The Astronomical Journal, 167:202 (38pp), 2024 May https://doi.org/10.3847/1538-3881/ad29f0

DIAmante TESS AutoRegressive Planet Search (DTARPS). I. Analysis of 0.9 Million Light Curves

Elizabeth J. Melton[1,2,3], Eric D. Feigelson[1,2,4], Marco Montalto[5], Gabriel A. Caceres[6], Andrew W. Rosenswie[1,7,8], and Cullen S. Abelson[1,9]

[1] Department of Astronomy & Astrophysics, Pennsylvania State University, University Park, PA 16802, USA
[2] Center for Exoplanets and Habitable Worlds, 525 Davey Laboratory, The Pennsylvania State University, University Park, PA 16802, USA
[3] Department of Physics and Optical Engineering, Rose-Hulman Institute of Technology, 5500 Wabash Avenue, Terre Haute, IN 47803, USA
[4] Center for Astrostatistics, 525 Davey Laboratory, The Pennsylvania State University, University Park, PA 16802, USA
[5] INAF-Catania, Osservatorio Astrofisico di Catania, Via Santa Sofia 78, 95123 Catania, Italy
[6] EY-Parthenon, 1540 Broadway, New York, NY 10036, USA
[7] Institut für Physik und Astronomie, Universität Potsdam, D-14476 Golm (Potsdam), Germany
[8] Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, D-14482 Potsdam, Germany
[9] Department of Physics and Astronomy, University of Pittsburgh, 100 Allen Hall, 3941 O'Hara Street, Pittsburgh, PA 15260, USA

Received 2023 February 13; revised 2024 January 11; accepted 2024 January 11; published 2024 April 9

Abstract

Nearly one million light curves from the TESS Year 1 southern hemisphere extracted from Full Field Images with the DIAmante pipeline are processed through the AutoRegressive Planet Search statistical procedure. ARIMA models remove lingering autocorrelated noise, the Transit Comb Filter identifies the strongest periodic signal in the light curve, and a Random Forest machine-learning classifier is trained and applied to identify the best potential candidates. Classifier training sets are based on injections of planetary transit signals, eclipsing binaries, and other variable stars. The optimized classifier has a True Positive Rate of 92.5% and a False Positive Rate of 0.43% from the labeled training set. The result of this DIAmante TESS autoregressive planet search of the southern ecliptic hemisphere analysis is a list of 7377 potential exoplanet candidates. The classifier had a 64% recall rate for previously confirmed exoplanets and a 78% negative recall rate for known False Positives. The completeness map of the injected planetary signals shows high recall rates for planets with 8–30$R_\oplus$ radii and periods 0.6–13 days and poor completeness for planets with radii $<2R_\oplus$ or periods <1 day. The list has many False Alarms and False Positives that need to be culled with multifaceted vetting operations (Paper II).

Unified Astronomy Thesaurus concepts: Exoplanet catalogs (488); Exoplanet detection methods (489); Light curve classification (1954); Period search (1955); Time domain astronomy (2109); Transits (1711)

Supporting material: machine-readable table

1. Introduction

1.1. Challenges in TESS Planet Discovery

With the 2018 launching of the Transiting Exoplanet Survey Satellite (TESS), scientists acquired a tool for in-depth analysis of rare phenomena such as transiting exoplanets, stellar superflares, and tidal disruption events (Ricker et al. 2015). TESS surveys the entire celestial sphere in month-long observations with four wide-field cameras with 21″ pixels^{-1}. During the prime mission (TESS Years 1 and 2), over 200,000 bright stars were pre-chosen to have a 2 minute observing cadence as prime transit targets, but millions of relatively bright stars are accessible from Full Frame Images (FFIs) with a 30 minute cadence.

The principal goal of the TESS mission is the identification of sub-Neptune ($R < 4\ R_\oplus$) transiting planets around stars sufficiently bright for follow-up characterization of the planets' physical characteristics, including atmospheric composition. Quantitative calculations prior to the mission by Barclay et al. (2018) predicted that ∼3100 transiting planets would be found from light curves of ∼6 million FFI stars acquired during prime mission; of these, ∼1100 would be sub-Neptunes. They predicted that ∼12,000 larger planets ($R > 4\ R_\oplus$) would be discovered in the FFI database. In a revised calculation, Kunimoto et al. (2022) predicted that ∼4000 planets would be detected in the prime mission using FFI images.

The predictions of Barclay et al. (2018) and Kunimoto et al. (2022) have been overly optimistic. The TESS Objects of Interest from the prime mission include 2241 Planet Candidates based on automated detection of a transit-like signal followed by review by the TOI Vetting Team (Guerrero et al. 2021). These include 1035 unique FFI stars obtained with MIT's Quick Look Processing (QLP) pipeline. Nearly half of these have been subject to some follow-up observations (the community-based ExoFOP-TESS enterprise) via which most (88% in the published 2021 catalog) have been redesignated False Positives. Thus, only a few hundred FFI stars—rather than thousands—have emerged to date as reliable hosts of transiting planets from the TESS prime mission.

We can point to a variety of plausible contributions to this large discrepancy between predicted and actual performance in TESS planet discovery:

1. Barclay et al. assume a Gaussian noise model for the light curves, although with a generous 7.3σ detection threshold and consideration of flux dilution by nearby stars blended in the large pixels. But the Kepler mission showed that many stars are variable with amplitudes greater than planetary transit depths (Gilliland et al. 2011), and even

after detrending, they exhibit autocorrelated structure (Caceres et al. 2019b). Stellar variability has multiple sources: “red noise” (Pont et al. 2006), rotational modulation of starspots (Boisse et al. 2011), magnetic reconnection flares (Davenport 2016), pulsations, supergranulation, stellar multiplicity, and other effects.
2. Perhaps the most pernicious source of contamination, eclipsing binaries blended into the large TESS pixel, can, after dilution by the target starlight, produce periodic variations that closely resemble planetary transits. In some cases, it can be very difficult, using either automated classifiers or human vetting, to distinguish planetary candidates from blended eclipsing binaries (BEB).
3. Vetting efforts to reduce BEB contamination often rely on pixel-based crowding and centroid analysis that can inadvertently eliminate valid planetary candidates in the Galactic Plane.
4. Instrumental problems are present in addition to stellar variability in TESS FFI light curves. These include short-term flux variations from the appearance or disappearance of nearby stars as satellite pointing settles after instrument turn-on, as well as “ghost” light from bright variable stars contaminating wide areas of the image. The latter effect is called “ephemeris matching” (Coughlin et al. 2014).
5. Methodological difficulties arise in the analysis of the light curves. These include the sparsity of observations during brief transits at longer periods that hinders Gaussian-based statistical measures, spurious periodogram peaks from deviations from a regular cadence (e.g., periodic instrument closures due to the 13.7 satellite orbit), and mathematical difficulties in reliably evaluating false-alarm probabilities in any periodogram (e.g., VanderPlas 2018; Delisle et al. 2020; Koen 2021).

It is challenging to design a detection procedure that effectively removes such a variety of different stellar variations and treats the instrumental and methodological problems, while maintaining the planetary transit signal. BEB contamination, in particular at low Galactic latitudes, can occur more frequently than true planetary signals. The QLP—and any other transiting detection procedure—must thus adopt conservative classification and vetting procedures to reliably identify planetary signals in the midst of these varied sources of contamination and methodological difficulties. As Barclay et al. (2018) did not consider these issues, it is not surprising that they overestimated the number of smaller, sub-Neptunian planets that can realistically and reliably be found in TESS FFI data.

This situation motivates the search for TESS FII planets using methodologies different from the QLP used to generate the official FFI TOI list.[10] It is quite possible that different statistical approaches to detrending, periodicity searching, automated classification, and human vetting will find some of the true planetary candidates predicted by Barclay et al. (2018) and missing from the TOI FII list.

Several efforts at TESS planet detection have been reported using methodologies independent of the TOI pipeline described by Guerrero et al. (2021) and Kunimoto et al. (2022). These include: the DIAmante pipeline (Montalto et al. 2020) identifying 252 new candidates; Osborn et al. (2020) identifying 200 TESS threshold-crossing events as new candidates; Olmschenk et al. (2021) finding 181 new candidates; the citizen science Planet Hunters TESS project finding 90 new candidates (Eisner et al. 2021); Rao et al. (2021) identifying 38 new candidates; Nardiello et al. (2020) finding 33 new candidates in stellar clusters; Feliz et al. (2021) finding 24 new candidates around M dwarfs; and Dong et al. (2021) identifying 19 new Warm Jupiter candidates. Together, these efforts add over 700 new planetary candidates.

One major consideration is the statistical procedure for detrending the light curves prior to searching for transit-like periodicities. The QLP pipeline uses a high-pass filter, outlier removal, and spline fits to detrend the light curve (Huang et al. 2020). Other procedures use well-known statistical procedures (such as Principal Component Analysis or Gaussian Processes regression) or more advanced signal-processing methods (such as Independent Component Analysis, correntropy, empirical mode decomposition, and Singular Spectrum Analysis).

Following detrending, most analyses seek transiting signals with Box Least Squares (BLS) periodograms (Kovács et al. 2002). In the QLP pipeline, BLS peaks with local signal-to-noise ratio >9 are flagged as threshold-crossing events. These are then inputted into a convolutional neural network classifier trained on human-labeled TESS light curves (Shallue & Vanderburg 2018; Yu et al. 2019). The promising cases are then subject to visual inspection by a vetting team to arrive at the TOI list (Guerrero et al. 2021).

1.2. AutoRegressive Planet Search

In this paper and its companions, we apply a transit detection procedure that shares some similarities with the QLP pipeline but also has significant differences. We adapt the four-stage AutoRegressive Planet Search (ARPS) developed by Caceres et al. (2019b) and applied to the 4 yr Kepler light curves by Caceres et al. (2019a). The ARPS process is outlined in Figure 1.

The ARPS procedure, presented in detail in Section 2, differs from the QLP pipeline in several respects. First, nonstationarity in the light curve are removed with a simple nonparametric algorithm called “differencing,” rather than more complicated semi-parametric detrending procedures like spline or Gaussian Processes regression. Differencing treats both stellar and instrumental variations in a single step, but it leaves behind sudden changes such as transit ingress and egress. Second, we fit low-dimensional parametric autoregressive moving average, or ARMA(p,q), models to remove short-memory autocorrelation in the detrended light curve. This crucial step is missing in other transit searching pipelines. Together, these procedures are known as the ARIMA or Box–Jenkins analysis that has dominated analysis of stochastic time series since the 1970s in fields such as econometric and engineering signal processing. The textbook by Box et al. (2015) has five editions and over 56,000 citations.

The differencing operation changes a sharp-edged box-shaped transit into a double-spike pattern representing the planet ingress and egress. This motivated the third innovation of the ARPS procedure, a new sensitive Transit Comb Filter (TCF) periodogram developed by Caceres et al. (2019a) to measure the amplitude of periodic double-spike patterns in the ARIMA residuals. This replaces the traditional BLS method applied to detrended light curves that do not involve differencing. A simulation study by Gondhalekar et al. (2023) shows in detail that the ARIMA-TCF sequence

[10] https://tess.mit.edu/toi-releases

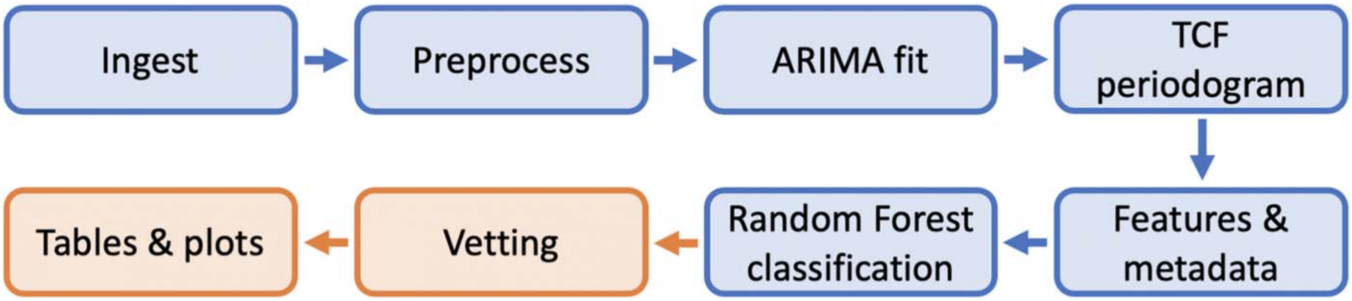


Figure 1. The AutoRegressive Planet Search process. Blue boxes represent steps in the ARPS analysis covered in this paper. Orange boxes represent steps in ARPS covered in Paper II.

outperforms a standard detrender followed by BLS in most TESS-like light curves. The TCF periodogram exhibits fewer spurious speaks, reduced heteroscedasticity (noise varying with period), and reduced trend compared to the BLS periodogram. Combined with ARIMA's removal of short-memory autocorrelation, the ARPS procedure is very sensitive to smaller transiting planets.

ARPS continues with tuning a Random Forest (RF) classifier based on dozens of features to select possible exoplanet transiting candidates from the vast number of non-transiting systems. Following Montalto et al. (2020), the classifier is trained toward a positive training set of light curves with injected simulated planetary transit signals, and it is trained away from a negative training set that includes simulated BEB signals.

The final stage of the ARPS procedure, like the QLP procedure, involves multifaceted visual inspection to reduce False Alarms and False Positives that passed over the Random Forest threshold. This stage is presented in Melton et al. (2024a, Paper II). ARPS is thus a comprehensive planetary transit analysis system, starting with extracted light curves cleaned of most instrumental effects and ending with a new list of planetary candidate transiting systems.

When the ARPS method was applied to ∼150,000 Kepler 4 yr light curves by Caceres et al. (2019b), though without the final vetting step, it recovered 97% of the Kepler Golden sample, provided the Kepler model had a signal-to-noise ratio (S/N) >20, and it identified 97 new Kepler exoplanet transit signals. Most of the new ARPS-identified Kepler planet candidates were (sub)Earths with periods $P < 20$ days orbiting faint stars that could not be readily confirmed with follow-up radial velocity spectroscopy. One case, a Mars-size planet orbiting an M star, is discussed by Cañas et al. (2022). The application to TESS data discussed here produces a collection of candidates that is much more accessible for follow-up study than the earlier Kepler ARPS study.

1.3. DTARPS-S: Application to TESS Year 1 Southern Ecliptic Hemisphere

In this paper, together with Paper II and Melton et al. (2024b, Paper III), we combine procedures from two pipelines for TESS FFI exoplanet detection: the DIAmante pipeline developed by Montalto et al. (2020, henceforth M20) for light-curve extraction and preprocessing, and the ARPS pipeline outlined above for candidate transiting-planet detection. M20 extract 0.9 million light curves from TESS Year 1 FFIs covering the southern ecliptic hemisphere. Using the BLS periodogram and an RF classifier, they proceed to identify 396 exoplanet candidates. The present study is based on the 0.9 million DIAmante-extracted and preprocessed light curves, but then diverts the analysis to the ARPS procedure for planetary transit identification. We call the combined effort the DIAmante TESS AutoRegressive Planet Search or DTARPS. Application to the TESS Year 1 southern ecliptic hemisphere data is called DTARPS-S. The Year 1 full-frame images are available at in MAST (STScI 2022), and the extracted DIAmante light curves are available as a High-Level Science product: doi:10.17909/t9-p7k6-4b32 (Montalto 2020).

The DTARPS-S study of TESS Year 1 FFIs covering the southern ecliptic hemisphere is presented in three papers. Paper I here describes the application of the ARPS method through the application of the RF classifier to the data (Figure 1), producing a list of 7377 stars that exhibit transit-like behavior. This stage is roughly comparable to the threshold-crossing events (TCEs) of the QLP TOI analysis (Guerrero et al. 2021). Paper II (Melton et al. 2024a) describes the rigorous multifaceted vetting procedure applied to the RF classifier results and presents the final list of 462 DTARPS-S candidates. Paper III (Melton et al. 2024b) provides an initial scientific analysis of these candidates.

Our papers are purposefully more detailed than most presentations of transiting exoplanet discoveries. For example, in the present study, we analyze the performance of the RF classifier with respect to planetary injections, astronomically confirmed planets, previously identified planetary candidates and False Positives in the full DIAmante data set. We present a thorough analysis of the performance of the ARPS methodology on injected planetary signals into the DIAmante TESS FFI light curves and the performance of the RF classifier on these synthetic planetary injections.

This level of detail has two benefits. First, it allows us to analyze, and seek to improve, each step of the DTARPS pipeline. Second, it provides a rigorous foundation for science results such as the first TESS-based planet occurrence rate calculation in Paper III.

The present paper is structured as follows. The ARPS methodology, updated from Caceres et al. (2019a), is presented in Sections 2.1–2.3. DIAmante extraction and preprocessing is outlined in Section 2.4. Section 3 discusses the ARIMA modeling and TCF periodogram for detrending and transit search. Sections 4 through 6 describe the RF classifier training set (including creating synthetic injections), the process of RF optimization, and the final RF classifier. The performance of the final RF classifier is presented in Section 6, producing a list of 7377 TESS stars exhibiting transit-like behavior (Section 7). Section 8 discusses the accuracy of the TCF routine for transit fitting. Section 10 compares the results of the RF classifier with other exoplanet surveys, and Section 9 discusses the completeness of the RF classifier. The principal product is the DTARPS-S Analysis List of 7377 TESS light curves in Section 7. The findings are summarized in Section 11, with motivation for the vetting stage described in Paper II. The Appendix describes external data sets used for comparison and validation of this effort.

2. DTARPS-S Methodology

2.1. Background: Detrending and Transit Identification

Any transiting exoplanet search must try to remove a wide range of stellar variability behaviors and variations due to instrumental effects. Detrending procedures used in transit detection outlined in the Appendix include the following: the NASA MIT Quick Look pipeline (Huang et al. 2020), which uses a high-pass filter and fitted splines to detrend the light curve; the `eleanor` pipeline (Feinstein et al. 2019), which cotrends the extracted light curves using Principal Component Analysis (PCA); photometry extraction with difference image analysis (Oelkers & Stassun 2018); the DIAmante pipeline (Montalto et al. 2020) with difference image analysis, PCA cotrending, and individual star spline fits; and for M-dwarf targets, NEMISIS (Feliz et al. 2021), which combines pixel-level decorrelation with an iterative smoother. In the search for transiting planets, a common choice for light-curve detrending is Gaussian Processes regression (e.g., Luger et al. 2016; Angus et al. 2018), but other methods have been tried, such as Independent Component Analysis (Waldmann 2012), correntropy (Huijse et al. 2012), empirical mode decomposition (Roberts et al. 2013), and Singular Spectrum Analysis (Greco et al. 2016).

The analysis pipeline used for transit detection of preselected TESS stars observed with rapid cadence is the NASA–Ames Science Processing Operations Center pipeline (Jenkins et al. 2017b; Guerrero et al. 2021). It involves a complex series of operations including autoregressive filling of gaps, removal of some instrumental systemic effects, whitening with power spectral density analysis, removal of multiscale temporal structures with wavelet analysis, and identification of transiting exoplanet signals with adaptive wavelet-based matched filters. A statistical bootstrap test is applied to the light curve and transit detection, to determine the probability of the event being a False Alarm (Jenkins et al. 2017a).

Following detrending in most analyses, transiting signals are sought with Box Least Squares (BLS) periodograms (Kovács et al. 2002). The BLS method models a transit signal as a simple box shape based on the fraction of duration (duration/period), the transit depth, and the epoch of the transit, and it relies on the anticipated rigid shape of the transit light curve to identify transiting exoplanets. For each period being searched for an associated transit signal, BLS utilizes a least squares algorithm to fit the other box-model transit parameters to the folded light curve and returns the signal residue as the "strength-of-fit" measure to create the periodogram. When handling a large number of observations, BLS bins the folded light-curve data into small bins with respect to the expected transit duration (Kovács et al. 2002). Variants of BLS include the transit least squares (TLS) algorithm that considers the effect of stellar limb darkening during planetary ingress and egress (Hippke & Heller 2019), and the fast BLS computational algorithm (Shahaf et al. 2021).

2.2. AutoRegressive Planet Search: ARIMA and TCF

Stellar activity is often classified as an autoregressive behavior, wherein future photometric values depend on current and past values (Caceres et al. 2019b). For example, the waiting times between X-ray flares have strong temporal autocorrelations on timescales of hours (Wheatland 2000; Aschwanden & McTiernan 2010). Standard nonparametric detrending procedures with a kernel or window—such as running medians, spline fitting, or Gaussian Processes regression—will not remove short-memory stochastic autocorrelated behaviors. But short-memory autocorrelation can be removed with low-dimensional parametric regressions such as ARMA models that are specifically designed to fit stochastic autocorrelated behaviors. ARMA-type modeling is well-established in many fields of time series analysis, with an extensive methodology described in textbooks such as Box et al. (2015), Chatfield & Xing (2019), and Hyndman & Athanasopoulos (2021). Feigelson et al. (2018) argue that these models can be effective for many time domain problems in astronomy.

For transiting planet identification, stationary low-dimensional linear autoregressive models are combined with a simple differencing operator to remove non-stationary trends arising from stellar and instrumental variations. These ARIMA models are flexible low-dimensional parametric models fit by maximum likelihood estimation without any free parameters. The best ARIMA model is chosen by balancing improvement in likelihood with model complexity. Caceres et al. (2019a) show that the shape of a transit is transformed from a box to a double-spike shape by ARIMA modeling, necessitating a new periodogram—called the Transit Comb Filter (TCF)—to identify periodic sequences of double-spike patterns in the cleaned, whitened ARIMA residual light curves. A brief overview of the ARPS methodology is given here; the interested reader will find more details on ARIMA and the TCF periodogram in Caceres et al. (2019b).

2.2.1. Autoregressive Modeling of the Light Curves

The autoregressive moving average (ARMA) model family is very broad and can treat an enormous variety of both stationary and non-stationary time series (Box et al. 2015). ARMA-type models are widely used in signal processing, econometrics, and voice recognition, among many other applications. ARIMA and its extensions like ARFIMA, GARCH, and HAR, are able to treat short-memory processes, long-memory processes, volatility, burstiness, and nonstationarity in the light curve without being subject to choices like a smoothing kernel function or bandwidth in Gaussian process regression or choices of a basis function and denoising threshold in wavelet analysis. ARMA-type models are fit by maximum likelihood estimation without any free parameters while balancing goodness of fit with model complexity.

