Identifying False Alarms in the Kepler Planet Candidate Catalog

F. Mullally¹, Jeffery L. Coughlin¹, Susan E. Thompson¹, Jessie Christiansen², Christopher Burke³, Bruce D. Clarke¹, and Michael R. Haas³

¹ SETI/NASA Ames Research Center, Moffett Field, CA 94035, USA; fergal.mullally@nasa.gov
² NASA Exoplanet Science Institute, California Institute of Technology, Pasadena, CA 91125, USA
³ NASA Ames Research Center, Moffett Field, CA 94035, USA

Received 2015 October 21; accepted 2016 February 9; published 2016 June 13

Abstract

We present a new automated method to identify instrumental features masquerading as small, long-period planets in the Kepler planet candidate catalog. These systematics, mistakenly identified as planet transits, can have a strong impact on occurrence rate calculations because they cluster in a region of parameter space where Kepler’s sensitivity to planets is poor. We compare individual transit-like events to a variety of models of real transits and systematic events and use a Bayesian information criterion to evaluate the likelihood that each event is real. We describe our technique and test its performance on simulated data. Results from this technique are incorporated in the Kepler Q1–Q17 DR24 planet candidate catalog of Coughlin et al.

Key words: methods: data analysis – planets and satellites: detection – binaries: eclipsing

Online material: color figures, machine-readable table

1. Introduction

One of the major goals of the Kepler mission (Koch et al. 2010) is to estimate the frequency of Earth-size planets in the habitable zones of solar-type stars. To that end, the spacecraft collected 4 years of near-continuous data on ~150,000 stars, searching for the faint signal of small transiting planets.

The Q1–Q16 catalog of Kepler planet candidates (Mullally et al. 2015) reported 554 new candidate planets. They noted an excess of candidates at periods longer than 100 days over what would be expected if planets were evenly distributed in orbital period. They identified two large populations of long-period false-alarm threshold crossing events (TCEs, or the statistically significant periodic events in a light curve that are vetted to produce the planet candidate catalog). One narrow peak centered around 372 days (the orbital period of the spacecraft) and a larger, broader peak spanning 200–600 days (see their Section 3.1). While their vetting process filtered out most of the false-alarm events from the narrow peak (leaving no excess of candidates at this orbital period), they concluded the observed excess of long-period planet candidates was more likely due to incomplete filtering of the broader peak than any super abundance of small planets in this period range (see their Section 9.1).

Burke et al. (2015) computed the occurrence rates of planets as a function of radius and period around solar analogs using the catalog of Mullally et al. (2015). They found a sharp rise in the computed frequency of Earth analogs for periods longer than 300 days. They traced the excess to 5 planet candidates with radii <1.2 R⊕ and periods of 450–550 days, a region of parameter space where such planets could only be detected around 1% of their sample (see Figure 14 in Burke et al.).

It is clear that understanding the reliability of the Kepler catalog (the fraction of candidates actually due to transiting planets) is an important precondition to measuring terrestrial planet occurrence rates. Much work has been done identifying astrophysical false positives due to eclipsing binaries and stellar multiplicity scenarios (e.g., Torres et al. 2011; Morton 2012; Colón et al. 2015; Santerne et al. 2012; Bryson et al. 2013; Désert et al. 2015), but less on instrumental artifacts. The statistical validation techniques used on Kepler planets (e.g., Torres et al. 2011; Rowe et al. 2014) compare the probability that a transit was due to a planet to various eclipsing-binary-type scenarios. However, small, long-period Kepler candidates have a non-negligible probability of being caused by an instrumental or processing artifact. Rowe et al. (2014) explicitly avoided low signal-to-noise (S/N) cases for this reason.

