Planet Detection Metrics:

Per-Target Flux-Level Transit Injection Tests of TPS for Data Release 25

KSCI-19109-002
Christopher J. Burke & Joseph Catanzarite
June 13, 2017

NASA Ames Research Center
Moffett Field, CA 94035

Prepared by: ______________________________ Date 6/13/17
Christopher J. Burke, Science Office

Prepared by: ______________________________ Date 6/13/17
Joseph Catanzarite, Science Operations Center

Approved by: Michael R. Haas Date 6/13/17
Michael R. Haas, Science Office Director

Approved by: ______________________________ Date 6/13/17
Natalie M. Batalha, Project Scientist

Document Control

Ownership
This document is part of the Kepler Project Documentation that is controlled by the Kepler Project Office, NASA/Ames Research Center, Moffett Field, California.

Control Level
This document will be controlled under KPO @ Ames Configuration Management system. Changes to this document shall be controlled.

Physical Location
The physical location of this document will be in the KPO @ Ames Data Center.

Distribution Requests
To be placed on the distribution list for additional revisions of this document, please address your request to the Kepler Science Office:

Michael R. Haas
Kepler Science Office Director
MS 244-30
NASA Ames Research Center
Moffett Field, CA 94035-1000
Michael.R.Haas@nasa.gov

The correct citation for this document is: Burke, C. J. & Catanzarite, J. 2017, *Planet Detection Metrics: Per-Target Flux-Level Transit Injection Tests of TPS for Data Release 25*, KSCI-19109-002

DOCUMENT CHANGE LOG

DATE	RELEASE	PAGES	CHANGES/NOTES
May 24, 2017	001	all	Original release
June 13, 2017	002	20	Corrected number of targets in KSOC-5006 run

Contents

1 Introduction

Quantifying the ability of a transiting planet survey to recover transit signals has commonly been accomplished through Monte-Carlo injection of transit signals into the observed data and subsequent running of the signal search algorithm (Gilliland et al., 2000; Weldrake et al., 2005; Burke et al., 2006). In order to characterize the performance of the *Kepler* pipeline (Twicken et al., 2016; Jenkins et al., 2017) on a sample of over 200,000 stars, two complementary injection and recovery tests are utilized:

1. Injection of a single transit signal per target into the image or pixel-level data, hereafter referred to as pixel-level transit injection (PLTI), with subsequent processing through the Photometric Analysis (PA), Presearch Data Conditioning (PDC), Transiting Planet Search (TPS), and Data Validation (DV) modules of the *Kepler* pipeline. The PLTI quantification of the *Kepler* pipeline's completeness has been described previously by Christiansen et al. (2015, 2016); the completeness of the final SOC 9.3 *Kepler* pipeline acting on the Data Release 25 (DR25) light curves is described by Christiansen (2017).

2. Injection of multiple transit signals per target into the normalized flux time series data with a subsequent transit search using a stream-lined version of the Transiting Planet Search (TPS) module. This test, hereafter referred to as flux-level transit injection (FLTI), is the subject of this document. By running a heavily modified version of TPS, FLTI is able to perform many injections on selected targets and determine in some detail which injected signals are recoverable. Significant numerical efficiency gains are enabled by precomputing the data conditioning steps at the onset of TPS and limiting the search parameter space (i.e., orbital period, transit duration, and ephemeris zero-point) to a small region around each injected transit signal.

The PLTI test has the advantage that it follows transit signals through all processing steps of the *Kepler* pipeline, and the recovered signals can be further classified as planet candidates or false positives in the exact same manner as detections from the nominal (i.e., "observed") pipeline run (Twicken et al., 2016, Thompson et al., in preparation). To date, the PLTI test has been the standard means of measuring pipeline completeness averaged over large samples of targets (Christiansen et al., 2015, 2016; Christiansen, 2017). However, since the PLTI test uses only one injection per target, it does not elucidate individual-target variations in pipeline completeness due to differences in stellar properties or astrophysical variability. Thus, we developed the FLTI test to provide a numerically efficient way to fully map individual targets and explore the performance of the pipeline in greater detail. The FLTI tests thereby allow a thorough validation of the pipeline completeness models (such as window function (Burke & Catanzarite, 2017a), detection efficiency (Burke & Catanzarite, 2017b), etc.) across the spectrum of *Kepler* targets (i.e., various astrophysical phenomena and differences in instrumental noise). Tests during development of the FLTI capability revealed that there are significant target-to-target variations in the detection efficiency.

Overall, both the PLTI and FLTI tests are instrumental in supporting planet occurrence rate science using the catalog of planet candidates (Thompson et al., in preparation) identified with the *Kepler* pipeline (Twicken et al., 2016). The *Kepler* pipeline completeness model for previous data releases is described in Burke et al. (2015) and updated for DR25 in Burke & Catanzarite (2017a,b). In principle, it is possible to perform $\sim 10^4$ FLTI trials for every *Kepler* target and fully map the detailed performance of TPS, thereby forgoing the need for a pipeline completeness model. However, even with the enormous efficiency gains achieved by FLTI, it is currently too computationally expensive to perform this brute force approach for all *Kepler* targets. Thus, the approach we have adopted is to perform a set of 'deep' FLTI runs with $\sim 10^5$ injections for a representative set of ~ 100 targets supplemented by 'shallow' FLTI runs with $\sim 10^3$ injections on a much larger set of $\sim 10^4$ targets. The complementary nature of the 'deep' and 'shallow' runs has enabled the development and validation of the pipeline completeness model described in Burke & Catanzarite (2017b).

The remainder of this document describes the FLTI techniques in Section 2, defines the output file format in Section 3, and details the various test runs in Section 5. Section 4 describes sample python code for calculating pipeline completeness using the FLTI results and provides some worked examples.

2 Flux-Level Transit Injection (FLTI) Description

This section describes the implementation details for FLTI. The first step is to take the standard input files (quarter-by-quarter Presearch Data Conditioning (PDC) time series; Stumpe et al., 2012; Smith et al., 2012; Jenkins et al., 2017) used for running TPS (Seader et al., 2015; Twicken et al., 2016) and reformat them for use by FLTI. Recall that TPS performs additional preprocessing on the PDC light curves before searching for transit signals by identifying and removing significant harmonic signals in the Fourier domain, normalizing the light curves between quarters, and filling gaps where data are missing with noise having properties similar to those observed in adjacent data. These 'Light Curve Preprocessing' steps are described in Chapter 9 of Jenkins et al. (2017) and illustrated in the block diagram of TPS in their Figure 9.2. Since these steps are time consuming, they are done once by calling TPS and saving the resulting "cleaned" and normalized flux time series for use in all subsequent FLTI searches.