The linear ARIMA model has three components: autoregressive (AR), integrated (I), and moving average (MA). We start our analysis with the "I" ("Integrated") component that treats nonstationarity; typically, nonstationarity arises from variations in the mean fluxes of the light curve. This operation is described by

$$(1 - B)^d x_t = \epsilon_t, \text{ where } Bx_t = x_{t-1}, \tag{1}$$

and d is the order of differencing, B is the backshift operator, ϵ is a Gaussian noise term, and t is the integer index of an observation in a regularly cadenced light curve. We chose the

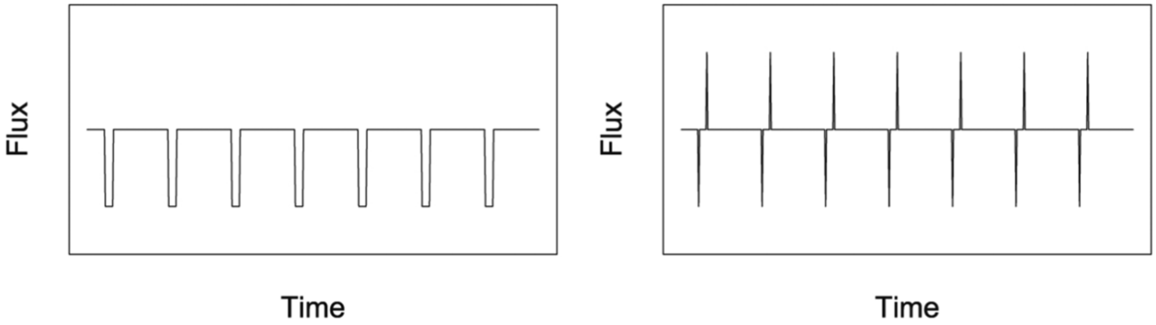


Figure 2. The transformation of the shape of the transit from the original light curve by the differencing step of the ARIMA processing into the double-spike shape (Caceres et al. 2019b).

simplest setting with $d = 1$; experimentation shows that higher-order differencing is not necessary for most TESS light curves. This essentially removes the narrowest possible median filter of the time series, equivalent to the transformation

$$x_t^{\rm dif} = x_t - x_{t-1}. \qquad (2)$$

When the ARIMA model is applied to the light curve, we separate the differencing step (Equation (2)) from the ARMA fit and apply it to the light curve before passing the differenced values to the ARMA function. By separating the difference step from the application of the ARMA model, the transits (if present) are guaranteed to be changed into the double-spike pattern and thereby become detectable by TCF. Caceres et al. (2019b) found that the differencing step greatly reduced the interquartile range (IQR) of light curves with intrinsic stellar variations (such as rotationally modulated starspots) but slightly increases the IQR for light curves without noticeable variation present.

The autoregressive AR(p) portion of the model represents how the stellar flux responds to recent previous values of the flux according to

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + \epsilon_t, \qquad (3)$$

where x_t is the value of the light curve at time t, p is the order of the AR component, and ϕ is a vector of unknown real-valued coefficients with length p. As with most regressions, the error term is assumed to be a homoscedastic Gaussian, $\epsilon = N(0, \sigma^2)$, where the variance σ^2 is another parameter of the model.

The moving average MA(q) portion of the model represents the effects of random shocks to the light curve by modeling the current flux value as

$$x_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}, \qquad (4)$$

where q is the order of the MA component, ϵ_t is the same error as in Equation (3), and θ is a vector of unknown coefficients with length q.

For every possible combination of p and q, the ϕ and θ coefficients are computed using a maximum likelihood estimator. In order to reduce computation time for this step, we restricted p and q to have a sum less than or equal to 10. In practice, this restriction has little effect on the solution. The best ARIMA(p,d,q) model is chosen to balance the accuracy of the ARIMA model compared to the difference light-curve data and the overall model complexity using the Akaike Information Criterion (Sakamoto et al. 1986), a penalized likelihood measure that balances the model complexity and accuracy of fit in a self-consistent manner. It is similar to, but with a different penalty than, the more commonly used Bayesian Information Criterion.

The temporal structure is examined using the nonparametric autocorrelation function for both the original light curve and its ARIMA(p,d,q) residuals. In practice, we find that most TESS ARIMA model residuals have little or no autocorrelation and are consistent with white Gaussian noise.

A critical question is whether the ARIMA model absorbs the planetary signal in addition to stellar and instrumental variations. Because an exoplanetary transit signal, if present, occurs during only a very small fraction of the observations and the number of time steps between transits is larger than our maximal p and q values, it is mostly ignored by the maximum likelihood estimator. We find a bias does occur in the depth of the deepest transits (e.g., inflated hot Jupiters), as the ARIMA model incorporates some of the transit signal. This bias is corrected in a later stage of analysis (Paper II).

2.2.2. Transit Comb Filter Periodogram

The difference step of the ARIMA processing in Equation (2) changes the shape of a planetary transit from a periodic box pattern to a double-spike pattern (Figure 2). The period, depth, duration, and phase of the transit are still available in the transformed light curve. Caceres et al. (2019b) developed the Transit Comb Filter (TCF), a matched filter algorithm that searches over a grid of durations and phases to find the strongest periodic double-spike patterns at a chosen trial period. For a time series of Gaussian white noise, the algorithm is equivalent to the maximum likelihood estimator. A periodogram is constructed from the strength of the matched filter fit to the ARIMA residuals for each period passed to the TCF. The code involves the same triple loop as the traditional Box Least Squares algorithm (Kovács et al. 2002). Gondhalekar et al. (2023) show that the TCF periodogram of ARIMA residuals is typically more sensitive to small planetary transits than the BLS periodogram of residuals from a standard smoother.

As with the BLS periodogram (Ofir 2014), the TCF periodogram can have systematic changes in mean as one passes from short to long periods. We remove this trend with a smoothed locally fitted, robust least squares regression polynomial function—the LOESS algorithm (Cleveland & Devlin 1988). The power of the TCF for a specified period is then measured from the TCF power above the LOESS curve. The peak with the highest S/N in a window around the peak, with respect to the LOESS curve, is chosen as the most likely fit for an exoplanet transit in the light curve (Caceres et al. 2019b).

2.2.3. ARIMAX Model

After the best TCF periodogram peak is found for a light curve, the transit parameters are used to fit a new ARIMA model to the light curve in order to jointly fit the transit and the autocorrelated noise in the light curve. In the parlance of ARMA modeling, this is an ARIMAX model where "X" refers to "exogenous" variables (Hyndman & Athanasopoulos 2021). A simple box transit mask is built for the best peak from the TCF periodogram using the transit period, phase, and duration from the periodogram strongest peak. The depth of the box transit mask is left as a free parameter of the exogenous variable. Therefore, when the ARIMAX model uses maximum likelihood estimation to jointly model the transit depth and the autocorrelated noise of the light curve, it also models a transit depth with a confidence interval (error value). Further details about the ARIMAX modeling are provided by Caceres et al. (2019b) and Caceres et al. (2019a).

We found that the ARIMAX depth tended to underestimate the depth of the transit necessitating astrophysical transit models to be fit to the candidates. Like the bias produced by ARIMA modeling (Section 2.2.1), this has to be corrected in the later stage of DTARPS-S analysis so reliable estimates of the planet radius can be obtained (Paper II).

2.3. Classification and Vetting to Identify Planet Candidates

While a prominent peak in the periodogram is a necessary indicator that a transit-like periodicity is present in a light curve, this alone is not a sufficient criterion. Caceres et al. (2019b) compared the results of classifying based on periodogram strength alone and a machine-learning classifier based on many features of the light curve and periodogram. They found that the machine-learning classifier performed better. The principal reason is not lack of sensitivity to planetary signals by the TC periodogram, but rather the capture of non-planetary signals, in particular BEBs. Classifiers are developed to maximize the number of planets in the identified planet candidates while at the same time minimizing the number of non-planetary objects. But reliance on automated classifiers without human vetting risks statistical False Alarms and a higher percentage of astronomical False Positives in the final data set (Burke et al. 2015).

Two main families of classifiers used in exoplanet transit identification are deep-learning classifiers and decision-tree-based classifiers (Jara-Maldonado et al. 2020), though other types are available. Deep-learning classifiers learn features automatically from training sets of light curves. While training the neural network, the parameters of the linear combination inputs into each hidden layer feature are tuned using a cost function to minimize the classification prediction error. Convolutional neural networks have been developed for Kepler and TESS planet discovery (Ansdell et al. 2018; Shallue & Vanderburg 2018; Yu et al. 2019); these are used in the QLP pipeline leading to the TOI list[1] (Guerrero et al. 2021). Other transit detection groups use decision trees based on extracted features rather than the light curves themselves (McCauliff et al. 2015; Coughlin et al. 2016; Armstrong et al. 2018).

2.3.1. AutoRegressive Planet Search: Random Forests and Human Vetting

The ARPS method utilizes a Random Forest (RF) decision tree classifier to identify the most promising exoplanet candidates from the full data set with a high recall rate, followed by human vetting to further refine the candidate sample. We carefully test different training sets, features, and classifier settings to optimize its performance.

RF machine-learning classifiers were developed by Breiman (2001) as an extension of his earlier Classification and Regression Tree procedure (CART; Breiman et al. 1984). CART classification uses a decision tree that has been grown based on the problem's training set in an iterative procedure. Each node produces two daughter nodes based on a break in a single data feature that best separates the classes according to some cost function. To reduce overfitting, the tree is pruned to a predetermined level. The main drawbacks of CART trees are that they tend to overfit the training data and they often use only a few of the possible features in classification. RF overcomes the disadvantages of CART classification by using multiple CART trees with randomized data subsets and feature subsets at each branching node. This "bagging" strategy avoids overfitting because each tree in the RF sees a different data set and avoids overemphasis on just a few features in the classifier. Whereas a single decision tree produces a "hard" prediction for each object in the test set, an RF gives "soft" or probabilistic predictions arising from votes of many trees. The RF prediction value is a pseudo-probability; higher predictive scores point to more likely exoplanet transiting candidates.

RFs are extremely versatile classifiers that can use data of different types (integer, categories, floating point numbers), units, and scales. RFs have been shown to be robust to imbalanced training set problems, performing well on training sets whose positive class comprises only 2% of the entire training set (Chen et al. 2004) and can handle small fractions of mislabeled data in the training set (Mellor et al. 2015). The contribution of each feature to the classifier allows for RFs to be partially interpretable, whereas most deep-learning classifiers are not readily interpretable.

Vetting the results of a machine-learning classification is necessary to remove lingering False Alarms and False Positives in samples that pass the RF classifier. Although this is time consuming, every object classified as a potential candidate by the RF is examined by human vetters. The vetting procedure employed here, described in Paper II, is a mixture of multifaceted automated vetting tests and subjective vetting by humans.

Due to the rarity of transiting planets in a random sample of stars, the RF classifier and the subsequent vetting procedures must strive to reduce the number of False Alarms and False Positives in the final catalog. For a RF classifier with a False Positive Rate $\gtrsim 1\%$ as measured with a validation set, the number of expected False Positives in sample of a million light curves would overwhelm the planet candidate sample, even if the classifier were to identify every true transiting planet.

2.4. Light-curve Input from the DIAmante Project

DIAmante is a pipeline for extracting and analyzing light curves from the TESS full-frame images (FFIs) developed by M20 and applied to TESS Year 1 FFIs (M20) and Year 2 FFIs (Montalto 2023). M20 used the DIAmante-extracted light

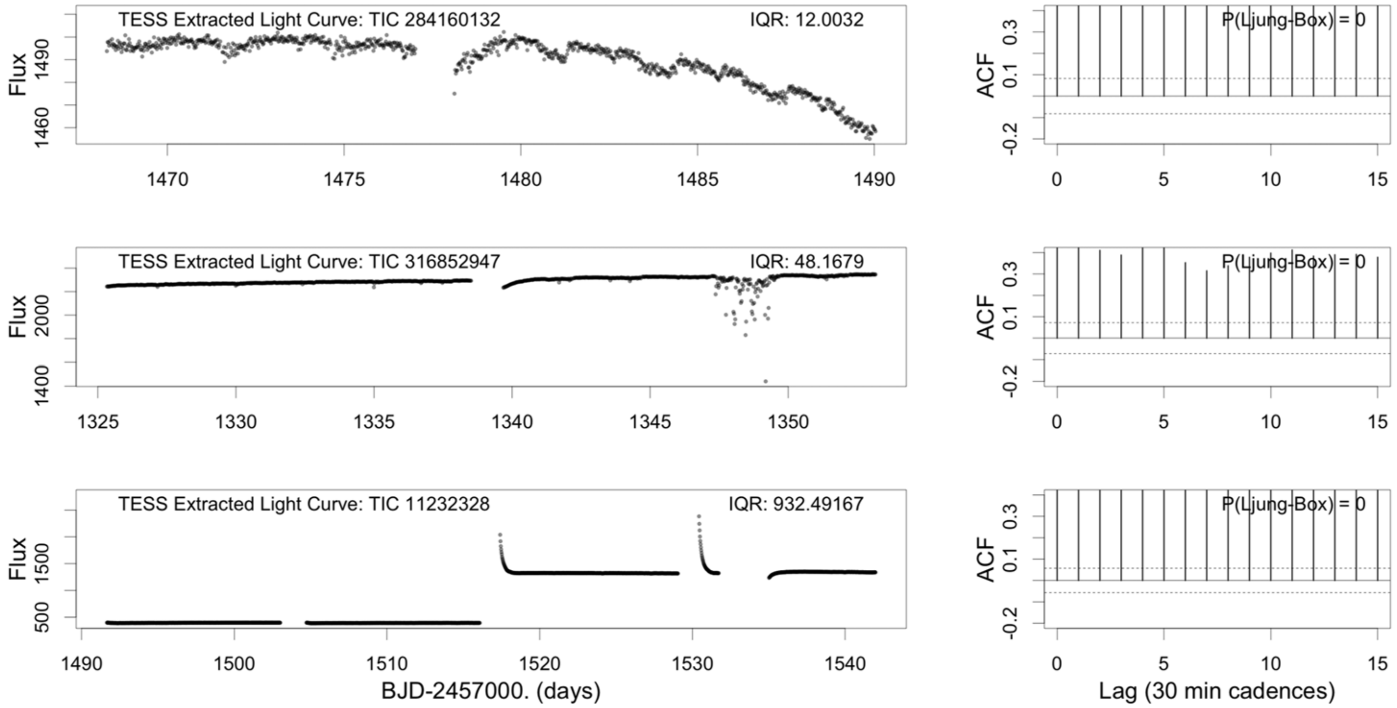


Figure 3. Left: The raw light curves extracted from TESS FFIs for three example stars in the DIAmante data that were later found to have DTARPS-S candidate transiting planets. Right: Plot of autocorrelation present in the light curve as a function of lag in units of the 30 minute FFI cadence. The p-value from the Ljung–Box test is 0 for all three light curves, indicating that the flux values of the light curves are autocorrelated.

curves to search for exoplanet transits and identified 396 exoplanet candidates. We applied the DTARPS-S method to the M20 light curves extracted and preprocessed with the DIAmante pipeline.

M20 defines a sample of 976,814 dwarfs and subgiants with spectral types F5 to M falling in the footprint of TESS sectors 1 −13 surveyed during Year 1 with identifications in the TESS Input Catalog (version 8; Stassun et al. 2019). FGK stars were restricted to $V < 13$ magnitude while M stars were restricted to $V \leqslant 16$ magnitude and distance $D < 600$ pc. The sample is further limited to dwarf and sub-giant stars with $\log g \geqslant 3$.

The DIAmante extraction was applied to the calibrated FFIs available from the Science Processing Operations Center (Jenkins et al. 2016; Tenenbaum & Jenkins 2018). These calibrated FFIs are reduced CCD images have already been processed with TESS instrument specific corrections. The DIAmante extraction pipeline is based on Difference Image Analysis (Alard & Lupton 1998) that reduces the impact of contaminants on the target photometry through the efficient subtraction of a reference image convolved with a kernel to separate the target and the background flux. Because TESS FFIs are known to have erratic background variations that depend heavily on the boresight angle between the camera, Sun, and Moon, a flux-conserving delta basis kernel was utilized to create a differential background model using a 20 pixel box smoothing region. After calibration to 250 standard stars in the reference images for each CCD, photometry was extracted from a circular aperture with a radius of one pixel.

The DIAmante light curves from each CCD for each camera and sector were processed with cotrending to remove systemic variations from the instrument. Principal Component Analysis was applied to the most highly correlated light curves to extract top eigenvectors to cotrend the light curves. After cotrending, individual stellar variations are further detrended with an 8 hr median filter and a B-spline interpolation. Outliers more than twice the interquartile range (IQR) from the median were iteratively removed from the original light curve. The final DIAmante light curves are the averaged values of the B-splines evaluated at each observation time.

Figure 3 shows the raw TESS extracted light curve for three stars in the DIAmante data set prior to any preprocessing. These examples were extracted using the Python Lightkurve package (Lightkurve Collaboration et al. 2018). DTARPS-S identified a new planetary candidate around each of these three examples that had not been identified to date (Paper II). Two of the light curves shown (top and middle panels) are the length of one TESS sector, as are most of the light curves from the DIAmante data set. The bottom light curve panel is a source observed in two sectors because it was in the overlap area between sectors.

The three panels to the right of the light curves in Figure 3 show strong autocorrelation present in the light curves between the 30 minute time steps for the TESS FFIs. The ARIMA fitting in the ARPS procedure is designed specifically to remove autocorrelation in light curves; the presence of autocorrelation is measured with the Ljung–Box test (Ljung & Box 1978). The p-value from Ljung–Box test is included in subsequent figures to indicate the presence or absence of autocorrelation in the light curves. A small p-value indicates that there is significant autocorrelation present in the light curve, and a p-value $\gtrsim 0.01$ means that the light curve is consistent with white noise without autocorrelation.

2.5. Additional Ramping and Outlier Removal

There is a well-known issue in TESS of erroneous flux variations lasting a few hours being introduced to the FFI light curves near the beginning of a sector, the end of a sector, and

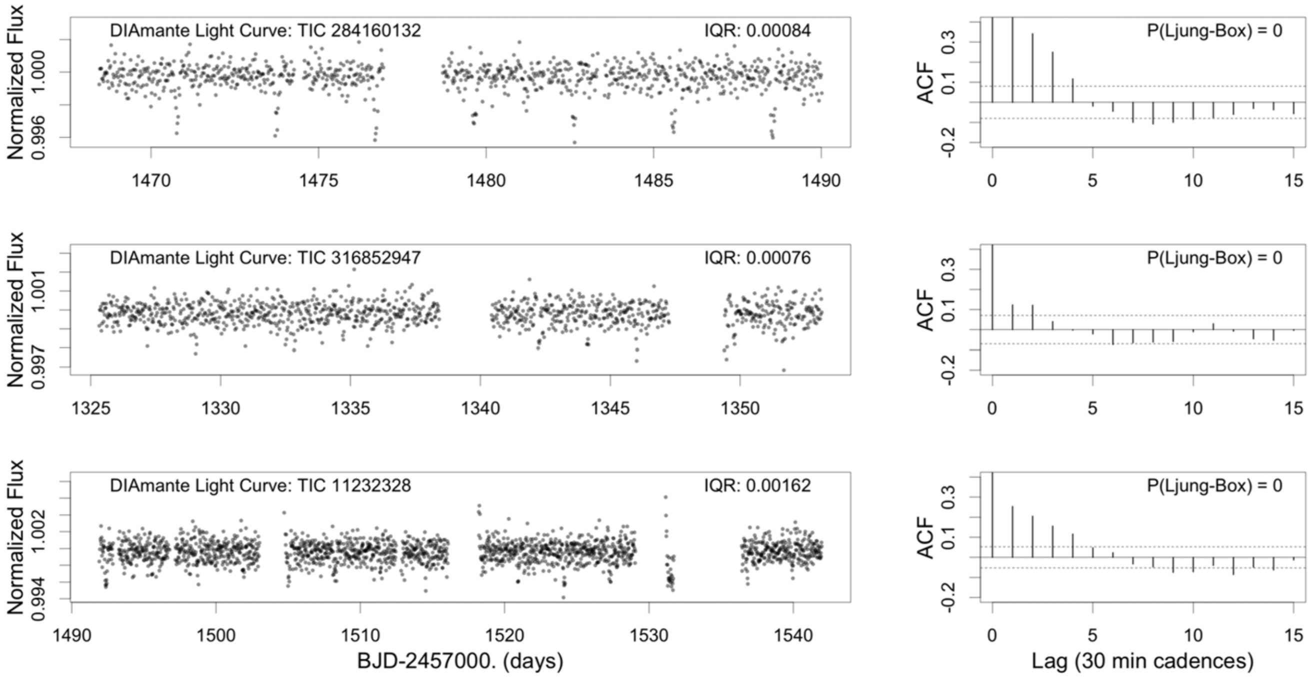


Figure 4. The DIAmante light curves for three example stars in Figure 3 after the removal of trends, ramping effects, and outliers (Section 2.5). The autocorrelation functions on the right show that autoregressive structure is still present in all three light curves.

the beginning and end of the mid-sector gap in the light curve for the data download (Ricker et al. 2015). Due to spacecraft jitter before the pointing settles, the target star can land on CCD areas with different quantum efficiency, or a neighboring star can enter or leave the field, changing dilution levels. The DIAmante pipeline was able to remove most—but not all—of the trends and the jitter effects. There is a weak ramp-up effect seen in the top and middle panels of Figure 3 after the mid-sector gap in the light curves as well as in the second sector of the bottom panel of Figure 3 after the second gap. The bottom panel light curve in Figure 3 has two strong ramps during the second sector. Most, but not all, of these ramps are removed by the DIAmante pipeline-extracted light curves.

We therefore add a preprocessing step to reduce remaining ramps around data gaps, flares, and other outliers that may be in the data. The clipping routine is characterized by the outlier threshold and the gap threshold. The outlier threshold defines the maximal distance between the median value of the light curve and a data point; we set the threshold at five times the standard deviation of the light curve. The gap threshold defines how large a gap of missing data can be, in time steps, before the clipping routine will examine the points on either side of the gap for evidence of ramping. We set the gap threshold to be 50 time steps, or 25 hr. With this gap threshold, the clipping routine removes erroneous ramping points from the beginning and end of the light curve, as well as points leading up to and away from a large gap in the light curve. After removal of a data point, the clipping routine is recursively applied to the modified light curve until no more points are removed as outliers.

Figure 4 shows the DIAmante light curves in Figure 3 after DIAmante detrending and our additional ramping and outlier reduction procedure. The ramping procedure removed most of the ramping in TIC 11232328 left by the DIAmante extraction pipeline. A weak residual ramping effect around Day 1531 remains (Figure 4). However, such brief and weak effects will have little effect on our transit search algorithm (Section 2.2.2). None of these three transiting objects have been identified previously in the literature, despite each of them having strong transit signals in the DIAmante-extracted light curves.

3. ARIMA Modeling and Periodogram Analysis

The right panels of Figure 4 illustrate that short-memory autocorrelation is often still present in the preprocessed light curves, even though the noise level is often reduced below 0.1%. This autocorrelated behavior can increase noise in periodograms such as Box Least Squares (BLS) and thus reduce sensitivity to weaker planetary transits (Gondhalekar et al. 2023). Some of autocorrelation variations may be due to planetary transits, but typically it is intrinsic to the star. Transits from large planets can be seen with the unaided eye in Figure 4.

3.1. Autoregressive Modeling of the Light Curves

When the ARIMA model is applied to the light curve, we apply the differencing operation (Equation (2)) first and then obtain fits from the ARMA model (Equations (3)–(4)). This guarantees that any real transits will be changed into the double-spike pattern and thereby be detectable with TCF. Figure 5 shows the differenced light curves for each of the three example light curves in Figure 4 along with the autocorrelation panels off to the right of the light curves showing the amount of autocorrelation in the light curve as a function of the lag in time steps. The p-value from the Ljung–Box test is 0, which is expected due to the negative correlation at lag = 1 induced by the differencing operation. Applying the full ARIMA model to the differenced light curve will remove this autocorrelation introduced at lag = 1 cadence.

