Probabilistic Model-Based Analysis of Kepler Transit Signal Locations

STEVE BRYSON[1], TIMOTHY D. MORTON[2].

[1]NASA Ames Research Center, [2]Princeton University

Abstract

A dominant source of false positives in Kepler data is stellar eclipses or planetary transits on background stars. An important method of identifying these background transit signal sources is the determination of the position of the source relative to the target star. Traditionally the Kepler project has used a simple "the distance from the source to the target star is greater than 3 sigma" threshold to declare when the transit source is unlikely to be on the target star. This approach is unsatisfactory in several ways, including:

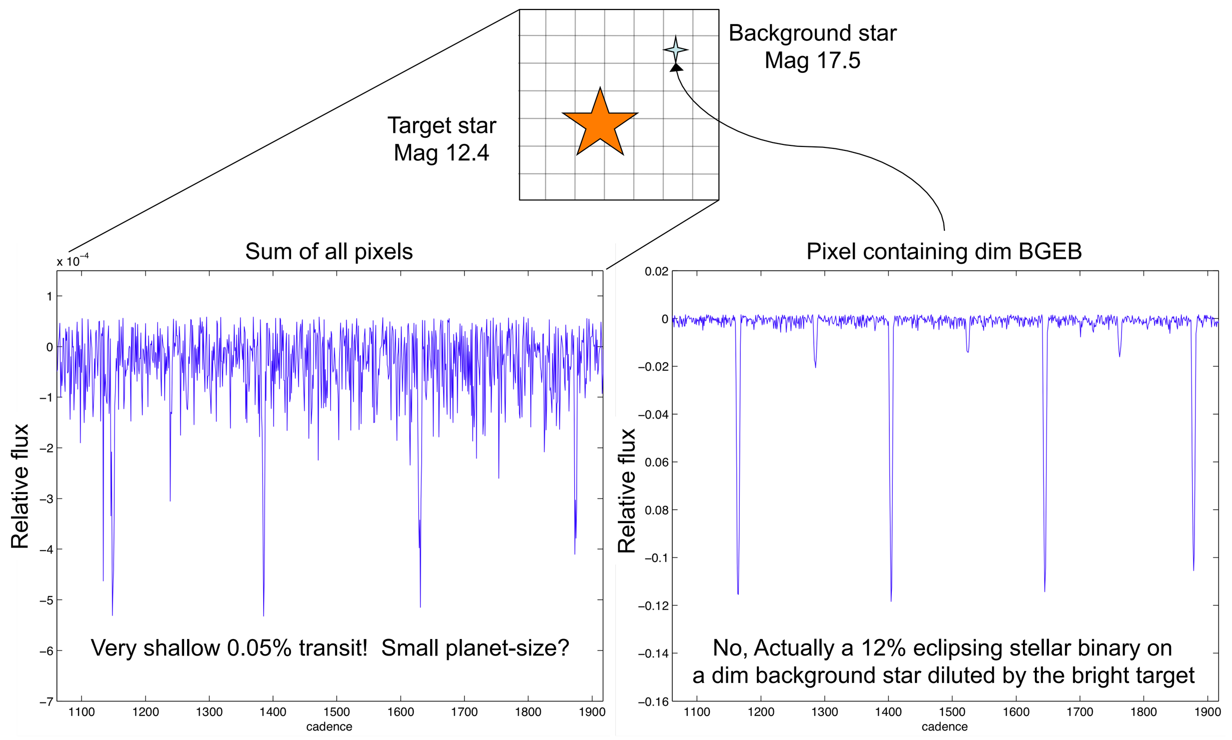
1) This simple threshold assumes that the transit signal location measurement error obeys Gaussian statistics;
2) the case of background stars within 3 sigma are not handled in an informative way;
3) the case of the transit source location measurement apparently coinciding with a known star is not handled differently from when there is no known star;
4) systematic error in the transit signal location measurement due to field crowding is not accounted for;
5) the Galactic-latitude-dependent diffuse background source density is not accounted for.