In this paper, we report on a new method for identifying and rejecting false-alarm candidates at long periods. This method, which we dub Marshall,⁴⁵ fits the individual transit events for candidates with models of transits and commonly found artifacts and evaluates which one is more probable, based on a Bayesian information criterion (BIC). Based on simulations, we find that known artifact types are rejected ≈60%–70% of

⁴ After a character in the animated series Paw Patrol.
⁵ Never let a three year old name your algorithms.
the time, while transits are preserved at the 95%, level as
discussed in Section 3.

Marshall is one component of the Kepler Robovetter, an
automated process for vetting planet candidates in Kepler
data, which includes the Flux Robovetter (Coughlin et al. 2016),
the Centroid Robovetter (F. Mullally 2016, in preparation),
the ephemeris matching of Coughlin et al. (2014), and the machine
learning technique of Thompson et al. (2015) that identifies
short-period variable stars mistakenly identified as planets. The
Robovetter replaces the manual vetting approach of previous
catalogs with an automatic, rules-based technique. In addition
to removing some of the inevitable subjectivity of the manual
process, the performance of the Robovetter can be tested
against large numbers of simulated transits.

Marshall can be easily applied to any transit search where
short-duration signals (such as instrumental artifacts) are
misidentified as transits. In any matched filter approach, such
as box least squares (Kovács et al. 2002), or the TPS algorithm
used by the Kepler pipeline (Seader et al. 2015), non-transit
signals will be occasionally mistaken for transits. If the false-
alarm signals can be modeled, the Marshall approach can be
used separate true transits from the false alarms. The Marshall
algorithm will be useful analyzing data from the upcoming
TESS and PLATO missions. We make a reference implementa-
tion in Python available at https://sourceforge.net/projects/
mmarshall.

2. Method

The key insight of our technique is that long-period
candidates have few enough events that there is sufficient
S/N in an individual event to discriminate between valid and
invalid transit shapes. By looking at individual events, we have
access to information about the transit shape that can be lost in
the folded transit event.

For each Kepler object of interest (KOI) in the Q1–Q17
DR24 catalog (Coughlin et al. 2016), we extract a snippet of
data 1.5 times the reported transit duration on either side of the
midpoint of each individual transit event (i.e., once per orbital
period of the candidate planet). We use the transit parameters
(orbital period, epoch, transit duration) from Seader et al.
(2015). We obtain the publicly available light curves from
MAST6 and use the co-trended light curve that corrects for
many instrumental features (available in the PDCSAP_FLUX
column of the light curve file; see Fraquelli & Thompson 2014;
Smith et al. 2012). We then fit each event with the following set
of models:

1. A parabola.
2. A parabola with a negative-going step-wise discontinuity
   (i.e., a step down) at the reported time of ingress.

3. A parabola with a positive-going step-wise discontinuity
   at the reported time of egress.
4. A parabola plus a sudden pixel sensitivity dropout event
   (SPSD; an SPSD is typically caused by a cosmic-ray hit
   on the CCD). We model the SPSD shape as

\[
\begin{align*}
0 & : t < t_{\text{ingress}} \\
-\lambda e^{-\tau(t-t_{\text{ingress}})} & : t > t_{\text{ingress}}
\end{align*}
\]

where \( \lambda \) and \( \tau \) are free parameters, and \( t_{\text{ingress}} \)
is the reported transit ingress time.

5. A parabola plus a box-shaped transit, defined as

\[
\begin{align*}
-d & : t_{\text{ingress}} < t < t_{\text{egress}} \\
0 & : \text{otherwise}
\end{align*}
\]

where \( d \) is a free parameter, while \( t_{\text{ingress}} \) and \( t_{\text{egress}} \)
are constrained so that the transit midpoint is fixed.

We include a parabolic term in each model to describe the
continuum flux (i.e., the flux we would expect if the proposed
transit were not present). The algorithm used for co-trending
(PDC; Smith et al. 2012) tries to preserve stellar variability, so
the continuum is often not flat even on the short timescales of
interest here.