In addition, when a transit signal is identified by TPS, the iterative multi-planet search algorithm in the *Kepler* pipeline identifies the cadences contributing to the transit signal and deweights the in-transit cadences for subsequent search iterations. Thus, for the targets with transit signals identified during the *Kepler* pipeline search, the FLTI input time series has been "cleaned" of identified transit signals. Consequently, by reusing the preprocessed time series from TPS, FLTI does not need to repeat these numerically costly steps during its search for each injected signal. The down-side is that the FLTI tests do not quantify the impact of these conditioning steps. Fortunately, these steps were investigated as part of the PLTI investigation (Christiansen et al., 2013, 2015). The inputs to FLTI are further augmented to specify a region of parameter space (i.e., orbital period, $P_{\rm orb}$, and planet radius, $R_{\rm p}$) for signal injection that is tailored to the stellar properties of each specific target under study.

In other words, FLTI uses a modified version of TPS that skips the light curve preprocessing steps as discussed in the previous paragraphs and begins by injecting a limb-darkened transit model (LDTM) for each trial (see Section 3.2.1 for the LDTM parameters and their definitions). Next, three separate transit searches are performed as follows:

1. Localized 2x2 Grid Search – Normally, TPS searches a fine grid over the full parameter space of transit duration, $P_{\rm orb}$, and ephemeris zero-point that is appropriate for the subject light curve (Note: for the Q1-Q17 DR25 search, the light curves typically contain the full four years of data obtained by the *Kepler* mission). In order to achieve high efficiency, the FLTI tests limit the search space to a 2x2 grid of transit duration and orbital period that brackets the injected signal. Section 3.2.2 provides definitions for the local grid search output.

2. Supplemental Targeted Search – In order to understand the outcome of each transit search in detail, a supplemental search over ephemeris zero-point is performed at the transit duration and orbital period of the injected signal. Output of this supplemental targeted search is described in Section 3.2.3.

3. Direct Evaluation – If the supplemental targeted search does not result in a detection, then several of the detection metrics are evaluated directly at the ephemeris zero-point, orbital period, and transit duration of the injected signal. If the direct evaluation is performed, its output overwrites the supplemental targeted search results as described in Section 3.2.3.

Once these searches are complete, the FLTI outputs are organized into a separate FITS file for each target.

3 FLTI Output Files

The FLTI results are stored in the FITS file format (Pence et al., 2010) and organized into run identification folders (see Section 5) with a separate FITS file for each target. The file names have the format: kplr<*identifier*>_<*release*>_<*ksoc*>_flti.fits, where <*identifier*> is the nine-digit *Kepler* identification number left-padded with zeroes, <*release*> is dr25 corresponding to DR25, and <*ksoc*> is the four-digit integer run identifier. For example, the FITS file featured in this document has the file name: kplr007702838_dr25_5008_flti.fits, indicating that it contains FLTI results for *Kepler* target 7702838 from the KSOC-5008 run. Each run is typically tailored to test a particular aspect of TPS performance. Some runs involve a few targets with $\sim 10^5$ injections per target and others involve many targets with $\sim 10^3$ injections per target (see Section 5).

3.1 Primary Header Contents

The key words in the primary FITS file header provide information about the individual target and light curve noise properties relevant to the FLTI test. They are:

1. KEPLERID: Kepler identification number.

2. RADIUS: Stellar radius assumed for the limb-darkened transit model (LDTM) in units of Solar radius.

3. TEFF: Stellar effective temperature assumed for the LDTM in units of Kelvin.

4. LOGG: Log base 10 of the stellar surface gravity assumed for the LDTM in cgs units.

5. FEH: Log base 10 of the stellar metallicity relative to solar abundance. This metallicity is used to generate the limb darkening parameters for the LDTM.

6. LIMBD1 - LIMBD4: Four parameter nonlinear limb-darkening-law coefficients 1 through 4 assumed for the LDTM. Coefficients are from Castelli & Kurucz (2004) or Claret & Bloemen (2011), and the LDTM is from Mandel & Agol (2002).

7. RCDP01P5 - RCDP15P0: RMS CDPP noise measurement for the flux time series in units of ppm. There are fourteen header values that correspond to the noise values for the fourteen transit durations searched for planet signals in TPS (i.e., 1.5, 2.0, 2.5, 3.0, 3.5, 4.5, 5.0, 6.0, 7.5, 9.0, 10.5, 12.0, 12.5, 15.0 hr). The last four characters of the variable names represent the transit durations with the obvious substitution of a lower-case p for the decimal point.

8. RPMIN - Minimum planet radius injected in units of $R_{\oplus}$ for the current FLTI test.

9. RPMAX - Maximum planet radius injected in units of $R_{\oplus}$ for the current FLTI test.

10. PERMIN - Minimum orbital period injected in units of days for the current FLTI test.

11. PERMAX - Maximum orbital period injected in units of days for the current FLTI test.

3.2 Data Table Contents

The FITS data table reports the input transit signal parameters as well as diagnostic output from TPS relevant to signal detection. The table has NAXIS2 (data table keyword) rows, where a row corresponds to a transit signal injection and recovery trial. The data columns are described in the following four subsections.

3.2.1 Limb-Darkened Transit Model (LDTM) Input Parameters

The columns that define the injected transit signal LDTM and significance of the signal are:

1. Period: Orbital period of the injected transit signal ephemeris in days.

2. Epoch: Mid-transit time of the injected transit signal ephemeris in the *Kepler* Barycentric Julian Date system (BJD-2454833).

3. t_depth: Depth of the LDTM at the transit mid-point in units of ppm.

4. t_dur: Transit duration of the LDTM from the first to fourth contact points in hr.

5. b: Impact parameter of the LDTM relative to the stellar host radius.