Caceres et al. (2019b) found that the differencing step greatly reduced the IQR of light curves with intrinsic stellar variations, but slightly increased the IQR for other light curves. In our case, the DIAmante pipeline removes most of the light

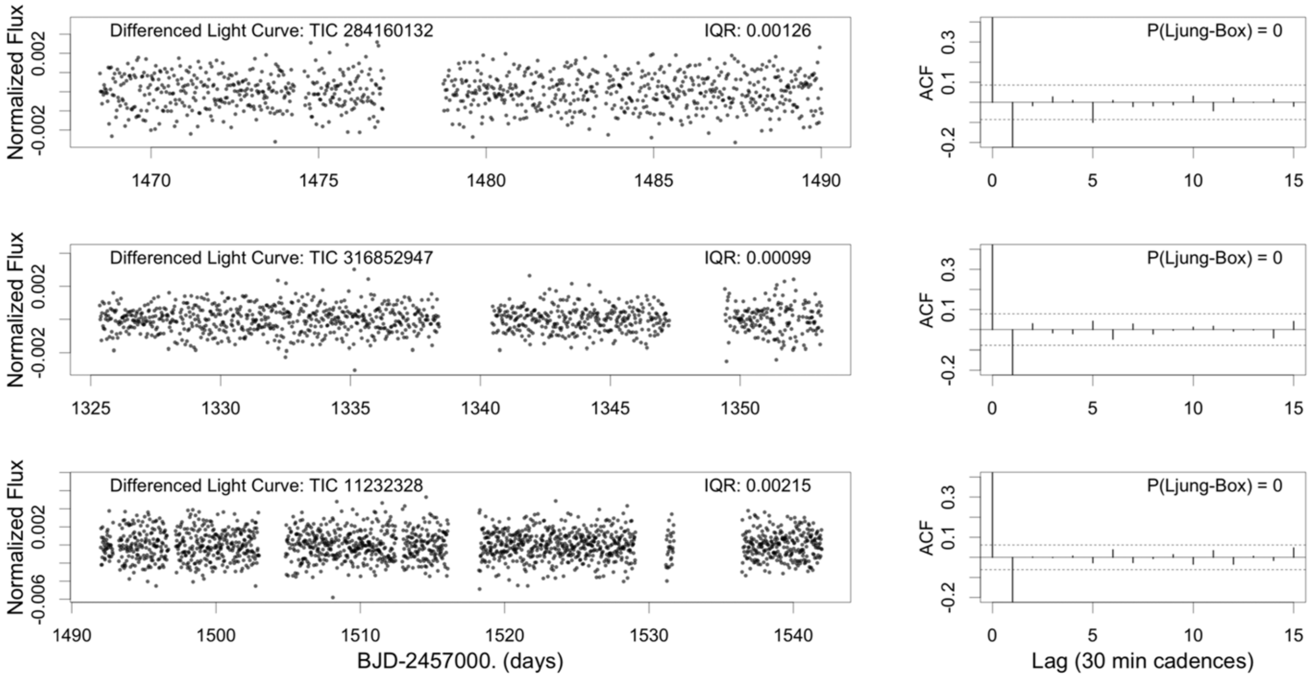


Figure 5. The differenced light curves for three example stars in Figure 3 with autocorrelation functions.

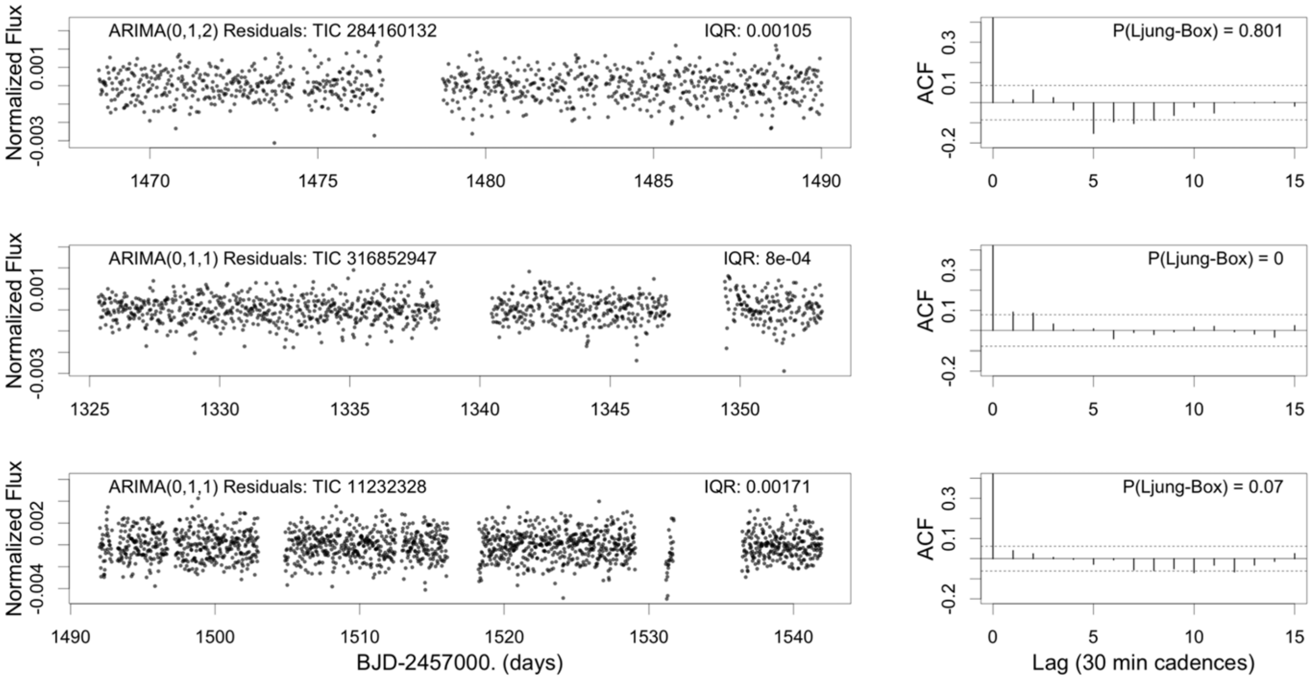


Figure 6. The residuals after the best-fit ARIMA model has been subtracted from the differenced light curves for three example stars in Figure 3. The three plots on the right show the amount of autocorrelation present in the light curve as a function of time step between points. The Ljung–Box test shows that two of three ARIMA residuals are consistent with Gaussian white noise.

curve trends before we receive them. Therefore, it is not surprising that the IQR of the light curve does not change much over the course of the ARIMA processing.

We implement the ARIMA model using the *auto.arima* function from the *forecast* package (Hyndman & Athanasopoulos 2021) in the statistical computing language R (R Core Team 2020). Figure 6 shows the residuals after the best-fit ARMA model has been subtracted from the differenced light curve for the three example light curves in Figure 3. The p and q values for the best-fit ARIMA models are given in each of the three panels where it states "ARIMA(p, d, q)." The autocorrelation function of the residuals is often consistent with Gaussian white noise with associated Ljung–Box test p-value >0.01. The improvement in Ljung–Box probabilities for the full DIAmante sample is shown in Figure 7: 46% of the light curves from the DIAmante pipeline have significant autocorrelation present in the light curve, while only 4% have autocorrelation after ARIMA modeling.

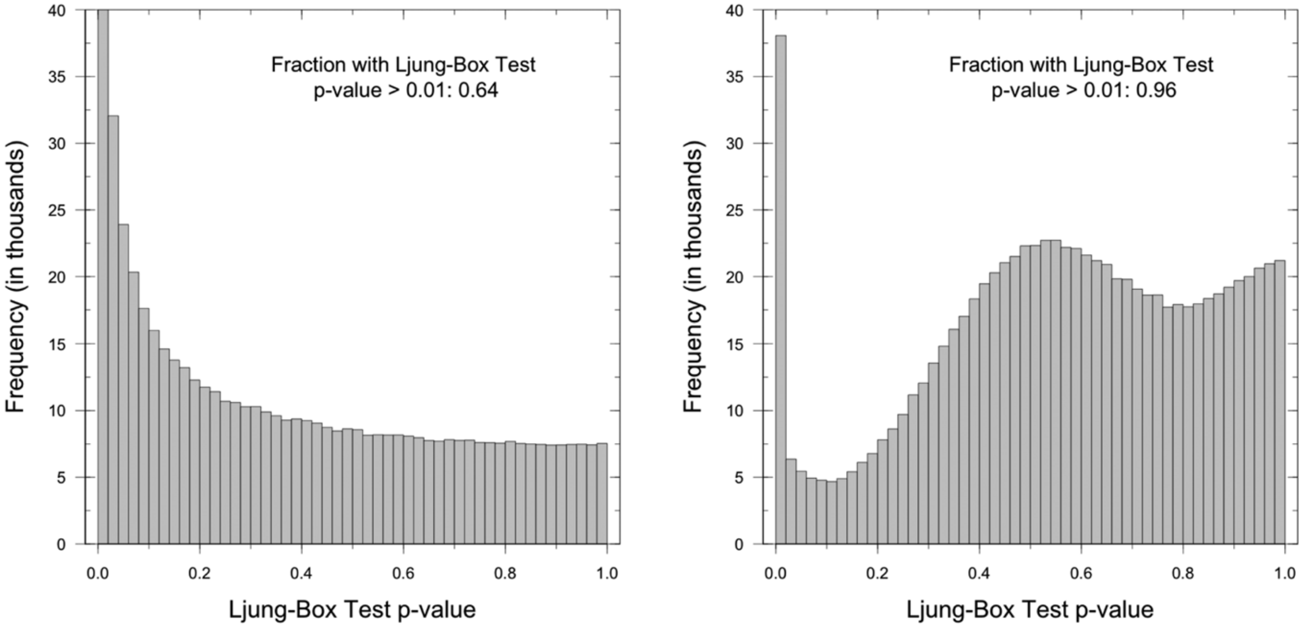


Figure 7. Distribution of p-values from the Ljung–Box test for the light curves extracted from the DIAmante data set (on the left) and for the ARIMA residuals (on the right). The first bin in the histogram of Ljung–Box test p-values in the DIAmante data (left) has a frequency of 318,227.

The residual light curves in Figure 6 now exhibit periodic double-spike patterns characteristic of a transiting planet but without stellar or instrumental autocorrelated behavior. In most cases, the DIAmante preprocessing, outlier removal, and ARIMA modeling together successfully remove the structure present in the TESS light curves, except for transits or other brief non-stochastic behaviors.

3.2. Transit Comb Filter Periodogram

The TCF algorithm described by Caceres et al. (2019b) is coded in Fortran for computational efficiency and is called from our *R*-based DTARPS pipeline. The Fortran program is available in the Astrophysics Source Code Library (Caceres & Feigelson 2022). The periods tried by TCF in its matched filter algorithm were restricted to periods between 0.2 and 30 days. The TCF algorithm finds the optimal phase, duration, and depth for each period passed to the algorithm, so the periods were restricted such that they covered the length of an entire sector with a 3 day pad so that longer-period planets could be identified in multisector light-curve data. Thirty days is long enough that the TCF periodogram would be able to fully characterize the shape of a high-powered peak near 27 days (the length of a single sector) and allow any peak near 27 days to be correctly sorted as either a genuine peak from a transit signal match in the data or noise from a multisector light curve being folded by the length of a sector. The lower limit of 0.2 days was chosen to facilitate the search for extreme ultra-short-period exoplanets. It is just larger than the shortest reported period for a confirmed planet in the NASA Exoplanet Archive (as of 2022 March 15), K2-137 b with a period of 0.179 days. The 354,982 periods passed to the TCF search algorithm were chosen to be evenly distributed in log-space. The durations looped over for each period were limited to a range from a minimum of 15% of the period to a maximum of 25 hr or 50 time steps (Caceres et al. 2019b).

Figure 8 shows the resulting TCF periodograms and phase-folded light curves for the three stars in Figure 3. The periodograms in the top panels show the power of the best transit fit in phase, duration, and depth as a function of the period investigated. The LOESS curve is plotted in red to allow removal of any trend in the periodogram noise values. The period that was used to fold the light curve and the ARIMA residuals was chosen by the peak in the periodogram with the best S/N in a window of 10,000 periodogram values to either side of the peak. (The S/N reported in the TCF periodogram refers to the S/N of the periodogram peak, not the S/N of the transit depth as is often seen elsewhere.) The typical best S/N of a TCF peak seen in stars without planetary transit signals is typically between 9 and 13, much lower than the peak S/Ns between 43 and 75 in Figure 8. The potential DTARPS-S candidates identified by the RF classifier have top peak S/Ns typically between 17 and 55 and the top peak S/Ns for our final DTARPS-S candidates are between 32 and 71.

The bottom two panels below each of the TCF periodograms in Figure 8 show the original DIAmante light curve and the ARIMA residual fluxes plotted modulo the TCF peak period. The phase is adjusted so the transit is centered at phase 0.5. The phase-folded ARIMA residuals shows the double-spike shape that the transit shape was transformed into due to differencing step of the ARPS processing. The double-spike shape is clearly seen in the folded ARIMA residuals. TCF ran its nested loop matched filter algorithm on the ARIMA residuals light curve shown in the bottom right panel, but the transit is more intuitively identified by human vetters as a box shape in the bottom left panel showing the folded DIAmante light curve.

Ancillary information from the TESS Input Catalog is provided in the TCF periodogram panels. TIC 284160132 (DTARPS-S 548 in Paper II) is an early-G star with $V = 12.2$ mag at a distance 384 pc. DTARPS-S identified periodic dips consistent with a gas giant with orbital period 2.96217 days (Paper II). Alias peaks at the double period (∼6 days) and half period (∼1.5 days) are easily seen in the TCF periodogram. TIC 316852947 (DTARPS-S 604) is a mid-F star with $V = 11.6$ at a distance of 341 pc. DTARPS identified a transit consistent with a Neptune-sized object with

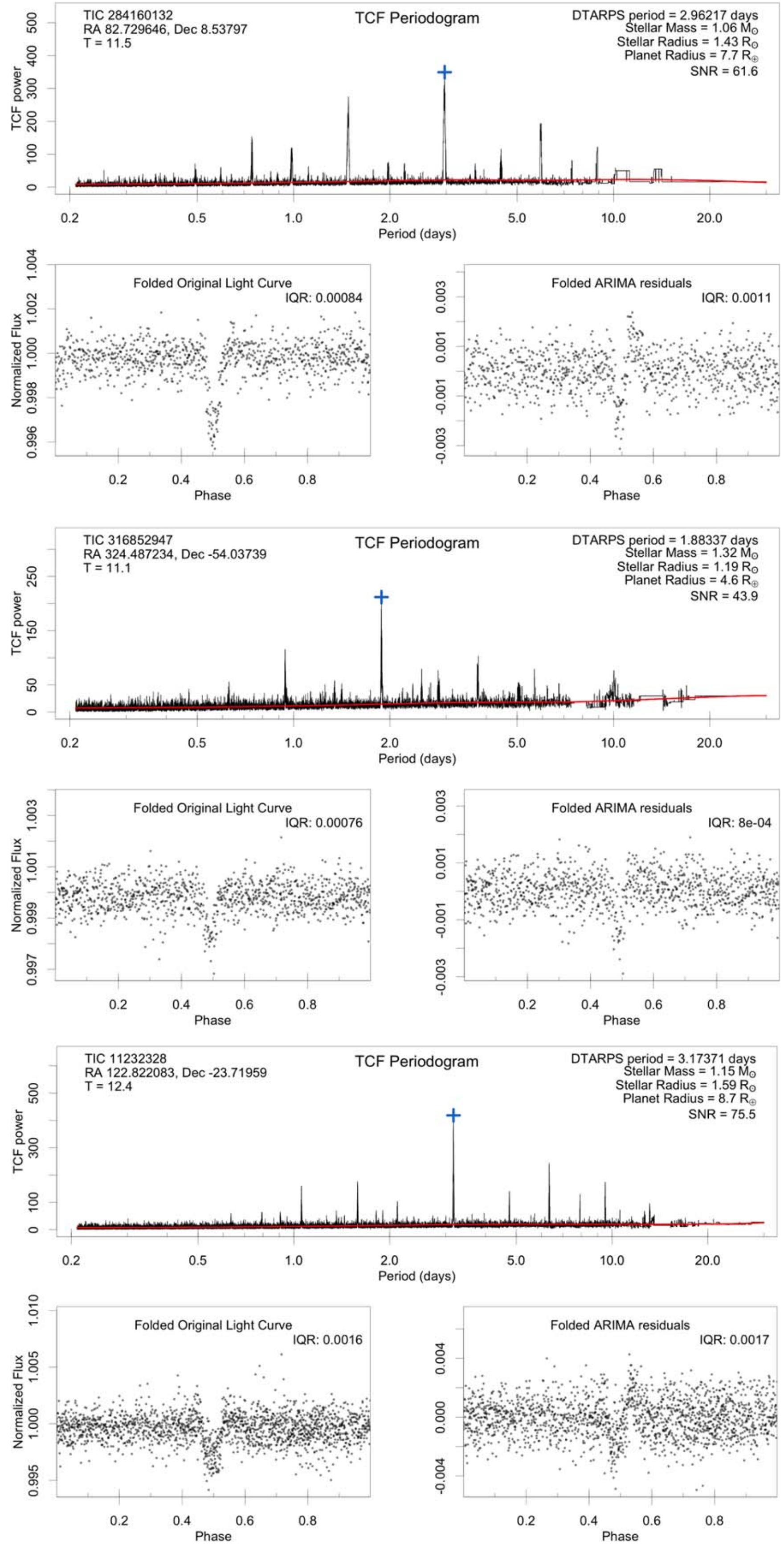


Figure 8. The TCF periodograms and best period phase-folded light curves for the stars in Figure 3. The red curve is the LOESS fit to trends in the median of the periodogram. The blue cross indicates the peak with the highest S/N over a window around that peak. The lower left panel shows the original light curve phase-folded on the best TCF period. The lower right panel shows the residuals after the best-fit ARIMA model has been subtracted from the light curve phase-folded on the parameters from the best TCF periodogram peak. It should be noted that the ordinate scales of the two folded light curves differ.

a period 1.88337 days. TIC 11232328 (DTARPS-S 25) is a fainter, late-F star with $V = 12.9$ mag at a distance of 714 pc. DTARPS-S identified a transit consistent with a Saturn-size planet orbiting with period 3.17371 days.

3.3. ARIMAX Model

The ARIMAX model is run using the *auto.arima* function from the R package *forecast* (Hyndman & Athanasopoulos 2021). We often found that the ARIMAX model underestimated depths for the TESS DIAmante light curves, probably because the ARIMA model incorporated some of the transit signal. We did, however, use the transit depth error as a feature in the RF classifier (Section 6).

4. Random Forest: Training Set

In human affairs, training sets for machine-learning procedures are often well-defined; e.g., photographs of "dogs" versus "cats." But in astronomical applications, there is often considerable flexibility in defining the training sets, and these choices can be a dominant contributor to classifier performance. For binary classification (planet versus non-planet), an RF classifier requires labels for both positive training examples (light curves with exoplanet transit signals) and negative training examples (light curves without exoplanet transit signals). Following M20, we introduce simulated "injected" False Positive signals into the negative training sets to steer the classifier away from BEBs. Section 4.1 describes the positive training sample with injected exoplanet transit signals and Section 4.2 explains the negative training sample with injected eclipsing binary (EB) signals. Section 4.3 describes how the two samples were combined into a training set and a validation set for Random Forest classification.

4.1. Kepler-based Planet Injections

The injected transit signals are drawn from the Kepler 4 yr mission exoplanets that can be considered to be an unbiased sample of the true planetary occurrence rate for the shorter-period exoplanets and higher radii that TESS is designed to identify during its prime mission. The Kepler 4 yr mission exoplanet sample was acquired from the NASA Exoplanet Archive (accessed 2021 March 14). We confine injections to stars with a Kepler identifier that have a confirmed planet with $P < 13.5$ days (allowing at least two transits during a 27 days TESS sector exposure). Following the finding of Caceres et al. (2019b) that roughly half of the "confirmed" Kepler Objects of Interest (KOIs) with low Kepler Model.SNR were not recovered with ARPS analysis, we removed KOIs with Kepler Model.SNR < 20. Of the 2356 Kepler confirmed planets, 949 are suitable for injections.

The left panels of Figure 9 show the distribution of transit parameters period, duration, and depth for these 949 confirmed Kepler planets in the left column. Most of the transit signal depths from the Kepler sample are below 1000 parts per million (ppm), corresponding to a planetary radius of $\sim 3.5 R_\oplus$ for a Sun-like star. Only 62 gas giant planets are among the 949 Kepler planet sample with planetary radii $> 8 R_\oplus$.

If only TESS-detectable planets are considered, this training sample is too small for viable training of an RF classifier for an imbalanced classification problem. We therefore augmented the sample of 949 Kepler planets with synthetic exoplanets sharing the same distribution of transit parameters. This allowed us to inject thousands of DIAmante light curves with planetary transit signals without reusing any set of transit parameters, preventing our training set from being biased by a few Kepler planets. A modified version of the widely utilized Synthetic Minority Oversampling Technique (SMOTE; Chawla et al. 2002) was applied to the 949 Kepler planets in order to create synthetic exoplanet transit sample. Specifically, we used the Adaptive Neighbor SMOTE (ANS) described in Siriseriwan & Sinapiromsaran (2017) with code implementation in the CRAN package *smotefamily* within the R statistical software environment (Siriseriwan 2019).

The SMOTE algorithm selects a random instance in the minority class and finds the k-nearest minority class neighbors in feature space. The parameter k is set by the user with a common default value of five. SMOTE then randomly selects one of the k-nearest neighbors and generates a synthetic minority class instance by randomly choosing a point along the line between the original minority class instance and the neighbor minority class instance in feature space (Chawla et al. 2002). ANS modifies SMOTE by removing the need for a user-set parameter k and finds an optimal k for each instance of the minority class based on the maximum distance between the minority class points and their nearest minority class neighbor in feature space, η. For each instance of the minority class, $_{pi}$, k_i is set to the number of minority class instances within η of $_{pi}$. The regular SMOTE method of generating synthetic minority class members is then executed using the assigned k_i value for each minority point until the desired number of synthetic minority class instances have been generated (Siriseriwan & Sinapiromsaran 2017).

We used the ANS SMOTE procedure to create a sample of periods, durations, and transit depths for 10,850 synthetic exoplanets shown in the right panels of Figure 9. The ANS algorithm preserves the distribution of transit parameters and does not extend the sample of points beyond the extreme values of the distribution. For instance, like the Kepler planet sample, $\sim 6\%$ of the synthetic exoplanet transits had depths consistent with gas giant planets.

A well-sampled exoplanet transit signal will also have an ingress and egress duration as a transit parameter. We assume the ingress/egress time for the exoplanets ranges from ~ 5 minutes for smaller exoplanets (Earths to sub-Neptunes) up to ~ 30 minutes for hot Jupiter exoplanets. Our injection model is a trapezoidal shape with straight-line ingress, duration, and egress. Given that the TESS Year 1 FFI cadence is 30 minutes, the ingress and egress are instantaneous in most cases.

We injected the 10,850 planetary signals into 6506 random light curves from the DIAmante light curve sample. Unlike other studies that inject planetary transit signals into the pixel data from the instrument (e.g., Christiansen et al. 2020), we inject the transit signal into the light curve data after DIAmante preprocessing but prior to any DTARPS-S processing. This focuses the analysis on the efficiency of DTARPS-S to identify planetary transit signals in the light curve data set without including instrumental and light-curve extraction effects in our analysis.