We show representative examples of the various models in
Figure 1. Models 2–4 represent the most common kinds of
artifacts we see in the data. Model 1 (the parabola) catches the
case where a single strong artifact triggers a detection of a
possible planet, and the other reported events show no strong
signal. A box (Model 5) is the crudest possible model for the

\footnote{http://archive.stsci.edu/kepler}

Figure 1. An example instrumental feature from Kepler data (Kepler Id
4575824, alias K06428.01), showing the functional forms of the models fit to
the event, offset vertically for clarity. The model in each case includes a
parabola to describe the shape of the continuum.

(A color version of this figure is available in the online journal.)
transit shape, but the second-order details of ingress shape or limb darkening are not expected to be visible in the low-S/N case of a single transit of a small planet.

The discontinuities in our models present a challenge for many optimization algorithms. We chose the Amoeba, or Nelder-Mead method (Nelder & Mead 1965), which does not require the first derivative of the model to be well behaved. We found the Amoeba often converged on a local minimum, so we repeated the fit for each model with a variety of initial conditions and selected the best-fitting result.

We occasionally found transits that lie very close to gaps in the data. The fit is often poor when some of the expected data are missing, so we imposed a stringent requirement that no more than 25% of the expected cadences in the selected snippet be missing or gapped before running the algorithm on that event.

We select the preferred model using the Bayesian information criterion (Schwarz 1978). This metric rewards models that fit the data with fewer parameters, and is given by

\[
\text{BIC} = -2 \ln \mathcal{L} + k \ln N
\]

where \(k\) is the number of free parameters in the model and \(N\) is the number of data points fit. \(\mathcal{L}\) is likelihood of the fit, defined as

\[
\mathcal{L} \propto \prod \exp \left[ -\frac{1}{2} \left( \frac{y_i - f(t_i|\theta)}{\sigma} \right)^2 \right]
\]

where \(y_i\) is the value of the \(i\)th data point, and \(f(t_i|\theta)\) is the value of the model given a set of parameters \(\theta\) and evaluated at the time of the \(i\)th data point. \(\sigma\) is the uncertainty assigned to each data point, calculated in a manner that is robust to outliers in the data

\[
\sigma = 1.4826/\sqrt{2} \times \text{median}(\text{abs}(y_{i+1} - y_i)).
\]

The normalization ensures the computed value of \(\sigma\) is consistent with the value expected if the values of the data points were drawn from a Gaussian distribution. The Gaussian assumption is frequently invalid for \textit{Kepler} light curves, but any mismeasurement of \(\sigma\) affects all model fits equally.

The model with the lowest value of the BIC is the preferred model. We define the fit score as the difference between the BIC value of the transit fit and the best-fitting artifact model. Positive scores mean that the artifact model is preferred over the box model. \textit{Kepler} catalogs adopt the principle of “innocent until proven guilty,” erring on the side of including suspected false positives instead of incorrectly rejecting planet candidates. We therefore only reject events as false alarms if they have a score \(\geq 10\). Kass & Raftery (1995) argue that a difference in a BIC value of 10 or more means the evidence against the disfavored model is very strong. However, the choice of threshold is arbitrary; in the next section, we describe how we tuned our algorithm on simulated transits to achieve the desired completeness (fraction of valid transits passed by the algorithm) at the chosen threshold, at the expense of high reliability (fraction of false alarms detected).

3. Performance

To evaluate the performance of our algorithm we ran it on the simulated transit set used to evaluate the completeness of the \textit{Kepler} planet finding pipeline as described by J. L. Christiansen (2016, in preparation) and based on the technique of Christiansen et al. (2015). They injected a range of physically realistic transits for a range of planet sizes, orbital periods, etc. into the pixel level time-series data of all \textit{Kepler} target stars using the method described in Christiansen et al. (2015). This end-to-end modeling means that the effects on transit shape of the light curve generation (PA; Twicken et al. 2010), and systematic removal (PDC; Smith et al. 2012) modules of the pipeline are properly accounted for in our simulation.

