Data validation in the *Kepler* Science Operations Center pipeline

Hayley Wu*[a], Joseph D. Twicken[a], Peter Tenenbaum[a], Bruce D. Clarke[a], Jie Li[a], Elisa V. Quintana[a], Christopher Allen[b], Hema Chandrasekaran[c], Jon M. Jenkins[a], Douglas A. Caldwell[a], Bill Wohler[b], Forrest Girouard[b], Sean McCauliff[b], Miles T. Cote[d], Todd C. Klaus[b]

[a]SETI Institute, NASA Ames Research Center, MS 244-30, P.O. Box 1, Moffett Field, CA, 94035;
[b]Orbital Sciences Corporation, NASA Ames Research Center, MS 244-30, P.O. Box 1, Moffett Field, CA, 94035;
[c]Lawrence Livermore National Laboratory, P.O. Box 808, L-478, Livermore, CA 94551;
[d]NASA Ames Research Center, MS 244-30, P.O. Box 1, Moffett Field, CA, 94035

ABSTRACT

We present an overview of the Data Validation (DV) software component and its context within the *Kepler* Science Operations Center (SOC) pipeline and overall *Kepler* Science mission. The SOC pipeline performs a transiting planet search on the corrected light curves for over 150,000 targets across the focal plane array. We discuss the DV strategy for automated validation of Threshold Crossing Events (TCEs) generated in the transiting planet search. For each TCE, a transiting planet model is fitted to the target light curve. A multiple planet search is conducted by repeating the transiting planet search on the residual light curve after the model flux has been removed; if an additional detection occurs, a planet model is fitted to the new TCE. A suite of automated tests are performed after all planet candidates have been identified. We describe a centroid motion test to determine the significance of the motion of the target photocenter during transit and to estimate the coordinates of the transit source within the photometric aperture; a series of eclipsing binary discrimination tests on the parameters of the planet model fits to all transits and the sequences of odd and even transits; and a statistical bootstrap to assess the likelihood that the TCE would have been generated purely by chance given the target light curve with all transits removed.

Keywords: photometry, data validation, *Kepler*, Earth-size planets

1. INTRODUCTION

The *Kepler Mission* is designed to detect (habitable) Earth-size planets transiting Sun-like stars[1]. The spacecraft was launched on 6 March 2009 into an Earth-trailing heliocentric orbit with a period of 373 days. Pointing of the *Kepler* photometer is maintained to support imaging of the same star field continuously over the lifetime of the mission (nominally 3.5 years for the primary mission). The *Kepler* photometer field of view is ~115 square degrees. Incident light is captured by 42 charge-coupled device (CCD) detectors comprised of 94.6 million total pixels on the focal plane assembly. Short exposures are integrated on board to produce one image every 29.4 minutes for over 150,000 long cadence (LC) targets and one image every 0.98 minutes for 512 short cadence (SC) targets. The spacecraft rolls 90° on a q*uarterly basis so that the solar panels continuously point toward the Sun. Flux from any given stellar target is, therefore, captured by a different CCD detector from one science data acquisition season to the next.

* hayley.wu@nasa.gov; kepler.nasa.gov

Copyright 2010 Society of Photo-Optical Instrumentation Engineers. One print or electronic copy may be made for personal use only. Systematic reproduction and distribution, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper are prohibited.

The *Kepler* Science Operations Center (SOC) Science Processing Pipeline (hereafter referred to as the Pipeline) is described in detail by Jenkins et al[2] and the *Kepler* SOC architecture is described in detail by Middour et al[3]. The Calibration (CAL) software component[4] calibrates pixel values for each cadence. The Photometric Analysis (PA) component[5] extracts raw flux light curves and computes target photocenters (centroids). The Presearch Data Conditioning (PDC) component[6] corrects data anomalies and systematic errors and removes excess flux due to crowding in the target apertures. The Transiting Planet Search (TPS) component[7] then subjects long cadence corrected flux light cuves to a search for transiting planets and returns a Threshold Crossing Event (TCE) for each target and trial transit pulse duration that exceeded the detection threshold. The TCE includes the time of the first transit, suspected orbital period, duration of the matched trial transit filter associated with the TCE, and relevant detection statistics.

The primary task of the Data Validation (DV) software component is to perform an automated validation of the many TCEs produced by TPS of the LC targets[7]. DV is provided with the TCE for each threshold crossing target corresponding to the maximum multiple event detection statistic over the set of trial transit pulse durations. A transiting planet model is fitted to the light curve for the given target to obtain model parameters for the initial planet candidate. The model fit is subtracted from the light curve and a search for additional transiting planets is performed on the residual light curve. If an additional detection occurs, the transiting planet model is fitted to the residual flux based on the new TCE. A suite of automated validation tests is performed when no additional planet candidates can be identified through the multiple planet search (or when the operator-configurable iteration limit is reached). The main purpose of the automated validation tests is to facilitate the identification of the true planet candidates from the large number of false positive transiting planet detections, astrophysical and otherwise.

The automated tests performed in DV are by no means the final validation of new planet discoveries by the *Kepler Mission*. In fact, DV is only the beginning of the vetting process for *Kepler* planet discoveries. Pipeline results from TPS and DV are exported to the *Kepler* Science Analysis System (KSAS). There, they are federated with prior results, and planet candidates are scored and ranked in accordance with a list of science criteria. Promising planet candidates are screened by the Threshold Crossing Event Review Team (TCERT) which is comprised of the *Kepler* Science Principal Investigator and selected members of the *Kepler* Science Office and Science Team. Very promising candidates suited to vetting from the ground are further investigated from ground based observatories through the Follow-up Observing Program[8] (FOP). Planet discoveries are announced only after extensive review and follow-up observation where applicable.

This paper describes the nature of the automated validation tests. Section 2 presents an overview of DV and data flow through this software module. Section 3 describes the transiting planet signal generator and limb darkening model. Section 4 describes the automated validation tests for centroid motion, eclipsing binary discrimination, and detection significance; conclusions are discussed in section 5. The fitting of the transiting planet model is described in a companion paper[9].

