MEASURING TRANSIT SIGNAL RECOVERY IN THE KEPLER PIPELINE. I. INDIVIDUAL EVENTS

JESSIE L. CHRISTIANSEN\(^1\), BRUCE D. CLARKE\(^1\), CHRISTOPHER J. BURKE\(^1\), JON M. JENKINS\(^1\), THOMAS S. BARCLAY\(^2\), ERIC B. FORD\(^3\), MICHAEL R. HAAS\(^1\), ANIMA SABALE\(^4\), SHAWN SEADER\(^1\), JEFFREY CLAIBORNE SMITH\(^1\), PETER TENENBAUM\(^1\), JOSEPH D. TWICKEN\(^1\), AKM KAMAL UDDIN\(^4\), AND SUSAN E. THOMPSON\(^1\)

\(^1\) SETI Institute/NASA Ames Research Center, M/S 244-30, Moffett Field, CA 94035, USA; jessie.l.christiansen@nasa.gov
\(^2\) Bay Area Environmental Research Institute/NASA Ames Research Center, M/S 244-30, Moffett Field, CA 94035, USA
\(^3\) Astronomy Department, University of Florida, 211 Bryant Space Sciences Center, Gainesville, FL 32111, USA
\(^4\) Orbital Sciences Corporation/NASA Ames Research Center, Moffett Field, CA 94035, USA

Received 2013 February 20; accepted 2013 June 1; published 2013 August 1

ABSTRACT

The Kepler mission was designed to measure the frequency of Earth-size planets in the habitable zone of Sun-like stars. A crucial component for recovering the underlying planet population from a sample of detected planets is understanding the completeness of that sample—the fraction of the planets that could have been discovered in a given data set that actually were detected. Here, we outline the information required to determine the sample completeness, and describe an experiment to address a specific aspect of that question, i.e., the issue of transit signal recovery. We investigate the extent to which the Kepler pipeline preserves individual transit signals by injecting simulated transits into the pixel-level data, processing the modified pixels through the pipeline, and comparing the measured transit signal-to-noise ratio (S/N) to that expected without perturbation by the pipeline. We inject simulated transit signals across the full focal plane for a set of observations for a duration of 89 days. On average, we find that the S/N of the injected signal is recovered at MS = 0.9973(±0.0012) × BS – 0.0151(±0.0049), where MS is the measured S/N and BS is the baseline, or expected, S/N. The 1σ width of the distribution around this correlation is ±2.64%. This indicates an extremely high fidelity in reproducing the expected detection statistics for single transit events, and provides teams performing their own periodic transit searches the confidence that there is no systematic reduction in transit signal strength introduced by the pipeline. We discuss the pipeline processes that cause the measured S/N to deviate significantly from the baseline S/N for a small fraction of targets; these are primarily the handling of data adjacent to spacecraft re-pointings and the removal of harmonics prior to the measurement of the S/N. Finally, we outline the further work required to characterize the completeness of the Kepler pipeline.

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

Online-only material: color figures, machine-readable table

1. INTRODUCTION

The Kepler mission is a NASA Discovery Program mission designed to examine the population of planetary systems using high precision photometric observations. The primary goal of Kepler is to measure \( n_p \), the frequency of Earth-size planets in the habitable zone of Sun-like stars. It was launched in 2009, and since then has been nearly continuously monitoring the brightness of \( \sim 160,000 \) stars in 29.4 minute integrations in a fixed field of view in the constellation Cygnus, looking for the periodic dimmings indicative of transiting planets (Borucki et al. 2011a, 2011b, hereafter B11; Batalha et al. 2013, hereafter B13).

In order to determine the real, underlying population of planets from a sample of planet candidates, we need to understand two properties of the sample: the false negative rate, or what fraction of the real planets that should have been detected are not included in the sample (also called the survey completeness or survey detection efficiency); and the false positive rate, or what fraction of the candidates in the sample do not represent real planets (also called the reliability of the sample). Initial analyses of the published Kepler planet candidate catalogs (e.g., B11; Youdin 2011; Howard et al. 2012; Dong & Zhu 2012; Fressin et al. 2013) use various estimates of the false positive and false negative rates to constrain the occurrence rate of planets, but as yet there is no definitive empirical measure of either value, in particular for the false negative rate.

The false negative rate of a given transit survey has typically been measured directly by the injection of simulated transit signals into the observed data and measurement of the subsequent rate of discovery. Studies by Burke et al. (2006), Croll et al. (2007), Weldrake et al. (2008), Ballard et al. (2010, 2011) used extensive Monte Carlo simulations of transit injection and recovery to constrain the likelihood of a given transit signal being recovered in their analyses. Thus far, there has been no direct measurement of the false negative rate in the Kepler planet candidate sample, so published analyses of the planet candidate lists have needed to estimate the detection efficiency. Differences in the assumptions made to arrive at an estimate can result in differences in the derived planet occurrence rates. B11 assumed a Gaussian probability of detection with unit variance around the stated detection threshold of 7.1σ, i.e., a 50% chance of detecting a transiting signal with a signal-to-noise ratio (S/N) of 7.1σ, an 84% chance of detecting an 8.1σ signal, a 97.5% chance of detecting a 9.1σ signal, and so forth, for a signal with a minimum of two transits, out to orbital periods of 138 days. Howard et al. (2012) assumed a 100% detection efficiency for transiting signals with a S/N > 10σ, for planets larger than 2\( R_{\oplus} \), and periods less than 50 days, based on the decreasing rate of detection of these objects with time and their readily apparent signals to the eye. Youdin (2011) makes the same assumption and extends the parameter space to planets down to 0.5\( R_{\oplus} \). In a similar fashion, Dong & Zhu (2012) assume a 100% detection efficiency for transit signals with a S/N > 8σ, out to...
periods of 250 days. Fressin et al. (2013) estimate the detection efficiency empirically from a comparison of the distribution of their modeled false positives with the planet candidates reported in B13, finding a linear increase in detection efficiency from 0% at 6σ to 100% at 16σ. These teams report significant variation in their derived planet occurrence rates. Although it is difficult to disentangle the impact of the various assumptions of detection efficiency from the assumptions about the false positive rate, they certainly impact the derived rates.