We selected for our study \(10^4\) individual simulated transit events for planets with periods >200 days and radii <5 \(R_\oplus\) that were recovered by the pipeline. We tuned our algorithm until >95% of these individual events were passed. We then measured the performance of this tuned algorithm against simulated artifacts.

As J. L. Christiansen (2016, in preparation) did not inject artifact signals in their simulations, we perform our own. We add an artifact model to the light curves before co-trending, as produced by the PA module of the pipeline (the SAP\_FLUX column of the MAST light curve file). We cotrend the simulated signal by fitting and removing the appropriate covariant basis vectors produced by the PDC module of the pipeline (Stumpe et al. 2012). These vectors represent the coarse systematic signals in \textit{Kepler} light curves. They are available at MAST\(^7\) and are described in Section 2.3.4 of the Archive Manual, and a tutorial on their use is given in Kinemuchi et al. (2012). PDC uses advanced techniques (Smith et al. 2012; Stumpe et al. 2014) to apply the vectors to the data in an effort to prevent overfitting of stellar variability. For our purposes, we capture the important behavior of PDC with the computationally simpler direct fits, so we apply those instead. We run \(10^4\) simulations over a range of targets and model parameters for each of the discontinuity and SPSD models.

We show our results in Figure 2. The thin blue and orange lines show our performance at identifying and rejecting systematic events in our simulations (discontinuities and SPSDs, respectively) as a function of the threshold value. The vertical magenta line indicates the adopted threshold value of 10 (i.e., the transit model has a BIC score at least 10 points higher than the most favored artifact model). At a threshold of 10, we reject 60%-70% of the injected events. The thick black line shows the number of injected transits that are preserved by

\(^7\) http://archive.stsci.edu/missions/kepler/cbv
the technique, which is $>96\%$. These results give us confidence that we can identify many of the false-alarm events in the real data, while preserving the signal of most of the real planets.

4. Application to the Q1–Q17 DR24 Catalog

The Q1–Q17 DR24 catalog of KOIs (Coughlin et al. 2016) incorporates results from Marshall. The KOI creation process is described in detail in Rowe et al. (2015). Briefly, a KOI number is assigned to a periodic signal in the light curve of a Kepler target that appears to be due to the transit or eclipse of an astrophysical body. KOI numbers are assigned based on a preliminary analysis and are sometimes assigned to other phenomena such as stellar variability or instrumental artifacts. Further vetting identifies some more of these artifacts as well as false-positive signals due to eclipsing binaries. Any KOI is marked as a planet candidate unless there is conclusive evidence that it is not.

A KOI incorporates at least three events equally spaced in time, and Marshall deals with individual events. To apply Marshall to KOIs, we need to choose a disposition (planet candidate or false positive) based on the combined scores of individual events. We adopt the following rules:

1. Count $N_{\text{good}}$, the number of transit events where $>75\%$ of the expected cadences in the appropriate interval were collected, and where the computed score is $<-10$.
2. Count $N_{\text{skip}}$, the number of events where some data, but less than $75\%$ of the expected cadences, were collected.
3. $N_{\text{good}} + N_{\text{skip}} \geq 3$, then the KOI passes, otherwise it fails. This rule is consistent with the mission requirement of needing at least three transits to claim the transit detection.

We apply this test to KOIs in the Q1–Q17 DR24 catalog with orbital periods $>150$ days. We seed our fits with the transit parameters (period, epoch, duration, etc.) from Seader et al. (2015) and available at the NASA Exoplanet Archive (Akeson et al. 2013). Although Marshall is tuned for low-S/N events and ignores second-order effects on transit shape (such as ingress shape and limb darkening), we find it correctly identifies high-S/N events as transits. Our treatment of $N_{\text{skip}}$ has a small effect on our results. From the set of 228 TCEs with periods $>150$ days and not already marked as false positives in the Q1–Q17 DR24 catalog, only 3 additional TCEs are marked as false positive if we require $N_{\text{good}} \geq 3$.