2. DATA VALIDATION OVERVIEW AND DATA FLOW

DV addresses only LC targets for which the transiting planet detection threshold is exceeded in TPS. The DV unit of work may include one or more targets to support load balancing on worker machines in the Pipeline cluster[10,11]. The duration of DV's standard unit of work for the initial software release is a single science data acquisition quarter (~93 days). A future version of DV will accommodate light curves spanning multiple LC target tables and quarterly spacecraft rolls. At that point, DV will likely be invoked quarterly with all data acquired since the beginning of Quarter 1 (12 May 2009) for targets with TCEs.

Figure 1 illustrates data flow through DV within the *Kepler* SOC Pipeline including the major fields in DV input and output structures, and DV's primary components. DV also automatically generates an extensive report in PDF (not shown in Figure 1) for each target processed and saves it to the *Kepler* Database[12] (*Kepler* DB) with other DV outputs when the Pipeline module is executed.

Data timestamps in the Pipeline are specified in Modified Julian Days (MJD), representing the start, middle and end of each LC observation aboard the spacecraft. Timestamps are adjusted for each target to the solar system barycenter to prevent modulation of the transit timing by the heliocentric orbit of the photometer. Sky coordinates of the individual

targets are obtained from the *Kepler Input Catalog*[13] (KIC) and NAIF SPICE kernels. The latter contains the reconstructed spacecraft trajectory and solar system ephemeris and are produced by the JPL Navigation organization. The timestamp corrections also include small offsets introduced by the multiplexed readout of the CCD array.

Ancillary engineering data (sensor data such as temperature and state of reaction wheels) from the spacecraft, ancillary pipeline data from other SOC software modules, and motion polynomials from PA[5] are utilized to detrend the target light curves for planet model fitting and for performing the DV Centroid Test. The ancillary data and motion polynomials are first synchronized to the LC timestamps.

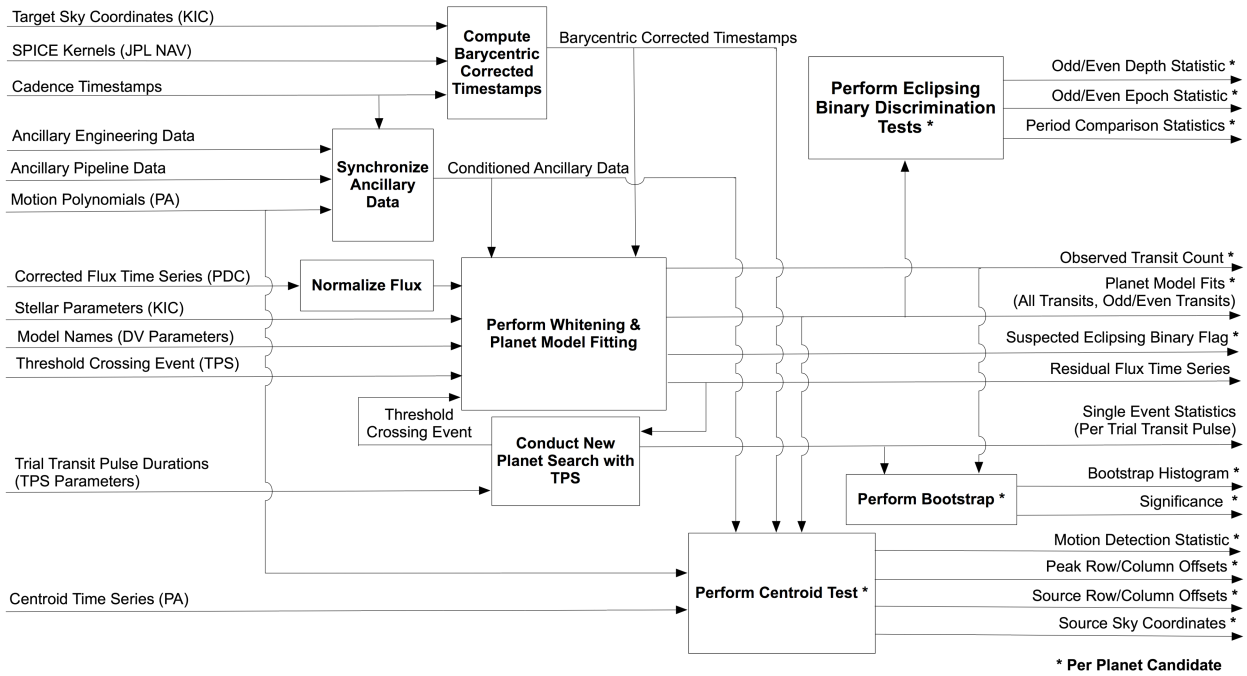

Figure 1. Data flow diagram for the Data Validation (DV) component of the Kepler SOC Pipeline. Processes performed for all planet candidates associated with each DV target and outputs produced for all planet candidates are shown with asterisks. Inputs are obtained from the *Kepler* DB and outputs are written to the *Kepler* DB.

The corrected flux light curve generated in PDC for each DV target is initially normalized such that the values inside the transit indicate the fractional transit depth and those out-of-transit are zero-valued. The normalized flux is whitened to account for stellar variability and fit with a transiting planet model in an iterative process. The transit model is also separately fit to sets of odd and even transits for further analysis of the planet candidate. The DV fit algorithms are sufficiently complex that they are described in detail in a separate article[9], but an overview of the DV transit model generator is described in Section 3.

After the fitting process is complete, the fitted transits are removed and TPS subjects the residual flux to a search for additional planets. The whitening and fitting process is repeated for a new planet candidate if an additional TCE is generated. The search for additional planets concludes when no additional TCEs are produced or an iteration limit is reached. After all planet candidates have been identified, the final residual flux time series, single event statistics for all trial transit pulses, parameter values, and associated covariance matrices for all model fits are saved to the *Kepler* DB, as is a flag for each planet candidate that the fitter suspects to be an eclipsing binary.