While the assumptions about the detection efficiency implemented thus far are well motivated, comparison of the catalogs released progressively by Kepler indicates that the simpler S/N thresholds are insufficient to describe the detection efficiency. B13 showed that the number and distribution of additional planet candidates detected in their sample, compared to the previous catalog, was significantly higher and systematically different from what would be expected from a simple S/N extrapolation of the planet candidates listed in B11. This pointed at both the incompleteness of the B11 sample and the intervening improvement in the pipeline detection efficiency. We expect that the B13 catalog also suffers from incompleteness. Indeed, several teams have identified planet candidates that were not included in that catalog using the same or shorter observing baselines (Fischer et al. 2012; Huang et al. 2013; Ofir & Dreizler 2013). This also highlights the role of the Kepler review team in the final catalog completeness—almost all of the planet candidates identified in the works cited were identified by the pipeline at one point, but were not subsequently promoted to planet candidate status.

As the Kepler catalogs grow more comprehensive and more consistent, and as the planet candidates become more interesting, pushing to longer periods and smaller planet sizes, the question of the detection efficiency of the Kepler pipeline becomes much more critical. The most interesting candidates that Kepler will find will have low S/Ns with a small number of shallow events, potentially very close to the 7.1σ pipeline detection threshold. In this regime, the reliability and completeness of the planet sample may represent large corrections and a significant source of uncertainty, as the assumptions described previously have not been rigorously tested. This work presents the first results of an experiment to directly measure the preservation of transit signals into the Kepler pipeline (i.e., the extent to which transit signals are preserved over the course of the data reduction). We investigate the validity of this assumption by injecting simulated transit signals into the Kepler pixels and processing the pixels through the pipeline in the same manner as the original data, then examining the measured S/N of the injected signals. In Section 2, we summarize the processes in the Kepler pipeline and discuss those that might impact transit recoverability. In Section 3, we describe the experimental design and execution; in Section 4, we present the results; and in Section 5, we discuss the implications. Section 6 summarizes the conclusions and describes future work.

2. THE KEPLER SCIENCE PIPELINE

The Kepler mission design, performance, and data products have been described in several papers (Koch et al. 2010; Borucki et al. 2010); we direct the reader there for a description of the concepts and terminology used in this paper. To understand the method by which we inject simulated transits and the processes that might perturb their recovery, it is important to first understand the data reduction pipeline. The pipeline has also been described in detail in a series of papers; for an overview see Jenkins et al. (2010a) and Figure 1 therein; we briefly review the processes here.

The Calibration (CAL) component, described by Quintana et al. (2010), calibrates the raw pixels from each CCD channel,5 including pixels from the target stars, background, and collateral data.

The Photometry Analysis (PA) component of the pipeline, described by Twicken et al. (2010), constructs the initial flux time series for each target star. After correction for argabrightenings, cosmic rays, and background flux, the flux time series (light curve) for each target is generated by summing, for each cadence, the corrected flux value in each pixel in the optimal aperture defined for that target (Bryson et al. 2010).

The Pre-search Data Conditioning (PDC) component, initially described by Twicken et al. (2010) and updated in Smith et al. (2012) and Stumpe et al. (2012), fits and removes systematic signals from the light curves that are common to many stars on a given channel. See Figure 1 from Stumpe et al. (2012) for examples of the types of systematics that are corrected, some of which are described in Sections 2.1 and 2.2 below. On a given channel, each light curve is correlated with every other light curve. For purely astrophysical signals, we have no expectation for correlations between light curves, which are much more likely to arise as a result of onboard systematics. We rank the light curves by their degree of correlation and select the first 50%, which are the most highly correlated. We then use singular value decomposition on the correlated light curves to create a set of cotrending basis vectors (CBVs). Using a Bayesian maximum a posteriori approach (see Smith et al. 2012 for details), we generate a correction for each light curve from the set of CBVs using priors based on the locations and brightness of neighboring targets.

The Transiting Planet Search (TPS) component, described by Jenkins (2002), Jenkins et al. (2010b), and Tenenbaum et al. (2012, 2013), searches the corrected light curves for periodic box-shaped signals. In order for a given target to be considered a threshold crossing event (TCE), the target must pass a series of tests, the most significant and straightforward of which is that the significance of the detection must be $\geq 7.1\sigma$. See Tenenbaum et al. (2012, 2013) for a detailed description of the test criteria used to generate TCEs previously. Once a target passes these tests, it is considered a TCE and is passed to the final component of the pipeline.

The Data Validation (DV) component, described by Wu et al. (2010), performs a suite of diagnostic tests on the set of TCEs that are detected by TPS. In this work, we are concerning ourselves only with the investigation of the detection statistics as reported by TPS; however, our long-term plan is also to quantify the performance of the DV tests, especially in the area of identifying false positives.

2.1. Signal Distortion

There are many processes by which a transit signal, present in the raw flux time series as measured by Kepler, could be distorted from its original shape and depth. One is aperture errors and losses due to the changing amounts of flux falling into the target aperture. As described above, we integrate the flux for a given cadence over a fixed set of pixels. Any change

5 Kepler began with 84 operating channels; after $\sim 8$ months of operation, one module (comprising four channels) failed, and we are currently operating with 80 channels.
in the location of the target or the telescope focus will cause a different fraction of the total flux from the target to fall within the fixed optimal aperture for each cadence. Changes in the location of the target may be due to jitter and drift in the pointing, plate-scale changes caused by focus changes, or differential velocity aberration. Changes in the telescope focus are caused by thermal changes on board the spacecraft, dominated by a yearly seasonal variation and punctuated by significant transients after spacecraft re-pointings.