The light curves that received planetary injections were selected randomly from the sample of ~ 0.9 million DIAmante stars. The following cuts were made using the TIC stellar radii and effective temperatures to avoid subgiants: $T_{\rm eff} < 4750$ K and $R < 1 R_\odot$, $4750\,{\rm K} \leqslant T_{\rm eff} < 5250$ K and $R < 1.125 R_\odot$,

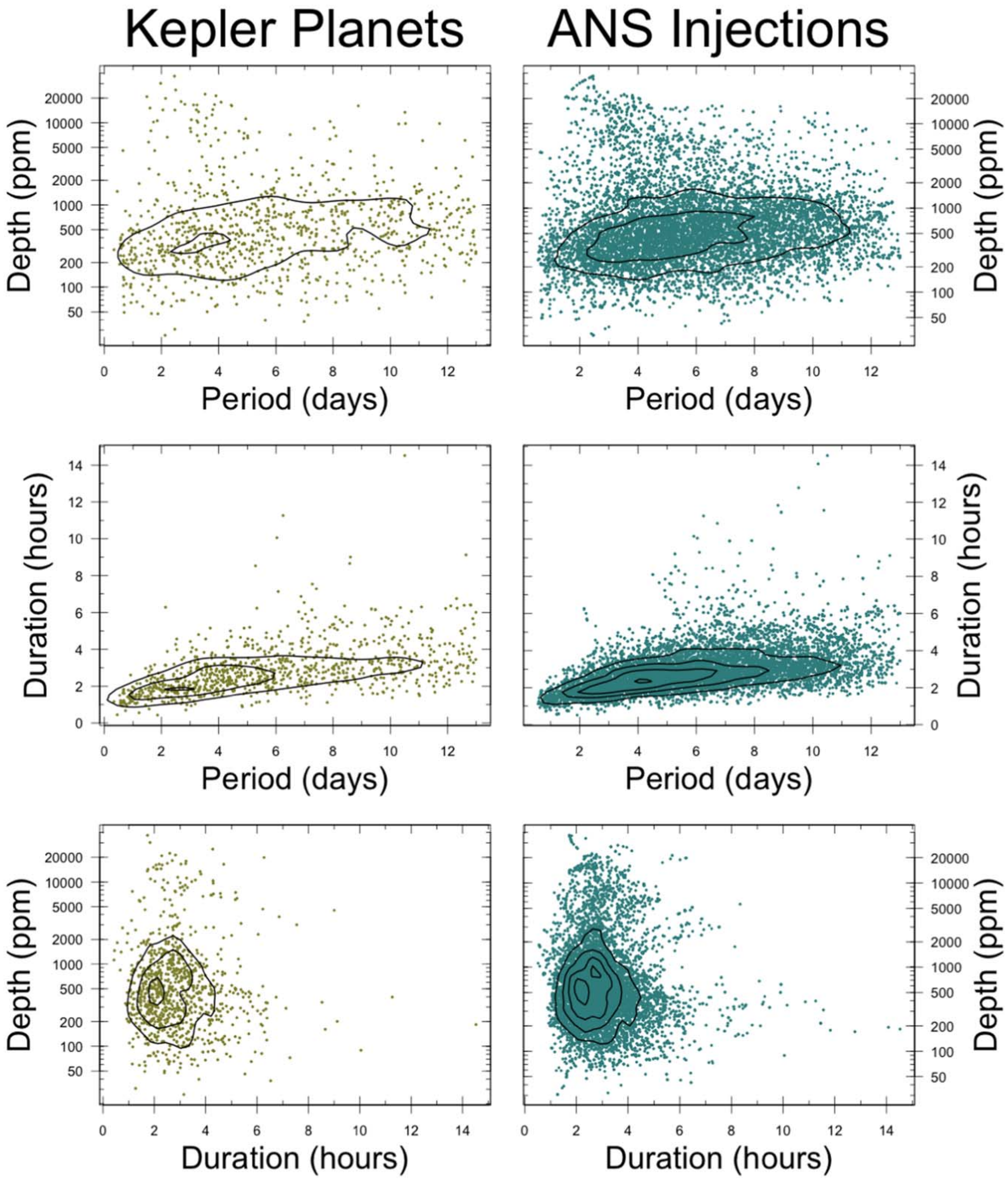


Figure 9. Scatter plots of the transit parameters of the 949 confirmed Kepler planets from the NASA Exoplanet Archive (NASA Exoplanet Science Institute 2022), accessed 2021 March 14, plotted in olive green in the left column. The distribution of the synthetically created exoplanet transit parameters created using Adaptive Neighbor SMOTE (ANS) on the 949 Kepler planets are plotted in teal in the right column. Smoothed contours have been included to better illustrate the distributions.

$5250\,\mathrm{K} \leqslant T_{\mathrm{eff}} < 5600\,\mathrm{K}$ and $R < 1.325 R_{\odot}$, $5600\,\mathrm{K} \leqslant T_{\mathrm{eff}} < 5900\,\mathrm{K}$ and $R < 1.45 R_{\odot}$, $5900\,\mathrm{K} \leqslant T_{\mathrm{eff}} < 6200\,\mathrm{K}$ and $R < 1.55 R_{\odot}$, $6200\,\mathrm{K} \leqslant T_{\mathrm{eff}} < 6500\,\mathrm{K}$ and $R < 1.65 R_{\odot}$, and $6500\,\mathrm{K} \leqslant T_{\mathrm{eff}}$ and $R < 1.7 R_{\odot}$.

Figure 10 illustrates the injection process for a planetary injection. The top panel shows the DIAmante-extracted TESS FFI light curve for TIC 398441407, a $V = 12.1$ G2V star radius of 1.1 $R_{\odot}$. The middle panel shows the modeled transit that was injected into each light curve, characterized by the transit period, depth, and duration. The phase of each transit was chosen randomly. Sometimes the depth of the injection signals may appear to vary due to the transit jittering with respect to the

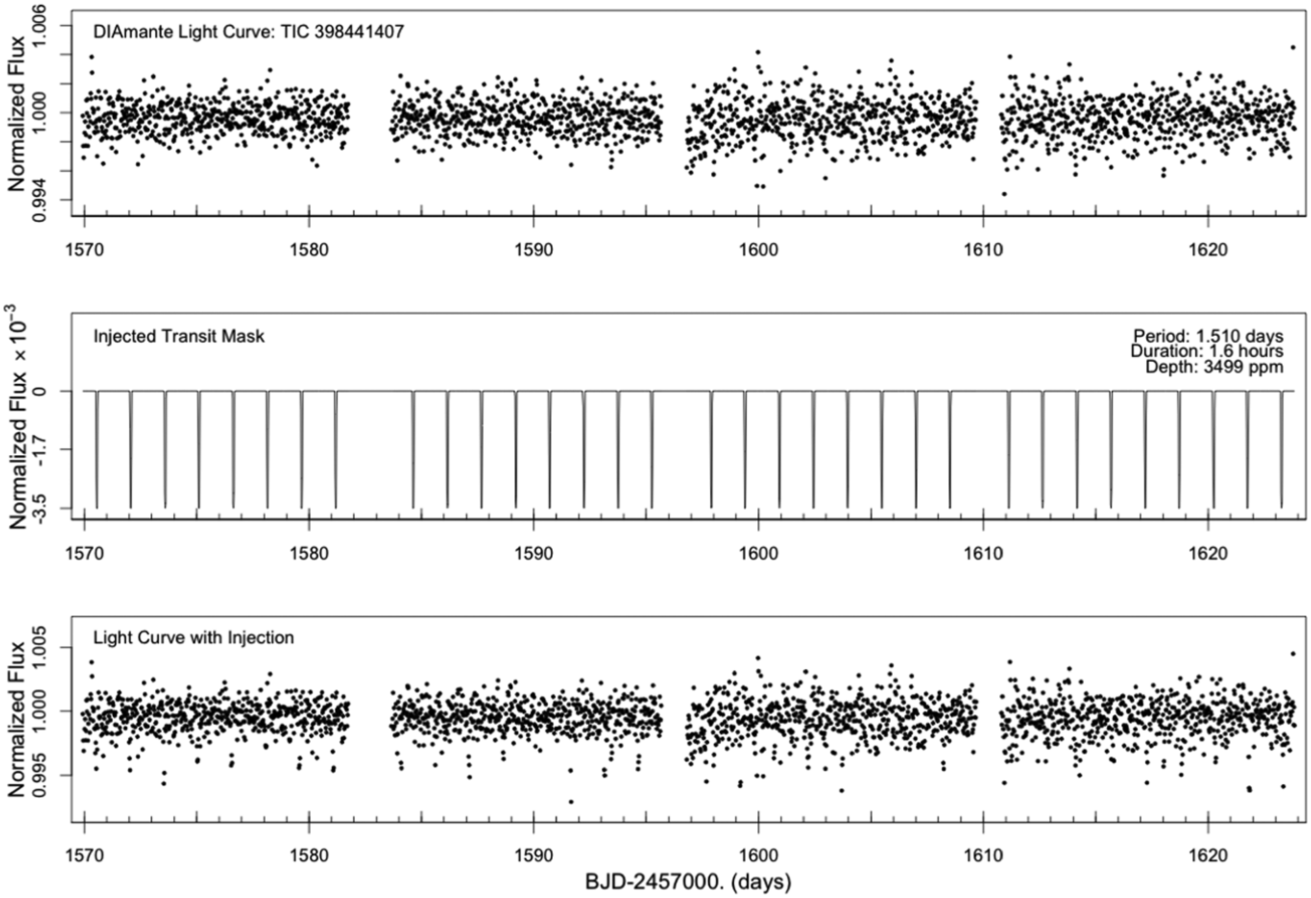


Figure 10. The steps for injecting a synthetic transit signal into a random DIAmante light curve. The top panel shows the original DIAmante light curve. The middle panel shows the transit mask created for a planetary injection. The bottom panel shows the final injected light curve.

30 minute time step cadence. The third panel shows the resulting light curve after the flux values associated with the modeled transit were added point-wise to the flux values of the DIAmante light curve.

After the injected light curves had been modeled with a best-fit ARIMA model and analyzed by TCF, the resulting TCF periodograms were vetted by visual inspection to identify injected transit signals that had been successfully recovered by TCF. The injected transit signal was only considered recovered if the orbital period of the peak TCF period matched the injected orbital period (or an integer ratio of the injected orbital period) within 1% and a transit dip in the folded light curve was visible. Smaller injected planets were more likely rejected by human vetting due to the lack of a visibly discernible transit. Of the 10,850 injections, 1327 synthetic injections were recovered for use as the positive training set for the RF classifier.

4.2. Negative Training Set

Identifying transiting exoplanets is inherently a highly imbalanced classification problem, as most planets have unsuitably inclined orbits or are too small for transit detection. However, the RF technique is well-adapted to this situation; Chen et al. (2004) showed that RF classifiers can perform well with a positive training set that is as small as 2% of the entire training set. We chose to a ~20:1 ratio of negative to positive training set sizes with 26,953 light curves without injected planetary signals compared to 1327 light curves with planets.

The negative training set of the RF classifier should be made up of light curves with no transiting exoplanet signals. However, it is infeasible to manually vet and remove transits from this large negative training set. From the Kepler survey, Howard et al. (2012) found that the expected planet occurrence rate of exoplanets with radii 2–32$R_\oplus$ and periods less than 10 days for GK dwarfs is 0.034 ± 0.003. Because most of these have inclined orbits without transits, an unvetted random sample of TESS light curves thus suffers negligible contamination from transiting exoplanets. In any case, RF classifiers have been shown to perform well when a small fraction of their training set has been mislabeled (Mellor et al. 2015).

BEBs comprise one of the largest sources of expected False Positives (FPs) in an exoplanet transit survey. In order to push the classifier away from labeling eclipsing binary (EB) transit signals as an exoplanet transit signal, we follow the procedure of M20 by supplementing the negative training with the injected FP light curves. The injected FP light curves in M20 are made up of injected EB transit signals corresponding to secondaries with radii larger than 2.5 R_J in both circular orbits and eccentric orbits. Short-period sinusoidal signals mimicking rotationally modulated spotted single stars are also injected. The entire set of injected FP was separated evenly between EBs with a circular orbit, eccentric EBs, and sinusoidal variables. The injected FP signals were not vetted to see if the FP signal was recovered after the ARIMA processing and the TCF analysis, because the classification and characterization of FPs is not the goal of the classifier. Therefore, all of the injected FP signals were included in the negative training set. We used 11,342 injected FP light curves and 15,611 random light curves in our negative training set.

4.3. Training and Validation Sets

The full set of labeled objects used for training is split into a training set for the RF classifier and a validation set to measure the performance of the RF classifier on a set of labeled data. We reserved 20% for the validation set, chosen at random, leaving a training set of 1048 injected exoplanet signals, 9095 injected FP signals and 12,475 random light curves. The validation set holds 279 injected exoplanet signals, 2247 injected FP signals, and 3136 random light curves.

5. Random Forest: Optimization

5.1. Training RF Classifiers

In order to optimize the final RF classifier, we trained thousands of trial RF classifiers with different parameters. Trials examined different combinations of feature selection, feature weights, and algorithmic options to maximize the performance of the RF classifier on the validation set. The number of features to try at each node was left at the default value of seven, but instead of testing a set number of splits in the data for the node features, the optimal split for each node was found. Because the training set is highly imbalanced (Section 4.3), the balanced Random Forest option is used (Chen et al. 2004; Ishwaran 2022). Balanced RFs compensate for an imbalanced training set by undersampling the majority class for each tree in the RF classifier so that each tree is grown using a balanced subsample of the full training set. The number of trees in the forest was varied from 500 to 1000. The RF analysis was performed using CRAN package *randomForestSRC* (Ishwaran 2022) implemented by public domain *R* statistical software environment (R Core Team 2020).

The feature selection and weights were the focus of our tuning parameters to build an RF classifier with the best possible performance. Over 100 features were gathered from every stage of the ARIMA and TCF analysis as in Caceres et al. (2019b). Features describing the light curve were extracted from the light curve, the differenced light curve, and the residuals of the light curve after the best-fit ARIMA model had been subtracted. Features from the TCF analysis included the features from the top 100 peaks of the TCF periodogram as well as features from the peak with the greatest peak signal-to-noise ratio (S/N). Features were also created for the light curve folded according to the parameters from the best TCF periodogram peak. Stellar metadata from the TIC v8 (Stassun et al. 2019) and the Gaia DR2 (Gaia Collaboration et al. 2016, 2018) were gathered for each light curve. Finally, two features that were of high RF feature importance in M20 were calculated for all of the light curves. We examined a wide variety of features because RF classifiers are data-driven classifiers, not physically motivated classifiers. The features that are useful to the RF classifier may not have physical interpretations (Genuer et al. 2010).

The optimal RF classifier was found by creating a large number of test RF classifiers, each with 500 trees, to test different combinations of features. The RF classifiers whose Area Under Curve (AUC) for the Receiver Operating Characteristic (ROC) curve was greater than 0.9 and the AUC for the Precision–Recall curve was greater than 0.85 were kept for further consideration. The features from the top-performing RF classifiers were then combined with different feature weights and a new batch of RF classifiers were grown. Altogether, approximately 20,000 classifiers were considered. The RF classifiers with the same criteria for the AUCs of the ROC and the Precision–Recall curves were retained and the optimization process was repeated three more times, creating a smaller number of RF classifiers each time in order to narrow down the feature and feature weight choices and thus find the optimal combination. In the final two rounds of optimization, the number of trees for each RF classifier was raised to 1000 trees.

In the last round of optimization, we added 133 random candidates from M20 to the validation set and used the True Positive Rate (TPR) of the M20 candidates to help make the final classifier decision. The final RF classifier that we chose had the highest AUC of the ROC and AUC of the Precision–Recall curve along with the highest recall rate of the 133 random M20 candidates at the threshold for the maximal Youden's J index.

5.2. Classification Metrics

To evaluate classifier performance, Akosa (2017) describes several classification metrics appropriate for machine-learning problems with an imbalanced training set. Criteria for selecting the best classifier are based on scalar classification metrics and the AUC for the ROC curve and the Precision–Recall curve.

The output of an RF applied to a new data point is a prediction value between 0 and 1. The prediction value is not a probability value, but it can be considered a pseudo-probability that the input to the RF belongs to the positive class (in this case, a light curve with an exoplanet transit signal). After the classifier is applied to a validation set, we must set a classifier threshold to convert this "soft" classification pseudo-probability to a "hard" classification to produce a confusion matrix of True Positives, False Positives, True Negatives, and False Negatives. The threshold of the classifier can be placed anywhere between 0 and 1.

There is no rule for guiding the choice of the threshold for a classifier other than *ex post facto* performance metrics like the Matthews Correlation Coefficient (MCC), Youden's J Index, and adjusted F-score (Powers 2011). These are defined as follows:

$$\mathrm{MCC} = \frac{\mathrm{TP}\times\mathrm{TN}-\mathrm{FP}\times\mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}}, \tag{5}$$

$$\text{Adjusted } F\text{−Score} = 5\sqrt{\frac{\mathrm{TP}\times\mathrm{TN}}{(5\,\mathrm{TP}+4\,\mathrm{FP}+\mathrm{FN})\times(5\,\mathrm{TN}+4\,\mathrm{FP}+\mathrm{FN})}}, \text{ and} \tag{6}$$

$$\text{Youden's J Index} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} - \frac{\mathrm{FP}}{\mathrm{TN}+\mathrm{FP}}, \tag{7}$$

where TP is the number of true positives (exoplanet injections above the RF threshold), TN is the number of true negatives (negative validation set objects below the RF threshold), FP is the number of False Positives (negative validation set objects above the threshold), and FN is the number of False Negatives (exoplanet injections below the RF threshold).

MCC is the correlation coefficient between the labeled test data set and the predicted labels for the validation set. It can have values between −1 and 1, with 1 corresponding to a perfect classifier, 0 indicating a random classifier, and -1 corresponding to the worst possible classifier. The adjusted F-score, ranging from 0 to 1, is an improvement to the normal F-score that balances classifier recall and precision for imbalanced classes. It gives a higher weight to the correctly classified positive instances in the test data set and a stronger weight against FPs than the traditional F-score. In Youden's J, also ranging from 0 to 1, the first term is the classifier recall rate or True Positive Rate (TPR) and the second term is the False Positive Rate (FPR). When evaluating a trial classifier, the threshold corresponding to the maximum Youden's J index was used. This is not the final threshold used for the final RF classifier, but a standard choice for performance comparison.

The ROC plots the TPR as a function of the FPR for every possible threshold value of the classifier. We used package ROCR (Sing et al. 2005) implemented with the R software (R Core Team 2020) to calculate the ROCs. The AUC for the ROC is a measure of classifier performance that does not depend on a single threshold choice. An AUC of the ROC of 1 indicates a perfect classifier, 0.5 indicates a classifier that performs no better than assigning random labels, and 0 indicates the worst possible classifier.

Closely related to the ROC, the Precision–Recall curve is used for imbalanced classification problems because both precision and recall focus on the correct classification of the positive class. The AUC for the Precision–Recall curve is another measure of classifier performance that does not depend on a threshold choice.

5.3. Feature Selection

Table 1 lists the 37 features and feature weights in the final optimized RF classifier. The weights indicate the probability of the feature being among the randomly chosen features for each node calculation. The table organizes the features by the stage of DTARPS-S analysis. The feature groups are described here:

Stellar properties: Stellar metadata from the TIC v8 (Stassun et al. 2019) and from the Gaia DR 2 catalog (Gaia Collaboration et al. 2018). Stellar properties tested include the effective temperature, mass, TESS T magnitude, Gaia parallax, Gaia G magnitude, G magnitude S/N, and others. The classifier optimization found that stellar radius, surface gravity, luminosity, and Gaia $G_{\rm BP}-G_{\rm RP}$ color index played significant roles in classification.

DIAmante light-curve properties: The reduced χ^2 measures the goodness-of-fit of the light curve to a constant median brightness. The tail range compares the range of the middle 96% of the light curve flux values with the range of the middle 50% of the light-curve flux values. The Positive Outlier Measure (POM) measures the most extreme positive outlier in the light curve with respect to the median. As discussed by Caceres et al. (2019a), the POM helps identify stars with strong flares that may cause spurious peaks in the TCF Periodogram. Skewness, the third standardized moment of the distribution, is a measure of the asymmetry of the distribution of light-curve flux values around the mean. Kurtosis, the fourth standardized moment of a distribution, is helpful to measure the strength of outliers with respect to a

Table 1
Scalar Features Used in the Optimized Random Forest classifier

Feature	Weight	Description
		Stellar Properties
tic_Radius	0.027	Radius of star [$R_\odot$]
logg	0.027	Gaia DR2 log(g) measurement
Gaia.color	0.027	Gaia DR2 color, $G_{\rm BP} - G_{\rm RP}$
Luminosity	0.027	Gaia DR2 luminosity of the star
		DIAmante Light-curve Properties
median.lc	0.027	Median value of the DIAmante light curve
Redchisq.lc	0.027	Reduced χ^2 value for a flat model
tail_range.lc	0.027	98th quantile – 2nd quantile divided by the IQR
POM.lc	0.027	Greatest positive outlier measure
skew.lc	0.027	Measure of Non-Gaussian asymmetry
kurt.lc	0.027	Measure Non-Gaussian outlier strength
P_autocor.lc	0.027	*p*-value from the Ljung–Box test for autocorrelation
		Differenced Light-curve Properties
quantiles.diff.01	0.027	1st quantile
quantiles.diff.90	0.027	90th quantile
POM.diff	0.027	Greatest positive outlier measure
		ARIMA Residuals Properties
quantiles.resid.10	0.027	10th quantile
quantiles.resid.99	0.027	99th quantile
POM.resid	0.027	Greatest positive outlier measure
IQR.resid	0.027	Interquartile range
Redchisq.resid	0.027	Reduced χ^2 value for a flat model
Prob_norm.resid	0.027	*p*-value from the Anderson–Darling test for normality
Prob_autocor.resid	0.027	*p*-value from the Ljung–Box test for autocorrelation
IQR.improv	0.108	Ratio of the IQR for the ARIMA residuals to the DIAmante light curve
Redchisq.improv	0.108	Ratio of the reduced χ^2 value of the ARIMA residuals to the DIAmante light curve
		TCF Periodogram Properties
LOESS_mnsnr	0.081	Mean S/N of top 100 periodogram peaks with respect to LOESS fit
TCFpeaks_mean	0.081	Mean raw power of the top 100 TCF peaks
LOESSpeaks_mean	0.081	Mean power above the LOESS fit of the top 100 TCF peaks
		Best TCF Transit Properties
TCF_power	0.135	Strength of best peak in the TCF periodogram
TCF_period	0.054	Period of the best TCF periodogram peak
TCF_depthSNR	0.135	S/N of the best TCF periodogram peak
TCF_shape	0.135	Mean of the in-transit folded light curve divided by the MAD out-of-transit folded light curve
Folded_AD	0.135	Anderson–Darling test on the phases of the folded light curve
arbox_deperr	0.135	Transit depth error from ARIMAX model
even.odd.p_value	0.135	*t*-test *p*-value for even and odd transit depths
trans.p_value	0.135	*t*-test *p*-value for in-transit and out-of-transit depths
snr.transit	0.135	$\delta/\sigma(1/\sqrt{N_{\rm in}} + 1/\sqrt{N_{\rm out}})$ (See Section 10.1.1 M20)
frac_dur	0.135	Fractional transit duration (See Section 10.1.8 M20)

Table 1
(Continued)

Feature	Weight	Description
planet_rad_tcf	0.054	Planet radius calculated from the TCF transit depth

Note. All features for the light curve have been calculated after the removal of outliers and ramping problems from spacecraft jitter.