6. Rp: Planet radius of the LDTM in $R_\oplus$.

7. a: Semi-major axis of the orbit for the injected transit signal in AU.

8. exp_mes: Expected multiple-event statistic (MES) calculated for the ideal match case in TPS. To first order, the MES is proportional to, and has the same dependence as, the basic equation for transit signal-to-noise ratio (SNR) in the limit of stationary noise if the transit signal is an identical match to the search template (Jenkins, 2002; Christiansen et al., 2012). However, in practice, MES differs from the simple SNR ideal case, and the exp_mes value provided is calculated by TPS using the signal template employed during the signal search and assuming the known ephemeris and duration of the injected signal.

9. exp_mesA: Expected MES as calculated by TPS including known effects that lead to a degradation of the transit signal such as local flux time series perturbations and finite coarseness of the search grid parameterized in terms of $P_{\rm orb}$, transit duration, and ephemeris zero-point.

3.2.2 TPS Output from Localized 2x2 Grid Search

The columns populated by the FLTI version of TPS as a result of the localized 2x2 grid search described in Section 2 are:

1. Recovered: A flag value of 1 (0) indicates that a TCE was (was not) created. In other words, a flag value of 1 indicates that a signal passing the MES = 7.1 threshold, three transit and three-transit-weight-check requirements, and the veto thresholds was identified during the search. A flag value of 0 indicates that no signal was identified that simultaneously passed all these criteria.

2. r_mes: Returned multiple-event statistic.

3. r_period: Returned orbital period in days.

4. r_epoch: Returned mid-transit time of the signal ephemeris in the *Kepler* Barycentric Julian Date system (BJD-2454833).

5. r_dur: Returned duration of the signal in hrs. The duration corresponds to one of the fourteen quantized durations (i.e., 1.5 ... 15.0 hrs) employed by the search algorithm.

6. r_ntran: Returned number of transit events after accounting for losses due to missing data.

7. r_threeTransitFail: A flag value of 1 (0) indicates that the final signal fails (passes) the three-transit weight check. If r_ntran $\neq 3$, then r_threeTransitFail has a flag value of 0 and is not evaluated. This implies that $0 <$ r_ntran < 2 situations will have r_threeTransitFail $== 0$, even though the signal fails the r_ntran ≥ 3 requirement for detection.

8. r_depth: Returned average depth of signal over the returned duration, r_dur, in ppm as computed during the robust statistic veto evaluation (see section 9.4.4.1 of Jenkins et al., 2017).

9. r_robstat: Returned value of the robust statistic veto (see section 9.4.4.1 of Jenkins et al., 2017).

10. r_chiVeto: Returned value of the Chi-squared statistic veto (see section 9.4.4.2 of Jenkins et al., 2017).

11. r_gofVeto: Returned value of the Goodness-of-fit statistic veto (see section 9.4.4.3 of Jenkins et al., 2017).

It is important to note that these variables take on different values, depending on the results of the transit search. If an injected signal is identified and passes all TPS criteria for detection (i.e., Recovered == 1), then the reported values correspond to that signal. On the other hand, if the injected signal is not detected (i.e., Recovered == 0), then the reported values correspond to the strongest signal (ranked by MES) that TPS identified during its search. This means that if no veto passing signal is detected (Recovered == 0), then the variables do not report on the status of the injected ephemeris and should not be used to infer pipeline detection efficiency. Nevertheless, these values are recorded for use in validating TPS or understanding its behavior in a particular situation. We also note that even when a veto-passing detection (Recovered == 1) is reported, it does not necessarily correspond to the injected ephemeris. However, such situations are rare, given the localized nature of the 2x2 grid search, so it is generally safe to ignore such cases when analyzing FLTI data.

3.2.3 Output for Supplemental Targeted Search or Direct Evaluation

Like the reported values in the last subsection, the contents of some columns populated by the second and third transit searches (see Section 2) have different meanings, depending on circumstance. Here, the FLTI version of TPS uses the exact period and transit duration of the injected planet signal to perform a supplemental targeted search. Since the period and duration are fixed, this search only varies the orbital ephemeris zero point. If this supplemental targeted search does not result in a MES threshold and veto passing detection (i.e., Recovered_AtPerAndDur == 0), then the direct evaluation step is performed. In this case, TPS is forced to compute the MES and veto values using the injected period, duration, and ephemeris zero-point. Note that the values for ralt_mes, ralt_ntran, ralt_ThreeTransitFail, ralt_depth, and ralt_robstat from this direct evaluation overwrite the values produced by the supplemental targeted search. In other words, the values for these five parameters come from the supplemental targeted search when the parameter Recovered_AtPerAndDur == 1 and from direct evaluation when Recovered_AtPerAndDur == 0.

1. Recovered_AtPerAndDur: A flag value of 1 indicates that a signal passing all thresholds and vetoes was identified during the supplemental targeted search over ephemeris zero-point. A flag value of 0 indicates that no signal was identified passing all criteria.

2. Recovered_Alt: A flag value of 1 indicates that a signal passed the three transit test, three-transit weight check, and the robust statistic veto during direct evaluation. Due to a bug in the direct evaluation only, the MES threshold, chi-squared veto, and goodness-of-fit veto were not evaluated properly, and the flag value does not accurately reflect these three thresholds. A flag value of 0 indicates that no signal was identified passing the three transit test, three-transit weight check, and robust statistic veto.

3. ralt_mes: MES for the final signal. When direct evaluation is performed, this value is replaced by an estimate of the MES determined by taking the square root of the individual transit SNRs added in quadrature.

4. ralt_ntran: Number of transits for the final signal that are not lost due to missing data.

5. ralt_ThreeTransitFail: A flag value of 1 (0) indicates that the final signal fails (passes) the three-transit weight check.

6. ralt_depth: Average depth of final signal over r_dur in ppm.

7. ralt_robstat: Value of the robust statistic veto for the final signal.

