There is much left to do. May 30, 2016
Preprint typeset using LaTeX style AASTeX6 v. 1.0

PLANETARY CANDIDATES OBSERVED BY *Kepler*. VIII. A FULLY AUTOMATED, UNIFORM CATALOG BASED ON QUARTERS 1 THROUGH 17, DATA RELEASE 25 TUNED FOR COMPLETENESS AND RELIABILITY.

Susan E. Thompson[1,2], Jeffrey L. Coughlin[2], Fergal Mullally[1,2], Jessie L. Christiansen[3], Christopher Burke[1,2], Michael Haas[1], Steve Bryson[1], Natalie Batalha[1], Joe Catanzarite[1,2], Kelsey Hoffmann[2], Jason Rowe, Catalog Generators, Interpreters of Candidates, Created List of TCEs, and Made Data Available

[1]NASA Ames Research Center, Moffett Field, CA
[2]SETI Institute, 189 Bernardo Ave, Mountain View, CA
[3]NExScI

ABSTRACT

Enlightening words about the purpose of this final Kepler Catalog and how it fits in with the other Kepler Planet Candidate catalogs. Then some details about how many KOIs we added and how many are planet candidates.

1. SCOPE

The purpose of this paper is to describe how we sifted through the DR25 TCE table and identified the planet candidates, the transiting candidates with significant secondaries, background eclipsing binaries, ephemeris matches and the false alarms. It then describes additional aspects of the KOI table that allow this table to be used to either follow-up certain targets, or calculate occurrence rates. The Completeness and Reliablity of the catalog should be measured and discussed.

Those items that specifically need to be discussed are:

- A Description of individual metrics that were not in Coughlin et al.
- A list of old metrics and how they were updated for DR25
- How thresholds for dispositions were set.
- Transit model fits done with MCMC analysis
- Discussion of the completeness of our catalog
- Discussion of the effectiveness of our catalog at removing known False Alarms and False Positives.
- How to make cuts on the catalog to get a catalog with very high reliability or very high completeness.
- A detailed list of interesting Candidates (read small HZ candidates or new multi-systems) in the catalog.
- How to interpret and use the reliability Scores
- How to estimate the reliability based on our effectiveness number, and a discussion of an error on that number.
- How federation was done.
- A list of known problems with the catalog.
- Give Details about locations of Robovetter codes, i.e. github

Items we may want to discuss in the paper but only if we have time:

- Measurement of effectiveness or completeness based on hand vetting.
- The effectiveness or completeness of the catalog as a function of planetary radius and semi-major axis. (ie derived from fits)
- A list of confirmed planets that now have a reason to be doubted (if any exist)
- A comparison to the CFP catalog
- A comparison to EB catalog
- Analysis of the injected EBs in terms of effectiveness

Items that do not belong in this paper:

- An estimate of eta Earth or the occurrence rate for any type of planet.
- Autovetter Results

susan.e.thompson@nasa.gov

- Discussion of metrics not used by the Robovetter.
- Complete description of the logic of the Robvetter (see Coughlin et al.)
- A Comparision to Planet Hunters Catalog

2. INTRODUCTION

1. What is Kepler and what is its mission?
2. How have we made catalogs in the past and why did we switch to a robovetter. What is the robovetter?
3. Introduce what is new in this paper, hint that the population of TCEs is different and that has created challenges in automatic vetting. Introduce how the goal is to get a sense of the completeness and the reliability of the catalog.
4. Introduce the concepts injection and inversion to test the robovetter.
5. Outline how the Robovetter creates Dispositions and the catalog also includes MCMC fits.
6. Create a road map of what is in this paper?

3. THE Q1–Q17 DR25 TCES

1. Summary of how this TCE population differed, especially compared to DR24 and Q1-Q16.
2. Basic stats of the distribution of TCEs in Period and maybe MES.
3. Describe what information comes from the pipeline.
4. There are way more TCEs than expected KOIs, the task at hand is to filter out the transit-like, astrophysical events and turn them into KOIs. Amongst those, identify which of those are likely planet candidates.

4. VETTING

4.1. *Robovetter Overview*

Describe here the idea of not-transit-like and how we look for evidence that the signal is due to an eclipsing binary. Include an updated Figure 2 from the previous paper. Introduce the major flags.

4.2. *Individual Metrics, updates and improvements*

1. LPP retrained
2. SWEET
3. ChASES
4. Centroid Robovetter
5. Odd Even
6. DMM
7. Shape Metric
8. Smear
9. Ghosting Metric
10. Rolling Band Flags
11. Ephemeris Matching improvements and results

4.3. *Evaluating Performance*

1. At our disposal we have two tools to test our metrics, a population of signals we know to be real (Injection) and a population of signals we know to be false (inversion).
2. we use these to gain a sense of the quality of our catalog.
3. a Quick summary of how injection was done and what we produce with injection. Show a plot of the distribution of recovered injected things in terms of period or MES.
4. A summary of how inversion was done and what we get from it.
5. In addition we have confirmed planets and confirmed false positives. However, these are not great tests for low MES or long Period.
6. Limitations of injection inversion or labeled data sets are that we do not probe all types of false alarms or false negatives.

