Draft version April 15, 2025
Typeset using LaTeX default style in AASTeX631

Transit-APP: A Centroid-Free Method of Identifying Background Transit False Positives

Steve Bryson [1] and Michelle Kunimoto [2]

[1] NASA Ames Research Center, Moffett Field, CA 94035, USA
[2] Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, BC V6T 1Z1, Canada

ABSTRACT

We present a centroid-free method to identify which star is the source of a known transit signal from pixel data. This method is based on observed and simulated difference images, and provides a relative probability of a star being the source of the transit compared to other known stars. This method is robust at low S/N compared to common centroid-based methods, and is well suited for space-based transit surveys such as Kepler and TESS.

1. INTRODUCTION

Background false positives (FPs) are a significant source of FPs in transit surveys (Brown 2003; Bryson et al. 2013), particularly space-based surveys like Kepler (Borucki et al. 2010) or TESS (Ricker et al. 2015). Identifying such FPs from survey data typically involves pixel analysis to detect a change of flux centroids correlated with the transit signal. A transit signal is typically determined to be an FP if such a centroid shift is statistically significant, indicating that the transit signal is not co-located with the target star. These centroid-based methods suffer when the transit has low S/N because the centroids will have lower precision and may fail in individual quarters/sectors.

An improvement over the "yes/no" determination of whether a transit signal is co-located with the target star is to estimate the probability that the signal originates from a background FP. This can be done by modeling the expected centroid shift caused by the transit signal on nearby stars and comparing these models to the actual observed centroid shift. This approach was implemented for Kepler data in Bryson and Morton (2017) and for TESS data in Hadjigeorghiou and Armstrong (2023). TRICERATOPS (Giacalone et al. 2021) computes the probability that a transit signal is consistent with a background source via light curve analysis, but does not compute the probability that the signal is located on a background star.

Centroids are essentially weighted averages over pixel values, so they lose information contained in the pixel data. Here we present a centroid-free method of computing the probability that a transit is due to a known background source. Because our method jointly analyzes all pixel values, we have found it to be more robust at low S/N than common centroid-based methods.

2. THE CENTROID-FREE METHOD

Our method is based on *difference images*, which are formed by subtracting average in-transit pixel images from average out-of-transit pixel images. For details, see Bryson et al. (2013). So long as the transit signal source is the only flux variation in the pixels, the difference image will be a star-like image at the location of the signal source. One difference image is created for each quarter (Kepler) or sector (TESS).

Given a transit signal with specified duration and depth, we model the difference images for each star near the target star on which the transit signal was observed. To jointly analyze all quarters/sectors, we form a collection D_{obs} of observed difference images, one per quarter/sector, and a similar collection of simulated difference images D_s for each nearby star s, including the target star. We also form the collection of formally propagated observed difference image uncertainties σ_{obs}. Jointly fitting the quarters/sectors leverages all the available information to determine which model is the best fit. D_{obs}, σ_{obs} and D_s are each converted into 1D arrays by flattening and concatenating the quarters/sectors, so the n^{th} element of each array refers to the same difference image pixel in the same quarter/sector.