We show our results in Figure 3 and Table 1. Of the 228 KOIs tested that were not otherwise failed by the Robovetter, 20 fail our test of having fewer than 3 valid transits (i.e., $N_{\text{good}} + N_{\text{skip}} < 3$) mostly at periods $>400$ days and radii $<3 R_\oplus$. Visual inspection confirms the artifact nature of all but two of these KOIs (K05805.01 and K02758.01). False-positive identification in the Q1–Q17 DR24 catalog is made entirely by rule, so these KOIs are marked as false positive in the final
KOIs Examined by Marshall

| KOI      | Kepler ID  | TCE | NTrans. | Period (days) | Radius (R$_\oplus$) | Score     |
|----------|------------|-----|---------|--------------|---------------------|-----------|
| K00998.01| 1432214    | 1   | 9       | 161.78000(36) | 30.140(49)          | $-2.1 \times 10^5$ |
| K01099.01| 2853093    | 1   | 9       | 161.52800(34) | 5.97(76)            | $-5.5 \times 10^5$  |
| K01788.02| 2975770    | 2   | 4       | 369.07900(26) | 2.7(2.4)            | $+2.7 \times 10^1$  |
| K03549.01| 2307206    | 1   | 7       | 204.029000    | 44.140000          | $-1.4 \times 10^3$  |
| K03674.01| 2446623    | 1   | 8       | 175.066000    | 28.6(7.1)           | $-4.2 \times 10^3$  |
| K03681.01| 2581316    | 1   | 7       | 217.832000    | 11.410(19)          | $-8.0 \times 10^3$  |
| K03709.01| 2576107    | 1   | 7       | 205.583000    | 21.0(2.6)           | $-2.8 \times 10^3$  |
| K04959.01| 2987433    | 1   | 4       | 420.929000    | 49.4(1.7)           | $-5.0 \times 10^3$  |
| K06254.01| 1433531    | 1   | 3       | 567.676(13)   | 4.27(46)            | $+3.1 \times 10^5$  |
| K06295.01| 2856960    | 1   | 7       | 204.304000    | 13.730(62)          | $-1.4 \times 10^3$  |

**Note.** Summary of planet candidate KOIs from Q1–Q17 DR24 with periods greater than 150 days examined by Marshall. *Kepler* ID is the unique identifier of the target star in the *Kepler* input catalog. TCE indicates the order in which this KOI was found around the target star by the *Kepler* pipeline in Q1–Q17. NTrans. refers to the number of observed transits as measured by the pipeline. The uncertainty in the last two digits of period and radius are given in parentheses. Period and radius values are taken from Seader et al. (2015) and may differ slightly from the final values in Coughlin et al. (2016). Score represents the confidence Marshall assigns to a KOI. Negative scores indicate high confidence that a KOI is due to a transit while positive scores indicate lower confidence. KOIs with scores >10 are deemed to be false positives. The score given in the table is equal to the score of the third strongest individual event for the KOI which indicates how close the KOI is to being marked as a false positive.

(This table is available in its entirety in machine-readable form in the online version of this article.)

The 4% of injected transits that were rejected, we inspected the data to understand why they failed. We show an example in Figure 4. The top panel shows the injected event, while the bottom panel shows the light curve after long-term trends were removed by PDC. In this case, the shape of the transit was deformed to make it look more like a systematic effect. For stars without rapid variations in the light curve, this is the most common reason why simulated planet candidates were misidentified. However, attempts to use the light curves without the detrending from PDC had significantly worse performance due to the more complicated structure in the out-of-transit flux.