We present an alternative approach that uses forward modeling and non-parametric reconstruction of the measurement error distribution from Kepler data to address these concerns. Specifically we produce estimated distributions of both the measured transit source position and the expected transit source position under the assumption that the transit is on each known star or the diffuse background. The normalized integral of the product of the observed distribution and each star's (or background) predicted distribution gives the relative probability that the transit occurs on that star (or background). The choice of method for reconstructing the error distribution is crucial. We describe several possibilities, and recommend a smooth bootstrap reconstruction, which combines a bootstrap analysis with kernel density estimation. We describe a table giving these probabilities for KOIs that have appropriate centroid data.

Funding for this mission provided by NASA's Discovery Program Office, SMD.

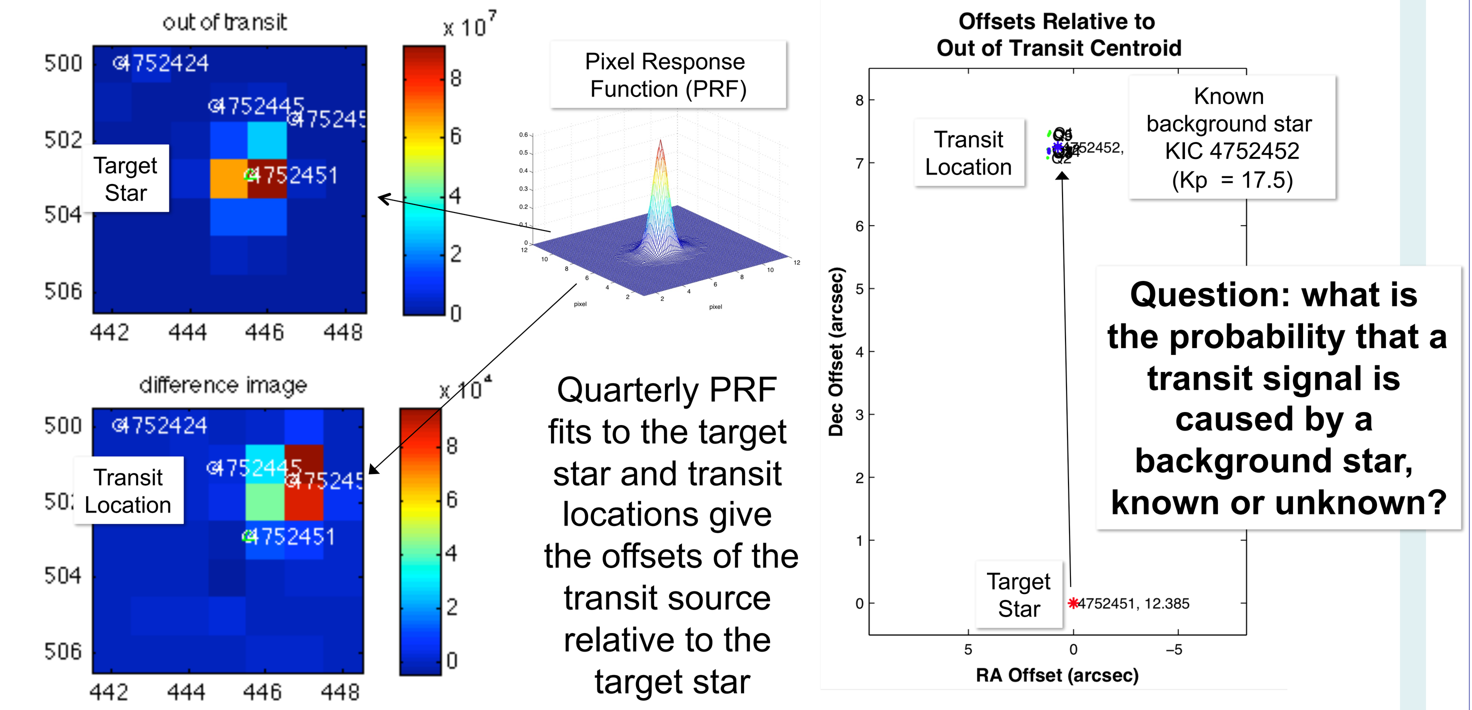
The Detection of Background Eclipsing Binaries (BGEBs) by Pixel Analysis