Gaussian distribution. The Ljung–Box test for autocorrelation applied to the light curve has a null hypothesis that the flux values are independently distributed and tests the alternative hypothesis that the flux values show correlation. A p-value $\gtrsim$0.01 indicates that the light curve is consistent with white noise.

Differenced light-curve properties: Statistics of the distribution of flux values for the differenced light curve including the 1st and 90th quantiles of the distribution. The POM, again, measures the most extreme positive outlier in the differenced light curve with respect to the median of the differenced light curve, which would identify a sharp transit brightening.

ARIMA residual properties: These features include: four statistics describing the residuals after the best-fit ARIMA model has been subtracted from the differenced light curve (10% and 90% quantiles, POM, and IQR); three statistical tests applied to the residual light curve (χ^2, Anderson–Darling and Ljung–Box); and two measures of the importance of the ARIMA fitting (IQR and χ^2 improvements).

These last two measures, with high feature weights, raised classifier performance more than most features; this demonstrates the importance of ARIMA modeling for planet detection. The ratio of the IQR of the ARIMA residuals to the IQR of the DIAmante light curve measures the improvement in the noise of the time series, with a smaller ratio indicating a greater effect of the ARIMA fit. The ratio of the reduced χ^2 values for a constant flux model of the ARIMA residuals to the DIAmante light curve measure how well the ARIMA model removed variation and trend from the light curve. The Ljung–Box test applied to the ARIMA residuals indicates how well the ARIMA model did at removing short-memory autocorrelation from the time series. The Anderson–Darling normality test is used here to determine if the ARIMA residuals fluxes follow a Gaussian distribution. A p-value $\gtrsim$0.01 indicates that ARIMA residuals are consistent with Gaussian noise.

TCF periodogram properties: Caceres et al. (2019a) found that the collective top periodogram peak properties, not just the results from the strongest TCF periodogram peak, improved the performance of the classifier for Kepler light curves. We find the same behavior for the TESS classifier. The mean S/N of the top 100 TCF peaks is used to identify periodograms with many noisy peaks in the periodogram, suggesting that no strong periodicity was identified. The mean power of the top 100 TCF peaks calculated from the raw TCF output and from the LOESS regression line are used to pick out periodograms with just a single or a few strong peaks. Having a low mean power for the top 100 peaks but a strong power for the best peak indicates that the best exoplanet transit peak is highly significant and does not arise from noise in the ARIMA residuals.

Best TCF transit properties: The 11 properties based on the best TCF period have high feature weights. These include: three properties of the highest TCF periodogram peak; six measures of the time series folded modulo the best period; and the significance of the depth derived from the parametric ARIMAX model.

The three properties of the strongest TCF peak are the period, TCF power, and S/N of the strongest peak in the TCF periodogram within a small window after subtracting the LOESS smoother. Caceres et al. (2019a) refrained from including the transit period and corresponding planetary radius from their RF classifier, to avoid biasing their candidate results. We include the transit period because it reduces the significance of spurious peaks with periods 13–15 days due to the TESS satellite orbit (Figure 13).

The shape parameter of the folded DIAmante light curve compares the mean value of the transit with the median absolute deviation (MAD) value for the other mean values of the non-transit sections of the phase-folded light curve. This measures the transit's flux difference compared with the rest of the light curve, and provides a subtle distinction between planets and EBs.

The Anderson–Darling test is applied to the distribution of phases for the observations in the phase-folded light curve, to test if the phases are consistent with an underlying uniform distribution. If not, then the TCF may have found a spurious periodic signal by aligning gaps in the light curve rather than identifying a true exoplanet transit. This is a crucial feature for identifying and reducing spurious periodogram peaks arising due to periodicities in cadence gaps.

The t-test, designed to quantify the difference in means of two Gaussian distributions, is applied to the even and odd transit light curve flux values and the in-transit and out-of-transit light curve flux values. A larger p-value is desired for the even–odd t-test to distinguish planet transits from EBs, while a smaller p-value is desired comparing the in-transit and out-of-transit fluxes states to show that the transit represents a statistically significant dip in flux. These tests rely on a transit mask to label points in the light curve that are in transit and out of transit, as well as to label even and odd transits.

The S/N of the primary transit feature describes how well the transit signal with the period and phase from TCF adds up over the phase-folded light curve. It is described in both M20 and Kovács et al. (2002).

The fractional transit duration is the ratio of the transit duration to the transit period. This feature was used by M20.

Inferred planet radius: The radius of an exoplanet is calculated from the depth of the transit from the best TCF peak and the stellar radius from the TIC. We include the planetary radius because the injected FP signals included in our negative training set (Section 4.2) allows us to train the RF classifier away from likely astrophysical FPs with very deep transit depths. However, it should be noted that this may reduce the DTARPS-S classifier sensitivity to very large, inflated gaseous exoplanets.

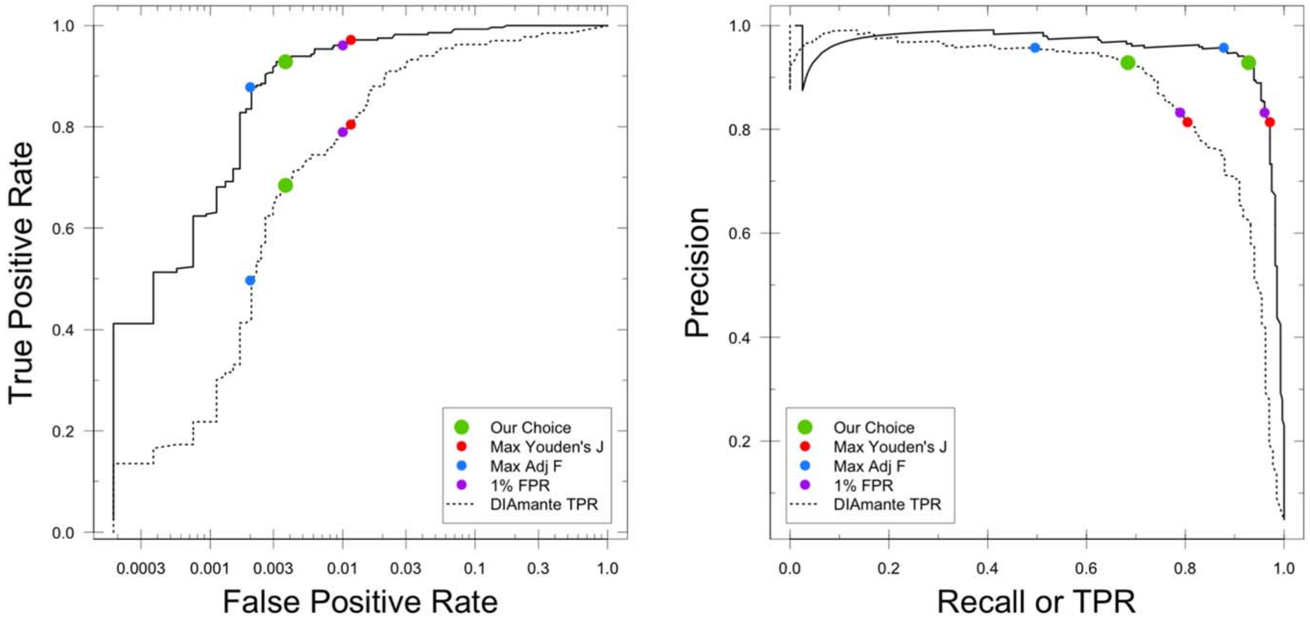


Figure 11. Performance of the final Random Forest classifier for every possible threshold choice shown with the Receiver Operator Characteristic (left) and Precision–Recall (right) curves. The solid lines derive from application to the validation set, and the dashed lines are the recall from 133 planet candidates of M20. Our choice of threshold $P_{\rm RF} = 0.300$ (shown in green) is compared with other possible threshold choices, including the 1% FPR choice used by M20 (purple).

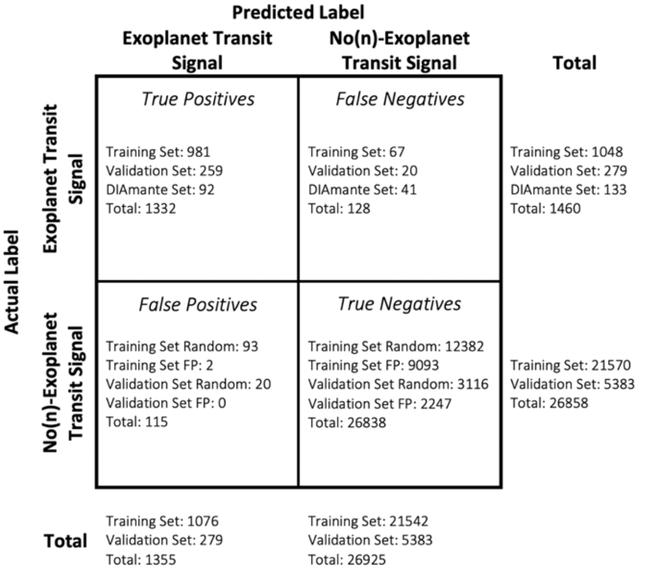
Table 2
Classification Metrics for the Validation Set

	$P_{\rm RF}$ Threshold	TPR	FPR	MCC	Adjusted F-score	Youden's J Index
Max Youden's J	0.155	0.9713	0.0115	0.8830	0.8869	0.9598
Max Adj F	0.481	0.8781	0.0020	0.9127	0.9682	0.8761
1% FPR	0.174	0.9606	0.0100	0.8884	0.8927	0.9505
Our Choice	0.300	0.9283	0.0037	0.9246	0.9283	0.9246

6. The DTARPS-S Final Classifier

Figure 11 shows the ROC and Precision–Recall curves for the final RF classifier. It is worth noting that the abscissa is logarithmically transformed to highlight small values of FPR needed for reliable transit discovery. The solid lines give the TPR and the FPR (or the TPR and the precision) for every possible threshold value between 0 and 1. The dashed lines give the recall rate for a random sample of 133 planet candidates from the M20 DIAmante study.

Our threshold choice of Random Forest probability $P_{\rm RF} = 0.300$ is shown with the larger green points. We chose this threshold to minimize the FPR as much as possible while maintaining high DIAmante survey recall and the TPR. Three other threshold choices are shown for comparison; M20 chose a threshold that gave an FPR of 1% that lies very close to the maximum Youden's J threshold. The final TPR and FPR values for our chosen threshold, as well as the comparison thresholds, are listed in Table 2 along with classification metrics described in Section 5.



		Predicted Label		
		Exoplanet Transit Signal	No(n)-Exoplanet Transit Signal	Total
Actual Label	Exoplanet Transit Signal	*True Positives* Training Set: 981 Validation Set: 259 DIAmante Set: 92 Total: 1332	*False Negatives* Training Set: 67 Validation Set: 20 DIAmante Set: 41 Total: 128	Training Set: 1048 Validation Set: 279 DIAmante Set: 133 Total: 1460
	No(n)-Exoplanet Transit Signal	*False Positives* Training Set Random: 93 Training Set FP: 2 Validation Set Random: 20 Validation Set FP: 0 Total: 115	*True Negatives* Training Set Random: 12382 Training Set FP: 9093 Validation Set Random: 3116 Validation Set FP: 2247 Total: 26838	Training Set: 21570 Validation Set: 5383 Total: 26858
Total		Training Set: 1076 Validation Set: 279 Total: 1355	Training Set: 21542 Validation Set: 5383 Total: 26925	

Figure 12. Confusion matrix for the final RF classifier with threshold $P_{\rm RF} = 0.300$. Values are based on the validation set and OOB predictions for the training set.

Further detail is given in Figure 12, showing the confusion matrix for the final RF classifier based on our chosen threshold $P_{\rm RF} = 0.300$. The confusion matrix shows how well the predicted labels from the RF classifier line up with the actual labels of the data in the training and validation sets. For the training set, we used the out-of-bag (OOB) RF prediction value to determine the predicted label. Each tree in the RF classifier uses a bootstrapped sample of the training set for construction, called bagging. OOB means that a data case from the training set was classified only by trees in the classifier where that data case was not used to "grow" or train those trees. OOB prediction values are calculated using only decision trees in the

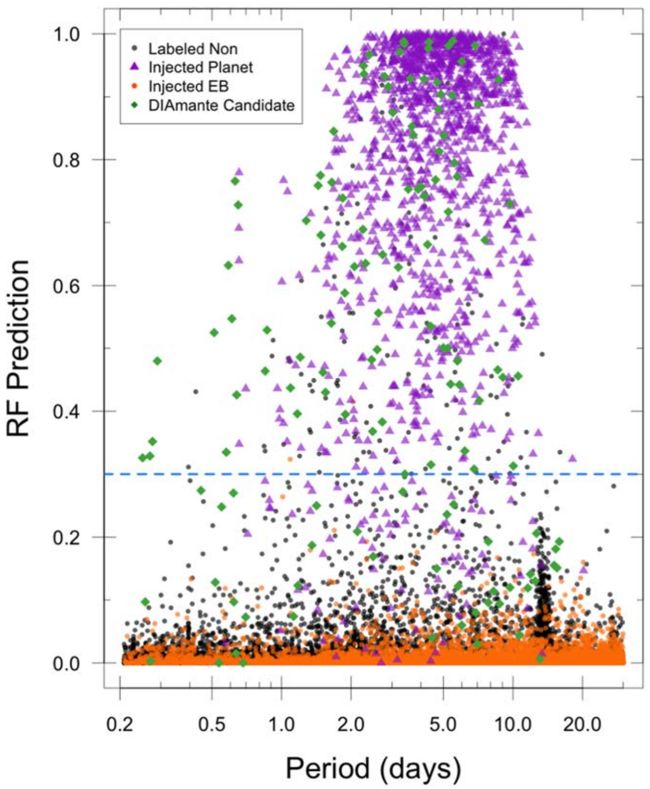


Figure 13. Distribution of the Random Forest pseudo-probability P_{RF} for the optimized classifier on the training and validation sets. The purple triangles represent light curves with injected transiting exoplanet signals that passed human vetting and were utilized in the positive training set (Section 4.1). The green diamonds represent the 133 candidates identified in M20 that assisted in optimizing the classifier. The black points indicate random light curves that were given negative class labels; most have P_{RF} values near zero. The orange points represent light curves with injected False Positive signals (Section 4.2). Our chosen threshold for the final classifier is shown as the blue dashed line at $P_{\mathrm{RF}} = 0.300$. The strong performance of the classifier is directly visible: purple and green points lie mostly above the threshold, while the black and orange points lie mostly below the threshold.

RF ensemble that were not grown using that training data case (Breiman 2001). We include the predicted labels for the random sample of 133 DIAmante candidates (M20) used to create the dashed TPR line in the ROC curve (see Figure 11, left plot). The $P_{\mathrm{RF}} = 0.300$ threshold for the RF classifier gave us a DIAmante TPR of 69% for the 133 randomly selected candidates.

For the data whose actual label is "non-exoplanet transit signal," we separately counted the negative data sets of randomly selected light curves and injected FPs. Perhaps the most important result here is the extraordinary effectiveness of the final RF classifier with respect to injected False Positive EBs: only 2 out of the 11,342 (0.02%) injected FPs used in the labeled data sets are falsely labeled as an exoplanet transit signal. Overall, the optimized Random Forest classifier attains a 92.5% True Positive recovery rate with 0.4% False Positive contamination with respect to injected exoplanet transits and simulated variable stars.

The performance of the final optimized classifier is shown visually in Figure 13, where the classification results are plotted as a function of period from the best TCF peak (Section 3.2) for the entire labeled data set (Section 4.3) and 133 randomly selected candidates identified in M20. The RF prediction values for labeled data set objects in the training set comprise the OOB prediction value.

The plot of P_{RF} against TCF best period in Figure 13 provides valuable insights into the classifier performance that are not revealed in the confusion matrix. The vast majority of injected planets are recovered with periods ∼0.7 to 11 days, and nearly all labeled negatives are rejected from 0.2 to 30 days. A small number of False Positives (with respect to the threshold) do not have preferred periods.

A strong spike of negative label points lying below the $P_{\mathrm{RF}} = 0.300$ threshold is present at periods 13–15 days. This arises from the TESS satellite's 13.7 day lunar-synchronous orbital period having a large gap in the middle of the FFI light curve in each sector. This leads the Transit Comb Filter algorithm, in the absence of a strong transit signal, to fold the data in half to line up the gaps in the data. Other period search procedures applied to TESS light curves are similarly affected (e.g., M20, Chakraborty et al. 2020). Many of the trial RF classifiers were less successful than the final classifier in pushing down the P_{RF} values for these spurious periodicities. However, in the final classifier, this spurious spike in Figure 13 has the indirect effect of causing a sharp drop in the RF prediction value of all objects with periods longer than ∼11 days. As a result, our classifier is insensitive to true exoplanet transits at longer periods. This might have been alleviated if our injected exoplanet training set extended to longer periods $\sim 15 - 25$ days.

In contrast, although the injected exoplanet periods do not go shorter than 0.625 days (because the injections were based on Kepler planets based on a transit search truncated below 0.5 days), the optimized RF classifier does not appear strongly biased against short-period transit signals. This is shown by the recovery of several DIAmante candidates in the 0.2–0.6 day regime.

Figure 14 shows the feature importance plot associated with the final RF classifier where input features are ordered by their importance to the classification. Feature importance is calculated by comparing the training set label accuracy from a perturbed OOB forest ensemble with the unperturbed OOB forest ensemble (Ishwaran 2022). For each feature, the label from the perturbed OOB forest is found by classifying each data case normally on the OOB trees in the forest for that data case, but whenever a node is encountered that is split using the feature for which the importance is being calculated, the opposite daughter node is used for classification. Therefore, the feature importance shows the improvement of the accuracy of the entire RF classifier when the correct classification path in the trees is used for a feature rather than the opposite classification path for that feature. The feature importance is calculated from the predictive success of the feature and often cannot be interpreted physically (Genuer et al. 2010).

The signal-to-noise ratio of the transit in the folded light curve is the most important feature, followed by the error on the ARIMAX fitted transit depth, the period, and the planet radius. Some features serve as positive discriminators of planet transits (such as snr.transit and arbox_deperr) while others serve to push away spurious effects (TCF_period helps remove 13–15 day periodogram peaks, and planet_rad_tcf helps remove deep EB eclipses). Of the top five most important features, three are also among the most important features in the DIAmante classifier derived by M20 (snr.transit, planet_rad_tcf, and frac_dur).

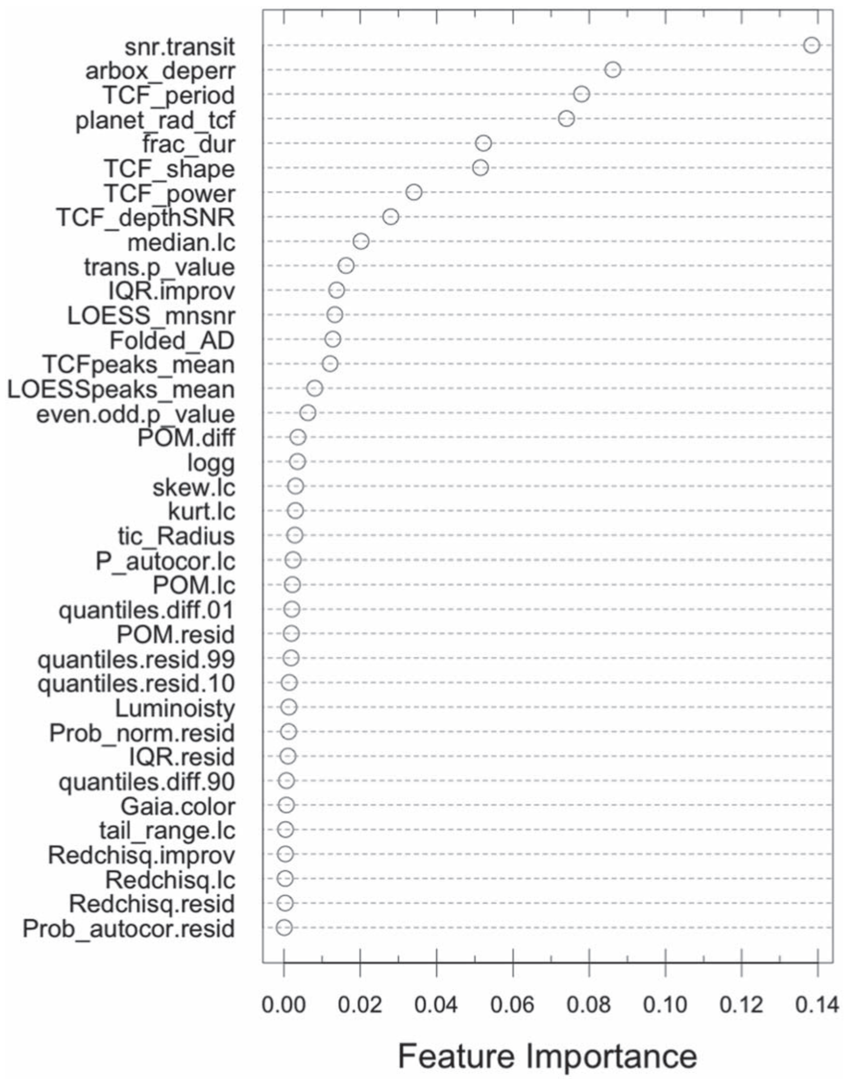


Figure 14. Feature importance for the final Random Forest classifier ordered by importance. Descriptions of the 37 features appear in Table 1.

7. The DTARPS-S Analysis List

The final RF classifier was applied to a test set of 823,099 DIAmante light curves. This is the full DIAmante collection of TESS Year 1 light curves, minus those with missing features (Table 1). Random Forest classifiers require full characterization of each object, and we did not attempt imputation of missing data. The classifier threshold of $P_{\mathrm{RF}} = 0.300$ was then applied. The result is 7377 DTARPS-S processed DIAmante light curves had an RF prediction value above the threshold. We call this the DTARPS-S Analysis List of TESS stars. This DTARPS-S Analysis List represents 0.9% of input light curves selected by uniform statistical procedures to have periodic transit-like features. It is roughly similar to the Threshold-Crossing Event (TCE) list obtained by the TESS official pipeline as a step toward producing their TESS Object of Interest (TOI) list (Guerrero et al. 2021), although their processing steps are quite different from the DTARPS-S analysis.

A small portion of the DTARPS-S Analysis List is shown in Table 3 with the full list available in machine-readable format from the electronic edition of this paper.

We emphasize that, while this DTARPS-S Analysis List of 7377 TESS stars has captured many transiting planets, it is still dominated by False Alarm and False Positive objects. The False Positive Rate of 0.0037 estimated from the combined labeled training and validation sets (Table 2) predicts that at least ∼3000 of the 7377 objects are not valid transiting planets. A rigorous vetting process to remove as many falsely labeled objects as possible is therefore needed to give a smaller catalog with much higher reliability (Section 11.3). In the parlance of machine-learning classification, the 7377 stars represents the maximum recall of the DTARPS-S analysis but with low sensitivity. The catalogs of 772 stars after several vetting procedures presented in Paper II represents the subset with high sensitivity but reduced recall rate.