3.2.4 Auxiliary Output

The variables defined in this subsection quantify the extent to which the injected transit signals overlap with deweighted or missing data cadences (see section 9.3.4 of Jenkins et al., 2017). In particular, the TPS algorithm has an iterative search process (i.e., 'remove feature') that identifies events in the flux time series that are deemed not transit-like in nature and deweights them. Thus, an increase in the number of deweighted in-transit cadences indicates that the injected transit signal was misidentified as a non-transit-like outlier and deweighted before the final search was conducted. In its final run configuration (i.e., SOC 9.3), TPS permitted two such events to be deweighted in a given light curve. The following parameters, which measure this 'remove feature' effect, are summed over all transit events:

1. nDeWeightedCadInit: The initial number of in-transit cadences that are deweighted by a factor of 0.5 or more.

2. fracDeWeightedCadInit: The initial fraction of in-transit cadences that are deweighted by a factor of 0.5 or more.

3. nDeWeightedCadFinal: The final number of in-transit cadences that are deweighted by a factor of 0.5 or more (i.e., after TPS is allowed to remove up to two events during its iterative search process).

4. fracDeWeightedCadFinal: The final fraction of in-transit cadences that are deweighted by a factor of 0.5 or more (i.e., after TPS is allowed to remove up to two events during its iterative search process).

4 Worked Examples

In this section we provide pseudo-code for the most common use cases for the FLTI output relevant to understanding pipeline completeness and calculating planet occurrence rates. The following worked examples use the output for the single target KIC 7702838. For this target we injected 603,360 realistic limb-darkened transit signals. In order to further illustrate the steps outlined in the following subsections, the python code (Kepler-FLTI.py) and data table (kplr007702838_dr25_5008_flti.fits) are provided so that users can reproduce these figures and examples (see https://github.com/nasa/Kepler-FLTI).

4.1 Calculating an Empirical Detection Contour

The FLTI study of KIC 7702838 involved $\sim 10^5$ transit signal injections. Having this large number of injections for a single target allows us to empirically measure the pipeline completeness that results from the TPS search alone. We show in Figure 1 the measured detection contour based upon the FLTI results for KIC 7702838. These detection contours measure the fraction of injected transit signals that are recovered over the chosen parameter space. The parameter space adopted for this example plots injected orbital period, $P_{\rm orb}$, vs planet radius, $R_{\rm p}$. These empirical detection contours serve as 'ground-truth' for use in validating our pipeline completeness model (Burke & Catanzarite, 2017b). The relevant steps in the calculation of these contours are as follows:

1. Read the FITS file data columns for the injected orbital period (*Period*), injected planet radius (*Rp*), and recovery flag from the TPS local 2x2 grid search (*Recovered*).

2. Define the two-dimensional grid of $P_{\rm orb}$ and $R_{\rm p}$ bins used to calculate the fraction of injected transit signals that are recovered. The user should determine a grid spacing that satisfies the scientific goals of their investigation, while balancing the trade-off between desired resolution and sufficient injection tests per bin.

3. Assign the number of injections attempted, nAll, to each bin in the $P_{\rm orb}$ vs $R_{\rm p}$ grid. Similarly, assign the number of attempted injections that were actually recovered (*Recovered == 1*), nRecvr, to each bin in the grid.

4. Calculate the detection fraction, nRecvr/nAll, for every bin in the grid. This represents the fraction of transit signals injected over a uniform distribution of impact parameters ($0<b<1$) that are recoverable by the TPS search algorithm (see Figure 1).

The empirical detection contours in Figure 1 illustrate the major features that the pipeline completeness model (Burke et al., 2015; Burke & Catanzarite, 2017b) attempts to recreate, so as to serve in the absence of FLTI results for all stars. The window function effect (Section 4.2) results in the drop off in completeness towards long $P_{\rm orb}$, and the detection efficiency (Section 4.3) results in the drop off in completeness towards small $R_{\rm p}$.

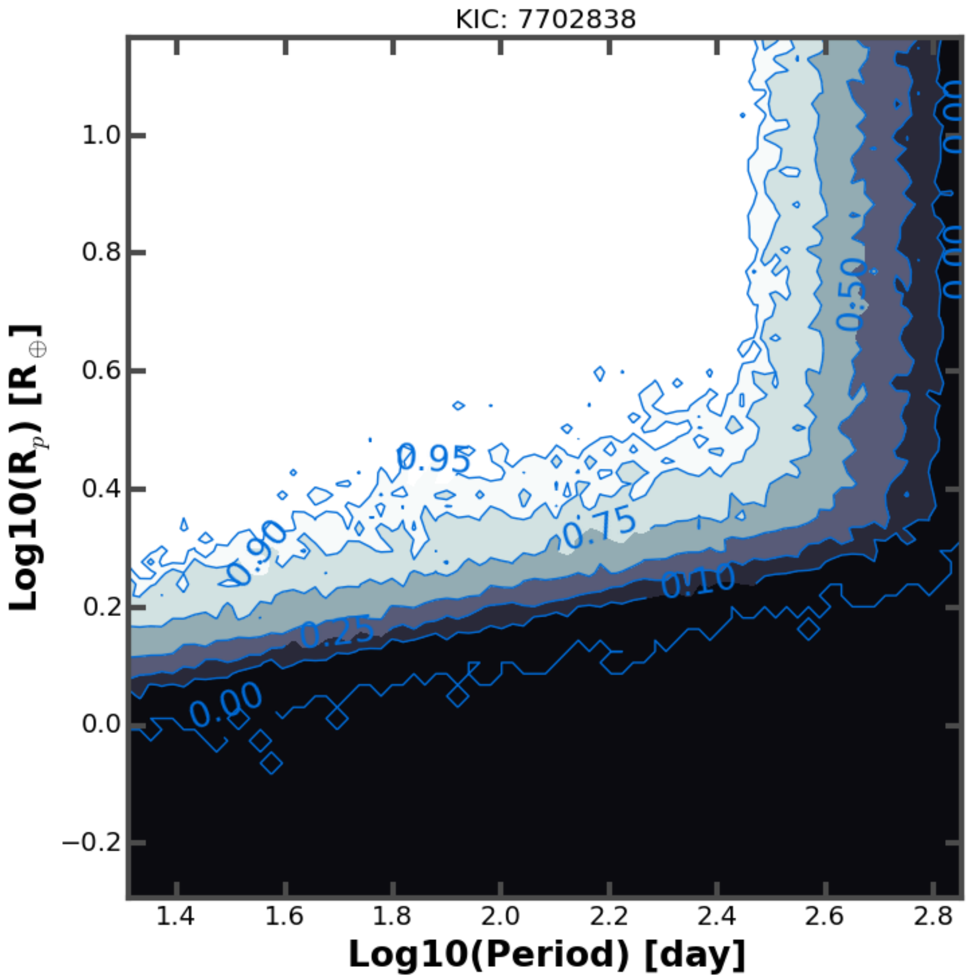

Figure 1: The empirical detection contours determined from the FLTI output for KIC 7702838. The labeled contour lines and shading represent the fraction of transit signals injected with a uniform distribution of impact parameter (0<b<1) that are recovered by the localized 2x2 grid search with TPS. The empirical detection contours serve as 'ground-truth' that can be used to validate a model for pipeline completeness.