A centroid motion test is performed on the centroid time series for each target to ascertain whether there is statistically significant motion of the aperture photocenter during the transits of the respective planet candidates. Centroid motion can be a strong indicator that the observed events may be due to a background eclipsing binary present in the stellar

aperture. Centroid motion alone cannot be used to rule out true planet candidates, however, as there will be motion of the centroid for a target with a legitimate transiting planet if there is any significant crowding in the aperture. The peak centroid row and column offsets during transit are determined, and the change in brightness during transit is utilized to determine the actual row and column offsets of the transit (or eclipse) source from the nominal out-of-transit centroid coordinates. The test is also intended to produce the celestial coordinates of the source. The barycentric corrected timestamps are also utilized when the centroid test is performed.

A series of eclipsing binary discrimination tests is conducted on key model fit parameters to determine if the planet candidate is statistically likely to be a true transiting planet or an eclipsing binary. The depths of the odd and even transit sequences for each planet candidate are compared statistically for equality. The timing of the first transit in the odd and even transit sequences are compared statistically for consistency with the period for all observed transits. In the cases of both depth and timing, equality is consistent with a true planet. Finally, the period for each of the planet candidates associated with a given target is compared statistically with the next shorter and next longer period of all planet candidates for the given target. Equality here is indicative that the candidate is *not* a true planet.

A statistical bootstrap test is performed for each planet candidate to determine the likelihood that the detection statistic reported in the transiting planet search would have been produced in the absence of any transits by noise alone. A histogram of multiple event statistics is populated based on the single event statistics computed from the final residual time series for the target and the number of observed transits for the given planet candidate. The probability of false detection (i.e. bootstrap significance) is given by the probability that the multiple event statistics represented in the histogram exceed the value of the maximum multiple event statistic for the TCE associated with the given planet candidate. The bootstrap results, in addition to the results of the other automated DV tests, are saved to the *Kepler* DB.

3. TRANSIT MODEL

The DV fitter performs iterative fits of a planet model to potential candidate light curves[9]. The planet model uses TCE parameters and stellar parameters obtained from the KIC (stellar radius, effective temperature, and surface gravity) to estimate limb darkening coefficients and compute light curves at barycentric-corrected cadence timestamps.

TCE parameters (duration of trial transit pulse, phase of first transit, orbital period, and maximum event detection statistic) are combined with KIC parameters to generate the following set of parameters to seed the fit: the transit epoch (time to first mid-transit), orbital eccentricity, longitude of periastron, minimum impact parameter, star radius, transit depth, and orbital period. Note that we assume central transits to seed the fit (minimum impact parameter = 0) and circular orbits throughout (eccentricity = 0 and longitude of periastron = 0). Once the initial planet model is generated and fitted, we compute and output the planet semimajor axis, planet radius, transit duration, and transit ingress time in order to obtain a complete set of model parameters. After the initial fit is performed, we use only physical parameters for subsequent fits, which include planet radius and semimajor axis instead of transit depth and orbital period.

Tables developed by Claret[14] supply nonlinear limb darkening coefficients for a range of stellar parameters. To estimate the coefficients for each TCE, we interpolate across the Claret tables using stellar surface gravity and effective temperature, turbulent velocity and stellar metallicity equal to zero, and values for the *Kepler* R-band. To generate the light curve, we first compute the orbit at the exposure times within each transit (using the *Kepler* CCD exposure time, read-out time, and number of exposures per cadence). The orbit is then rotated to obtain the desired minimum impact parameter projected onto the plane of the sky. The Mandel-Agol[15] methodology is used to compute the light curve at the exposure times from the time-dependent impact parameters, limb-darkening coefficients, and ratio of the planet/star radii. If the normalized radius of the eclipsing body is less than ~0.01, a small-body approximation is implemented which speeds up the algorithm. In this approximation, it is assumed that the surface brightness of a star is constant under the disk of the eclipsing object, and the semimajor axis is large compared to the size of the star so that the orbit is essentially a straight line. The time series that is output from the transit model represents the integral of the transit signal over each LC and is normalized to zero during out-of-transit for use by the fitter. The DV inputs are designed to accommodate multiple planet and limb-darkening models, but we currently only support the Mandel-Agol[15] analytic models and the Claret[14] limb darkening tables.

4. VALIDATION TESTS

4.1 Centroid test

We describe the planned implementation of the DV Centroid Test. Note that this test as implemented in the *Kepler* Science Processing Pipeline as of Spring 2010 does not yet meet the full intent of the test as described in this section.

4.1.1 Overview

The purpose of the DV Centroid Test is to assess correlations between variations in the centroid (photocenter) time series and fitted transit signatures in the corrected flux time series. If the centroid variations are uncorrelated with a transit signature, the transit signature is likely due to variations in flux from the target star and not from a background source within the target aperture. One possible source of such variations is a planetary transit of the target star.

If the centroid variations are highly correlated with a transit signature, the transit signature may be due to a background source such as a faint eclipsing binary. A high correlation does not necessarily rule out a planetary transit of the target star, however, if the target aperture is crowded, the centroid shift may, in fact, be due to a planetary transit. It is therefore necessary to follow up detected correlations with an estimate of the location of the centroid perturbing source.

An estimate of the source location in row and column coordinates on the focal plane may be obtained from the fractional depth of the transit feature in the flux time series, the absolute offset of the corresponding feature in the centroid time series, and the nominal out-of-transit centroid value. These row and column coordinates may then be converted to celestial coordinates and compared to the known location of the target star and other nearby background stars.

A measure of the correlation—the detection statistic—is obtained by applying a matched filter to the whitened row and column centroid time series and adding in quadrature. The relevance of the measured correlation—the significance—is developed assuming the detection statistic is a Chi-squared variable with two degrees of freedom. The significance has a value between zero and one. It is the likelihood that a detection statistic at least as large as the one calculated would be obtained from uncorrelated data containing only random statistical fluctuations. For the DV Centroid Test, a reported significance of zero indicates high confidence that transit features in the flux time series are correlated with features in the centroid time series and implies that transit features in the flux time series may be due to a background source. A reported significance of one indicates low confidence of correlation and implies that transit features in the flux time series may be due to a planetary transit of the target star.