8. Effect of ARIMA Modeling on Injected Signals

As the injected signals are the basis of the RF classifier's ability to identify true planetary transit signals, it is important to understand how the ARIMA modeling affects the injected signals, TCF periodogram, and features that drive the classifier. The ARPS procedure described in Sections 2.2.1–2.2.3 is designed for sensitive and reliable detection of planetary transits, and it may not produce an accurate *characterization* of planetary properties such as orbital period and planet radius. In this section, we examine the limitations and biases present in

Table 3
DTARPS-S Analysis List

Designations								Star			Light-curve and ARIMA Residuals					Transit Comb Filter			Classifier
TIC	DTARPS-S	NEA_{name}	NEA_{disp}	TOI	TOI_{disp}	DIAm	Refs	R.A.	Decl.	T	N_{lc}	IQR_{lc}	$P_{LB,lc}$	IQR_{AR}	$P_{LB,AR}$	Period	Depth	S/N	P_{RF}
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
4711	⋯	⋯	PC	3129.01	⋯	F	⋯	218.91878	−26.17928	10.9	849	0.0020	−6.0	0.0022	−0.1	2.33034	0.0045	33	0.73
17361	1	⋯	APC	3127.01	⋯	T	⋯	219.33632	−24.95848	11.3	849	0.0009	−6.0	0.0012	−0.1	3.60412	0.0072	83	0.94
44870	⋯	⋯	⋯	⋯	PC	F	⋯	220.19607	−29.25022	11.4	922	0.0014	−6.0	0.0014	−0.3	3.44101	0.0017	12	0.33
113636	⋯	⋯	⋯	⋯	⋯	F	⋯	222.35362	−28.42375	11.6	922	0.0010	−6.0	0.0010	−0.2	10.75002	0.0029	40	0.42
153687	⋯	⋯	⋯	⋯	⋯	F	⋯	223.58956	−27.16942	11.6	922	0.0047	−6.0	0.0032	0.0	0.34148	0.0104	324	0.31

Notes. The full table of 7377 light curves exceeding the DTARPS-S Random Forest threshold is available in the electronic version of the paper. Column descriptions: (1) TIC: TESS Input Catalog identifier. (2) DTARPS-S: DTARPS-S identifier from Paper II. Sequence number 1–467 from the DTARPS-S Candidate Planet catalog (Paper II, vetting Levels 1 and 2). GP indicator for Galactic Plane DTARPS-S list with reduced vetting (Paper II, vetting Level 3). (3) NEA_{name}: Name, NASA Exoplanet Archive (NASA Exoplanet Science Institute 2022). (4) NEA_{disp}: Disposition, NASA Exoplanet Archive (combined Confirmed Planets and TOI[1] list; NASA Exoplanet Science Institute 2022). CP includes CP (Confirmed Planet and KP (Known Planet); PC includes APC (Ambiguous Planet Candidate) and PC (Planet Candidates); FP includes EB (Eclipsing Binary); FA (False Alarm); FP (False Positive). ... = previously unidentified by NEA (accessed 2022 March 15). (5) TOI: TESS Object of Interest[1] (accessed 2022 February). (6) TOI_{disp}: TOI Disposition[1]. (7) DIAm: Flag for DIAmante planet candidate (M20). (8) References: Identified as a planet candidate or other object of interest by other studies: Co = Collins et al. (2018); Dr = Dressing et al. (2019); Do = Dong et al. (2021); Ei = Eisner et al. (2021); Fe = Feinstein et al. (2019); Ko = Kostov et al. (2019); Kr = Kruse et al. (2019); Ma = Mayo et al. (2018); Ol = Olmschenk et al. (2021); Sc = Schanche et al. (2019); Tu = Tu et al. (2020); vB = von Boetticher et al. (2019); Yu = Yu et al. (2019). ((9)–(10)) R.A., Decl. = R.A. and decl. from Gaia DR2 catalog. (11) T = TESS band magnitude. (12) N_{lc}: Number of measurements in TESS FFI light curve. (13) IQR_{lc}: InterQuartile Range of normalized light-curve fluxes. (14) $P_{LB,lc}$: Log probability of autocorrelation in light curve from the Ljung–Box test. Values above −2 are consistent with white noise while values below −4 are not. (15) IQR_{AR}: InterQuartile Range of ARIMA residuals. (16) $P_{LB,AR}$: Log probability of autocorrelation of ARIMA residuals from the Ljung–Box test. (17) Period: Transit period (day) from TCF periodogram. (18) Depth: Transit depth from TCF. (19) S/N: Signal-to-noise of peak power in TCF periodogram. (20) P_{RF}: Pseudo-probability of planet classification from Random Forest classifier.

(This table is available in its entirety in machine-readable form.)

the recovery of injected objects and their properties. We use the injected populations to investigate the effect of DTARPS-S processing on the physical parameters of the injected signals. The low recovery rate of the injected planetary transit signals (Section 4.1) is also explained. We return to the injected population in Section 9 for a third issue: estimating the completeness of the DTARPS-S Analysis List.

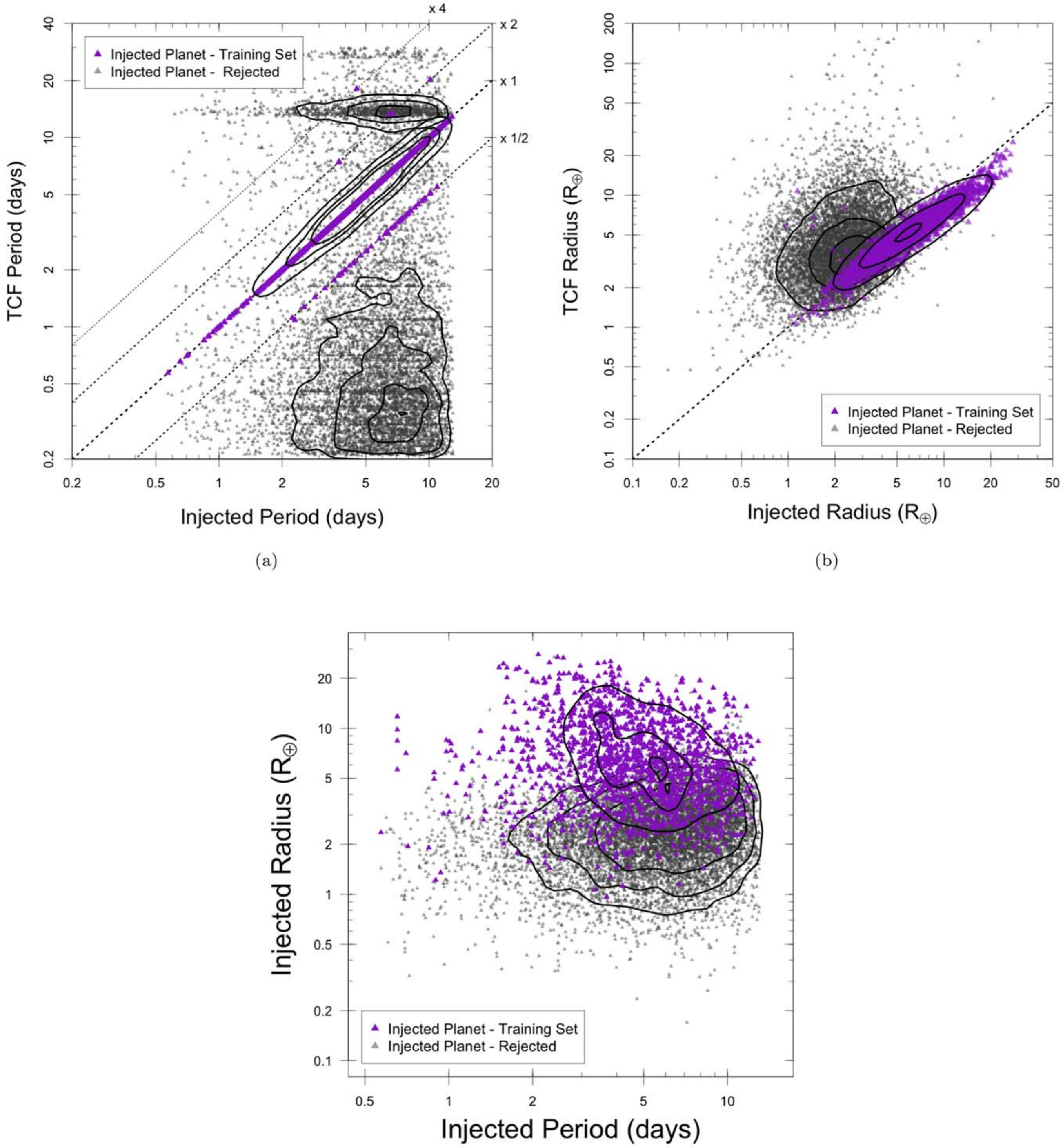
8.1. Recovered Planet Properties

The results of the RF classifier depend heavily on the orbital parameters obtained from the TCF algorithm for the best period. The top eight most important features of the RF classifier (Figure 14) are either extracted from the best TCF peak and TCF periodogram or are computed on the light curve phase-folded at the period identified by the best TCF transit model. However, of the 10,850 synthetic planet injections, only 1327 (12%) had TCF orbital periods that were close enough to the injected synthetic period (or an integer ratio) to be recovered with human vetting.

Panels (a) and (b) of Figure 15 compare the injected orbital parameters with the orbital parameters from the TCF analysis for the full set of synthetic planetary injections. The synthetic planetary signals whose best TCF peak orbital period matched the injected orbital period (or an integer ratio) are shown as purple triangles. Spurious periodicities with periods of 13–15 days arise from the 13.7 day orbital period of the TESS satellite. It is not surprising that TCF would align the two halves of the light curve and find spurious double-spike associated with ramping problems that escape our data-cleaning procedure (Section 2.5). The pileup of identified TCF periods between 13 and 15 days and near the extreme limit of the TCF period search of 27 days are both expected and easily removed by the RF classifier and by vetting.

The tendency of TCF to identify much shorter periods than the injected period for injected planetary signals was not expected. This is shown as the cloud of gray points in the lower right of Figure 15. Just over half of the 10,850 synthetic injected planetary transit signals were assigned periods <1 day by the TCF algorithm, while only 3% of the injected FP signals had a spurious period found by TCF to be shorter than one day. This difference can be attributed to the injected transit depth: injected FP signals radii exceeded 38 $R_\oplus$ while injected planet radii were less than 28 $R_\oplus$ and often much smaller. This suggests that the DTARPS-S procedure will have difficulty identifying shallow transit signals, especially at shorter periods.

This issue is further elucidated in Figure 15(b), which compares the injected radius and the radius from the best TCF peak. The injected planets included in the positive training set shown in purple are concentrated along a locus that falls just below the desired 1:1 line. When TCF identifies the correct period for a planetary signal, it gives a slightly smaller radius than the radius of the underlying signal; the effect is more pronounced for radii



(a)

(b)

(c)

Figure 15. Comparison of the orbital parameters found from the best TCF peak with the injected orbital parameters for the synthetic injected planets. Recovered injections are shown as purple triangles. (a) Injected and recovered orbital periods. Dotted lines show integer ratios between the injected and TCF periods. (b) Injected and recovered planet radii. (c) Distribution of the injected planets in period–radius space.

$\gtrsim$10 $R_\oplus$. This bias has multiple causes. First, the ARIMA model incorporates some of the transit signal with the stellar variability (Caceres et al. 2019b). This effect can also occur with other detrending statistical procedures such as wavelet analysis and Gaussian Processes regression. Second, for longer periods, the ARIMA residuals have only a few points in the ingress and egress spikes and the TCF-matched filter has difficulty correctly fitting the extreme values of the spike shape. This partially accounts for the paucity of recovered large 10–20 $R_\oplus$ planets at long periods in Figure 15(c). Third, the ingress and egress will often be split between two TESS cadence slots, so neither one captures the full height of the spike in the ARIMA residuals. Stellar limb darkening may further slow ingress and egress, weakening the spike. We mitigate this radius bias in Paper II, both by fitting likelihood-based astrophysical models to the stronger transits and by visual correction of transit depths for weaker transits.

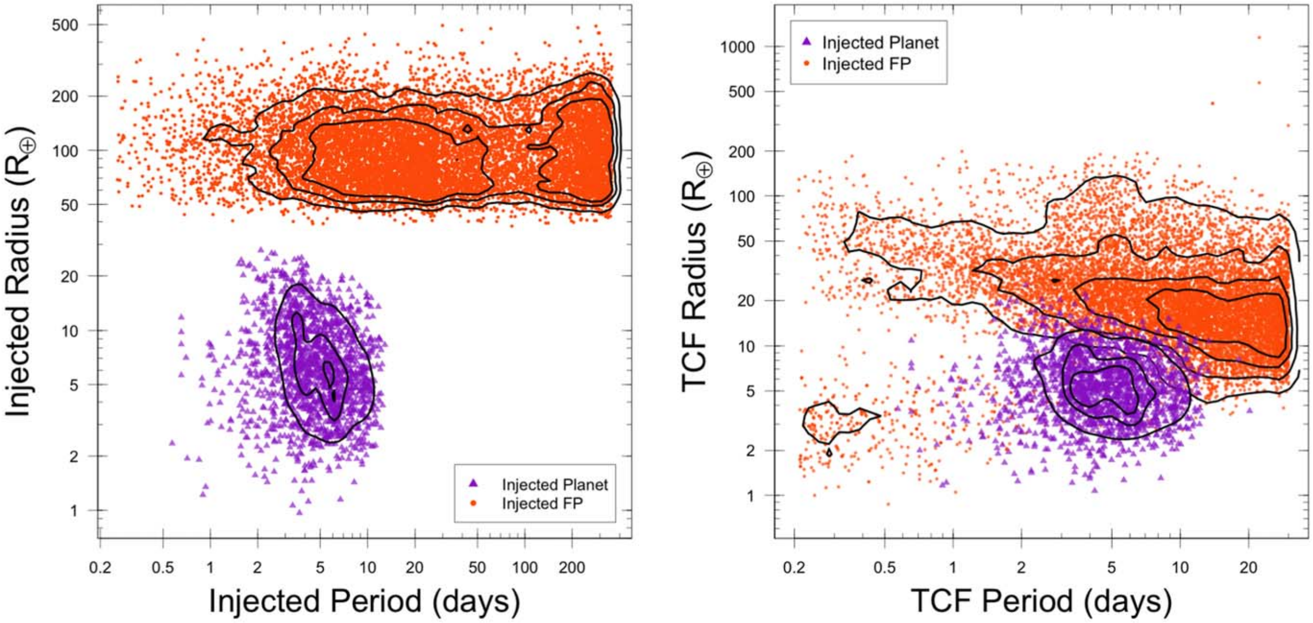


Figure 16. Comparison of the synthetic orbital parameters that were injected into random light curves for the positive (purple triangles) and negative (orange points) training sets with the orbital parameters from the best TCF transit model. Injected FPs were used supplement the negative training set and only represent about half of the negative training set (see Section 4.2). The top panel shows the distribution of injected orbital parameters, and the bottom panel shows the distribution of the best orbital parameters from TCF analysis. Smooth contours are plotted on both panels to aid the eye in seeing the underlying distribution.

The gray points in Figure 15(b) reveal another bias: when TCF fails to identify the correct orbital period for a planet, it also tends to overestimate the planet radius. This overestimation of the TCF radii is strongest for smaller planets. For injected planets rejected from the positive training set, the TCF radius is more than twice the true radius for 43% of injections with $R < 4\ R_\oplus$, compared to 14% of injections with $R > 4\ R_\oplus$.

The distribution of the injected planets in Figure 15(c) shows that the ability of TCF to recover the injected orbital parameters does not depend stringently on the injected period, but it is moderately conditional on the injected planet radius. Both distributions cover the same range of injected periods from 0.5 to 13.5 days. The distribution of recovered injected planets is centered at a higher injected radius than the distribution of rejected planetary injections, indicating that ARPS has an easier time identifying planetary signals with $R \gtrsim 5\,R_\oplus$. This effect will be quantified in Paper II with completeness curves.

8.2. Injected False Positives

Figure 16 shows the injected and TCF-recovered radius–period distribution for both injected planets and injected False Positive (FP) signals. The injected FPs had periods from 0.2 to 357 days, i.e., the length of the longest extracted TESS FFI light curve in the DIAmante data set. Because DTARPS-S does not seek to correctly characterize the orbital parameters of FPs, we chose to include in the negative training set the full range of FP signals that may be present in the DIAmante data set.

The tendency of DTARPS-S to underestimate the depth of the transit signal for large injected radii signals (Figure 16(b)) causes DTARPS-S to greatly reduce the radii of injected FPs, as can be seen by comparing the orange points in the two panels of Figure 16. This undoubtedly is due to incorporation of deep EB transits into the ARIMA model. The center of the distribution of injected FP radii is reduced from $\sim$100 $R_\oplus$ to $\sim$20 $R_\oplus$. The FP and planet-injection distributions overlap in the TCF parameters, while they are fully separated in the injection parameters. The group of injected FPs with radii $\lesssim$10 $R_\oplus$ and periods <1 day are not injected EB signals, but rather the injected sinusoidal signals simulating rotationally modulated variable stars.

The predilection of TCF to report ultra-short periods when it fails to find the correct period, combined with the much smaller TCF radius value, created a sample of $\sim$400 injected FP signals with TCF periods <2 days and TCF radii <10 $R_\oplus$. Although the injected FP signals include short-period sinusoidal variable signals (Section 4.2), none of them have a TCF radius smaller than 15 $R_\oplus$. Only injected FP signals with spuriously identified TCF periods <2 days have TCF radii <2 $R_\oplus$. There are only $\sim$100 injected planetary signals in the positive training set in that region. The erroneously characterized FPs completely dominate the shortest TCF periods with 0.2–1 day. This means, even though TCF tends to identify spurious short periods when it cannot correctly identify a transit signal (Figure 15), the RF classifier will be unlikely to identify them as potential DTARPS-S Candidates because that region of TCF period–radius space is dominated by injections from the negative training set. The classifier is also less likely to recover true planetary signals with periods less than 2 days. The RF classifier is more complicated than drawing boxes in the regions of TCF period–radius space, as it includes influence of over 30 other variables. But given that the TCF period and radius are the third- and fourth-most important features in the RF classifier (Figure 14), this mischaracterization of injected FP signals will affect the final classification.

In addition to the population of injected FPs with short TCF periods and TCF radii consistent with planetary objects, there are a large number of injected FPs with TCF radii consistent with a Jovian planet ($\sim$10–20 $R_\oplus$), in particular at longer periods $\gtrsim$10 days. The presence of this population of FPs in the negative training sample may cause DTARPS-S to be less sensitive to Jovian planets and long-period planets. The large number of injected FPs with TCF radii consistent with Jovian planets irrespective of orbital period may make it more difficult to find Jovian planets, despite TCF's ability to better recover the correct orbital period for larger planetary signals.

8.3. Conclusions from Injected Populations

Only 12% of the injected planets are reliably recovered by the DTARPS-S procedure. This low fraction is not a surprise, as the distribution of injected radii is drawn from the more sensitive Kepler mission dominated by planets too small for TESS detection (Figure 9, top right panel). The rate of capture of planetary injections will be examined in our completeness analysis below (Section 9).

For the correctly identified injected planets, the TCF periods are mostly accurately recovered from 0.5 to 13 days. For a small fraction, the half-period harmonic is preferred by the TCF. We use harmonic, here and later in the dissertation, with the definition used in time-series analysis: a frequency (period) is an integer ratio of another frequency (period) in the time series (light curve). TCF-derived radii, on the other hand, are underestimated for correctly identified injected planets, in particular for large-radius injections. This bias is understood as a combination of ARIMA and TCF behaviors. We correct this bias for astronomically interesting candidates via manual intervention or astrophysical modeling in Paper II.

The DTARPS-S analysis also recovers a small fraction of injected False Positive signals as potential planetary candidates. The response to astronomical False Positives is complicated and is discussed in Section 10.5. False Positive (and False Alarm) contamination motivates the strictness of our vetting procedures in Paper II. In that study, we remove nearly 90% of the objects in the DTARPS-S Analysis List when creating the DTARPS-S Planet Candidate catalog. This reduces the completeness (in statistical parlance, the "recall") of the listing in this study, but it greatly improves the reliability ("precision") of the DTARPS-S planet candidates.

9. Random Forest Classifier Performance for Planet Injections

The recall rate of the injected planetary signals across the range of the injected period and radii can quantitatively measure the ability of the RF classifier to recover planets in the DIAmante data set. It is important to understand how the classifier performs across the planetary radius–period distribution, to evaluate the completeness of our intermediate DTARPS-S Analysis List with 7377 objects and the smaller DTARPS-S Planet Candidate catalog produced in Paper II. The completeness (or recall rate) of the RF classifier for different bins in planetary period–radius space is measured using the full set of synthetic planetary injections based on the Kepler planet sample (Section 4.1). The analysis is based on 7751 of the 10,850 synthetic planetary injections that were processed by the RF classifier; the remaining objects were omitted due to missing features.

Analysis of recall rates for synthetic injections gives a more reliable view of the completeness of the DTARPS-S analysis for a physical population of planets than does the comparison with other surveys (Section 10). The latter are subject to the different, and often poorly known, limitations and biases of TESS light-curve analysis and telescopic follow-up by many research groups.

Comparison with injections by (Christiansen et al. 2020, and earlier studies) has proved invaluable for deriving true planetary occurrence rates from the Kepler 4 yr survey. However, we emphasize that planetary occurrence rates cannot be reliably estimated from the analysis at this stage, because the DTARPS-S Analysis List is dominated by non-planetary signals. Occurrence rates will be estimated in Paper III after vetting has greatly improved the "sensitivity" of the sample, though with reduced "recall" rates.

9.1. Survey Completeness

Figure 17 shows the completeness for the 7751 synthetic planet injections. The completeness for each period–radius bin is the number of injected planets in the bin with an RF prediction value greater than the $P_{\rm RF} = 0.300$ threshold divided by the total number of injected planets in the bin. The bins are distributed evenly in log-space for the injected planetary radius and the injected orbital period. We chose to analyze the completeness of the RF in the injected orbital parameter space, rather than in the TCF identified period–radius space, because the planetary injections were created to have a realistic distribution in the injected period–radius space that is not preserved in the TCF period–radius space (Figure 15).

The distribution of the underlying points in Figure 17 roughly follows the distribution of points in Figures 15(c); 15(c) contains all of the injected planetary signals, whereas Figure 17 only uses injected planetary signals that were classified by the RF classifier. The completeness in Figure 17 appears to correspond with the recovered planetary injection signals in Figure 15(c), but only because we trained a highly effective classifier with an OOB recall rate of 92.5% for the training set of recovered planetary injections. Not all of the recovered planetary injections in Figure 16(c) had an OOB RF prediction value above the threshold, and not all of the rejected planetary injections were rejected by the RF classifier. There is a subtle but important distinction: namely, Figure 16(c) shows the distribution of the recovered injected planetary signals used in the positive training set of the classifier, whereas Figure 17 shows the completeness of the DTARPS-S methodology on the whole set of injected planetary signals.

Figure 17 shows poor completeness (<10%) for radii less than 2 $R_\oplus$ across all periods. There is also poor completeness for periods less than 1 day and radii <5 $R_\oplus$. In these regions, the classifier fails to recover enough planets to make meaningful statements about the exoplanet population. The classifier has low completeness (10%–25%) for planets with radii between 2 and 4 $R_\oplus$, and high completeness (70%–100%) for periods between 0.6 and 13 days for planetary radii between 8 and 30 $R_\oplus$. In the latter region, the classifier essentially captures the full exoplanet population. At a given planet radius, the DTARPS-S classifier achieves somewhat higher recall rates for periods around 2–4 days than around 7–13 days, producing a tilt in the heat map. The outlying bins of the distribution in Figure 17 are sparsely populated bins where recall rates are uncertain. For example, the pale yellow bins with radii larger than 10 days do not represent an inability of DTARPS-S to recover planets, as only a single injected object is present in those bins.

