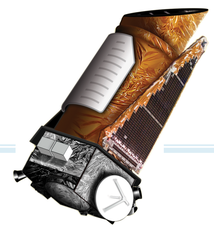
Characterization and Validation of Small HZ Planet Candidates in the Kepler SOC Pipeline



Joseph D. Twicken[1], B.D. Clarke[1], F. Girouard[2], T.C. Klaus[2], J. Li[1], S.D. McCauliff[2], E.V. Quintana[1], S.E. Seader[1], P. Tenenbaum[1], J.M. Jenkins[1], N.M. Batalha[3], S.T. Bryson[3], C.J. Burke[1], D.A. Caldwell[1], J.F. Rowe[1]

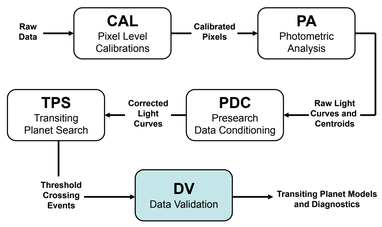
[1]SETI Institute, [2]Orbital Sciences Corporation, [3]NASA Ames Research Center

Abstract

The Transiting Planet Search (TPS) and Data Validation (DV) components of the Science Operations Center (SOC) Processing Pipeline were run with the SOC 9.1 codebase on the Q1-Q16 *Kepler* data set in August 2013. This represents nearly all of the photometric data that were acquired before science data collection was halted due to the loss of a second reaction wheel in May 2013 (Q17). Pipeline results (Threshold Crossing Event details and Data Validation products) will be delivered to the Exoplanet Archive at the NASA Exoplanet Science Institute (NExScI) following a review by the *Kepler* Data Analysis Working Group. Long cadence light curves for the vast majority of *Kepler* targets are searched for transiting planet signatures in the TPS component of the SOC Pipeline. Those targets for which the detection threshold is exceeded are subsequently processed in the DV Pipeline component. The primary functions of DV are to (1) characterize planets identified in the transiting planet search, (2) search for additional transiting planet signatures in light curves after transit signatures have been removed, and (3) perform a powerful suite of diagnostic tests to aid in discrimination between true transiting planets and false positive detections. We describe the characterization and validation of transiting planets in the Q1-Q16 data set with emphasis on small Habitable Zone (HZ) planet candidates that were not identified in earlier Pipeline runs with shorter data sets. We present planet model fit results for selected candidates and examples of the diagnostics utilized to validate them. We also present an example of the type of false positive detection most common in the small planet/long period regime. The Exoplanet Archive is located at http://exoplanetarchive.ipac.caltech.edu. Funding for the *Kepler Mission* has been provided by the NASA Science Mission Directorate.

Kepler Data Processing Pipeline

Data Flow

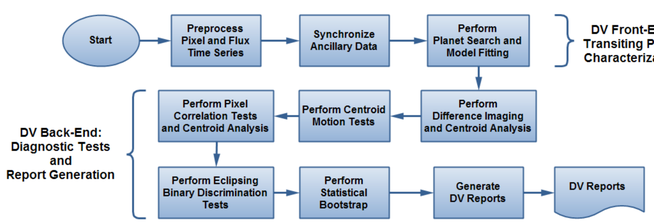


Science data from over 160,000 long cadence (LC) and 512 short cadence (SC) stellar targets per quarter are processed in the Data Processing Pipeline. LC data are processed from Calibration (CAL) through Data Validation (DV). SC data are processed through Presearch Data Conditioning (PDC) only.

Components and Primary Functions

CAL	Calibrate science pixels (collateral, background and target) for each long or short cadence
PA	Extract raw flux and compute photocenter (centroid) for each target and cadence from associated target pixels
PDC	Correct systematic and other errors in raw light curves, remove excess flux due to aperture crowding, and condition light curves for the transiting planet search
TPS	Perform transiting planet search and return Threshold Crossing Events (TCEs) for planet detections
DV	Fit transiting planet model to light curves with TCEs, search for additional transiting planets, and perform diagnostic tests to validate planet candidates

Data Validation (DV)



DV is run on the NASA Advanced Supercomputing (NAS) Pleiades cluster. Front-end processing for each target star involves fitting a transiting planet model to its error corrected light curve and searching the residual for additional planet candidates. The process is completed when the transit model has been fitted for all planet candidates. Back-end processing includes a suite of diagnostic tests to aid in discriminating between true planets and false positives. Model fit and diagnostic test results are included in a DV Report generated in PDF format for each target. A one-page Report Summary PDF is also produced for each candidate. DV diagnostics include:

- Difference imaging to separately identify the locations of target star and transit source
- Centroid motion test to quantify in-transit centroid motion and locate transit source
- Pixel correlation test to identify the location of the transit source
- Eclipsing binary discrimination tests to identify presence of primary and secondary eclipses
- Statistical bootstrap to assess the false alarm probability of the transiting planet detection