4.1.2 Implementation

The DV Centroid Test processes one target at a time. The planned implementation is shown on the data flow diagram in Figure 2a. Inputs are the corrected flux time series (from PDC), residual flux time series, fitted transit models, barycentric timestamps, row and column centroid time series, and motion polynomials (latter two from PA).

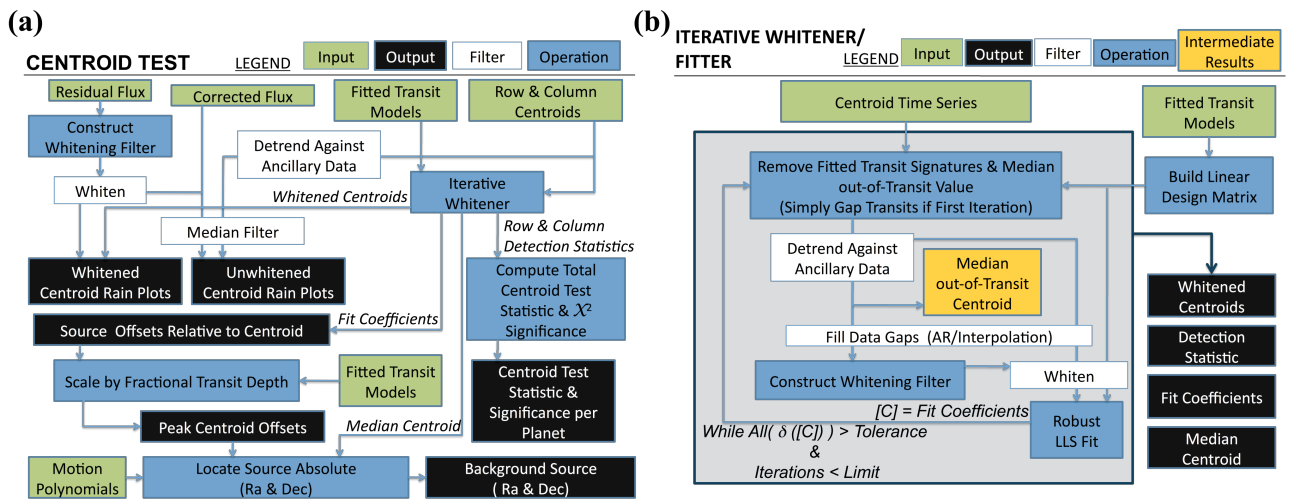

Figure 2. (a) Data flow diagram for the DV Centroid Test. (b) Data flow diagram for the iterative whitener used by the DV Centroid Test.

First, the residual flux time series is used to construct a whitening filter for the corrected flux time series as shown in Figure 2b (note that the whitener in the centroid test is performed independently of the whitener in the fitter[8]). Then the corrected flux time series is median-filtered and whitened. The length of the median filter is selected to preserve features with time scales on the order of the shortest fitted transit identified previously in DV by the fitter. Next, the row and column centroid time series are detrended against ancillary data to remove systematic variations correlated with known sources such as differential velocity aberration and temperature variations of the CCD readout electronics. These are then passed through the same median filter as the corrected flux time series. The row and column centroid time series and the fitted transit models are passed to an iterative whitener which produces the row and column whitened centroid time series, median row and column out-of-transit centroid value, the background source offset from the median centroid (in row and column) and the row and column centroid detection statistic.

The median-filtered corrected flux is plotted as a function of the detrended median-filtered centroids. This "cloud" or "rain" plot is used as a qualitative diagnostic to check for correlations between the flux and centroid time series. Figure 3 shows such plots in the unwhitened domain. The strength and direction of any "wind" observed in these plots indicate the magnitude and sign of any correlations. This plot is generated in the whitened domain as well.

The outputs of the centroid test are the following: centroid detection statistic, significance of the centroid statistic, maximum centroid offset in row and column, location of the background source relative to the nominal target centroid in row and column, celestial location of the background source in right ascension and declination.

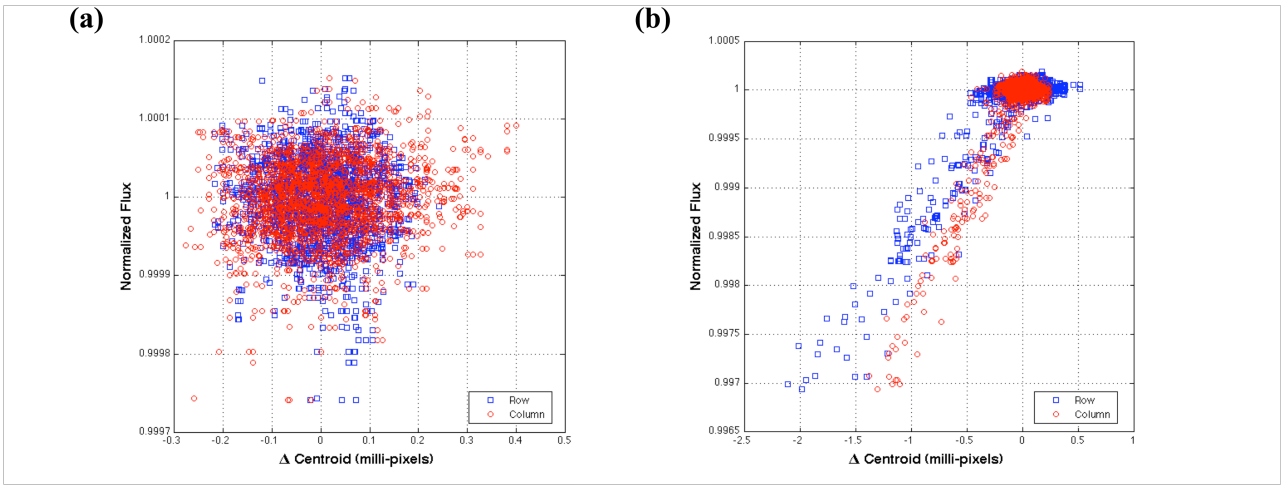

Figure 3. Flux weighted centroids "rain" or "cloud" plot in the unwhitened domain for two synthetic targets. (a) The transit signatures of a planet candidate consists of noise and there is little or no correlation between transit signatures and centroid shifts. (b) A background eclipsing binary within the aperture of a target of interest; transit signature is above noise and correlated with up to 2 milli-pixel centroid shift, hence the "rain" with "wind."