9.2. Weakest Signal Recovery

While Figure 17 shows the gradual deterioration of how well DTARPS-S recovers smaller planets, the sensitivity of a transit

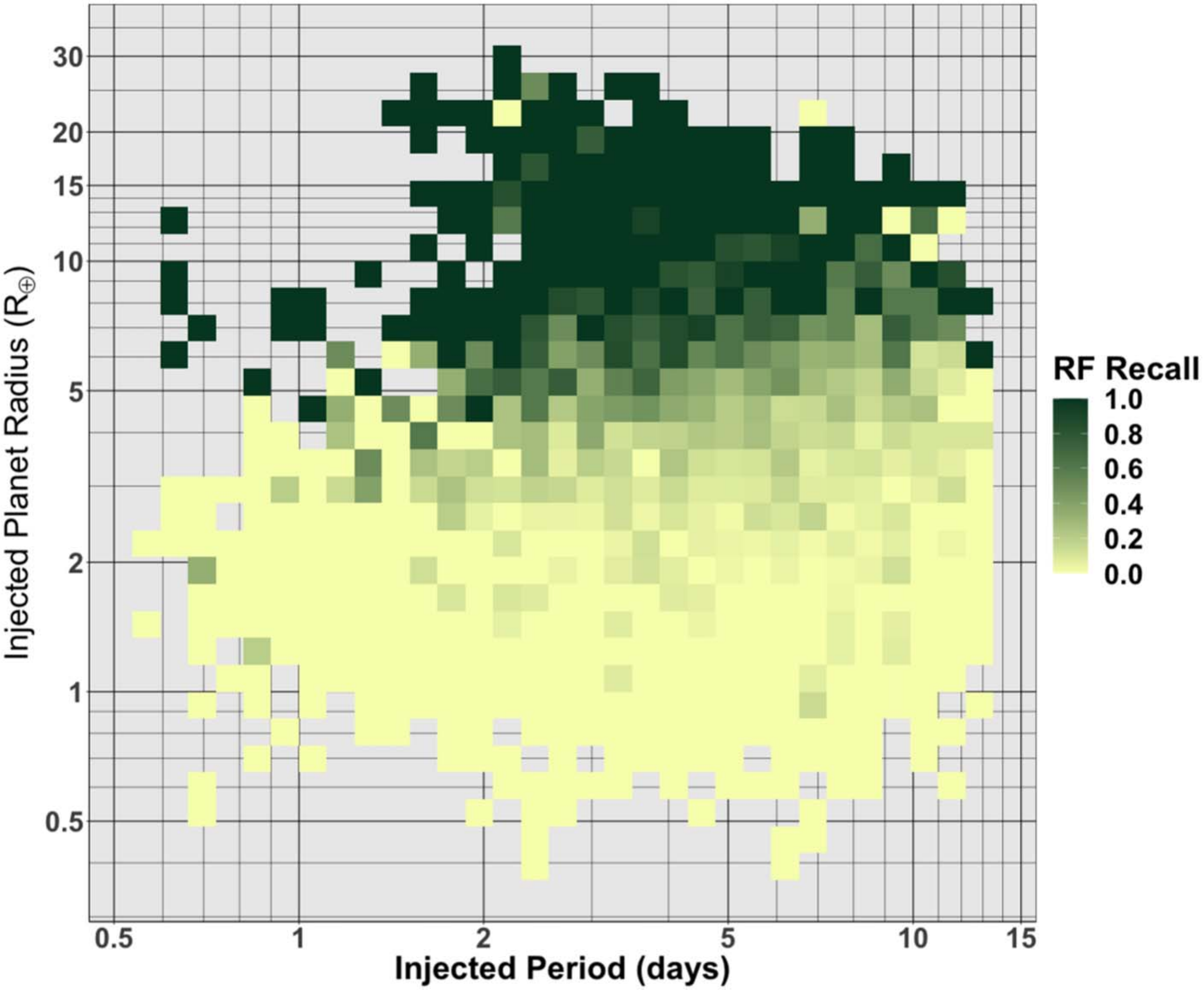


Figure 17. Heat map of recall rates of the Random Forest classifier for the synthetic planetary injections as a function of injected period and radius bins.

survey can also be examined in terms of signal strength. A common measure of the ability to recover a synthetically injected transit signal is the effective signal-to-noise ratio ($S/N^{\rm eff}$) of a transit signal (Kovács et al. 2002; Howard et al. 2012; Christiansen et al. 2013, 2016). The $S/N^{\rm eff}$ of a transit signal is the depth of the transit divided by the standard deviation of the average of all measurements in the transit.

Using features already calculated for use in the RF classifier, we calculate an $S/N^{\rm eff}$ for the injected transit signal and the periodic signal associated with the best peak in the TCF periodogram based on Equation (1) in Howard et al. (2012). Here, the effective S/N of the transit is

$$S/N^{\rm eff} = \frac{\delta}{\rm IQR}\sqrt{\frac{n_{\rm pts}\, T_{\rm dur}}{P}}, \qquad (8)$$

where δ is the depth of transit, IQR is the InterQuartile Range of the light curve from which the transit depth is being measured, $n_{\rm pts}$ is the number of points in the light curve, and $T_{\rm dur}/P$ is the fractional duty cycle of the transit. We substitute the IQR of the original light curve (for the injected signal) and the IQR of the ARIMA residuals (for the best TCF transit) instead of standard deviations. Both the IQR and the standard deviation of a distribution are measures of the spread of the distribution, but the IQR is more robust against non-Gaussianity.

Figure 18 compares the $S/N^{\rm eff}$ for the injected planet signal and the $S/N^{\rm eff}$ for the strongest periodic signal in the ARIMA residuals from the TCF periodogram. This is shown for two subsamples: the injections that are successfully recovered by DTARPS-S processing (purple) and the injections that were rejected (gray). The orange dashed lines in Figure 18 show the approximate lower boundaries for the $S/N^{\rm eff}$ for the recovered injected planet signals used in the positive training set. The boundaries were set at $S/N^{\rm eff} = 6$ for both the $S/N^{\rm eff}$ of planetary signals injected into the DIAmante-extracted light curve and for the $S/N^{\rm eff}$ of the best TCF periods obtained from ARIMA residuals.

About 58% of the injected planetary signals that were rejected from the positive training had a TCF $S/N^{\rm eff}$ less than 6 and 93% of the rejected injected planetary signals lie to the left or below the boundaries in Figure 18. The recovery rate of the injected planetary signals with effective S/Ns above the

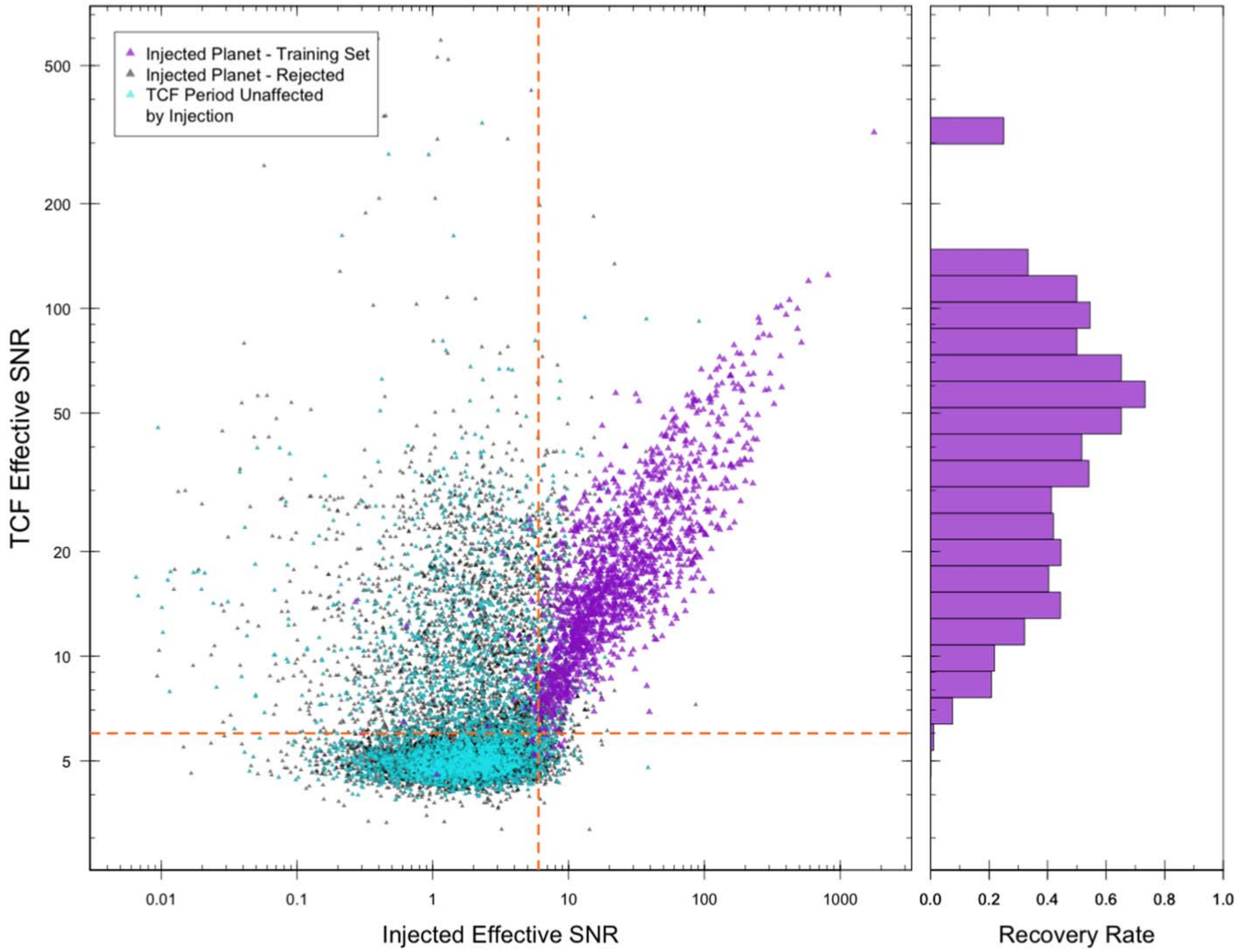


Figure 18. Effective S/N of the injected planetary signals. (Left) Comparison of the effective S/N (based on IQR rather than standard deviation) of the injected planetary transit to the effective S/N of the best transit from TCF. The injected planetary signals that were used in the training set are plotted as purple triangles, while injected signals whose period was not recovered by the TCF are gray triangles. Light curves whose TCF peak periods were unaffected by the injected planetary signal are plotted in turquoise. The dashed orange lines show approximate lower bounds of injected planets used in the training set. (Right) Fraction of injected planetary signals that were recovered given the spread of effective S/N of the best transit from TCF.

$S/N^{eff} = 6$ boundaries is 71%, much larger than the overall recovery rate of the injected planetary signals (Section 4.1). The histogram on the right side of Figure 18 shows the fraction of injected planetary signals used in the training set (the TCF periodogram peak recovered the injected transit period) for the S/N^{eff} of the TCF periodic signal in even, logarithmically spaced bins. The decline in the recovery rate of the injected planetary signals above a TCF S/N^{eff} of ∼60 is due to the rejected planetary injections whose injected transit S/N^{eff} was overwhelmed by noise or inherent periodic signals in the light curve.

The low injected S/N^{eff} values often arises from very shallow or short-period injections. Here, the paucity of TESS observations—only ∼1000 points in single-sector observations—hinders detection of small planets with small periods. Similar difficulties probably affect other TESS transit analysis systems: only 6% of TOI candidates have periods <1 day and only 5% have radii $<2\ R_\oplus$.

We can compare this situation to the earlier application of ARPS methodology to the Kepler 4 yr mission where Caceres et al. (2019a) identified 97 new mostly Earth- and Mars-size candidates. Gondhalekar et al. (2023) finds that application of ARIMA detrending of light curves and TCF periodograms is usually more sensitive to small planets than traditional detrending methods and BLS periodograms. However, the DTARPS-S completeness is low for small planets with shallow transits (Figure 17 and the histogram in Figure 18). This apparent discrepancy stems from the enormous difference in durations of Kepler and TESS FFI light curves. While both have time steps ∼30 minutes, the longer baseline of the Kepler four-year light curves meant that the Kepler light curves had ∼77,000 flux measurements while the TESS FFIs have on average ∼1000 flux measurements. Even in the continuous viewing zone for the Year 1 TESS data, the stitched-together light curve has a duration of only one year as opposed to four years from the Kepler mission. The increased number of points in the Kepler light curves increases the effective signal-to-noise ratio (S/N^{eff}) of the transit (Equation (8)).

The three panels of Figure 19 show a simulation of the recall rate of the injected planet signal given the S/N^{eff} from the histogram in Figure 18 for the planet injections for TESS-like surveys with three hypothetical durations: one month (left), one

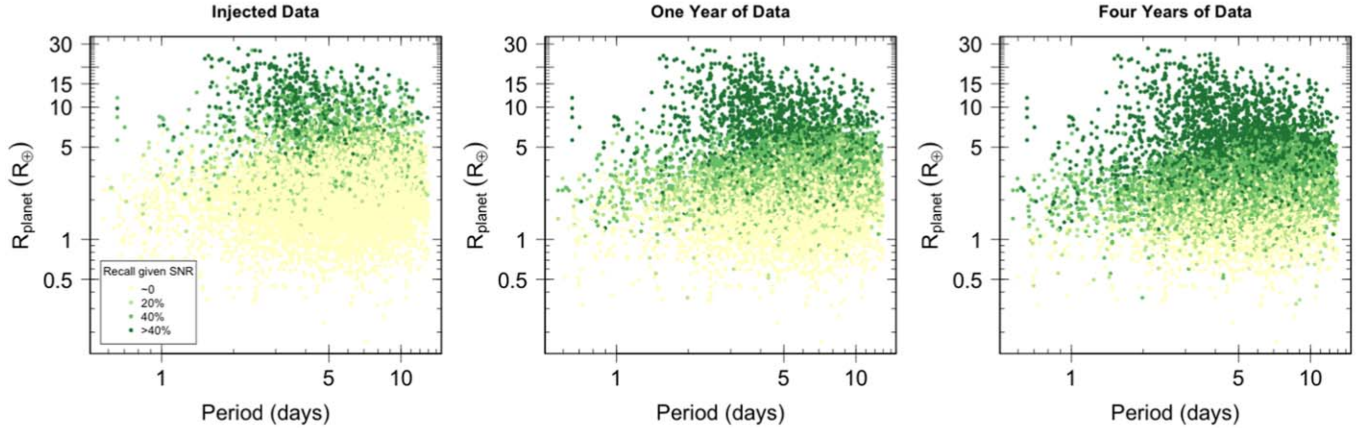


Figure 19. Radius–Period plots of the injected planets (Section 4.1) colored by the recall rate for the injected planet, given only the injected transit TCF signal S/N^{eff} for TESS with three different hypothetical observational durations: one month, one year, and four years. Green indicates a high recall rate for the injected planet, and yellow indicates a low recall rate for the injected planet.

year (middle), and four years (right). It shows that TESS could detect substantial fractions of Earth-size planets if four continuous years of observation were available. This sensitivity could be approached by the TESS extended missions near the ecliptic poles.

10. Relationship to Other Surveys

In addition to considering recall rates for the synthetic planetary injection sample, we can compare the RF classifier performance to previous studies. The main difficulty in making inferences is that other surveys may suffer incompleteness and erroneous (False Alarm or False Positive) exoplanet identifications. The "Confirmed Planets" from the NASA Exoplanet Archive (NASA Exoplanet Science Institute 2022; NEA; accessed 2022 March 15) is likely to have the fewest errors, though its listings are culled from a heterogeneous collection of studies. The TESS Objects of Interest (TOI) list, the community TOI (cTOI) list (NASA Exoplanet Archive 2022; both accessed 2022 March 15), and the M20 DIAmante analysis candidates are specifically derived from TESS data, but their reliability is unknown. The NEA and TOI efforts also list "False Positives" that are useful for comparison with DTARPS-S results.

We matched the DIAmante data set and the potential candidate transits in the DTARPS-S Analysis List with lists from 15 previous studies on exoplanet surveys or False Positives, such as low-mass eclipsing binaries, flare stars, and stellar rotation. The exoplanet survey studies utilized here are Mayo et al. (2018), Dressing et al. (2019), Feinstein et al. (2019), Kostov et al. (2019), Kruse et al. (2019), Yu et al. (2019), (Montalto et al. 2020, M20), Dong et al. (2021), Eisner et al. (2021), and Olmschenk et al. (2021). The False Positive studies are Affer et al. (2012); Collins et al. (2018); Schanche et al. (2019); von Boetticher et al. (2019), and Tu et al. (2020). The Appendix gives a brief description of each of the external surveys and their corresponding entries in the DTARPS-S Analysis List. Where TIC numbers were not available for matching DIAmante light curves with reported objects in these external studies, we used the best match between the R.A. and decl. coordinates of the objects, with a search radius of 5″. The periods reported in the external surveys (when available) are compared with the period from the best TCF peak. We consider the period to be matched when the TCF peak period is within a 1% fractional difference of the reported period (or a harmonic of the reported period).

10.1. M20 DIAmante Candidates

Of the 394 planet candidates identified by the M20 DIAmante analysis (some of which were later confirmed as planets or found to be False Positives), 364 were processed through the entire DTARPS-S procedure. The best DTARPS-S period matched the M20 reported period for 333 (91%) of the M20 candidates. The TCF periodogram and BLS periodogram thus emerge with identical results for nearly all DIAmante candidate planets. Of the 31 M20 candidates where TCF failed to recover the correct period, 17 had M20 periods greater than 13.5 days and were thus outside the range of our injected planet training set. TCF often identified the $1/2$ or $1/3$ harmonic period of long-period M20 candidates; it is not clear which period is correct in these cases.

When the $P_{RF} = 0.300$ threshold for the RF classifier is applied (DTARPS-S Analysis List), 213 of the 364 M20 candidates are captured, giving a 59% recall rate for the M20 study. The main difference between DTARPS-S and DIAmante results is thus attributable to the classification stage. All of these recovered M20 candidates had a TCF period matching the reported M20 period.

Figure 20 (top panel) shows the recall rate of the RF classifier for different bins in radius–period space from the best TCF peak. The RF classifier has the strongest recall rates for the candidates whose planetary radii from TCF are between 2 and 10 $R_\oplus$ and whose TCF orbital periods are between 1 and 10 days. The RF classifier only has a 20% recall rate for M20 candidates with TCF periods greater than 10 days and a 47% recall rate for TCF periods less than 1 day. The recall rate for the M20 candidates also falls off at larger TCF planet radius, likely due to the many injected False Positive signals in the negative training set with planetary-consistent radii.

In the bottom panel of Figure 20, the colored points show the recovered M20 candidates and the black points show the unrecovered M20 candidates with the RF classifier. The distribution of M20 candidates closely follow the recall distribution of synthetic planet injections.

The candidates reported in M20 mostly have radii $>7\ R_\oplus$, with moderate coverage around 3–7 $R_\oplus$. Neither the DTARPS-S nor the M20 samples cover well the region with small radii; only 16 candidates have radii $<3\ R_\oplus$. Despite this, DTARPS-S does a good job at recovering the M20 candidates with radii less than 5 $R_\oplus$ (77%).

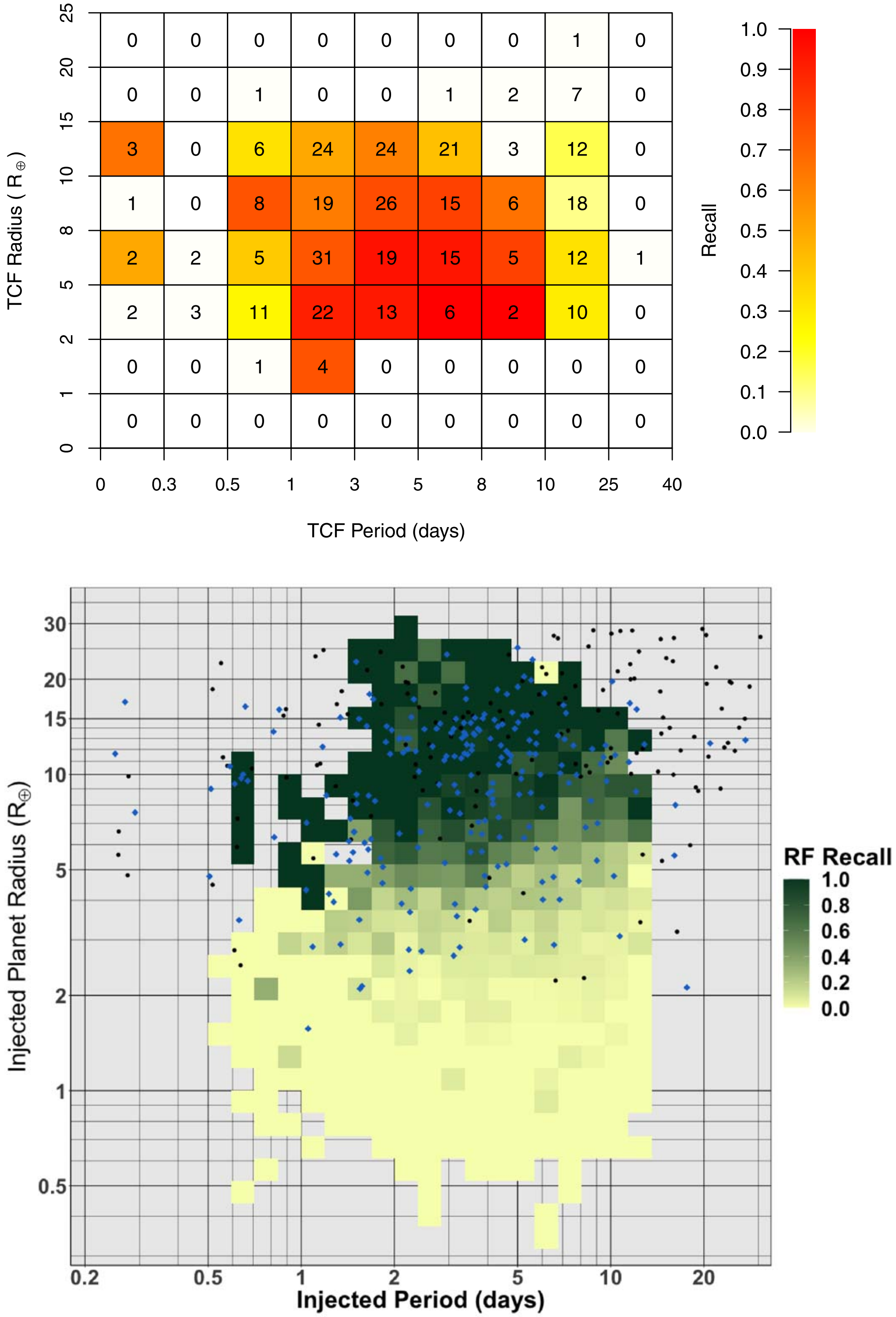


Figure 20. Recall rates by the DTARPS-S RF classifier for 364 M20 DIAmante candidates. (Top) Heat map as a function of planet radius and orbital period found with TCF. The colors give the recall rate, and the numbers are the number of M20 candidates in each period–radius bin. (Bottom) Candidate planets from M20 superposed on the DTARPS-S heat map for injected planets from Figure 17. Blue diamonds are recovered objects, and black points are not recovered.

