Name: Jon Jenkins
Email: Jon.Jenkins@nasa.gov
Institution: SETI Institute
Title: Likely Planet Candidates Identified by Machine Learning Applied to Four Years of Kepler
Type: Poster
Session: Exoplanet Statistics, False Positives, and Completeness Corrections
Abstract: Sean McCauliff, Joseph Catanzarite, Joseph Twicken, and Jennifer Campbell

Over 3200 transiting planet candidates, 134 confirmed planets, and ~2,400 eclipsing binaries have been identified by the Kepler Science pipeline since launch in March 2009. Compiling the list of candidates is an intensive manual effort as over 18,000 transit-like signatures are identified for a run across 34 months. The vast majority are caused by artifacts that mimic transits. While the pipeline provides diagnostics that can reduce the initial list down to ~5,000 light curves, this effort can overlook valid planetary candidates. The large number of diagnostics (~100) makes it difficult to examine all the information available. The effort required for vetting all threshold-crossing events (TCEs) takes several months by many individuals associated with the Kepler Threshold Crossing Event Review Team (TCERT).

We have developed a random-forest classifier that classifies each TCE as `planet candidate', `astrophysical false positive', or `non-transiting phenomena'. Ideally the algorithm will generate a list of candidates that approximates those generated by human review, thereby allowing the humans to focus on the most interesting cases. By using a machine learning-based auto-vetting process, we have the opportunity to identify the most important metrics and diagnostics for separating signatures of transiting planets and eclipsing binaries from instrument-induced features, thereby improving the efficiency of the manual effort.

We report the results of applying a random forest classifier to four years of Kepler data. We present characteristics of the likely planet candidates identified by the auto-vetter as well as those objects classified as astrophysical false positives (eclipsing binaries and background eclipsing binaries). We examine the auto-vetter's performance through receiver operating characteristic curves for each of three classes: planet candidate, astrophysical false positive, and non-transiting phenomena.

Funding for this mission is provided by NASA's Science Mission Directorate.