4.1.3 Estimating the location of the background source

The flux in the target aperture is the sum of the flux from the target star plus the flux from all other sources in the aperture, i.e., the background. The change in the centroid due to a background eclipse event is the ratio of the actual distance between the background binary and the target on the CCD to the change in brightness over the target aperture. The centroid shift in terms of the brightness contributed by the target star (B) and the brightness contributed by the background source (b) is:

$$\delta x = \frac{b\Delta x}{B+b} - \frac{(b-\delta b)\Delta x}{B+b-\delta b} \tag{1}$$

Where b = brightness of the background binary, B = brightness of the target star, δb = change in brightness of the background binary during eclipse, Δx = spatial offset of the background binary from the target centroid, and δx = change in centroid during the transit feature.

If the fractional transit depth is small compared to unity, a background binary eclipse can mimic a planetary transit of the target. For all planetary transit candidates identified by the fitter, it is the case that the apparent fractional transit depth is small compared to unity, e.g. $\delta b/(B + b) << 1$. With this approximation, the offset of the background source relative to the nominal out-of-transit centroid is:

$$\Delta x = \delta x \big/ \delta b/(B+b) \tag{2}$$

Standard propagation of errors (assuming independence) gives the variance of Δx as:

$$\sigma^2_{\Delta x} = \left(\delta b/(B+b)\right)^{-2}\left[\sigma^2_{\delta x} + (\Delta x)^2 \sigma^2_{\delta b/(B+b)}\right] \tag{3}$$

As implemented in the DV Centroid Test, the background source offset from the target centroid (Δx) and its variance ($\sigma^2_{\delta x}$) are determined directly within the iterative whitener from a fit of the whitened centroid data to a linear

combination of whitened transit models. The resulting fit coefficients (e.g. model scale factors) are in fact the background source offsets in the unwhitened domain. The corresponding centroid shift (δx) and its variance ($\sigma^2_{\delta x}$) in the unwhitened domain are then calculated using Equations (2) and (3). Adding Δx to the median out-of-transit centroid gives the background source location in row and column on the CCD. Inverting the motion polynomial for the cadence associated with the median out-of-transit centroid gives the celestial source location in right ascension and declination.

4.1.4 Generating the detection statistic

The detection statistic provides a measure of the relevance of the Linear Least Squares (LLS) fit results by comparing the size of the fitted signal to the nominal noise level in the data. It is calculated in the whitened domain as the inner product of the data and the candidate signal normalized by the nominal standard deviation of the data and by the norm of the candidate signal:

$$l = \frac{b \cdot s}{\sigma \sqrt{s \cdot s}} \tag{4}$$

Where l = detection statistic, b = raw data, s = signal to detect, and σ = nominal standard deviation.

A separate detection statistic is calculated for the row and column centroid time series for each transit signature modeled by the fitter. The sum of the squares of the row and column detection statistics form the total centroid detection statistic. There is one total detection statistic produced for each fitted planet candidate. The row and column detection statistics are assumed to be independent Chi-square variables making the total detection statistic a Chi-square variable as well, but with two degrees of freedom. Evaluating the Chi-square cumulative distribution function (CDF) for the total detection statistic value and two degrees of freedom yields the probability of producing a statistic less than or equal to the one observed given uncorrelated data containing only random statistical fluctuations (the null hypothesis). In the DV Centroid Test, the row and column centroid detection statistics are easily determined within the iterative whitener from the output of the last iteration. For each planet candidate, the detection statistic is the square root of the Chi-squared of the corresponding scaled whitened transit model.

According to the DV convention, a significance of one shall be consistent with the detection of a planet and a significance of zero shall be consistent with no planet detected. We therefore report the complement of the Chi-square CDF result as the significance of the detection. The reported statistical significance is a value between zero and one where zero indicates high correlation (the transit feature in the flux time series may be due to a background source) and one indicates no correlation (the transit feature in the flux time series may be due to a transit of the target star).

4.2 Eclipsing binary discrimination tests

The eclipses of an eclipsing binary system and the transits of a planet around a star may appear similar in a flux time series. To discriminate between them, we have designed and developed several tests based on their different characteristics. This section describes the tests, which collectively are called the *Eclipsing Binary Discrimination (EBD) Tests.* The EBD tests are statistical hypothesis tests on the consistency of key transit parameters. The fitter provides the parameters for each TCE[9]. The EBD tests consist of the following: Odd/Even Transit Depth Test, Odd/Even Transit Epoch Test, and Orbital Period Test.

The depths of multiple transits of a planet are ideally the same, and the transits of a planet are evenly spaced in time. In contrast, the depths of primary and secondary eclipses of an eclipsing binary system are generally different due to the difference in size and brightness of the two stars. The difference in the epoch times of the primary and secondary eclipses is usually not equal to half of the orbital period of the eclipsing binary system, since the orbit of two stars moving around their gravitational center is generally not circular. The *Odd/Even Transit Depth Test* and the *Odd/Even Transit Epoch Test* are designed to distinguish the flux time series of an eclipsing binary system whose primary and secondary eclipses are identified as one TCE. For each TCE, the transits are divided into odd and even sets, and the depths and the epoch times of the odd and even transit sequences are estimated separately in the fitter. The null hypothesis of the Odd/Even Transit Depth Test is that the estimated transit depths of odd/even transit fits of the TCE are consistent, and the null hypothesis of the Odd/Even Transit Epoch Test is that the difference of the epoch times of the odd/even transits is consistent with the average period of the transits. A small significance level leads to the rejection of the null hypothesis—i.e., the TCE is unlikely to be due to a planet.