A second potential source of distortion is the PDC component of the pipeline. As described above, in PDC we generate a set of CBVs for each channel and use linear combinations of these CBVs to remove the systematics from the target light curves. The CBVs can contain features that mimic transit signals either from large astrophysical signals in the sample of highly correlated light curves (although these should be mitigated by the entropy cleaner in PDC; Smith et al. 2012) or more commonly from systematics that mimic or distort transits. In practice, the entire CBV is being fit to the entire light curve, therefore, scaling the CBV to match local signals (such as transits) in the light curve will typically degrade the fit elsewhere, a less favorable solution. However, the presence of a signal in the CBV at the same time as a signal in the light curve will perturb the final fit and could ultimately decrease the depth of the signal, however slightly.

A third source of signal distortions are short-duration deviations in individual flux time series, which are not common to multiple light curves and therefore are not amenable to being removed with the CBVs and are not corrected elsewhere in the pipeline. These include unidentified cosmic rays and the occasional cadence misidentified as cosmic rays, which leads to an erroneous flux correction. See Figure 1 for an example of the latter. Unidentified or improperly corrected sudden pixel sensitivity dropouts (SPSDs) occurring during a transit will also alter the shape and detectability of that transit.

**2.2. Signal Masking**

Besides distorting transit signals, uncorrected systematics can register as significant detections when compared to a box-shaped trial pulse. In the first three catalogs released by the *Kepler* project, if the strongest signal found on a given target by TPS did not pass the additional tests, as is likely if it was caused by a systematic (see Tenenbaum et al. 2012, 2013 for further details), then TPS would not continue searching that target for weaker signals. Therefore, if there was a real transit signal in the flux time series that was weaker than the systematic, but stronger than $7.1\sigma$, then TPS would not have detected it. We can see evidence of this when examining the detection statistics of the planet candidates in B13—Figure 7 of that paper shows that, instead of increasing toward $7.1\sigma$ as expected, the distribution of signal strengths turns over at $10\sigma$. This points to a population of systematics masking some fraction of real transit signals between $7.1\sigma$ and $10\sigma$.

---

6 Sudden pixel sensitivity dropouts are sudden decreases in the sensitivity of a given pixel due to damage from a high-energy radiation event such as a cosmic ray. After the event, the sensitivity of the pixel is permanently degraded.
In the fourth sample of planet candidates (C. J. Burke et al., in preparation) and future catalogs, this problem is largely mitigated by a redesign of the pipeline. Now, instead of discontinuing the search if the strongest signal in a given flux time series does not pass the tests in TPS, we mask out the cadences that contributed to the spurious signal and perform a new search. We iterate this process until the strongest signal returned by the search passes all the tests or is under the 7.1σ threshold, which should greatly improve the rate of recovery of real transit signals above the threshold. We caution that for completeness analyses, it will be important to take into account the reduced number of cadences that ultimately contribute to the detection. For a given orbital period of interest, we also expect that signal masking will decrease with the increasing baseline of cadences, since real signals should increase in significance with more events. In contrast, spurious signals caused by systematics should decrease in significance because they are not likely to continue lining up in the same way. In this work, we do not attempt to constrain the rate of signal masking in the released catalogs, but we include this discussion here for clarity and to guide future work.

3. EXPERIMENT DESIGN

The goal is to assess the recoverability of a given transit signal in a given target light curve. Ideally, we would measure this by injecting the target flux time series with a grid of simulated transit signals over a set of planet parameters of interest (e.g., size, orbital period) and process the light curve for each signal through the Kepler pipeline and directly measure the detection statistics. However, due to the number of Kepler targets (∼225,000 with at least some observations) and the number of observations per target (53,000 and growing), this is computationally infeasible and we need to find a better solution.

Thus, we assess the recoverability of a given transit signal in an average target light curve in order to decrease the number of tests required by several orders of magnitude. To achieve this, instead of running multiple simulations on each target, we inject each target light curve with a different transit signal arising from a randomly generated single planet candidate, described below, and then measure the recovered detection statistics on average.

Another way to reduce the computational burden is to decrease the observation baseline, since the number of searches performed by the pipeline increases as $N^2$ for $N$ observations. Therefore, we perform this initial test using a single quarter of data—Quarter 3, which spans 89 days from 2009 September 18 to December 16. Since we are largely concerned with the average signal distortion introduced by the processes in the pipeline, we can treat each separate transit event in the light curve as an independent statistical test of the distortion. This allows us to place the transits unphysically close together in the flux time series and therefore to sample more of the systematics in a given flux time series as long as the transit events are separated well enough in time so as to not mutually influence each other’s measured S/N. The variance window over which TPS calculates the noise, in order to determine the significance of a detection, is 30 times the duration of the box pulse being tested. Outside of this window, the local noise properties should not be influenced by neighboring transits. For this investigation, we separate the injected transit epochs by 50 times their duration. In Sections 4 and 5, we discuss pipeline processes that might be influenced by this relatively close separation of transit epochs.

3.1. Transit Model Construction

We generate our injected transits using the Mandel & Agol (2002) model. The parameters of the model are constructed from three observable parameters: (1) the S/N is randomly drawn from a uniform distribution between 2σ and 20σ. This range allows us to determine any correlation between the distortion and the initial baseline S/N; (2) the signal duration is randomly drawn from a uniform distribution between 1 and 16 hr (in the pipeline, we search for transit pulses with durations from 1.5 to 15 hr); and (3) the phase of the first injected transit is randomly drawn from a uniform distribution between 0 and 1. This phase is used with the separation of the transit events to set the initial epoch. For further details on choices made in the generation of the injected transits, see the Appendix.