4.2 Calculating an Empirical Window Function

The window function is the component of the pipeline completeness model that describes the fall-off in transit signal detectability towards long orbital periods. It arises for *Kepler* from the requirement for at least three transit events and, in the limiting case of exactly three transit events, that none strongly overlap with missing or deweighted data. The window function effect for the example target KIC 7702838 is visible as the vertical fall-off in signal recoverability for $P_{\rm orb} > 250$ days in the high signal-to-noise (SNR) regime, $R_{\rm p} > 3R_{\oplus}$. A description for calculating a numerical window function tailored for all *Kepler* targets is described in Burke & Catanzarite (2017a). The numerical window function is validated against the empirical window function that we calculate from the FLTI output. The steps involved in that calculation are as follows:

1. Read the FITS file data columns for the injected orbital period (*Period*), injected planet radius (*Rp*), injected impact parameter (b), injected transit duration (*t_dur*), number of transits when evaluated during the supplemental targeted search (*ralt_ntran*), and the result of evaluating whether there are three transit signals that are not too deweighted during the targeted search (*ralt_threeTransitFail*). For the last two data columns, it is important to use quantities from the targeted search (i.e., *ralt_ntran* and *ralt_threeTransitFail*), rather than the similar quantities generated by the local 2x2 grid search (i.e., *r_ntran* and *r_threeTransitFail*) because the local grid search is not guaranteed to evaluate these quantities where the signal was actually injected.

2. Retrieve the stellar radius (*RADIUS*) and the rmsCDPP noise values (*RCDP01P5 ... RCDP15P0*) for the fourteen transit durations from the primary FITS header.

3. Estimate $R_{\rm p,lim}$, the injected planet radius that defines the lower limit to the high SNR region where recoverability of transit signals is high (i.e., > 90%) by:

   (a) Estimating a typical rmsCDPP noise value over a period range where the window function is transitioning from high to low completeness (note that this transition occurs when *ralt_ntran* == 3). Select the subsample of injections for which *ralt_ntran* == 3 and then find the median injected transit duration (*t_dur*) for this three-transit subsample.

   (b) Using nearest neighbor interpolation, find the rmsCDPP that corresponds to the median *t_dur* from the three-transit subsample.

   (c) Approximate the transit signal as $\mathrm{SNR} = (\Delta\sqrt{N_{tran}})/\sigma$, where $\Delta = (Rp/Rstar)^2$ is the geometric transit depth for a planet of radius *Rp* around a uniformly bright star of radius $Rstar$, N_{tran}=3 is the number of transits, and σ is the rmsCDPP noise estimate for orbital periods where 3 transits are expected.

   (d) Adopting SNR = 13 as the high SNR limit, solve the SNR equation to obtain $R_{\rm p,lim}$, the limiting planet radius that is valid for calculating the window function.

4. Identify a valid window function subsample by selecting injections that have planet radii $Rp > R_{\mathrm{p,lim}}$ and low impact parameter (i.e., $b < 0.7$) to use in the empirical window function calculation.

5. Identify the injections in the valid window function subsample that pass the *Kepler* window function criteria. Specifically, those with *ralt_ntran* > 3 pass, while those with *ralt_ntran*==3 only pass when *ralt_threeTransitFail*==0.

6. Define a grid of bins in P_{orb} space for use in calculating the empirical window function. The user should determine a grid spacing that satisfies the scientific goals of their investigation, while balancing the trade-offs between adequate resolution and sufficient injection tests per bin.

7. Assign to each bin in P_{orb} space the number of injections attempted, nAll. Similarly, assign to each P_{orb} bin the number of injections that pass the window function tests, nPass.

8. Calculate the pass fraction, nPass/nAll, for each bin. This represents the empirical window function estimate, which is shown for KIC 7702838 in Figure 2.

4.3 Calculating an Empirical Detection Efficiency

The detection efficiency represents the component of the pipeline completeness model that describes the fall-off in transit signal detectability towards smaller radii, lower SNR transit signals. It arises for *Kepler* from the requirement that transit signals meet the adopted MES threshold of 7.1, as well as three other criteria described in Seader et al. (2015) and Twicken et al. (2016). The detection efficiency for the example target KIC 7702838 is visible in Figure 1 as the roughly horizontal fall-off in signal recoverability for $P_{\mathrm{orb}} < 250$ days in the low SNR regime, $R_{\mathrm{p}} < 3R_{\oplus}$. Detection efficiency is measured in terms of MES, because this is the primary parameter controlling detectability by the TPS search algorithm (Jenkins, 2002; Christiansen et al., 2013; Burke et al., 2015). The process of creating a detection efficiency model for a large sample of *Kepler* targets is the subject of Burke & Catanzarite (2017b). One can calculate the empirical detection efficiency for a single target using the FLTI results as follows:

1. Read the FITS file data columns for the injected orbital period (*Period*), injected transit duration (*t_dur*), expected MES of the transit signal (*exp_mes*), recovery flag from the local 2x2 grid search (*Recovered*), number of transits evaluated during the targeted search (*ralt_ntran*), and result of evaluating whether there are three transits that are not too deweighted during the targeted search (*ralt_threeTransitFail*). For the last two data columns, it is important to use quantities from the targeted search (i.e., *ralt_ntran* and *ralt_threeTransitFail*), rather than the similar quantities from the local 2x2 grid search (i.e., *r_ntran* and *r_threeTransitFail*), because the local grid search is not guaranteed to evaluate these quantities where the signal was actually injected.