Although the artifact types we test for create most of the artifacts, we are aware of for small, long-period planet candidates, there are presumably other sources of systematics not yet accounted for that decrease the sample of planet candidates still further. In addition, our simulations suggest we only find two-thirds of our injected false alarms This suggests as many as 10 more false alarms of the kind we tested for remain in the Q1–Q17 DR24 catalog. There are 37 planet candidates with periods longer than 300 days in the catalog. If, similar to the identified false alarms, the uncaught systematics also have periods >300 days, then we expect no more than 27 (or 73%) of these long-period candidates are actually
transits. This places a rough upper bound on the reliability of the catalog for small, long-period planets.

The *Kepler* Robovetter emphasizes completeness over reliability. Ensuring that as many planets as possible are included as candidates in the catalog is a higher priority than removing as many false positives as possible. Users of the catalog who place a higher value on reliability may prefer to set their rejection threshold at a lower BIC score, at the expense of a decrease in the completeness of their sample. To this end, we provide the Marshall score for each KOI tested in Table 1.

Finally, we note that the score given in Table 1 is not the probability a KOI is due to a transit. Even KOIs with strong BIC scores may be due to non-planetary signals, such as the highly diluted signal of an eclipsing binary system with small angular separation from the target.

### 6. Conclusion

We present a new technique to identify systematic signals in *Kepler* data masquerading as planet candidates. The algorithm looks at each individual transit event and decides if the shape of that event is more likely a transit or one of a few known artifacts. We apply the algorithm to the DR24 planet catalog of Coughlin et al. (2016) and find it rejects 20 small, long-period KOIs otherwise marked as planet candidates by the Robovetter.

Funding for this Discovery mission is provided by NASA’s Science Mission Directorate. All of the data presented in this paper were obtained from the Mikulski Archive for Space Telescopes (MAST). STScI is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555. Support for MAST for non-*HST* data is provided by the NASA Office of Space Science via grant NNX13AC07G and by other grants and contracts. 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.

*Facilities: Kepler.*

---

### References

Akeson, R. L., Chen, X., Ciardi, D., et al. 2013, *PASP*, 125, 989
Bryson, S. T., Jenkins, J. M., Gilliland, R. L., et al. 2013, *PASP*, 125, 889
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
Colón, K. D., Morehead, R. C., & Ford, E. B. 2015, *MNRAS*, 452, 3001
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
Fraquelli, D., & Thompson, S. E. 2014, *Kepler* Archive Manual (KDMC-10008-005), [http://archive.stsci.edu/kepler/documents.html](http://archive.stsci.edu/kepler/documents.html)
Kass, R. E., & Raftery, A. E. 1995, *J. Am. Stat. Assoc.*, 90, 773
Kinemuchi, K., Barclay, T., Fanelli, M., et al. 2012, *PASP*, 124, 963
Koch, D. G., Borucki, W. J., Basri, G., et al. 2010, *ApJL*, 713, L79
Kovács, G., Zucker, S., & Mazeh, T. 2002, *A&A*, 391, 369
Morton, T. D. 2012, *ApJ*, 761, 6
Mullally, F., Coughlin, J. L., Thompson, S. E., et al. 2015, *ApJS*, 217, 31
Nelder, J. A., & Mead, R. 1965, *Comput. J.*, 7, 308
Rowe, J. F., Bryson, S. T., Marcy, G. W., et al. 2014, *ApJ*, 784, 45
Rowe, J. F., Coughlin, J. L., Antoci, V., et al. 2015, *ApJS*, 217, 16
Santerne, A., Díaz, R. F., Moutou, C., et al. 2012, *A&A*, 545, A76
Schwarz, G. 1978, *Ann. Stat.*, 6, 461
Seader, S., Jenkins, J. M., Tenenbaum, P., et al. 2015, *ApJS*, 217, 18
Smith, J. C., Stumpe, M. C., Van Cleve, J. E., 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., Van Cleve, J. E., et al. 2012, *PASP*, 124, 985
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
Twicken, J. D., Clarke, B. D., Bryson, S. T., et al. 2010, *Proc. SPIE*, 7740, 774023