3.2. Pipeline Processing

We inject the generated transit model into the calibrated pixels for the target in question. These modified pixels are then processed through the pipeline as normal. The only departure from standard operations is that the motion polynomials (used for calculating the location of the target) and the CBVs (used in the correction of systematic errors) are generated from a “clean” pipeline run. This is to avoid corruption from the presence of the injected transits since the motion polynomials and CBVs are generated from the data themselves. In general, the motion polynomials and CBVs should not be affected by the presence of real transits in a small fraction of the light curves, although as described before, these transits can be affected by the CBVs during correction.

In summary, the final order of processing is that we ran the calibrated pixels (the output of CAL) of Quarter 3 through PA, PDC, and TPS, without any modification to generate the motion polynomials, the CBVs, and the rms combined differential photometric precision (CDPP) for each target. We then inject the simulated transit signals into the calibrated pixels, one planet for every target in the full focal plane (84 channels), and re-run the modified pixels through PA, PDC, and TPS utilizing the previously generated information as described.

4. RESULTS

Of the 84 channels in the field of view, 80 channels completed the processing described in Section 3.2. Four channels (channels 23, 45, 62, and 67) failed for reasons unrelated to the transit injection and are not included in the subsequent analysis.

Our goal is to quantify the impact of the pipeline processing on the original injected signal. To achieve this, we inject the simulated transit signals in the pixel-level data and compare the measured S/N (MS) to the expected S/N for that model. Note that when we inject transits into the real data, local noise will result in individual transit events having a range of measured detection statistics, independent of distortions caused by the pipeline processes, so we cannot simply compare the measured S/N to the simulated model S/N. To isolate the impact on the transit signal due to the pipeline from local correlated noise, we instead take the unmodified flux time series from the first pass through the pipeline and, immediately prior to the measurement of the S/N, inject the identical set of transits as those that were processed all the way through the pipeline. This then generates the baseline S/N (BS), which represents how the signal would have been measured in the flux time series without any perturbation by the pipeline.
Figure 2. Measured and baseline S/N for the 1130 targets on channel 1. Left panel: each cadence that was modified by the injection of a transit signal is plotted. All points associated with the same target are plotted in the same color. Right panel: density plot of the same data as the left panel, showing a very tight distribution around the 1:1 correlation (note the log scale in density).

(A color version of this figure is available in the online journal.)

| Channel | # Stars | $a$ (±Δ$a$) | $b$ (±Δ$b$) | HWHM$_{Gauss}$ | HWHM$_{Lor}$ | $N_{supp}$ | $N_{aug}$ |
|---------|---------|-------------|-------------|----------------|--------------|------------|----------|
| 1       | 1130    | −0.0218(15) | 0.9970(03)  | 0.0315         | 0.0356       | 13         | 0        |
| 2       | 768     | −0.0175(16) | 0.9936(03)  | 0.0322         | 0.0372       | 9          | 1        |
| 3       | 815     | −0.0156(20) | 0.9965(03)  | 0.0315         | 0.0358       | 10         | 1        |
| 4       | 946     | −0.0266(16) | 0.9962(03)  | 0.0313         | 0.0352       | 9          | 0        |

Notes. # Stars is the final number of targets used in the analysis. $a$ and $b$ are coefficients of the fit $MS = a \times BS + b$, with 1σ errors shown for the last two significant digits. HWHM$_{Gauss}$ and HWHM$_{Lor}$ are the half-width at half-maxima of Gaussian and Lorentzian fits to the residuals as described in the text. $N_{supp}$ and $N_{aug}$ are the numbers of targets showing significant suppression or augmentation of their detection statistics, respectively, as defined in Sections 4.1 and 5.2.

(This table is available in its entirety in a machine-readable form in the online journal. A portion is shown here for guidance regarding its form and content.)

The impact of the pipeline processes is determined by comparing the measured S/N to the baseline S/N. In the comparison, we only consider the cadences that were modified by the injection of transits. We are also only considering targets for which the pipeline did not react strongly to the presence of many additional transits in the flux time series. This seems counterintuitive, since we are investigating how the pipeline impacts transits; however, in this study we are concerned with individual transit events, not the collective transit signal train. The motivation for the unphysical spacing of transits is to increase the number of statistical tests, not to affect the behavior of the pipeline. There are two ways the pipeline can react to the presence of many transits for which we filter here. The first is the SPSD detection algorithm and we therefore exclude targets where the SPSD detections differed between the two runs. The second effect is on the target variability measured by PDC, which treats targets differently based on this measurement. We exclude targets where the measured target variability changed by >0.2. This typically led to the exclusion of several hundred targets per channel (20%–30% of the total number of targets); the final numbers for each channel are given in Table 1. Our long-term goal is to examine recoverability of transit signal trains, as compared to single transits, at which point we will separate the transit events with realistic spacing and investigate the impact of these effects more thoroughly.

4.1. Individual Channel Results

Figure 2 shows the measured S/N plotted against the baseline S/N for the 1130 targets analyzed on channel 1. The left panel shows each cadence plotted as a separate point. All points for a given target have the same color (although the same color can be used for multiple targets). The width of the scatter around the 1:1 correlation in this visualization is misleading due to the number of cadences being plotted—the right panel shows the same data as a density plot, with a log scale in density, showing that the large majority of cadences are measured with a S/N extremely close to the baseline S/N. A robust linear fit of the form $MS = a \times BS + b$ was performed, where $MS$ is the measured S/N of the transit signal processed though the pipeline and BS is the baseline S/N, both in units of sigma. The data in the left panel yield coefficients of $a = 0.9970 ± 0.0003$ and $b = −0.0218 ± 0.0015$, indicating extremely high fidelity on average for the measured S/N compared to the baseline S/N. The fit is an iteratively re-weighted least-squares fit using a bi-square weighting function.