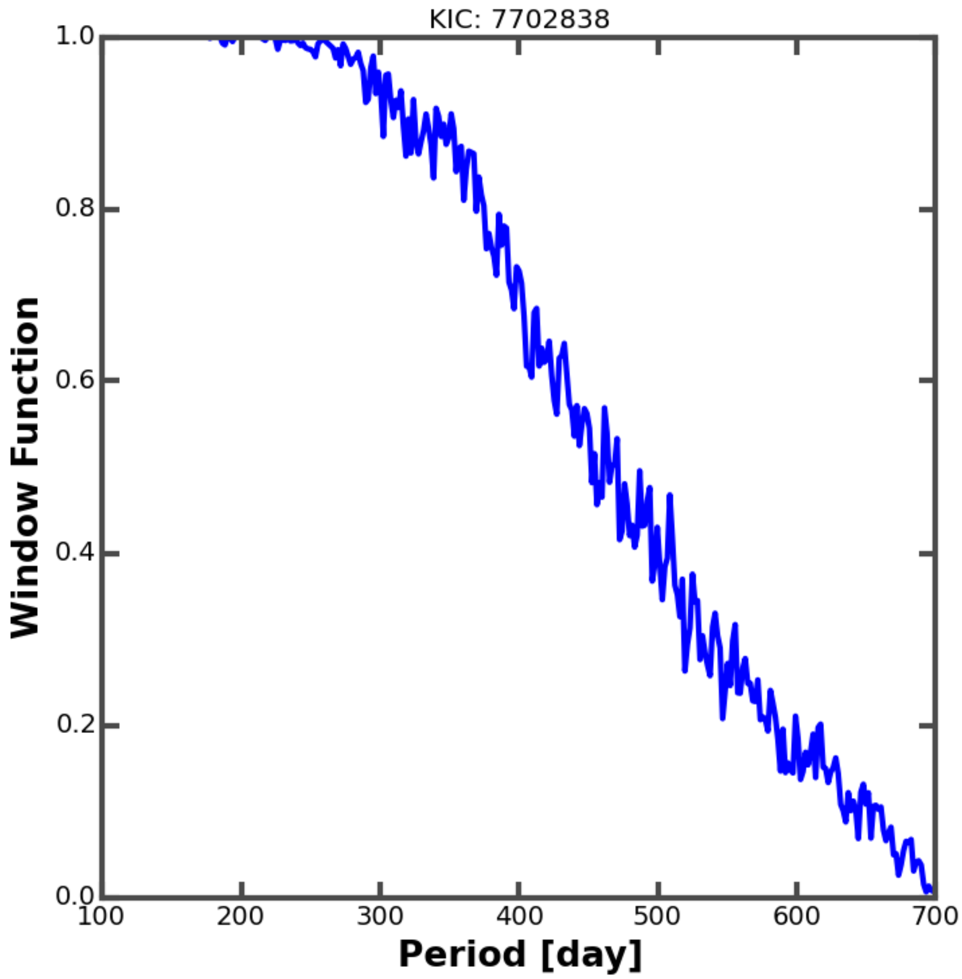

Figure 2: The empirical window function as determined from the FLTI output for KIC 7702838. The window function represents the fraction of injected signals as a function of orbital period that pass the *Kepler* requirements for a transit event. Namely, there must be at least three transits and, in the case of only three transits, they must not fall on missing or deweighted data. This empirical window function serves as the 'ground-truth' that was used to validate the numerical window function tailored to every target as described in Burke & Catanzarite (2017a).

2. Isolate detection efficiency effects from window function effects by selecting a subsample of injections that pass the *Kepler* window function checks. Specifically, those with *ralt_ntran* > 3 pass, while those with *ralt_ntran* $== 3$ only pass when *ralt_threeTransitFail* $== 0$.

3. Limit the subsample to transits with injected durations less than 15 hours, as it has been determined that the TPS search algorithm strongly suppresses transit signals with durations longer than the longest duration employed in the search (i.e., 15 hours). In addition, our adopted pipeline completeness model begins to break down for longer transit durations, though this effect has not been sufficiently studied. Users that choose to include injections with durations *t_dur* >15 hours should carefully assess the impact of their decision.

4. Define a grid of MES bins for use in calculating the empirical window function. The user should determine a grid spacing that satisfies the scientific goals of their investigation, while balancing the trade-off between adequate resolution and sufficient injection tests per bin.

5. Assign the number of injections attempted, nAll, to each MES bin. Similarly, assign the number of attempted injections that were recovered (*Recovered == 1*), nRecvr, to each MES bin.

6. Calculate the recovered fraction, nRecvr/nAll, for each MES bin. This represents the empirical detection efficiency estimate, which is shown for KIC 7702838 in Figure 3.

5 Description of Runs

The FLTI algorithm was run numerous times to validate various aspects of the *Kepler* pipeline completeness model (Burke & Catanzarite, 2017b). Each run has an internal identifier that is used as the folder name for the released files. In each folder there is a separate FITS file for each target included in that particular run. The following subsections describe the various FLTI runs, state their goals, and describe their characteristics. Each aimed for a particular number of transit injections per target, but the number of injections achieved varies due to run time differences. Users are strongly encouraged to read the details for every run, as they are all unique and the text exposes important differences.

5.1 KSOC-5004

This run includes 40 *Kepler* targets with a goal of ∼600,000 injections per target. The targets were selected from a subsample of all *Kepler* targets that were observed for the entire mission (i.e., Q1-Q17), had no Threshold Crossing Events (TCE) identified in the DR25 pipeline run (Twicken et al., 2016), and were below the 90$^{\text{th}}$ percentile for photometric noise. Previous work identified that the slope in rmsCDPP at the longest transit durations is correlated with