4.4. *A First Attempt at Vetting: Using the DR24 Robovetter*

1. We applied the DR24 robovetter to the DR25 TCEs.
2. We found that we needed to do better in the following areas.

Our first attempt at vetting the DR25 TCE list was done by applying the DR24 Robovetter (**?**). To do this, we calculated the metrics described in **?**, such as the LPP Metric, the Marshall Metric, the Model-Shift Uniqueness Test, the Significant Secondary Test, the Centroid Offset and the Ephemeris matching. We were then able to evaluate the performance of this version of the robovetter using the pixel-level transit injection and inversion data.

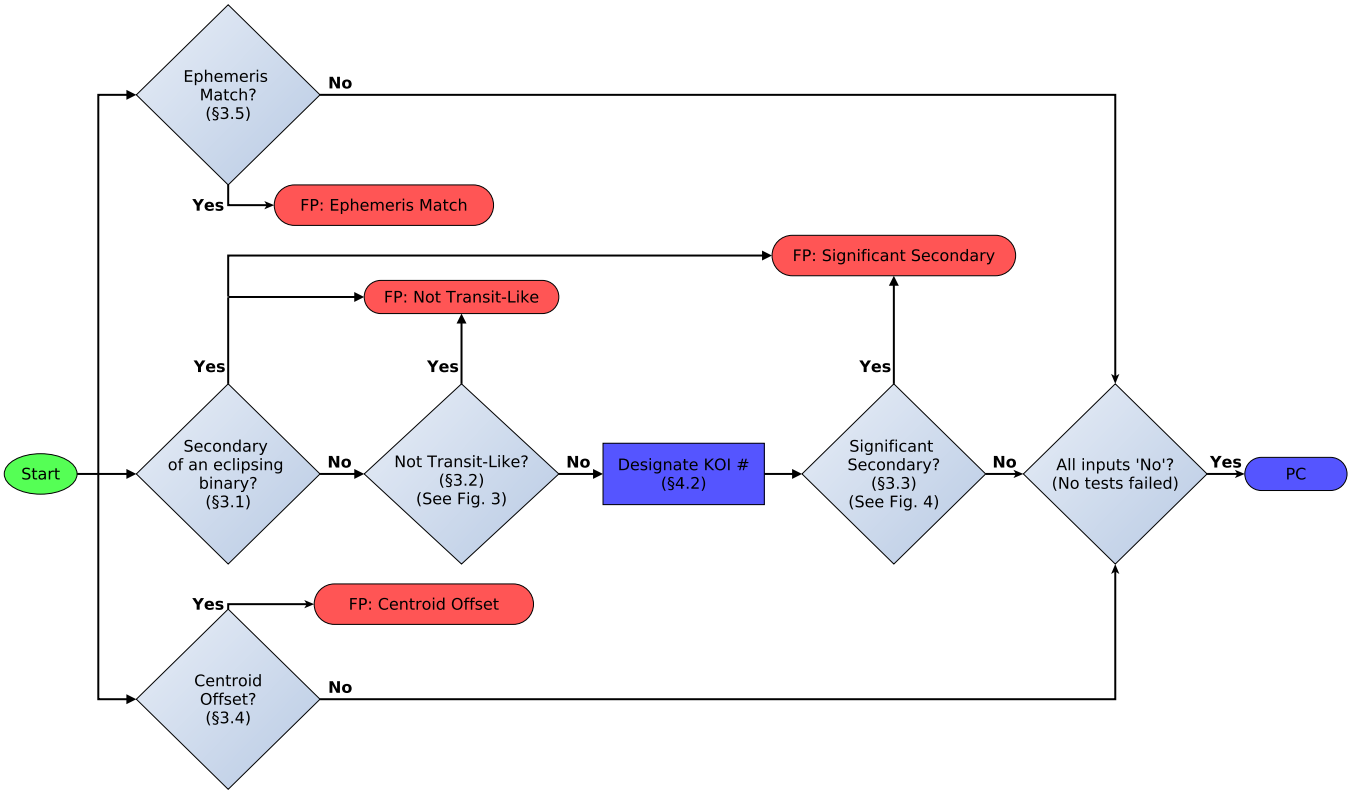

Figure 1. An Overview of the Robovetter logic. This is from DR24, so needs to be updated

Clearly, these results indicate that the TCE population had changed enough to require improvements to the robovetter in order to achieve a highly complete and highly reliability catalog.

4.5. *Optimization*

How we tuned all the metrics to do the best we can. Discuss any weaknesses with our methodology.

5. PUTTING TOGETHER THE CATALOG

5.1. *Federating to known KOIs*

5.2. *Model Fits using MCMC*

5.3. *Reporting Dispositions and Metrics*

6. EVALUATE THE CATALOG

6.1. *Catalog Completeness*

6.1.1. *Evaluate recovery of Pixel-level Injected Transits*

6.1.2. *Evaluate recovery of previous KOIs and Confirmed Planets*

6.1.3. *Completeness in Period and*

6.2. *Catalog Reliability*

6.3. *Evaluate the the failure rate of verified FPs*

6.3.1. *Use Inversion to evaluate Effectiveness*

6.3.2. *Estimate Reliability*

6.4. *What does this mean for occurrence rate calculations?*

7. INTERESTING PLANET CANDIDATES

7.1. *New Multi-planet systems*

7.2. *Potentially Rocky Planets in the Habitable Zone*

8. DISCUSSION

8.1. *Strengths and Weaknesses*

8.2. *Use of this catalog for occurrence rates.*

9.

APPENDIX

A. LIST OF ACRONYMS

B. ROBOVETTER MINOR FLAGS

REFERENCES