TESS Science Memo No. 11, Version 1, 2014 January 24

Detection Threshold for Statistical False Positives

Joshua N. Winn

Department of Physics, and Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139

ABSTRACT

I estimate the detection threshold (a certain number of σ) that would produce of order one statistical false positive transit detection over the course of the TESS mission.

1. Statement of the problem

What should be the detection threshold for finding transiting planets using TESS data? Transiting planets will be detected by searching for statistically significant periodic dips in flux. There will also be random fluctuations in flux, as well as quasi-random fluctuations due to astrophysical variability and instrumental artifacts. Setting too high a threshold will cause us to miss potentially interesting planets. Setting too low a threshold will cause us to drown in false positives.

Once a light curve has been "pre-whitened" or "de-trended" it is hoped that it can be described as a time series that has a value of unity everywhere except possibly for transits, where the flux values are lower. In addition to this signal, each data point has some added noise.

The basic test for a flux dip is to hypothesize a certain period P, first time of transit t_1, and transit duration or "width" W; identify all the data points that correspond to "in-transit" according to those parameters; and compute the mean flux $\bar{f}$ of the in-transit data points. The difference $1 - \bar{f}$ is then compared to the effective noise level σ of the in-transit data points. For white Gaussian noise, σ is the uncertainty in each flux datum divided by the square root of the number of in-transit data points. Finally, if $(1 - \bar{f})/\sigma > N_\sigma$, then the flux dip is deemed significant. The question posed above can now be restated: what should we choose for N_σ?

For *Kepler* the most troublesome types of false positives have proven to be due to time-correlated noise sources such as astrophysical false positives (eclipsing binaries) and instrumental artifacts. However, in principle there are also false positives produced by independent random noise fluctuations such as photon-counting noise. These will probably not be the most worrisome or numerous types of false positives, but they are at least amenable to formal mathematical analysis under some idealized assumptions, unlike the astrophysical and instrumental false positives which are more difficult to model. By considering statistical false positives under optimistic assumptions, we may at least determine a *lower limit* for N_σ. That is the goal of this memo.

2. Number of independent tests

Our goal will be to determine the value of N_σ such that we expect of order *one* statistical false positive after thoroughly searching the entire TESS dataset.

Because the transit parameters are not known *a priori* we must try many combinations of P, t_1, and W, checking in each case for a significant flux dip. The larger the number of trials, the more likely it is that random noise will produce a flux dip of a certain size. The basic problem is to determine how many effectively independent tests are performed for each star, i.e., how many independent samples of the detection statistic $(1-\bar{f})/\sigma$ are computed during the grid search. Then, assuming the detection statistic obeys a Gaussian PDF with zero mean and unit variance, we can compute the value of N_σ such that each star has a probability $N_\star^{-1}$ of producing a detection statistic that exceeds N_σ, and therefore that the entire ensemble of $N_\star$ stars will yield approximately one statistical false positive.

This exact problem was considered at length by Jenkins et al. (2002). This memo largely reflects the author's attempt to understand the basic idea of that paper, and may also reflect his very incomplete understanding of that paper. Much of the complexity of the paper seems to be about the possibility of time-correlated noise in the data, which is a genuine concern—however for this memo we will idealize the time series for each star as consisting of data points with independent Gaussian noise, and a time sampling fine enough to resolve even the shortest-duration transits that will be searched.

The grid search may be reduced from three dimensions to two dimensions by restricting W to be a typical value for a transit, given the hypothesized orbital period. In other words W will be set to a deterministic function of P. In general the transit duration for a circular orbit is (Winn 2011, Eqns. 18-19)

$$W = (13~\mathrm{hr})\left(\frac{P}{1~\mathrm{yr}}\right)^{1/3}\left(\frac{\rho_\star}{\rho_\odot}\right)^{-1/3}\sqrt{1-b^2} \tag{1}$$

where b is the impact parameter, and $\rho_\star/\rho_\odot$ is the host star's mean density divided by that of the Sun. For this memo we will adopt "typical" values $b = 0.5$ and $\rho_\star/\rho_\odot = 2$ (appropriate for an $M = 0.7~M_\odot$ dwarf), giving

$$W(P) = (1.3~\mathrm{hr})\left(\frac{P}{1~\mathrm{day}}\right)^{1/3}. \tag{2}$$

In reality one might want to search over a relatively small range of W values, to allow for variations in b and $\rho_\star$, as well as the possibility of eccentric orbit. However, for given values of P and t_1, such tests with differing W will be strongly correlated since they involve a large fraction of data points in common. We will assume that they do not add significantly to the total number of independent tests, which is dominated by the large number of values of P and t_1 that must be tested.