Selected Q1-Q16 Pipeline Candidates

Planet Characteristics

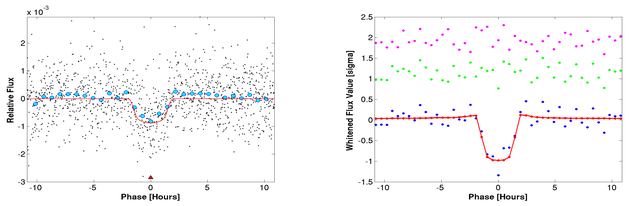
	KIC	KOI	Period (Days)	Radius (Earth Radii)	Equilibrium Temperature (K)	Comment
(a)	5868793	4290	4.838	0.8	320	**KOI 4290.01**; unclassified star with new parameters for Q1-Q16
(b)	6444896	3138	8.689	0.6	210	**KOI 3138.01**; unclassified star with new parameters for Q1-Q16
(c)	11497958	1422	34.142	1.3	240	Third HZ candidate and fifth overall on KOI 1422
(d)	5184911	2719	106.262	2.0	300	Second candidate on KOI 2719
(e)	9935983		161.650	1.4	310	
(f)	10055126	1608	232.045	1.8	320	Third candidate on KOI 1608
(g)	11037818		259.345	1.7	280	
(h)	9002278	701	267.284	1.5	200	**KOI 701.04**; no TCE in Q1-Q12 run
(i)	7219825	238	362.995	1.7	280	Third candidate on KOI 238
(j)	10905641		460.778	1.7	250	

This table is by no means complete. The Q1-Q16 results remain to be vetted by the *Kepler* TCE Review Team (TCERT). Three of these candidates were known KOIs prior to the Q1-Q16 run. **Uncertainties in radius may exceed 100%. Assumed albedo = 0.3.** Science community members are invited to participate in TCERT activities (contact: *michael.r.haas@nasa.gov*).

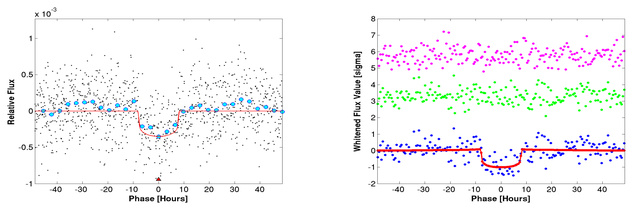
Stellar Characteristics (Huber et al. 2013, in prep)

	KIC	KOI	Kepler Magnitude (Kp)	Radius (Solar Radii)	Effective Temperature (K)	Log Surface Gravity (cm/sec/sec)
(a)	5868793	4290	17.1	0.16 ± 0.16	3100 ± 300	5.2 ± 0.5
(b)	6444896	3138	17.1	0.12 ± 0.12	2700 ± 300	5.3 ± 0.5
(c)	11497958	1422	15.9	0.37 ± 0.06	3500 ± 100	4.9 ± 0.1
(d)	5184911	2719	15.1	0.8 ± 1.4	4800 ± 200	4.5 ± 0.8
(e)	9935983		13.8	0.8 ± 0.5	5900 ± 200	4.6 ± 0.4
(f)	10055126	1608	13.8	1.1 ± 0.2	5900 ± 100	4.4 ± 0.1
(g)	11037818		14.8	0.9 ± 0.5	5900 ± 200	4.5 ± 0.3
(h)	9002278	701	13.7	0.65 ± 0.04	4800 ± 100	4.6 ± 0.1
(i)	7219825	238	14.1	1.1 ± 0.2	6100 ± 100	4.4 ± 0.1
(j)	10905641		13.5	1.0 ± 0.8	6100 ± 200	4.5 ± 0.4

Model Fitting



A limb-darkened Mandel-Agol transiting planet model is employed to characterize each planet candidate in DV. The fit is performed in a whitened domain where transits and model are distorted in a similar fashion. The fit results are shown above for the new 34.1-day candidate on **KIC 11497958 / KOI 1422**. The model fit is overlaid on unwhitened median detrended flux values on the left. The model fit is overlaid on binned and averaged whitened flux values on the right; residuals are shown in green with a vertical offset. The fit results are displayed below for the new 259.3-day candidate on **KIC 11037818**.

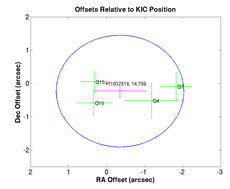


Transiting planet model fit results for the Pipeline candidates are displayed in the table below.