Corresponding author: Steve Bryson
steve.bryson@nasa.gov

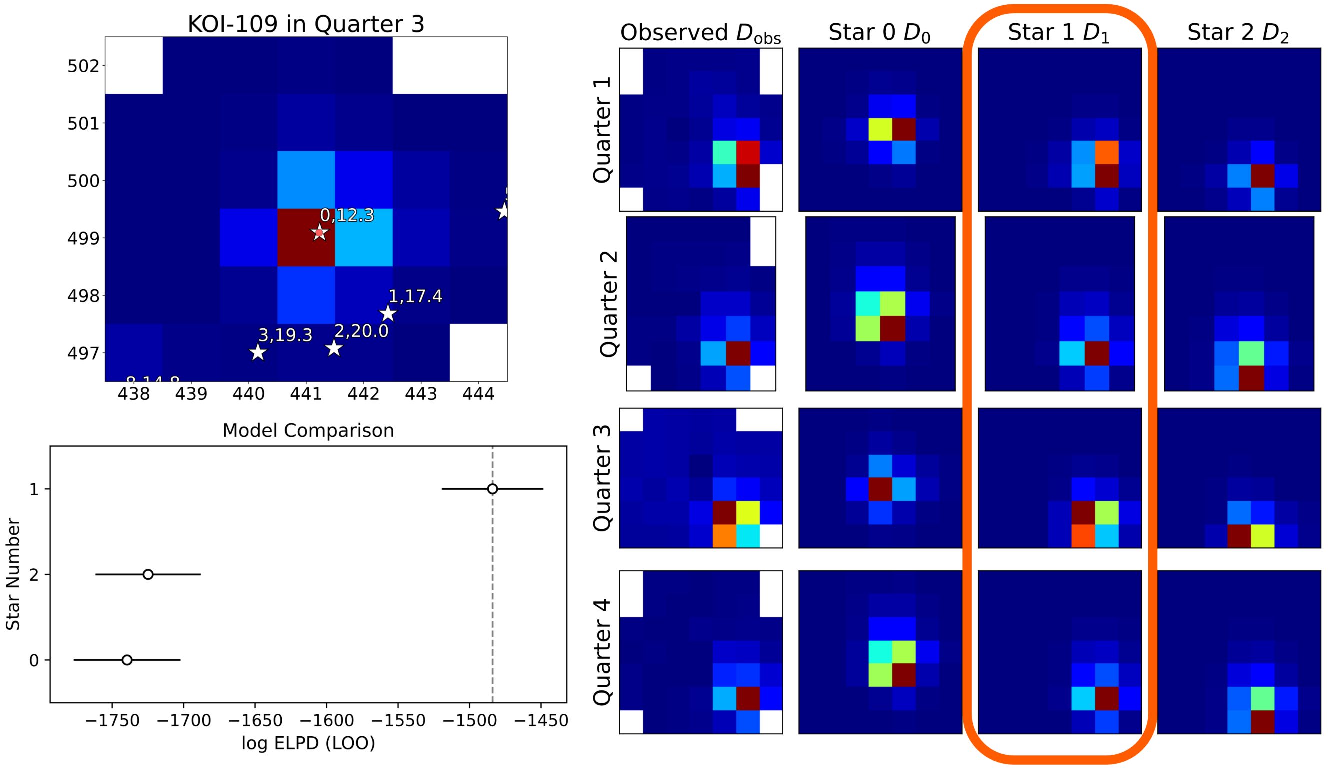

Figure 1. An example of the centroid-free method applied to the known Kepler background false positive KOI-109.01. **Upper Left:** the scene around the star KOI-109 in quarter 3, with stars from the Gaia DR3 catalog indicated by white stars and their star index and g-magnitudes. The red dot is the star KOI-109. **Right:** The observed and simulated difference image collections, with the observed difference images D_{obs} in the first column and each star's simulated difference images D_s in the other three columns. Only stars 0, 1, and 2 are shown for clarity. The rows are different quarters, with only quarters 1-4 shown for clarity. The orange outline shows the star whose simulated difference images most closely resembles the observed difference images. **Lower Left:** The expected log pointwise predictive density (ELPD) values and error bars for each star, showing that star 1 is overwhelmingly more likely to be the source of the transit signal than than star 0 (KOI-109) or star 2.