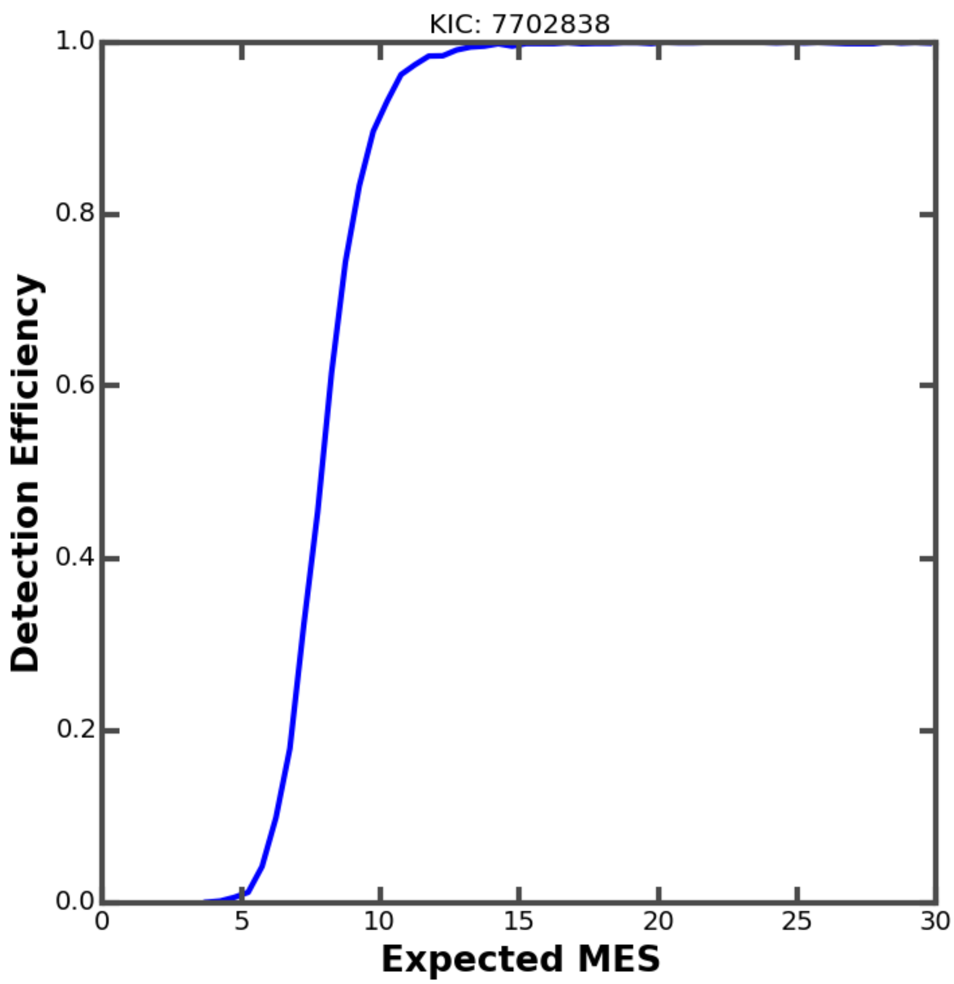

Figure 3: The empirical detection efficiency as determined from the FLTI output for KIC 7702838. The detection efficiency represents the fraction of injected signals as a function of expected MES that pass the *Kepler* requirements for detection. This single-target empirical detection efficiency serves as 'ground-truth' for the validation of the per-target detection contour model described in Burke & Catanzarite (2017b).

variations in detection efficiency. The slope in rmsCDPP is described in Burke & Catanzarite (2017b), but in summary, the rmsCDPP slope measures the rate of change in the rmsCDPP noise values as a function of transit duration. The white Gaussian noise expectation is for rmsCDPP to decrease with increasing transit duration (negative slope). Targets with increasing rmsCDPP slope have increasingly non-Gaussian noise properties and empirically possess less sensitive detection efficiencies.

The targets for this run were carefully selected to uniformly sample the range of long-duration rmsCDPP slopes spanned by all *Kepler* targets. This group contains targets with stellar parameters such that injected transit durations >15 hrs are present, which has not been studied in sufficient detail. The injections on each target were performed uniformly in logarithmic $P_{\rm orb}$ and $R_{\rm p}$ (as well as uniformly in impact parameter) over the region $20 < P_{\rm orb} < P_{\rm max}$ days and $0.5 < R_{\rm p} < 15R_{\oplus}$, where $P_{\rm max}$ is the maximum $P_{\rm orb}$ that can result in observing three transits for a target. A unique feature of this run is that the 'remove feature' algorithm which removes individual events from a light curve when the strongest MES signal does not pass all vetoes was enabled during the supplemental targeted search. The result of having 'remove feature' enabled is that the values for *ralt_ntran* and *ralt_threeTransitFail* are inaccurate, since some strong events may have been removed. Calculations and investigations of the empirical window function behavior for targets in this run may be corrupted due to having 'remove feature' enabled. Subsequent FLTI runs have 'remove feature' disabled during the supplemental search and accurately report the number of transits. Despite this corrupted window function behavior, this run is suitable for detection efficiency and detection contour investigations.

5.2 KSOC-5006

This run includes 32,316 *Kepler* targets with a goal of ∼2,000 injections per target. The targets were selected from a subsample of *Kepler* targets with data span > 300 days and duty cycle > 0.3. Data span measures the duration in days from the first to the final validly observed cadences, and duty cycle measures the fraction of cadences over the data span which contain valid observations (Burke & Catanzarite, 2017a). From this subsample, 30,000 random targets were selected for this test, as well as all M dwarfs with $2400 < T_{\rm eff} < 3900$ K and $\log g > 4.0$ in the DR25 stellar catalog (Mathur et al., 2017). Injections were performed uniformly in logarithmic $P_{\rm orb}$ and $R_{\rm p}$ over the region $20 < P_{\rm orb} < P_{\rm max}$ days and $0.5 < R_{\rm p} < 15R_{\oplus}$, where $P_{\rm max}$ is the maximum $P_{\rm orb}$ that permits three transits per target. Twenty percent of the targets did not achieve 2,000 injections. Due to the sparse nature of injections over the two-dimensional space of $P_{\rm orb}$ and $R_{\rm p}$, the individual targets in this run are only suitable for determining the detection efficiency (which is a single dimension in MES). However, given the wide variety of light curve noise, stellar parameters, and achieved number of injections, one should determine whether sufficient injections are available on a star-by-star basis for the MES region being investigated. The results for individual targets in this run are not suitable for determining two-dimensional detection contours (see section 4.1) or window functions (see section 4.2), but could be computed for small groups of targets. This run contains targets with stellar parameters that permit injections with transit durations >15 hrs; their frequency and impact have not been studied in sufficient detail.