	KIC	KOI	Observed Transits	Fit SNR	Transit Depth (ppm)
(a)	5868793	4290	70	13.4	2325 ± 226
(b)	6444896	3138	89	15.5	2259 ± 196
(c)	11497958	1422	30	11.4	876 ± 82
(d)	5184911	2719	14	10.2	503 ± 54
(e)	9935983		8	9.9	293 ± 33
(f)	10055126	1608	5	11.8	284 ± 23
(g)	11037818		5	11.1	349 ± 32
(h)	9002278	701	4	13.5	456 ± 35
(i)	7219825	238	4	9.5	243 ± 26
(j)	10905641		3	9.7	261 ± 28

Diagnostics

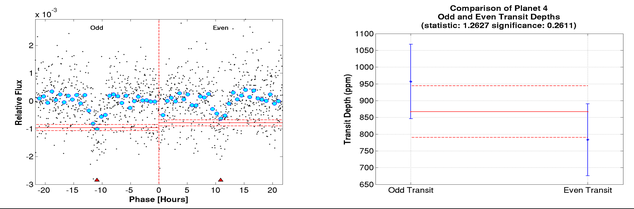
Difference Image Centroid Offsets



Quarterly difference image centroid offsets are computed separately with respect to out-of-transit image centroids and KIC position. The offsets represent estimates of the position of the transit source relative to the target star. Quarterly offsets are weighted and robustly averaged to produce a single estimate over the full data set. Difference image centroid offsets for the **KIC 11037818** candidate are shown here. The mean offset is displayed in magenta with 3σ uncertainty radius in blue. The magnitude of the mean offset (0.42 arcsec) is statistically insignificant. Centroid offsets are computed for all candidates identified in the Pipeline and are a powerful tool for identifying false positives due to background sources.

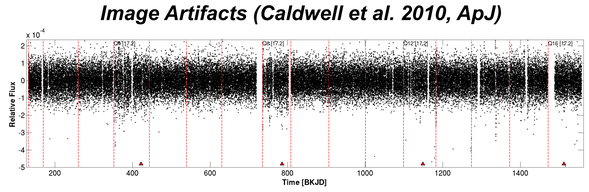
Eclipsing Binary Discrimination Tests

A series of diagnostic tests are performed in DV to identify the presence of primary and secondary eclipses. The depths and timing of the odd and even transit sequences are compared statistically to identify circular or near-circular binaries. The periods of all candidates are also compared to identify eclipsing binaries. The odd/even transit depth comparison results are shown below for the 34.1-day candidate on **KIC 11497958 / KOI 1422**. The odd and even transit depths for this candidate differ at the 1.1σ level which is not considered significant.



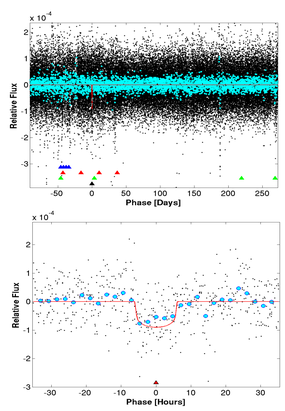
False Positives

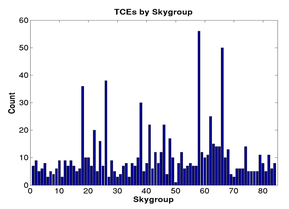
Image Artifacts (Caldwell et al. 2010, ApJ)



A relatively small number of Kepler CCD channels exhibit time and temperature dependent artifacts (so-called *rolling bands* and *Moiré pattern noise*) that originate in the detector electronics. The problem is compounded because four groups of targets rotate through the image artifact channels each year due to the quarterly roll of the spacecraft. Image artifacts are the dominant source of false positive transiting planet detections in the small planet/long period regime. The light curve displayed above is for **KIC 8374741**. Four "transit" events are marked with a period of 363.5 days; all fall on module output 17.2 which is the most severe image artifact channel. Image artifact noise is clearly visible during the quarters in which the events were observed. In fact, four false positive planet detections were generated for this star with periods ranging from 314 to 367 days. All would have been characterized as HZ candidates with radii ranging from 0.9 to 1.3 Earth radii and temperatures from 240 to 253 Kelvin.

The full phase folded light curve for the 363.5-day candidate on **KIC 8374741** is displayed in the upper figure on the right. Bin averages are shown in cyan. Event triangles identify the transit times for all of the false positives associated with this target. The bulk of these events were acquired on the same image artifact CCD channel. The lower figure on the right is zoomed on the phase folded transit event. This might appear to be a credible model fit with transit depth = 90 ± 13 ppm and SNR = 7.3. The detection, however, is not valid.





The histogram on the left displays the Q1-Q16 transiting planet detections versus sky group for candidates with period > 150 days, radius < 2.5 Earth radii and SNR > 7.1. The spikes represent those stars that rotate through known image artifact channels. These artifacts are the major source of false positive detections for small HZ candidates with long periods.

Discovery Mission 10 - Launched 2009 - http://Kepler.NASA.gov