Characterizing the width and shape of the distribution of the residuals around this correlation allows us to determine the likelihood of recovering a measured S/N for a given baseline S/N. Figure 3 shows a histogram of the fractional change in the measured S/N ($(MS − BS)/BS$) from the baseline...
S/N for each cadence modified by transit injection. The core of the distribution (%(MS − BS)/BS < 0.06) is well fit by a Gaussian form (shown in red) with a half-width at half-maximum (HWHM) of 0.0315. However, there are significant wings to the distribution, so we also fit a Lorentzian form out to %(|MS − BS)/BS < 0.4 (shown in green) with a slightly wider HWHM of 0.0356. Therefore, on average, the measured S/N of the injected transit is generally recovered to within 3%–4%.

To examine the dependence of the distribution on the baseline S/N, Figure 4 shows the same histogram of scaled residuals as a function of the baseline S/N. The five panels show data from the highest baseline S/N values (15σ−20σ) to the lowest (−5σ−0σ), in bins of 5σ. We again show the Gaussian fit to the core of the distribution (%(|MS − BS)/BS < 0.06) and the Lorentzian fit to the core+wings (%(|MS − BS)/BS < 0.4). The higher baseline S/N bins are well fit by the Gaussian form, but the distribution widens toward lower baseline S/N values and we see that the aforementioned wings in Figure 3 are dominated by the contribution from lower S/N values. At high values of the baseline S/N (>10σ), the HWHM of the Gaussian and Lorentzian functions are similar (0.0281 and 0.0287, respectively, for BS > 15σ and 0.0276 and 0.0285, respectively, for 15σ > BS > 10σ). At low values, the HWHM of the Lorentzian function is slightly larger (0.0390 for the lowest baseline S/N bin, compared to 0.0333 for the Gaussian fit). While the distribution widens from ∼3% at higher values to ∼4% at lower values, we see no significant change in the center of the distribution from high to low baseline S/N.

Besides the tight 1:1 correlation, we note two other populations in Figure 2: those lying below this line, having the detection statistic of the processed signal suppressed when compared to the original signal, and those lying above, conversely having the detection statistic augmented. The latter population will be discussed further in Section 5.2.

There are 17 targets on this channel with significant suppression; “significant” suppression being defined as having more than one cadence with a baseline S/N of ≥10σ and a measured S/N of <5σ for convenience in identifying outlier targets. For each channel, the number of targets with significant suppression is listed in Table 1 as N_{supp}. Of the 17 suppressed targets on channel 1, 11 have at least one transit injected near a data gap caused by a spacecraft re-pointing. Figure 5 shows an example for target KID 8055837. There are four significant transit events in the flux time series shown in the upper panel; these are the injected transits. The second injected transit falls immediately before the first long data gap, which is due to the monthly data downlink from the spacecraft. As described in Tenenbaum et al. (2012), the data that fall immediately before and after a spacecraft re-pointing are fit and corrected with exponential terms to mitigate the impact of the changing thermal environment of the spacecraft. When a transit event occurs close to a re-pointing, the exponential fit will attempt to include the transit event in the correction, with the typical result that the transit signal strength is suppressed. On further examination of this event and those on other channels, we find empirically that this behavior occurs for transits that fall within 15 hr of the data gap caused by the spacecraft re-pointing. We also note that this suppression is almost exclusively found for transits with durations longer than ∼8 hr (i.e., transits with shorter durations do not affect enough of the cadences near the data gaps to influence the exponential fit). The suppression of transit signal strength near long data gaps is by far the dominant cause of suppression that we have noted and is discussed further in Section 5.1.

The transits in 5 of the 17 targets were suppressed by the harmonic fitter in TPS. Before the flux time series is whitened and searched for periodic signals, a sinusoidal harmonic filter is applied to remove periodic stellar activity, allowing the pipeline to search for variable stars for transit signals (Tenenbaum et al. 2012). In a small number of cases, the presence of transits affects the fitted harmonics, resulting in a different set of detection statistics. Figure 6 shows an example of the harmonic fitter fitting high-frequency components to the transits and attempting to filter them out, reducing their significance in the resulting flux time series; two of the five targets showed this behavior. We see this type of behavior largely for variable stars where the flux time series already contains high-frequency components with a characteristic period ∼2–4 times the transit duration. For the remaining three targets, the harmonic fitter either fits no harmonics or a significantly reduced set of harmonics in the presence of the injected transits compared to the “clean” light curves, which had the effect of increasing the measured CDPP and therefore decreasing the significance of the detection statistics (which are normalized by CDPP).
Figure 4. Distribution of scaled residuals \((M_S - B_S)/B_S\) from Figure 3 in bins of decreasing baseline \(S/N\) from the top panel \((B_S > 15\sigma)\) to the bottom panel \((B_S < 0\sigma)\) in bins of \(5\sigma\).

(A color version of this figure is available in the online journal.)

For the last of the 17 targets, the outlier detection algorithm in PDC identified a different set of cadences as being “outliers” in the presence of injected transits, including several within 0.5 days of one of the injected transits. These outliers ultimately resulted in a measured \(S/N\) for this target which was suppressed compared to the baseline \(S/N\). This was the only instance of the outlier detection algorithm resulting in a significantly suppressed signal and therefore we do not believe this behavior to be endemic in the pipeline.

We examined three additional channels (channels 14, 37, and 64) in detail and they were found to behave very similarly to channel 1 in terms of numbers of suppressed targets and the dominant causes for the suppression (transits injected near data gaps and harmonic removal). We identified two additional causes for suppression that affected single targets. One target on channel 37 was affected by the cosmic-ray detection algorithm in the pipeline as described in Section 2.1, with two transits being “corrected” and therefore having their depth reduced. The second target was a variable star showing spot modulation. The detection statistics for this target were ultimately impacted by the PDC correction as described in Section 5.2. Briefly, in this case, the weight given to the priors in PDC for the correction to the flux time series with transits injected was much lower than the weight given to the priors for the “clean” flux time series. This leads to a larger amount of high-frequency noise being introduced into the flux time series with transits, increasing the measured CDPP and lowering the significance of the transit detection statistics. This was the only target for which this behavior was observed; Section 5.2 describes the handful of cases discovered showing the opposite behavior.

