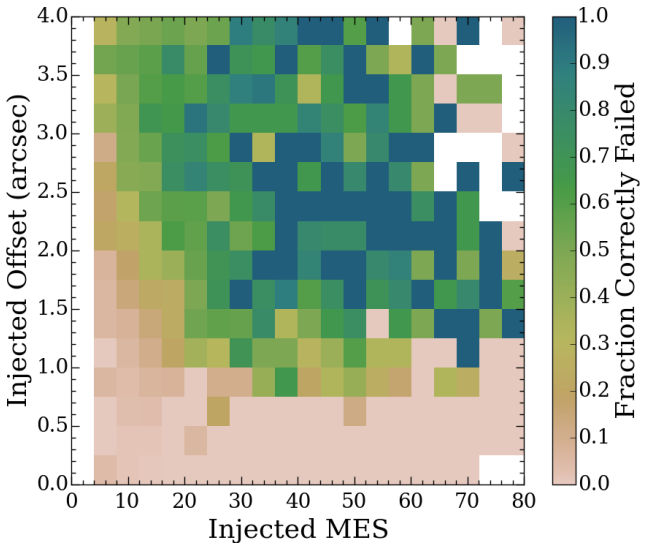

Figure 6: Fraction of transits injected with an offset from the target star that were correctly labeled as false positive, as a function of injected offset and MES. Bins with MES > 30 typically have fewer than 10 injections per bin. The detection rate doesn't rise much above 50% even for offsets approaching 4" (i.e., one *Kepler* pixel) and injected MES of 20 (i.e., the typical signal from three transits of a 1.7 $R_\oplus$ planet around a bright, quiet solar-radius star).

events in that region of parameter space.

We note that transits were injected using the PRF models of Bryson et al. (2010b) that were created using commissioning data (Bryson et al., 2017). The true PRF is known to vary by up to 10% from these models (Van Cleve et al., 2016). In this regard, our simulations probably overestimate our performance.

With these caveats in mind, we can summarize the results of the transit injection test by stating that 99% of on-source transits are preserved by the algorithm. For MES > 20 we correctly identify >50% of cases where the source of the transit is 1.5–4" from the source, and almost none of the cases where the source of the transit is between 0–1".

5.3 DR24 Minor Flag Names

We list here the mnemonics used in the DR24 minor flags table to describe the decision tree for the Centroid Robovetter, and provide a brief explanation of their intended meaning. The mnemonics help understand how the final decision on the disposition of a KOI was determined. Combinations of flags are often used to document a decision. For example, the flags FP, EYEBALL, KIC_OFFSET and SIGNIF_OFFSET in combination indicate a star in a crowded

field for which a significant offset between the KIC position and the difference-image centroid was detected.

Bit 0 Value 1, FP: The KOI is a false positive because the transit did not occur on the target star. In Coughlin et al. (2016), this flag is ignored if Bit 1 is also set.

Bit 1 Value 2, EYEBALL: The disposition of this KOI is uncertain and warrants further scrutiny. In Coughlin et al. (2016), no KOIs are marked as FP if this flag is set.

Bit 2 Value 4, KIC_OFFSET: The centroid offset was measured relative to the star's recorded position in the KIC, not the OOT centroid. The KIC position is a less accurate estimate of the stellar location than the OOT centroid in sparse fields, but more accurate in crowded fields. If this is the only flag set, there is no cause for concern for the KOI. This flag does not mean the KIC offset is significant, only that the KIC offset was used in preference to the OOT offset.

Bit 3 Value 8, SIGNIF_OFFSET: FP flag was set because there was a statistically significant shift in the centroid during transit.

Bit 4 Value 16, CLEAR_APO: FP flag was set because the transit occurs on a star that is spatially resolved from the target.

Bit 7 Value 128, INVERT_DIFF: One or more difference images were inverted, meaning the difference image claims the star got brighter during transit. This is usually due to a problem with the generation of the difference image due to variability of the target star. When this flag is set, the KOI is marked as requiring further scrutiny.

Bit 10 Value 1024, SATURATED: Star is saturated. The assumptions of the Centroid Robovetter break down for saturated stars, and all such KOIs are marked as requiring further scrutiny.

Bit 11 Value 2048, TOO_FEW_QUARTERS: Fewer than three difference images with sufficiently high SNR are available, and very few tests are applicable to the KOI. If set in conjunction with Bit 4 (CLEAR_APO), the source of the transit may be on a star clearly resolved from the target.

Bit 12 Value 4096, FIT_FAILED: Transit fit failed to converge in DV and no difference images were created. This flag is typically set for very deep transits of eclipsing binaries. If this flag is set, the KOI is passed due to lack of evidence.

Bit 13 Value 8192, CROWDED_DIFF: More than one potential stellar image found in the difference image. The EYEBALL flag is always set in conjunction with the CROWDED_DIFF

flag.

Bit 14 Value 16384, TOO_FEW_CENTROIDS: The PRF fit does not always converge, even in high SNR difference images. This flag is set if there are more than three high SNR difference images detected, but centroid offsets are recorded for fewer than three. If this flag is set, the KOI is passed due to lack of evidence.

Bit 15 Value 32768, CENTROID_SIGNIF_UNCERTAIN: The uncertainty in the offset significance is enough that the algorithm can not confidently say that the significance is either above or below the threshold. This flag typically gets set for KOIs with only three or four recorded centroids (see § 3.5).