We determine each star's probability of being the source of the transit signal via Bayesian model selection using the pymc package[1]. For star s, we apply a Gaussian noise model to each pixel of the simulated difference images by using the likelihood $\prod_{\mathrm{pixels}} \mathcal{N}\left(a_s D_s, b_s \sigma_{\mathrm{obs}}\right)$ to be compared with D_{obs}, where the product is over all pixels in the collections D_{obs}, σ_{obs} and D_s, and a_s and b_s are the parameters of our MCMC fit. a_s is a scale factor that provides a rough estimate of the true transit depth, and b_s scales the formal propagated uncertainties to best match the observed uncertainties. We use uninformative uniform priors on the interval $[0, 1000]$ for a_s and b_s. The expected log pointwise predictive density (ELPD) of the resulting MCMC chains are used, via the compare() function of the arviz python package[2], to determine the most likely model (Yuling et al. 2018; Spiegelhalter et al. 2002). The model with the highest ELPD is the best fit to the data. The weight returned for each star s by the compare() function can be interpreted as the probability of that star being the source of the the transit signal relative to other stars (see Yuling et al. (2018) for details).

3. RESULTS

With rare exception, the highest probability star selected by our centroid-free method is consistent with the results of centroid shift analysis, difference image analysis and the probability results of Bryson and Morton (2017) and Hadjigeorghiou and Armstrong (2023), when the transit S/N is high enough for those latter methods to give good results. When the S/N is high, our method can distinguish sources separated by about 0.2 pixels, ≈ 0.8 arcsec for Kepler and ≈ 5 arcsec for TESS. For lower S/N the results of our new method yields plausible results when the

[1] https://www.pymc.io/

[2] https://python.arviz.org/

centroid-based methods often fail. Like centroid-based methods, our method is not appropriate for saturated stars, because there is no sufficiently accurate model of saturation at the pixel level.

Similar to centroid-based methods, however, there are rare circumstances where our new method can result in incorrect results. The most common of these rare circumstances is the difference images being dominated by a nearby bright variable star whose variations are correlated with the transit period and duration. This problem is most commonly encountered in TESS data due to its larger pixels. We therefore do not advise using our method blindly.

4. CONCLUSIONS

We have presented a centroid-free method to determine the probability that a particular star is the source of the transit signal. In principle, jointly fitting all difference image pixel data leads us to expect our method to be more robust than common centroid-based methods, particularly at low transit S/N. These expectations are consistent with our experience. More formal study of this robustness will be the subject of future work. We present the algorithm here because it is currently being incorporated into TRICERATOPS and is being used to vet new transiting exoplanet discoveries.

A prototype implementation of our method can be found on github at https://github.com/stevepur/transit-APP.

5. ACKNOWLEDGEMENTS

We thank Jon Jenkins and Douglas Caldwell for helpful discussions. This work was supported by TESS GI Grant #20-TESS20-0042.

REFERENCES

Borucki, W., Koch, D., Basri, G., et al. 2010, Science, 327, 977 https://doi.org/10.1126/science.1185402

Brown, T. M. 2003, ApJ, 593, L125 https://doi.org/10.1086/378310

Bryson, S. T., Jenkins, J. M., Gilliland, R. L., et al, PASP, 125, 889 https://doi.org/10.1086/671767

Bryson, S. T. and Morton, T. M., 2017, NExScI Exoplanet Archive, KSCI-19108-001 https://doi.org/10.26133/NEA29

Giacalone, S., Dressing, C. D., Jensen, E. L. N., et al. 2021, AJ, 161, 24 https://doi.org/10.3847/1538-3881/abc6af

Hadjigeorghiou, A., Armstrong, D. J. 2023, MNRAS 527, 2 https://doi.org/10.1093/mnras/stad3286

Ricker, G., Winn, J., Vanderspek, R., et al. 2015, Journal of Astronomical Telescopes, Instruments, and Systems, 1, 014003 https://doi.org/10.1117/1.JATIS.1.1.014003

Spiegelhalter, D. J, Best, N. G., Carlin, B. P, Van Der Linde, A. 2002, Journal of the Royal Statistical Society Series B: Statistical Methodology, 6, 236 https://doi.org/10.1111/1467-9868.00353

Yuling, Y., Vehtari, A., Simpson, D., Gelman, A. 2018, Bayesian Anal. 13, 3 https://doi.org/10.1214/17-BA1091