Now consider a time series of duration D. Suppose we are searching for transits of period P. There will be $N_t = D/P$ trial transits. We may place t_1 anywhere within the time range from $t = 0$ to $t = P$. For effectively independent tests, the trial values for t_1 should be separated by $\Delta t_1 = W(P)$, so that a completely different collection of data points is examined during each test. Therefore for the period P, the number of independent tests we perform is $P/\Delta t_1 = P/W(P)$.

What should be the next period that is chosen for searching? If we increment the period from P to $P' = P + \Delta P$, then the nth trial transit is shifted forward in time by $n\Delta P$. The last trial transit (number $N_t = D/P$) is shifted the most. For very small values of ΔP, the new tests (at period P') and the previous tests (at period P) are not independent, because they largely involve the same data points. As ΔP is allowed to grow larger, eventually the last trial transit is shifted by a full transit width, compared to its position when the trial period was P. This condition is $N_t\Delta P = W(P)$ or $\Delta P = PW(P)/D$.

We will consider this to be the approximate condition for the size of ΔP that leads to the next independent set of tests. This is not strictly true, because many of the "in-transit" data points for the P'-tests will have substantial overlap with the "in-transit" data points that were considered in the P-tests. In this sense, choosing $\Delta P = PW(P)/D$ is expected to give an overestimate of the actual number of independent tests, and therefore the detection threshold we compute will be conservative (i.e., higher than it strictly needs to be, for the maximum desired false positive rate).

We have written a code that accepts as input D, $P_{\rm min}$, and $N_\star$. Beginning with $P_{\rm min}$, the code computes the number of effectively independent trials $P/W(P)$ for that period. Then it increments the period by $PW(P)/D$, and calculates the number of effectively independent trials for the new period, adding it to the running total. This period search is repeated until the trial period is D (i.e., single-transit detections). The code then reports N_P, the total number of periods searched; $N_{\rm EIT}$, the number of effectively independent trials per star; and $N_{\rm EIT,tot} \equiv N_{\rm EIT} N_\star$. Finally, it computes N_σ such that

$$\int_{-\infty}^{N_\sigma} \sqrt{\frac{1}{2\pi}} e^{-x^2/2} dx = N_{\rm EIT,tot}^{-1}. \qquad (3)$$

For TESS we will adopt $N_\star = 5 \times 10^5$ stars, $D = 28$ days, and $P_{\rm min} = 4$ hr (essentially the Roche limit). This gives $N_P = 2300$, $N_{\rm EIT} = 72000$, $N_{\rm EIT,tot} = 3.6 \times 10^{10}$ and $N_\sigma = 6.6$.

For *Kepler* the parameters are more like $N_\star = 2 \times 10^5$ stars, $D = 4.5$ yr, and $P_{\rm min} = 0.5$ day. These give $N_P = 1.1 \times 10^5$, $N_{\rm EIT} = 1.9 \times 10^7$, $N_{\rm EIT,tot} = 3.8 \times 10^{12}$ and $N_\sigma = 7.2$.

Given that there is some ambiguity in the choices of Δt_1 and ΔP, because it is not obvious how to set those precisely for an "effectively independent" test, this should all probably be validated with Monte Carlo simulations. We can produce a simulated well-sampled, noisy light curve (with no transits) lasting duration D, and search it for transiting planets with a very fine grid in P and t_1, and look at the distribution of the detection statistic. Then we can repeat a few million times and check on the level N_σ that produces only one false positive per $N_\star$ stars.

REFERENCES

Jenkins, J. M., Caldwell, D. A., & Borucki, W. J. 2002, ApJ, 564, 495

Winn, J.N. 2011, in *Exoplanets*, ed. S. Seager, University of Arizona Press, ISBN 978-0-8165-2945-2, p. 55-77 [arxiv:1001.2010]

This preprint was prepared with the AAS LaTeX macros v5.2.