BGEBs can mimic planetary transits



Find BGEB in the Pixels via Pixel Response Function (PRF) fit

Difference image = out-of-transit – in-transit pixels, shows transit location



From Modeling to Probability Distributions

Modeling Offsets

We simulate the transit on each star in the target star's pixels. For each star, we model the transit on that star by creating synthetic pixel images. These synthetic in- and out-of-transit images are created using the PRF and Kepler Input Catalog (see poster 1-125). The modeled transit depth is chosen to match the observed transit depth after dilution.

These synthetic images are analyzed just like the observed images to give quarterly offsets of the transit from the target. The resulting offsets include bias due to scene crowding.

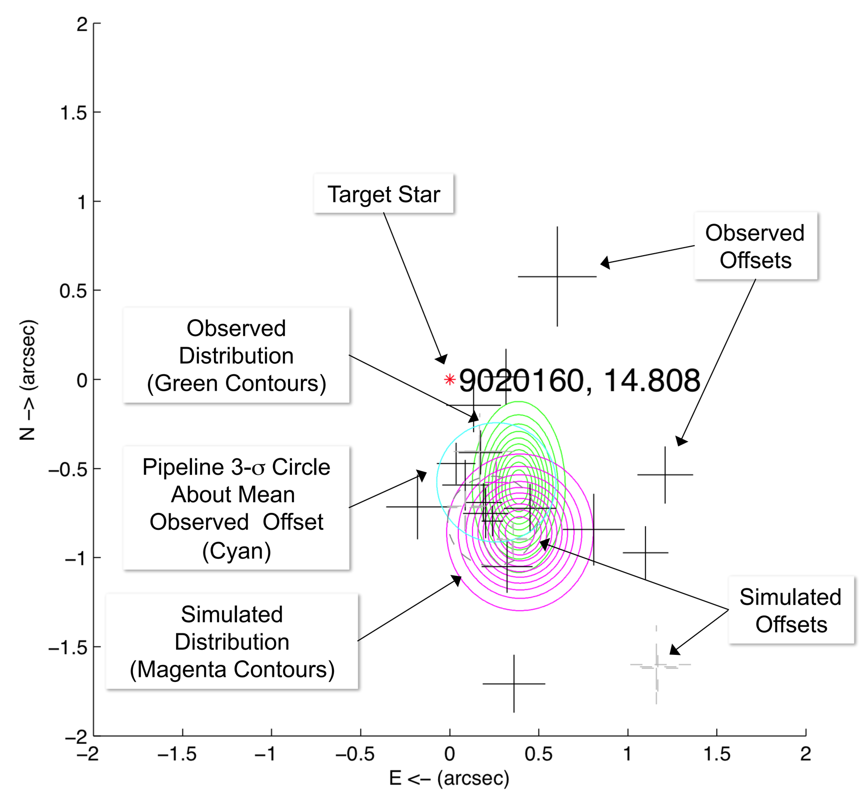
When the simulated offsets for a star "overlap" with the observed offsets we say that it is likely that this star is the source of the transit. We quantify this with normalized probability distributions.