Two nearly equal size planets cannot be in a stable orbit at the same period around a star[16]. However, in an eclipsing binary system, two stars move in one orbit around a common center of gravity with a single orbital period. If the primary and secondary eclipses are identified as two TCEs, the observed periods will be the same. The *Orbital Period Test* is designed to distinguish between the flux time series of a star with two transiting planets and that of an eclipsing binary system with primary and secondary eclipses, reported as two separate TCEs. The null hypothesis is that the orbital periods of the two TCEs are consistent. A small significance level leads to the rejection of the null hypothesis—i.e., the two TCEs are unlikely to be the primary and secondary eclipses of an eclipsing binary system.

In a general case, the consistency check of N independent measurements of a parameter, denoted as $\{x_i\}$, i=1, …, N, associated with uncertainties $\{\sigma(x_i)\}$, can be modeled as a statistical test with the null hypothesis: $\{x_i\}$ are drawn from N independent Gaussian distributions with the same mean and standard deviations equal to $\{\sigma(x_i)\}$. The statistic and significance level are determined as

$$s = (x_1 - \bar{x})^2 / \sigma^2(x_1) + \ldots\ldots + (x_N - \bar{x})^2 / \sigma^2(x_N) \tag{5}$$

$$p = \Pr\{\chi^2(N-1) > s\} \tag{6}$$

where $\bar{x}$ is the weighted mean of the measurements, and the weight of x_i is inversely proportional to $\sigma^2(x_i)$. $\chi^2(N\text{-}1)$ denotes a Chi-squared distribution with N-1 degrees of freedom and $Pr\{\cdot\}$ denotes "probability of." A small significance level (typically less than 0.05) leads to the rejection of the null hypothesis, i.e., the measurements are inconsistent.

Figure 4 shows the normalized flux time series of two target stars for second quarter flight data. In Figure 4a, one TCE is reported by TPS, and the transits of the TCE are labeled with dash-dot lines. The difference between the odd/even transit depths are much larger than the uncertainties, resulting in a large statistic (~1800) and a small significance level (~0) of the odd/even transit depth test. Therefore, the null hypothesis is rejected with high confidence—i.e., the light curve shown in Figure 4a is unlikely to be due to a planet. In Figure 4b, two TCEs are reported by TPS, and the transits of the first and second TCEs are labeled with dash-dot lines and dash lines, respectively. The calculated statistic (~0) and significance level (~1) of the orbital period test can be verified using the following observation: the estimated orbital periods of the two TCEs are almost equal, suggesting that the two TCEs are due to primary and secondary eclipses of an eclipsing binary system.

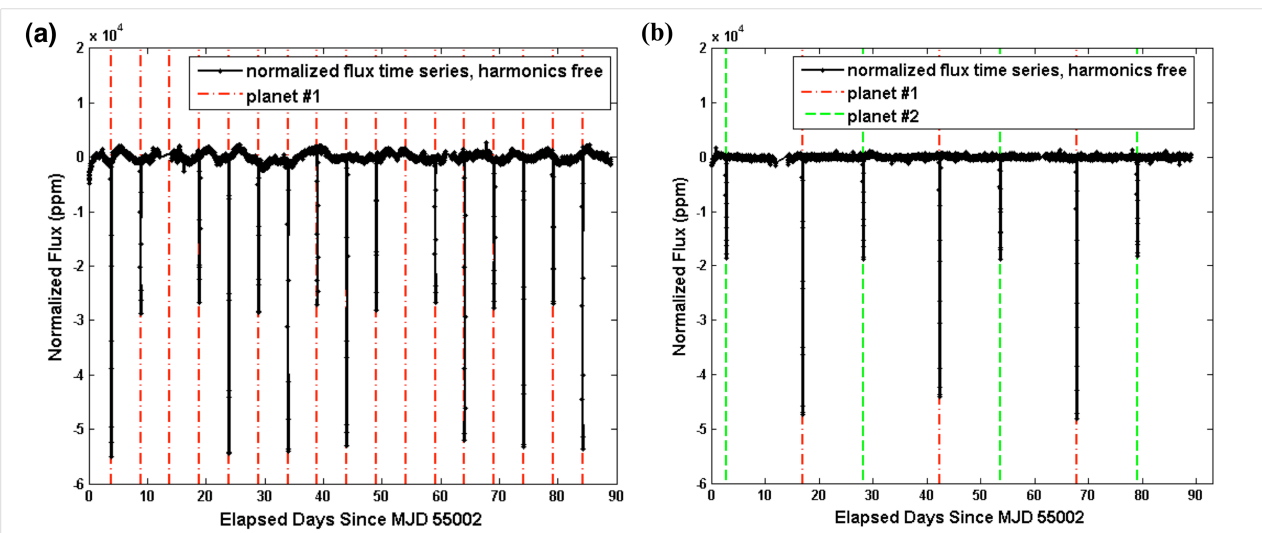

Figure 4. Normalized flux time series of target stars for second quarter flight data. (a) one TCE, (b) two TCEs.

4.3 Bootstrap test

4.3.1 Overview

In the search for transiting planets, the multiple event statistics[7] of a planet candidate with SNR=8 (typical for four transits of an Earth-size planet orbiting a 12th magnitude Sun-like star) are represented by the alternative hypothesis, H1, as depicted in Figure 5a. If all transit-like features are removed from the flux time series, a subsequent search for transits will generate the null multiple event statistics as depicted by H0. The DV Bootstrap Test seeks to evaluate the likelihood that a TCE produced under H1 could alternatively be generated from the null event statistics by chance alone, i.e., it seeks to evaluate the cumulative sum of the probabilities from the detection statistic that triggered the event to the end of the tail in H0. DV Bootstrap first constructs a histogram of the tail end of the null multiple event statistics starting from the search transit threshold, η. From this histogram, it obtains the probabilities at each detection statistic, then computes the cumulative sum of the probabilities to obtain the complementary cumulative distribution function (CCDF). The false alarm of a planet candidate is evaluated from the CCDF at the TCE detection statistic by either interpolating or extrapolating.