4.2. Overall Signal Recovery Rate

Figure 7 shows the density plots of the measured \(S/N\) compared to the baseline \(S/N\) for the first six channels, illustrating
that the overall response of each channel is largely the same. We examined the density plots for the remainder of the channels and see no departure from this pattern. Table 1 summarizes the fits for the 80 channels, showing the coefficients of the robust linear fit and the numbers of targets analyzed in each case. If the signal were being perfectly preserved by the pipeline, then we would expect a 1:1 correlation between the measured and baseline S/N; in fact, we are extremely close to this for every channel. The median recovery rate across all channels is

\[ MS = 0.9973(\pm 0.0011) \times BS - 0.0151(\pm 0.0049). \] (1)

This corresponds to an average measured S/N of 9.96σ for a 10σ signal, a ∼0.4% reduction. Across the 80 channels, the linear term ranges from 0.9936 to 1.0004 and the constant term from −0.0266 to 0.0012. The width of this distribution is also shown for each channel in Table 1 for the Gaussian and Lorentzian fits to the core and core+wings, respectively. The median Gaussian HWHM across all channels is 0.0264 with individual channels ranging from 0.0246 to 0.0360. The median Lorentzian HWHM across all channels is 0.0256 with individual channels ranging from 0.0220 to 0.0461. Therefore, across the full focal plane simulated transit signals are typically recovered to within 2%–4% of their expected S/N.

5. DISCUSSION

5.1. Signal Suppression

We examined a set of targets where the measured signal strength was suppressed compared to the expected value. Any systematic reduction in signal strength by the processing pipeline would have significant consequences when determining the population of planets detectable in the Kepler data. We find that ∼0.9% of targets experience significant suppression of at least one transit with more than one modified cadence having a measured S/N of <5σ with an expected baseline S/N of >10σ.

In the large majority of cases (∼80%), the cause of the suppression was that the target had a long-duration (>8 hr) transit injected near (within 15 hr) a long data gap. For these targets, the remaining injected transits were recovered as normal (i.e., even for the targets showing significant suppression on small number of transits, the majority of the transits are not suppressed). The current algorithm in TPS that corrects for thermal systematics after spacecraft re-pointing is impacted by the presence of transits in the flux time series window used in the correction. In Q3, there were six “edges” to long data gaps, at the start and end of each month. It is
Figure 6. Upper panel: a small section of the Q3 flux time series for target KID 7915515 on channel 1. The blue points are the PDC-corrected flux time series without the injected transits; the red points are the PDC-corrected flux time series with injected transits. The target exhibits high-frequency variability with an amplitude comparable to the depth of the injected transit, which is centered at MJD−55000 = 101.65. Middle panel: the blue points show the difference in the harmonic fits to the flux time series in the upper panel. The points marked with red circles show the cadences modified by transit injection; these are the cadences with the largest differences in the harmonic fits. Lower panel: the resulting detrended flux time series. The blue points show the baseline signal and the red points show the processed signal. The processed transit depth is significantly shallower than the baseline depth.

(A color version of this figure is available in the online journal.)

Figure 7. Density plot of the measured and baseline S/N for each of the first six channels. Note the log scale in density.

(A color version of this figure is available in the online journal.)
straightforward to re-calculate the correlations described in Section 4 after removing cadences that fall within 15 hr of these edges. Overall, we find an improved recovery rate of $MS = 0.9972(\pm0.0012) \times BS - 0.0036(\pm0.0011)$; the slope is essentially identical and the offset is significantly reduced, indicating that a large fraction of the original offset is due to these suppressed transits. On average, the width of the distribution of the scaled residuals did not change when the cadences near the gaps were removed. The median Gaussian and Lorentzian HWHM across all channels changed from 0.0264 and 0.0256, respectively, to 0.0263 and 0.0254, respectively.

In quarters with additional spacecraft re-pointings, due to safe modes for instance, there may be additional “edges.” We encourage readers to account for these windows of reduced signal recovery when determining the detectability of a given planet candidate by using a reduced observation baseline when calculating phase coverage. Transits of Earth-size planets around Sun-like stars average $\sim 10$ hr and thus could experience this effect if they fall close to a long data gap. The *Kepler* pipeline applies a set of de-emphasis weights to points falling within two days of a data gap when calculating the significance of signals in TPS. Most spacecraft events that do not involve re-pointing typically produce short gaps (such as losses of fine point, reaction wheel momentum desaturations, etc.) that do not impact signal strength recovery.

In a further $\sim 10\%$ of cases, injected transits had their recoverability impacted by the harmonic fitter in TPS, either by being directly fit and removed by the fitter or by the change in fitted harmonics increasing the measured CDPP relative to the “clean” light curves. This effect was largely seen for stars that already exhibited high-frequency variability (with periods of hours to tens of hours). An important note here is that the injected transit epochs have an artificial separation of 50 times the transit duration representing a “duty cycle,” or fraction of the phased light curve occupied by the transit, of 0.02. We expect and have observed the behavior of the harmonic fitter to be highly dependent on the duty cycle of the transit signal; the higher the duty cycle, the more likely the harmonic fitter will fit the transit with a Fourier series. Although a duty cycle of 0.02 does represent real physical parameter space for planets in short-period orbits, at longer and more interesting periods the duty cycle falls by an order of magnitude and we expect the harmonic fitter to have a significantly smaller impact. We will examine this behavior in a subsequent analysis where we inject transits with realistic separations into longer observation baselines.

### 5.2. Signal Augmentation

A small number of targets experienced systematic augmentation of their detection statistics during processing; only 13 targets across the full 80 channels had more than one modified cadence with baseline $S/N$ of $\leq 5\sigma$ and a measured $S/N$ of $>10\sigma$ ($N_{\text{aug}}$ in Table 1). We find that these are almost all variable stars showing evidence of spot modulation. In these cases, PDC incorrectly gave the priors a very low weight when processing the clean light curves, introducing high-frequency noise in the course of reducing the overall root-mean-square deviation, as described by Stumpe et al. (2012). This leads to an increase in the measured CDPP for the original flux time series and a correspondingly lower detection statistic. This behavior was corrected in the final version of SOC 8.3 and we do not expect augmentation of signals in future pipeline processing.