From Offsets to Probability Distributions

There are several ways to infer a probability distribution from a set of quarterly offsets (observed or simulated). We investigated three approaches:

- **Gaussian** with a mean set by the robust average of the offset positions, and a one-sigma uncertainty propagated through that robust average. This method is appropriate only when the distribution underlying the offset positions is in fact Gaussian.
- **Kernel Density Estimation (KDE)**, where the distribution is the sum of Gaussians placed at each quarterly offset position, with width given by an optimal formula from Silverman (1986) based on the standard deviation of the re-sampled means.
- **Smooth Bootstrap** (= bootstrap + KDE): As in a bootstrap computation, the quarterly offsets are randomly re-sampled (with replacement) (# of quarters)^2 times, and the means of these re-samplings are replaced by a Gaussian as in the KDE example. The distribution is the sum of these Gaussians

We choose the smooth bootstrap as the method that best reproduces the observed distribution. Separate distributions are created for the observed and simulated offsets.



The figure on the left shows example observed and simulated smooth bootstrap distributions for KOI-582, along with the observed and simulated quarterly offsets. The length of the crosses in the quarterly offsets are proportional to their 1-σ uncertainty.

When the observed (green) and simulated (magenta) distribution contours overlap, we say that the model transit on this star is consistent with the observations.

This example is particularly interesting because the robust mean of the observed offsets (cyan circle) are more than three-sigma from the target star according to the SOC pipeline, indicating a background false positive. But modeling reveals that field crowding (due to a bright star to the upper left, off this image) biases the PRF fits, and modeling the transit on the target star predicts that these biased offsets will roughly coincide with the observed offsets. This indicates that the diagnosis as a background false positive is likely incorrect.

From Probability Distribution to Relative Probability

Quantify "Overlap" as the Integral of the Product of the Observed and Simulated Distributions

Because the distributions are sums of Gaussians, the integral of the product of the observed and simulated normalized distributions is given by the explicit formula

$$\frac{1}{2\pi Q^2\sqrt{(w_{\mathrm{RA}}^2+v_{\mathrm{RA}}^2)\,(w_{\mathrm{DEC}}^2+v_{\mathrm{DEC}}^2)}}\sum_{q=1}^{Q}\sum_{r=1}^{Q}e^{-\gamma_{\mathrm{RA},q,r}}e^{-\gamma_{\mathrm{DEC},q,r}},\qquad \gamma_{q,r}=\frac{(\Delta_q-\Gamma_r)^2}{2\,(w^2+v^2)}$$

Where Δ_q are the means of each bootstrap re-sample q of the observed offsets, and w is the the Gaussian width. Γ_r and v are the same for the r re-sampled simulated offsets. Q = (# of quarters)^2 is the number of re-samplings.

From a Bayesian perspective, this overlap integral is the **likelihood** L_s of a transit on star s being observed in the same location the observed transit. We compare with the likelihood of another star t by computing the Hypothesis Ratio $H_{s,t} = L_s/L_t$.

We normalize the hypothesis ratio as $P_s = \frac{H_{s,t}}{\sum_k H_{k,t}} = \frac{L_s}{L_t}\frac{1}{\sum_k \frac{L_k}{L_t}} = \frac{L_s}{\sum_k L_k}$, which does not depend on the star t!

We call P_s the Relative Probability of the star s. The sum of P_s over all hypotheses (stars + background) equals 1.

What About Background Stars?

Given a local Background Density b, what is the relative probability that the transit is due to an unknown background star?

The background density of a possible transit source depends on Galactic latitude and the dilution in the aperture (Morton and Johnson 2011, Bryson talk)

We take the background density b to be locally constant on the scale of a target aperture. Then because the observed probability distribution is normalized, the overlap integral = the likelihood $L_{\mathrm{background}}$ is just b.

Results

Model-based Relative Probability

- Helps validate planet candidates by assigning a probability that the observed transit is consistent with the hypothesis that the transit is on the target star compared to other stars or the background.
- Reduces miss-identified false positives due to field crowding.
- Provides an automated way to identify transit signals with known background sources.

A table giving these relative probabilities for all KOIs will be published soon (Bryson and Morton, 2013, maybe 2014).