To construct H0 in the traditional way for a search consisting of T transits, we need to form the multiple event statistics for all combinations of N single event statistics, i.e., H0 consists of N^T multiple event statistics with all transits removed. For example, in a flux time series of 4500 cadences (~1 quarter) of data with 5 transits, there are $4500^5 = 2 \times 10^{18}$ null multiple event statistics that can be formed; for 4500 cadences of data with 6 transits, there are $4500^6 = 8 \times 10^{21}$ null multiple event statistics. The computational burden scales by the number of single event statistics to the power of number of transits, and generating the null distribution in this manner is computationally prohibitive. The solution to this problem lies in realizing that to compute the false alarm probability of a planet candidate, only the tail of the distribution of H0 above η is of interest.

In statistics, *bootstrapping*[17] is a method for estimating the sampling distribution of an estimator by repeatedly sampling, with replacement, from the original sample. We take a "modified" bootstrap approach by realizing that we are only interested in the upper tail portion of H0, and that the number of ways that a multiple event statistic can be formed from the single event statistic is known[18]. In this "modified" bootstrap approach, a counter, representing the indices that form the multiple event statistics, is used to index, obtain combinations, and update a histogram for the construction of H0.

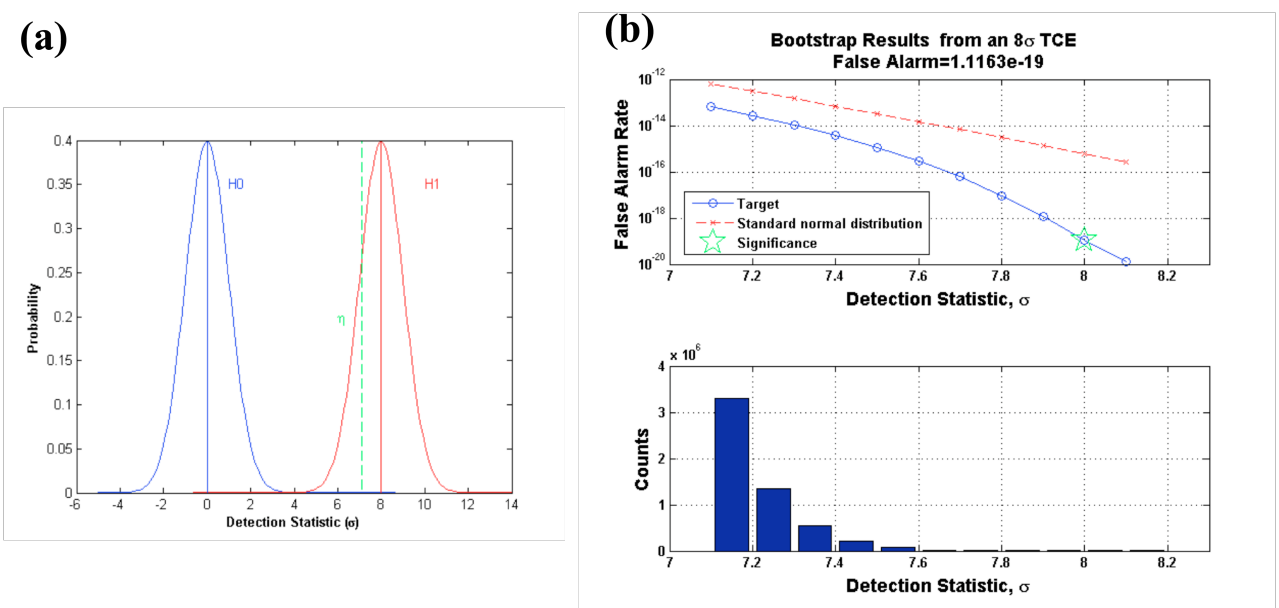

Figure 5. (a) Probability distribution for null detection statistics (H0), and for detection statistics of a transiting planet with SNR= 8 (H1). Search transit threshold is designated by η, and is set at 7.1σ. (b) Bottom: tail end of the histogram formed following the bootstrap algorithm. Top: cumulative sum of the probabilities (derived from the histogram below) from upper tail to η; false alarm probability is indicated by the star.

4.3.2 Algorithm

In the case of a flux time series containing four transits for a duration of 4 years (~72000 cadences) with all transits removed, the algorithm is described as follows. First, sort the transit-free single event statistics in descending order, preserving the "numerator" and "denominator"[7,18] so that the multiple event statistics can be formed. Set up a histogram by choosing a bin width of 0.1σ in which the minimum is at or below η and the maximum bin is the highest null multiple event statistic, rounded up to the nearest bin. Begin with a counter set at [1, 1, 1, 1], form the multiple event statistic from the associated single event statistics and compute the number of combinations for these digits (1). Add 1 count (number of combinations for [1, 1, 1, 1]) to the histogram bin corresponding to this detection statistic. Increment the counter to [1, 1, 1, 2], form the multiple event statistic, compute the number of combinations (4) and update the histogram with 4 counts in the bin with the corresponding detection statistic. This process is repeated many times and the digits in the counter are increased monotonically. If the formed multiple event statistic falls below η, the adjacent digit is incremented to 2 so that the counter reads [1, 1, 2, 2]. The procedure stops when successive multiple event statistics formed fall below η or when the counter reads [72000, 72000, 72000, 72000]. If there are multiple searches with different trial transit pulse widths, as is the case in DV (3, 6, 12 hour searches), then a histogram is formed for each. To obtain the probabilities, the histogram counts for each trial transit pulse are summed and divided by the total number of events, and the false alarm rate is calculated as the cumulative sum of the probabilities from left to right. Finally, a logarithmic robust linear fit is performed on the curve, and the false alarm probability for the TCE is evaluated.