### 5.3. Correlations of Signal Recovery with Injected Transit Properties

As noted in the previous discussion, there are some regions of parameter space where pipeline processing can modify transits, such as when long-duration transits falling near long data gaps or when targets have high-frequency variability. Here we examine the overall dependence of the signal recoverability on the injected transit attributes, in particular transit duration and transit $S/N$.

For a given channel, we find the robust linear correlation between the measured detection statistics and the baseline detection statistics for each target individually, again of the form $MS = a \times BS + b$. We then measure the robust linear correlation between the coefficients of those fits, $a$ and $b$, and the injected transit duration and injected transit $S/N$ of the target. We measure the correlation of $b$ with duration for each channel and find the correlations range from $-0.0057$ to $0.0003$, with a median correlation of $-0.0028$, a standard deviation of 0.0011 across all channels, and an average error per channel of 0.0014. For the linear coefficient, $a$, we find even less significance in the correlation, with correlations across the channels ranging from $-0.00058$ to 0.00067, a median correlation of 0.00010, a standard deviation of 0.00020 across all channels, and an average error per channel of 0.00080. Therefore, we conclude that there is no significant correlation between the fidelity of a processed signal and its duration. The long-duration transits injected near long data gaps are clearly rare enough as to not influence the average detection statistics, which is useful to note for planet completeness requirements.

We perform the same analysis for the injected transit $S/N$ and again find no significant (>3$\sigma$) correlation between the fidelity of the measured $S/N$ and the baseline $S/N$ (i.e., low-significance transits are no more likely to be impacted by pipeline processing than high-significance transits, between $2\sigma$ and $20\sigma$). Formally, the median correlation of $b$ across all channels is 0.00230, with a standard deviation of 0.00089 and an average error per channel of 0.00087. The median correlation of $a$ across all channels is $-0.00028$, with a standard deviation of 0.00016 and an average error per channel of 0.00060. This is also qualitatively evident from Figure 4, where the scaled residuals are shown for different bins of baseline $S/N$—there is no shift in the center of the distribution with decreasing baseline $S/N$. There is, however, a widening of the distribution with lower baseline $S/N$, indicating that although there is no systematic change in the average measured $S/N$, the probability of measuring a $S/N$ significantly different from the baseline $S/N$ increases at lower baseline $S/N$.

### 6. CONCLUSIONS AND FURTHER WORK

For a given transit signal, the *Kepler* pipeline has extremely high fidelity in reproducing the expected single event detection statistics. The average recovery of an individual transit event is $MS = 0.9973(\pm0.0012) \times BS - 0.0151(\pm0.0049)$, where $MS$ is the measured $S/N$ and $BS$ is the baseline, or expected, $S/N$. We have characterized the pipeline for single transit events from the summation of the pixels into a flux time series, through the systematic error correction, to the final detrending and whitening before the periodic signal search.

This initial analysis was limited in scope to a single quarter of observation. This necessitated injecting the simulated transit signals with smaller separations between transits than would be physically observed in order to maximize the number of statistical tests that could be performed per target. Although
we attempted to mitigate the influence of this artificially close separation on the results, several effects that we have discussed (i.e., the SPSD detector, the behavior of the harmonic fitter, and the influence of transits on the weight assigned to the priors in PDC) may not be reproducing the real performance of the pipeline. In the latter two cases, we expect the incidence rate of transit suppression to be lower with wider separations, in which case what we report here should be considered the worst-case scenario.

We therefore conclude that the pipeline does not systematically perturb the signal strength of individual transit events. This has two positive implications: (1) when calculating the uncertainty on a derived planet size, the contribution of the uncertainty on the transit depth due to pipeline perturbation is very small compared to the typical contribution from uncertainty in stellar parameters and (2) for teams performing independent searches for periodic transit signals in the flux time series produced by the pipeline, this initial work provides sufficient information to move forward with determining the completeness of their own generated planet candidate set. For analysis of the Kepler planet candidate set, this work provides reassurance that no systematic reduction in transit signal strength is introduced by the pipeline and that any significant incompleteness produced by the pipeline would therefore be localized to the periodic transit search.

In order to confirm our expectations, and more importantly to examine the performance of the pipeline through the entire analysis including the periodic transit search and the validation of the signal detection in DV, we need to perform an analysis similar to the one described in this work, but with realistic, physical separations of the transits in time. To accommodate this wider separation in time, our next analysis will be performed on a longer observation baseline. We also plan to examine the performance of DV in recognizing false positive signals. The ability to inject the simulated transit events at the pixel level allows us to introduce spatial offsets in the source of the transit signal. In this fashion, we can mimic background eclipsing binary stars, or small biases in the measurements of the position of the targets.

Funding for the Kepler Discovery Mission is provided by NASA’s Science Mission Directorate.

APPENDIX

TRANSIT MODEL GENERATION

From a few starting assumptions and knowledge of the target star, we can use the observable parameters selected in Section 3.1 to determine the planet parameters required to generate the model. For this test, we assumed circular orbits (eccentricity of zero) and central-crossing transits (impact parameter of zero). Eccentricity has only a very slight impact on the transit shape and therefore recoverability, and assuming circular orbits allows us to easily calculate the orbital period of the injected planet from the selected signal duration. On the other hand, the impact parameter has a significant effect on the shape and depth of the transit. However, we are primarily concerning ourselves with the question of distortion in the measured detection statistics: for an initial signal with 2σ–20σ significance, what is the typical final detection statistic measured by the pipeline? Since the transit depth is a function of both the injected planet radius and the impact parameter, we need only allow one of these to be a free parameter to test the correlation. In addition, the processes described in Section 2.1 should not affect differently shaped transits in systematically different ways. The possible exception is the potentially higher rate of cosmic-ray misidentification in more v-shaped transits, although as shown in Figure 1 this occurs in flat-bottomed transits as well. A recent update to the cosmic-ray identification algorithm is expected to improve this behavior when the data are re-processed.