5.3 KSOC-5008

This run includes 25 *Kepler* targets with a goal of ∼600,000 injections per target. The primary purpose was to sample the variety of data span and duty cycle combinations encountered among the *Kepler* targets in order to validate the numerical window functions calculated by Burke & Catanzarite (2017a). The *Kepler* targets were binned into two-dimensional regions of common data span and duty cycle. The data spans and duty cycles of the 25 bins most populated by *Kepler* targets were identified. From each of these data span and duty cycle bins, a single representative 'well-behaved' target was chosen. The criteria for being a 'well-behaved' target were no TCEs identified in the DR25 pipeline run, photometric noise below the 60^{th} percentile, negative rmsCDPP slope values lying in the region of parameter space with negligible detection efficiency variations (Burke & Catanzarite, 2017b), stellar dwarf $\log \mathrm{g} > 4.0$, and no time out issues (Twicken et al., 2016). Injections were performed uniformly in logarithmic P_{orb} and R_{p} space over the $20 < P_{\mathrm{orb}} < P_{\mathrm{max}}$ day and $0.5 < R_{\mathrm{p}} < 15 R_{\oplus}$ region, where P_{max} is the maximum P_{orb} that permits three transits per target.

5.4 KSOC-5104

This run includes 5436 *Kepler* targets with a goal of ∼2,000 injections per target. These targets were originally part of the KSOC-5006 run, but did not achieve sufficient injections to quantify the detection efficiency over the $5.5 < \mathrm{MES} < 12.5$ transition from low to high detectability. In addition, the targets were chosen to have positive rmsCDPP slopes, indicating that strong detection efficiency variations were expected. The parameter space for injection was tailored in P_{orb} and R_{p} for each target so as to achieve a higher density of injections over the important $5.5 < \mathrm{MES} < 12.5$ transition region. In addition, if the DR25 stellar parameters permitted injections with durations >15 hr (assuming b=0 and a circular orbit) at the longest orbital periods, then the stellar parameters were forced to Solar values to avoid this issue. The primary header for each FITS file provides the custom injection region and the adopted stellar parameters for that target. Ten percent of these targets still did not achieve sufficient injections to quantify the detection efficiency transition region because of their extremely long run times. The targets in this run are only suitable for determining detection efficiencies for individual targets. Since the injections were performed over a non-standard P_{orb} and R_{p} parameter space, grouping targets to compute average detection contours or window functions is not recommended.

5.5 KSOC-5125

This run includes 50 *Kepler* targets with a goal of ∼600,000 injections per target. The targets were selected from a subsample of all *Kepler* targets that met minimum data criteria (data span > 700 days and duty cycle > 0.6), were below the 75^{th} percentile for photometric noise, did not suffer time out issues during the DR25 pipeline run, had expected transit durations < 15.5 hr at the longest orbital periods under consideration assuming circular orbits and DR25 stellar parameters. Previous work identified parameters associated with the strength of the whitening filter employed in TPS (Jenkins, 2002) and the amount of data removed during the iterative multi-planet search as potential indicators for variations in detection efficiency. These three parameters were employed to uniformly select 10 targets across each parameter for a total of 30 targets (see Burke & Catanzarite (2017b) for a discussion of their analysis). Another set of 10 targets was chosen to uniformly sample the $T_{\rm eff}$ range of *Kepler* targets and another set of 10 targets was chosen randomly. Injections were performed uniformly in logarithmic $P_{\rm orb}$ and $R_{\rm p}$ space over the $20 < P_{\rm orb} < P_{\rm max}$ day and $0.5 < R_{\rm p} < 15 R_{\oplus}$ region, where $P_{\rm max}$ is the maximum $P_{\rm orb}$ that permits three transits per target.

References

Burke, C. J., Gaudi, B. S., DePoy, D. L., & Pogge, R. W. 2006, AJ, 132, 210

Burke, C. J., Christiansen, J. L., Mullally, F., et al. 2015, ApJ, 809, 8

Burke, C. J., Catanzarite, J. 2017a, Planet Detection Metrics: Window and One-Sigma Depth Functions for Data Release 25 (KSCI-19101-002)

Burke, C. J., Catanzarite, J. 2017b, Planet Detection Metrics: Per-Target Detection Contours for Data Release 25 (KSCI-19111-002)

Castelli, F., & Kurucz, R. L. 2004, arXiv:astro-ph/0405087

Christiansen, J. L., Jenkins, J. M., Caldwell, D. A., et al. 2012, PASP, 124, 1279

Christiansen, J. L., Clarke, B. D., Burke, C. J., et al. 2013, ApJS, 207, 35

Christiansen, J. L., Clarke, B. D., Burke, C. J., et al. 2015, ApJ, 810, 95

Christiansen, J. L., Clarke, B. D., Burke, C. J., et al. 2016, ApJ, 828, 99

Christiansen, J. L. 2017, Planet Detection Metrics: Pixel-Level Transit Injection Tests of Pipeline Detection Efficiency for Data Release 25 (KSCI-19110-001)

Claret, A., & Bloemen, S. 2011, A&A, 529, A75

Coughlin, J. L. 2017, Planet Detection Metrics: Robovetter Completeness and Effectiveness for Data Release 25 (KSCI-19114-001)

Gilliland, R. L., Brown, T. M., Guhathakurta, P., et al. 2000, ApJ, 545, L47

Jenkins, J. M. 2002, ApJ, 575, 493

Jenkins, J. M. 2017, *Kepler* Mission Data Processing Handbook (KSCI-19081-002)

Mandel, K., & Agol, E. 2002, ApJ, 580, L171

Mathur, S., Huber, D., Batalha, N. M., et al. 2017, ApJS, 229, 30

Pence, W. D., Chiappetti, L., Page, C. G., Shaw, R. A., & Stobie, E. 2010, A&A, 524, A42

Seader, S., Jenkins, J. M., Tenenbaum, P., et al. 2015, ApJS, 217, 18

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

Twicken, J. D., Jenkins, J. M., Seader, S. E., et al. 2016, AJ, 152, 158

Weldrake, D. T. F., Sackett, P. D., Bridges, T. J., & Freeman, K. C. 2005, ApJ, 620, 1043