The Centroid Robovetter algorithm presented here was used to help identify false positives in the planet catalog of Coughlin et al. (2016). A slightly earlier version was used in the catalog of Mullally et al. (2015). In Mullally et al. (2015), KOIs marked as needing further attention were scrutinized by two or more human vetters who made the final decision as to the disposition. Coughlin et al. (2016) is an entirely automated catalog, and we instead relied entirely on the automated decision. In keeping with the principle of innocent until proven guilty, KOIs marked as possible false positives, but needing further attention, were kept as planet candidates (but other parts of the Robovetter may fail these KOIs for other reasons). Although our injection tests show this is the correct thing to do for stars with small centroid offsets, some clearly resolved background eclipsing binaries will have a disposition of planet candidate. These possible false positives can be identified by a bit string value of 19 (FP, EYEBALL, CLEAR_APO) or 2067 (FP, EYEBALL, CLEAR_APO, TOO_FEW_QUARTERS) in the DR24 KOI catalog.

6 Conclusions

Our automated method of vetting KOIs meets our goal of reproducing the result of the manual TCERT approach with high fidelity. In addition, it allows us to test the algorithm against simulated transit events, something which would be difficult and time consuming to do otherwise. We find the approach has high completeness, in that it fails <1% of all simulated on-target events, but a lower effectiveness, in that $\approx$ 50% of off-target injections with injected MES $\sim$ 30 and offsets of 1–4" are correctly identified as such. This lower effectiveness is a consequence of our design choice to maximize completeness. Although half the background objects will be missed by the algorithm, the probability of there being a nearby background eclipsing binary must also be factored into the estimated false positive rate (Torres et al., 2011; Morton, 2012). Our approach will likely be applicable to other transit searches such as TESS (Ricker et al., 2014) or Plato (Rauer et al., 2014).

7 Acknowledgements

We are deeply indebted to Y. Shah, J. L. Christiansen, S. E. Thompson, J. L. Coughlin, G. A. Esquerdo, F. R. Girouard, M. R. Haas, C. E. Henze, D. W. Latham, A. Ofir, V. Van Eylen, & Ji-Wei Xie for manually vetting many objects to assess and improve the performance of the Centroid Robovetter. We thank Joe Twicken for constructive comments on an early draft of this manuscript. Funding for the *Kepler* mission is provided by NASA's Science Mission Directorate. This research has made use of the NASA Exoplanet Archive, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program.

References

Batalha, N. M., Rowe, J. F., Bryson, S. T., et al. 2013, *ApJS*, 204, 24
Borucki, W. J., Koch, D. G., Basri, G., et al. 2011a, *ApJ*, 728, 117
Borucki, W. J., Koch, D. G., Basri, G., et al. 2011b, *ApJ*, 736, 19
Brown, T. M., Latham, D. W., Everett, M. E., & Esquerdo, G. A. 2011, *AJ*, 142, 112
Bryson, S. T. 2017, Commissioning Data for Measurement of the Pixel Response Function and Focal Plane Geometry (KSCI-19100-001)
Bryson, S. T., Jenkins, J. M., Gilliland, R. L., et al. 2013, *PASP*, 125, 889
Bryson, S. T., Jenkins, J. M., Klaus, T. C., et al. 2010a, *Proc. SPIE*, 7740, 77401D
Bryson, S. T., Tenenbaum, P., Jenkins, J. M., et al. 2010b, *ApJL*, 713, L97
Burke, C. J., Bryson, S. T., Mullally, F., et al. 2014, *ApJS*, 210, 19
Burke, C. J., Christiansen, J. L., Mullally, F., et al. 2015, *ApJ*, 809, 8
Christiansen, J. L., Clarke, B. D., Burke, C. J., et al. 2015, *ApJ*, 810, 95
Christiansen, J. L., Clarke, B. D., Burke, C. J., et al. 2016, *ApJ*, 828, 99
Coughlin, J. L., Mullally, F., Thompson, S. E., et al. 2016, *ApJS*, 224, 12
Coughlin, J. L., Thompson, S. E., Bryson, S. T., et al. 2014, *AJ*, 147, 119
Désert, J.-M., Charbonneau, D., Torres, G., et al. 2015, *ApJ*, 804, 59
Fressin, F., Torres, G., Charbonneau, D., et al. 2013, *ApJ*, 766, 81
Haas, M. R., Batalha, N. M., Bryson, S. T., et al. 2010, *ApJL*, 713, L115
Jenkins, J. M. 2017, *Kepler* Mission Data Processing Handbook (KSCI-19081-002)
Koch, D. G., Borucki, W. J., Basri, G., et al. 2010, *ApJL*, 713, L79
Morton, T. D. 2012, *ApJ*, 761, 6
Mullally, F., Coughlin, J. L., Thompson, S. E., et al. 2015, *ApJS*, 217, 31
Mullally, F., Coughlin, J. L., Thompson, S. E., et al. 2016, *PASP*, 128, 074502
Rauer, H., Catala, C., Aerts, C., et al. 2014, *Experimental Astronomy*, 38, 249
Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2014, *Proc. SPIE*, 9143, 914320
Rosenfeld, A., & Pfaltz J. L. 1966, *J. ACM*, 13, 471
Rowe, J. F., Coughlin, J. L., Antoci, V., et al. 2015, *ApJS*, 217, 16
Thompson, S. E., et al. 2017, *in preparation*
Thompson, S. E., Mullally, F., Coughlin, J., et al. 2015, *ApJ*, 812, 46
Torres, G., Fressin, F., Batalha, N. M., et al. 2011, *ApJ*, 727, 24

Van Cleve, J. E., Christiansen, J. L., Jenkins, J. M., et al. 2016, *Kepler* Data Characteristics Handbook (KSCI-19040-005)
Wu, H., Twicken, J. D., Tenenbaum, P., et al. 2010, *Proc. SPIE*, 7740, 774019