To generate the model, we use the stellar parameters (surface gravity, effective temperature, stellar radius, and metallicity) from the Kepler Input Catalog (KIC; Brown et al. 2011) and default to solar values for unclassified stars. The pipeline generates a noise estimate for use in measuring the S/N of a putative transit signal called the CDPP (Christiansen et al. 2012). For a given cadence and box-pulse duration, it is the depth of the box at that cadence that would produce a S/N of 1. We use the rms CDPP previously calculated by the pipeline for the target flux time series prior to the injection of simulated planets for the duration that is closest to our model signal duration. We calculate the model transit depth, δ, in σ from the product of the rms CDPP and the model signal strength. From the transit depth, we can calculate the planetary radius, \( R_p \), from \( \delta = \left( \frac{R_p}{R_\star} \right)^2 \), where \( R_\star \) is the stellar radius.

We estimate the orbital period from Equation (1) of Gilliland et al. (2000). We then calculate the semimajor axis, \( a \), from Newton’s modification of Kepler’s third law and the geometric ratio \( a/R_\star \). We can then use the Mandel & Agol (2002) analytic transit model formalism to generate a physical transit model centered at the calculated starting epoch and repeated along the light curve at our pre-defined transit separation, which is 50 times the model transit duration, as described earlier. The model is generated at a sampling rate 30 times higher than the flux time series and then re-sampled onto the final time stamps. We use the KIC magnitude for each target star to convert relative depth, as calculated by the model, into the total number of photopixels that need to be subtracted from the light curve. We subtract absolute numbers of photopixels instead of relative numbers to account for the fact that the pixels comprising a given light curve may contain flux contributions from multiple targets. The simulated transit model is then used twice: first, it is injected into the calibrated pixels, which are then processed through the pipeline; and second, it is injected into the original “clean” flux time series at the end of the pipeline to generate a set of baseline or expected detection statistics for a signal unmodified by the pipeline processing.

REFERENCES

Ballard, S., Christiansen, J. L., Charbonneau, D., et al. 2010, ApJ, 716, 1047
Ballard, S., Christiansen, J. L., Charbonneau, D., et al. 2011, ApJ, 732, 41
Batalha, N. M., Rowe, J. F., Bryson, S. T., et al. 2013, ApJS, 204, 24
Borucki, W. J., Koch, D., Basri, G., et al. 2010, Sci, 327, 977
Borucki, W. J., Koch, D. G., Basri, G., et al. 2011a, ApJ, 732, 117
Borucki, W. J., Koch, D. G., Basri, G., et al. 2011b, ApJ, 736, 19
Brown, T. M., Latham, D. W., Everett, M. E., & Esquerdo, G. A. 2011, AJ, 142, 112
Bryson, S. T., Tenenbaum, P., Jenkins, J. M., et al. 2010, ApJL, 713, L97
Burke, C. J., Gaudi, B. S., DePoy, D. L., & Pogge, R. W. 2006, AJ, 132, 210
Christiansen, J. L., Jenkins, J. M., Caldwell, D. A., et al. 2012, PASP, 124, 1279
Croll, B., Matthews, J. M., Rowe, J. F., et al. 2007, ApJ, 671, 2129
Dong, S., & Zhu, Z. 2012, arXiv:1212.4853
Fischer, D. A., Schwamb, M. E., Schwinski, K., et al. 2012, MNRAS, 419, 2900
Fressin, F., Torres, G., Charbonneau, D., et al. 2013, ApJ, 766, 81
Gilliland, R. L., Brown, T. M., Guhathakurta, P., et al. 2000, ApJL, 545, L47
Howard, A. W., Marcy, G. W., Bryson, S. T., et al. 2012, ApJS, 201, 15
Huang, X., Bakos, G. Á., & Hartman, J. D. 2013, MNRAS, 429, 2001
Jenkins, J. M. 2002, ApJ, 575, 493
Jenkins, J. M., Caldwell, D. A., Chandrasekaran, H., et al. 2010a, ApJL, 713, L87
Jenkins, J. M., Chandrasekaran, H., McCauliff, S. D., et al. 2010b, Proc. SPIE, 7740, 10
Koch, D. G., Borucki, W. J., Basri, G., et al. 2010, ApJL, 713, L79
Mandel, K., & Agol, E. 2002, ApJ, 580, 171
Ofir, A., & Dreizler, S. 2013, A&A, 555, 580
Quintana, E. V., Jenkins, J. M., Clarke, B. D., et al. 2010, Proc. SPIE, 7740, 64
Smith, J. C., Stumpe, M. C., Van Cleve, J. E., et al. 2012, PASP, 124, 1000
Stumpe, M. C., Smith, J. C., Van Cleve, J. E., et al. 2012, PASP, 124, 985
Titenbaum, P., Christiansen, J. L., Jenkins, J. M., et al. 2012, ApJS, 199, 24
Titenbaum, P., Jenkins, J. M., Seader, S., et al. 2013, ApJS, 206, 5
Twicken, J. D., Chandrasekaran, H., Jenkins, J. M., et al. 2010, Proc. SPIE, 7740, 62
Twicken, J. D., Clarke, B. D., Bryson, S. T., et al. 2010, Proc. SPIE, 7740, 69
Weldrake, D. T. F., Sackett, P. D., & Bridges, T. J. 2008, ApJ, 674, 1117
Wu, H., Twicken, J. D., Tenenbaum, P., et al. 2010, Proc. SPIE, 7740, 42
Youdin, A. N. 2011, ApJ, 742, 38