The procedure described above is still computationally intensive. To ameliorate this, we have implemented a skip count feature for targets with many transits. If the detection statistic formed is above η, the counter is incremented by a fixed deviate. The minimum histogram bin is chosen to be conservatively below η to account for inaccuracies when skip counts are implemented. After histograms have been generated, their counts are scaled by 1 + skip count.

4.3.3 Example

We apply the procedure above using a series of 72000 normally distributed random numbers to simulate a transit-free single event statistic time series over the course of a 4-year mission. We assume that the TCE was triggered from an 8σ event. We then evaluate the likelihood that the TCE was caused by chance alone via our bootstrap method. Figure 5b illustrates the bootstrap results: tail end of the null distribution is represented by the histogram in the lower panel, its cumulative sum is shown by the circles on the top panel. We interpolate and compute a false alarm probability of 1.1×10^{-19} as indicated by the star, or ~ 1 in 9×10^{18} that this observation was produced by chance alone.

4.3.4 Limitations

In certain cases, the bootstrap algorithm cannot be used, e.g., hot Jupiters with periods on the order of a few days that generate 20 or more transits per quarter; these cases cannot be bootstrapped because calculating the combinations of the digits in the counter depends on calculating and/or representing factorial of the number of transits. For most computers, 20 is the maximum factorial that can be calculated accurately. To prevent halting downstream processes in DV, we have implemented a limit to the number of iterations that bootstrap will run before aborting. Perhaps the biggest limitation to the bootstrap test is the assumption that all transit signatures have been removed and the null single event statistic are comprised purely of noise. The DV Bootstrap Test is most useful with low numbers of transits that result in small changes in transit depths that trigger TCEs in the vicinity of η —i.e., TCEs that suggest Earth analogues. Hot Jupiter-like planets that exhibit deep transits, with periods on the order of days that trigger TCEs much greater than η (e.g., >1,000σ), will yield false alarm values of nearly zero if bootstrapped. In this respect, the bootstrap test is especially effective at flagging TCEs triggered by Earth-like planets where analyses have not accounted for all transit-like features, which results in non-Gaussian statistics and higher false alarm probabilities.

5. CONCLUSIONS

We have presented a suite of statistical validation tests performed in DV, which consists of a test to assess correlations between centroid shifts and transit signatures, eclipsing binary discrimination tests, and a false alarm bootstrap test. DV test performance was evaluated using a set of 70 simulated[19] targets in which the ground-truth was known. These targets consisted of synthetic Earths, Jupiters, eclipsing binary systems, and background eclipsing binary systems and a combination of these. Results of this exercise are described in the companion paper[8].

ACKNOWLEDGEMENTS

The authors would like to thank all the members of the *Kepler Mission* without whom DV results and this paper would not have been possible. Funding for the *Kepler Mission* has been provided by the NASA Science Mission Directorate.

REFERENCES

[1] Koch, D.G., et al., "*Kepler* Mission Design, Realized Photometric Performance, and Early Science," Astrophysical Journal Letters, 713, L79-L86 (2010).

[2] Jenkins, J.M., et al., "Overview of the *Kepler* Science Processing Pipeline," Astrophysical Journal Letters, 713, L87-L91 (2010).

[3] Middour, C., et al., "*Kepler* Science Operations Center Architecture," Proc. SPIE 7740, in press (2010).

[4] Quintana, E.V., et al., "Pixel-level calibration in the Kepler Science Operations Center pipeline," Proc. SPIE 7740, in press (2010).

[5] Twicken, J.D., et al., "Photometric analysis in the *Kepler* Science Operations Center pipeline," Proc. SPIE 7740, in press (2010).

[6] Twicken, J.D., et al., "Presearch Data Conditioning in the *Kepler* Science Operations Center pipeline," Proc. SPIE 7740, in press (2010).

[7] Jenkins, J.M., et al., "Transiting planet search in the *Kepler* pipeline," Proc. SPIE 7740, in press (2010).

[8] Gaultier, T.N, et al., "The *Kepler* Follow-up Observation Program," arXiv:1001.0352v1 [astro-ph.EP], in press (2010).

[9] Tenenbaum, P. et al., "An algorithm for fitting of planet models to *Kepler* light curves," Proc. SPIE 7740, in press (2010).

[10] Klaus, T.C., et al., "The *Kepler* Science Operations Center pipeline framework extension," Proc. SPIE 7740, in press (2010).

[11] Klaus, T.C., et al., "*Kepler* Science Operations Center pipeline framework," Proc. SPIE 7740, in press (2010).

[12] McCauliff, S., et al., "The *Kepler* DB, a database management system for arrays, sparse arrays, and binary data," Proc. SPIE 7740, in press (2010).

[13] www.cfa.harvard.edu/kepler/kic/kicindex.html

[14] Claret, A., "Non-linear Limb-darkening Law for LTE Models," Astronomy & Astrophysics, 363, 1081 (2000).

[15] Mandel, K. and Agol, E., "Analytic Lightcurves for Planetary Transit Searches," Astrophysical Journal, 580, L171-L176 (2002).

[16] de Pater, I. and Lissauer, J.J, [Planetary Sciences], Cambridge University Press, Cambridge, 24 (2001).

[17] Good, P.I., [Resampling Methods: a Practical Guide to Data Analysis], Birkhauser, Boston, 8 (2005).

[18] Jenkins, J.M., "The Impact of Solar-like Variability on the Detectability of Transiting Terrestrial Planets," Astrophysical Journal, 575, 493-505 (2002).

[19] Bryson, S.T., et al., "The *Kepler* end-to-end model: creating high-fidelity simulations to test *Kepler* ground processing," Proc. SPIE 7738, in press (2010).