However, for periods shorter than 0.5 day, DTARPS-S has only moderate recovery of DIAmante candidates. This is likely due to concentration of injected False Positives with TCF periods in this region and TCF radii consistent with planetary objects (Figure 16) that bias the classifier against short-period planet transit signals. DTARPS-S also has poor recovery of the M20 candidates with periods >10 days and radii >10 $R_\oplus$, for reasons explained in Section 8.1.

10.2. NASA Exoplanet Archive Confirmed Planets

Of 3,616 "Confirmed Planet" or "Known Planet" systems in the NASA Exoplanet Archive (NASA Exoplanet Science Institute 2022) or TOI[1] lists (accessed 2022 March 15), 202 were in the DIAmante data set classified by our RF classifier. TCF correctly matched the period for 166 (82%) Confirmed Planets. Of the 36 Confirmed Planets that failed the DTARPS-S selection criteria, nine have periods >13.5 days. But seven other Confirmed Planets with periods 13.5–30 days were recovered, and the 1/3 harmonic of one Confirmed Planet with period >30 days was found. Even though the injected planetary signals have periods restricted to <13.5 days, DTARPS-S is still capable of matching the periods of long-period planets. The $P_{\rm RF} = 0.300$ threshold of the DTARPS-S RF classifier identified 130 of the 202 Confirmed Planets, giving a 64% recall for known exoplanets. All have correctly matched periods.

Figure 21 shows DTARPS-S has a strong recall rate (>50%) for periods between 1 and 10 days, up to radii of 10 $R_\oplus$, similarly to the DIAmante planet candidates. DTARPS-S has poor recall rates for Confirmed Planets in the lower left bins, with periods <1 day and radii <5 $R_\oplus$, likely due to the overpopulation of short TCF periods (Section 8.1). Of the 22 Confirmed Planets in these bins, only three have TCF periods that match the reported period. Recall exceeds 60% for periods 0.5–1.0 day and radii 5–15 $R_\oplus$, but often the recovered periods are incorrect. DTARPS-S thus has poor recovery overall of Confirmed Planets with periods less than 1 day.

10.3. TESS Objects of Interest

The recall coverage of the TESS TOI confirmed planets and planet candidates are included in Sections 10.2 and 10.4, so they are not presented here again. The DTARPS-S sample contains 846 TESS TOIs, of which 140 have reported periods >13.5 days. DTARPS-S matches the periods for 566 of 706 of the remaining TOI planets and candidates (80%). The recall rate of the TOI confirmed planets and planet candidates is 51% with 433 of 846 TOIs.

10.4. Candidate Planets in Other Surveys

Figure 22 shows recall dependencies for the planet candidates in the surveys/lists listed in Section 10 above. The DIAmante data set contains light curves corresponding to 1,042 stars in these candidate planet samples, and it has an overall recall rate of 41% for previously identified planet candidates. The results are similar to the Confirmed Planets and DIAmante sample. DTARPS-S has strong recall for TCF periods 1–10 days and for TCF radii 1–10 $R_\oplus$. The recall rate drops to 13% for TCF periods <1 day. At large planet radii, the recall rate drops to 46% (10–15 $R_\oplus$) and 4% (>15 $R_\oplus$). We note that a handful of planets are recovered with 15–30 days periods despite the absence of injected planet training for these long periods.

As these reported candidates are not yet confirmed by spectroscopic observations, these recovery rates do not reflect true planetary populations. Rather, the value of DTARPS-S is to add confidence to the reality of candidates it confirms, and to cast some doubt on those it does not confirm. Appendix A gives more detail on the overlaps between DTARPS-S and these external surveys/lists.

10.5. False Positives

A list of False Positives was created by combining False Positives from the TOI list[1] available from the NASA Exoplanet Archive (NASA Exoplanet Science Institute 2022; accessed 2022 March 15), the cTOI list[2] available from the TESS Follow-Up Program website (accessed 2022 March 15), and independent studies conducted by Affer et al. (2012), Collins et al. (2018), Mayo et al. (2018), Dressing et al. (2019), Feinstein et al. (2019), Kostov et al. (2019), Kruse et al. (2019), Schanche et al. (2019), von Boetticher et al. (2019), Tu et al. (2020), Eisner et al. (2021), Olmschenk et al. (2021), and Yu et al. (2019). Objects labeled as False Alarms, flare stars, flare stars with a nearby M companion, False Positives, giant stars, probable eclipsing binaries, probable False Positives, pulsing stars, rotating stars, and planetary candidates with no discernible corresponding radial velocity signal were all considered as False Positives for this analysis. If an object was labeled as a planetary candidate by one study and a False Positive by another, then it was considered to be a False Positive object.

There are 513 previously identified False Positives in the full DIAmante data set, of which 390 (76%) have a TCF best period matching the reported period. The DTARPS-S RF classifier correctly identified 400 of the 513 (78%) False Positive signals as non-exoplanet candidates. In the parlance of statistical classification, this 78% is the specificity or the True Negative Rate, i.e., the probability of obtaining a negative test result given that the object is truly negative.

Figure 23 shows the specificity of the RF classifier for different bins in transit signal radius–period space. DTARPS-S has excellent specificity for most periods and radii, even at the extreme values: 90% for periods <1 day, 97% for periods >10 days; and 93% for strong signals with associated radii >15 $R_\oplus$. DTARPS-S thus effectively removes known False Positives over a wide range of parameters.

The specificity in Figure 23 is similar to a hypothetical inverted Figure 22 heat map. DTARPS-S has high specificity where Figure 22 has low recall rates and vice versa. The poorer specificity for periods between 1 and 8 days and radii between 2 and 15 $R_\oplus$ is likely due to the abundance of injected planetary signals in this region (Figure 16(c)). There are a number of Confirmed Planets and previously identified planet candidates in this region as well, confirming the concentration of planetary signals. Although the RF classifier utilizes many features during classification, both the radius and orbital period are important features; they affect the specificity in this region, as well as the recall rate of Confirmed Planets and planetary candidates at the extreme values.

10.6. Random Forest Recall Dependence on ARIMA and TCF

The overall recall rates reported in Sections 10.1–10.4 and the completeness heat map in Figure 17 are dependent on two aspects of the DTARPS-S process: the ability of the TCF periodogram to recover the previously identified signal from ARIMA residuals and the ability of the RF classifier to recognize the planetary transit signal. The RF classifier was trained only using injected planet signals whose TCF peak period matched the injected period. So the RF classifier is dependent (as with all planetary classification systems) on the TCF peak period.

As shown in Section 8, the accuracy of the best TCF periodogram peak orbital parameters depends heavily on the best TCF peak period matching the injected (or in these cases, the existing) period. The ability of the RF classifier to identify Confirmed Planets and previously identified planet candidates can, therefore, be more accurately reported by calculating the recall rate only for Confirmed Planets and planet candidates whose TCF peak periods match their reported periods. Table 4 compares the recall rates reported in Sections 10.1–10.4 with recall rates calculated using only objects in each category with TCF periods that matched their reported period. The biggest

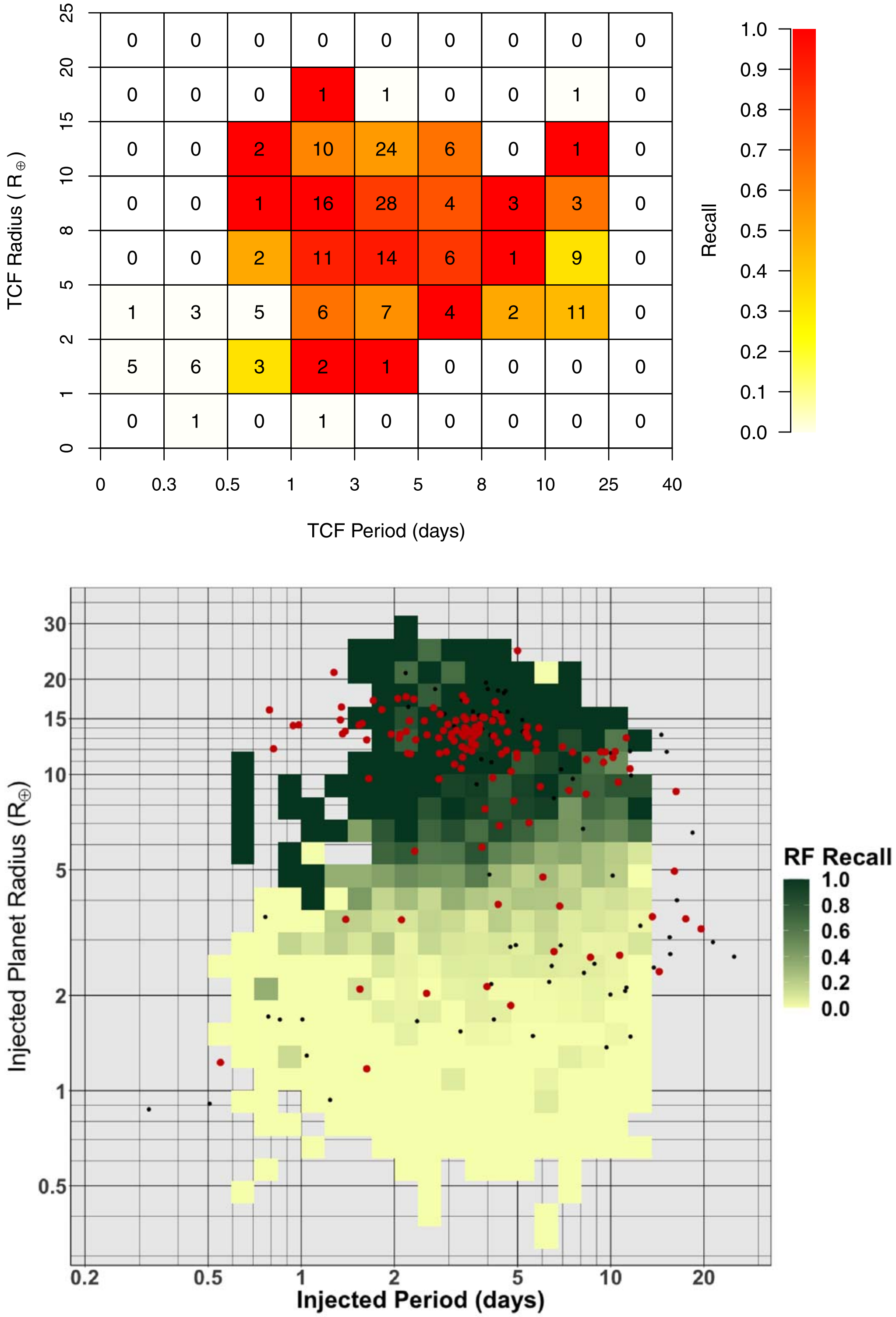


Figure 21. Recall rates by the Random Forest classifier for 202 Confirmed Planets in the NASA Exoplanet Archive (NASA Exoplanet Science Institute 2022). (Top) Heat map as a function of planet radius and orbital period found with TCF. (Bottom) Confirmed planets from the NASA Exoplanet Archive superposed on the DTARPS-S heat map for injected planets from Figure 17. Red circles are recovered objects and black points are not recovered.

improvement in the recall rate was the recall rate for the Confirmed Planets and the planet candidates in the TOI list. The recall rate for the full set represents what DTARPS-S can recover from the DIAmante data set, and the recall rates for the matched periods represent what DTARPS-S can recover from the processed DIAmante data whose best TCF peak corresponds to real periodicity in the light curve.

11. Discussion

11.1. Overview of the Study

This study is based on the premise that existing searches for transiting exoplanets—even those conducted by the TESS Science Office producing the TOI lists—have not identified the full detectable population of planetary systems in TESS FFF light

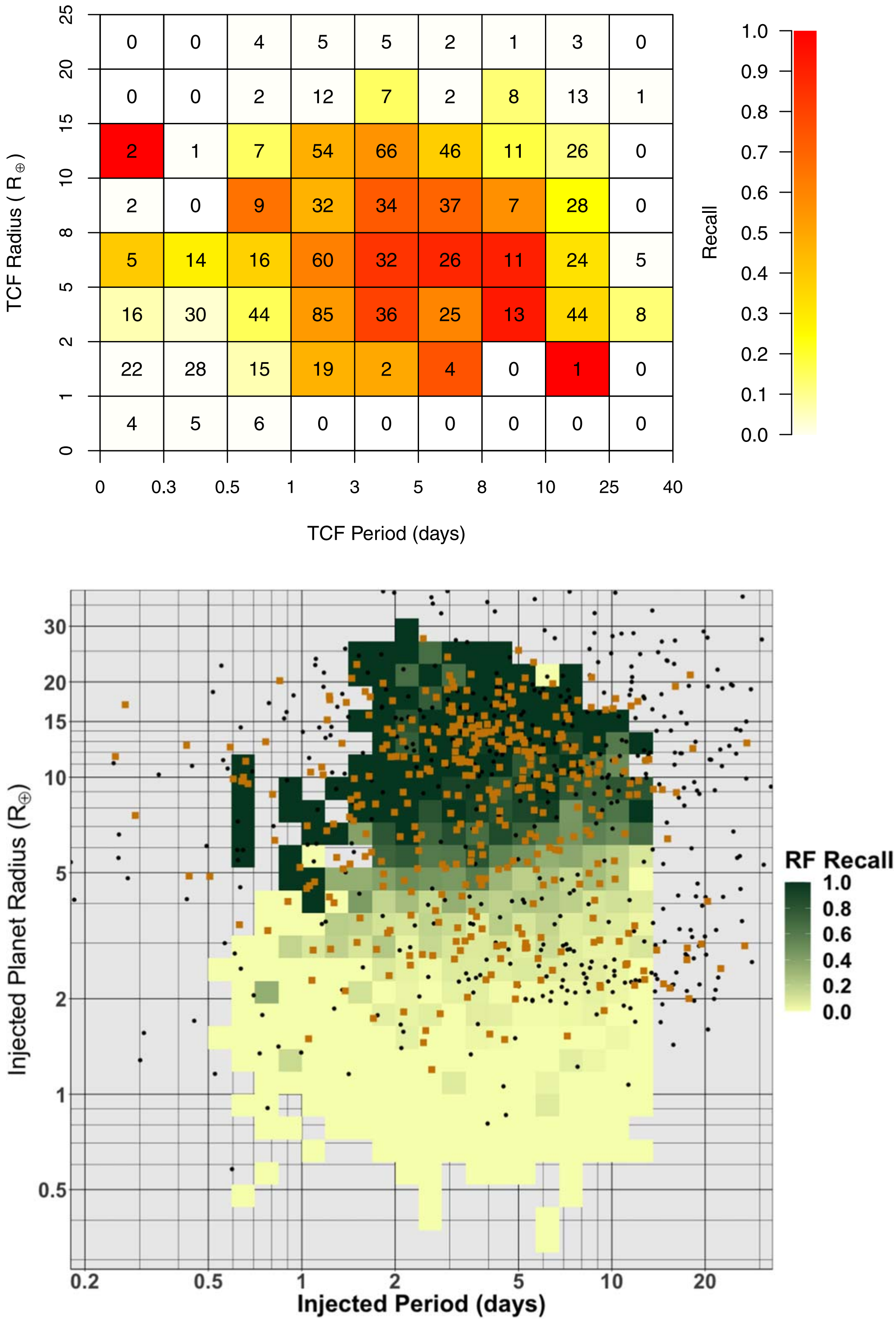


Figure 22. Recall rates by the Random Forest classifier for 1,042 previously identified planet candidates from various surveys. (Top) Heat map as a function of planet radius and orbital period found with TCF. (Bottom) Candidate planets from various surveys superposed on the DTARPS-S heat map for injected planets from Figure 17. Gold squares are recovered objects and black points are not recovered.

curves (Section 1.1). The sensitivity and reliability of transit search depends critically on the development and refinement of statistical methodology focused on the complexities of this scientific problem. There are many challenging issues: a wide variety of contaminating stellar and instrumental signatures in the light curves; a highly imbalanced classification problem with imperfect training sets; and limited telescope time to validate the resulting planetary candidates.

We adopt and refine the AutoRegressive Planet Search (ARPS) procedure developed by Caceres et al. (2019b) in an effort called DIAmante TESS AutoRegressive Planet Search for the southern ecliptic hemisphere, or DTARPS-S. It combines the time series extraction and preprocessing from the DIAmante project (Montalto et al. 2020, M20), Box–Jenkins analysis (ARIMA modeling) for light-curve detrending, our Transit Comb Filter (TCF) periodogram for

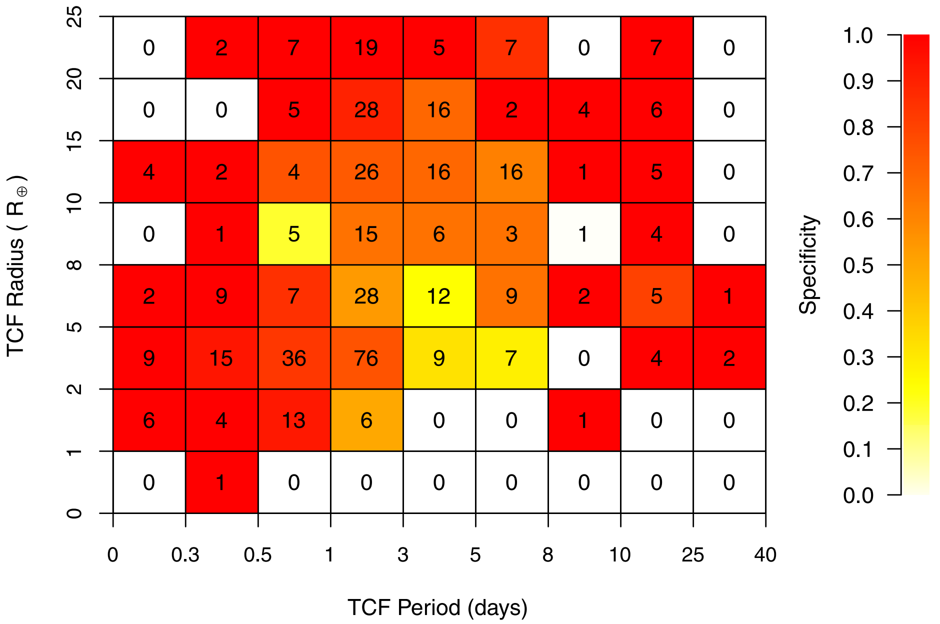


Figure 23. Specificity of DTARPS-S RF classifier for 513 previously identified False Positives. A red colored bin represents a high percentage of False Positives with correct classification as a non-planetary signal. The numbers of False Positives in each bin are labeled.

Table 4
Recall Rates for the DTARPS-S (Full Set) and the RF Classifier Alone (Matching Period)

	M20 Candidates	Confirmed Planets	TESS Objects of Interest	Combined Candidates[a]
Recall for Full Set (%)	58.5	64.4	51.2	41.2
Recall for Matched Periods Only (%)	64.0	78.8	68.2	54.7

Note.
[a] Planet Candidates from the TOI list[1], cTOI list[2], Mayo et al. (2018), Dressing et al. (2019), Feinstein et al. (2019), Kostov et al. (2019), Kruse et al. (2019), Yu et al. (2019), M20, Dong et al. (2021), Eisner et al. (2021), and Olmschenk et al. (2021).

transit discovery, and machine learning with Random Forest for optimizing True Positive and minimizing False Positive classifications. We apply the procedure to ∼0.9 million TESS Year 1 Full Field Image (FFI) light curves for brighter stars in the southern ecliptic hemisphere (Section 2.4). Best-fit ARIMA models, fitted by maximum likelihood estimation with optimized model complexity, are subtracted from the DIAmante light curves (Section 2.2.1). The TCF periodogram is then calculated to identify and characterize periodic transit-like behavior in the light curves (Section 3.2). Statistical virtues of TCF have been eludicated by Gondhalekar et al. (2023). The most likely transit signal period was chosen as the TCF periodogram peak with the greatest robust signal-to-noise ratio after detrending the periodogram.

Considerable effort was expended to tune a Random Forest (RF) machine-learning classifier that identifies exoplanet transit candidates while reducing various sources of contamination (Section 4). The positive training set was constructed from synthetic planetary injections augmented from confirmed planets in the 4 yr Kepler survey, and the negative training set of random light curves was supplemented with synthetic eclipsing binary (EB) and short-period variable injections from M20. Several dozen features drawn from every stage of the analysis were combined with stellar metadata to construct aN RF classifier; the final classifier has 37 features with different weights (Section 6).

After choice of a threshold of RF prediction value, we produce a list of 7377 objects called the DTARPS-S Analysis List (Section 7). While it has high recall of True Positives, it is particularly optimized to have a small False Positive Rate. The classifier performance is summarized in Figures 11, 12, and 13. The list has a True Positive Rate of 92.5% with respect to injections of simulated planets, and False Positive Rate of 0.43% with respect to simulated astrophysical False Positives (mostly EBs) and random light curves (Section 10.2). We compare the DTARPS-S Analysis List to other southern hemisphere samples: NASA Exoplanet Archive Confirmed Planet (NASA Exoplanet Science Institute 2022), TESS Objects of Interest, and other transit surveys (Section 10, Appendix).

Our classifier has imperfections. Smaller injected planets are not recovered by the TCF fitting algorithm, and the TCF transit depth (scaling to planet radius) is somewhat underestimated. This effect stems from overfitting by the ARIMA modeling and underfitting by the TCF-matched filter procedure (Section 8). In addition, large-radii stellar companions are fitted with planetary-radii signals, thereby biasing the RF classifier against longer-period (>8 days) Jovian planets and short-period (<1 day) planets.

The DTARPS-S completeness heat map of the injected planetary signals for the RF classifier (Figure 17) shows that the DTARPS-S method has poor recall for radii $<2\ R_\oplus$ or periods <1 day, low completeness for planets with radii between 2 and 4 $R_\oplus$, and high completeness for planes with radii between 8 and 30 $R_\oplus$ and periods between 0.6 and 13 days. The distribution of the recall for the confirmed planets, M20 candidates, and previously identified candidates generally follow the results of the injected planet completeness map (Section 9).

The principal product of this paper, the DTARPS-S Analysis List (Table 3, available as a machine-readable table), optimizes recall (completeness) at the expense of precision (acceptance of False Positives). It serves three purposes:

Potential transiting planets for spectroscopic follow-up: This would proceed with the understanding that more than half are likely to be False Alarms (no real periodicity) and False Positives (non-planetary periodicity). It may be particularly useful for subsets such as very bright host stars.

Intermediate list ready for vetting: Vetting will increase precision, greatly reducing False Alarms and False Positives, but with reduced recall (completeness). This how we proceed in Paper II; see Section 11.3 for discussion.

Support for other surveys: If a star in the DTARPS-S Analysis List was independently found to be an unconfirmed planetary candidate in the TOI list or another transit search procedure, then confidence in its planetary nature is increased.

11.2. Improvements to DTARPS-S Methodology

The statistical issues arising in reliable planetary transit identification are complex and differ with each survey. The success of our DTARPS-S effort is based on the ARPS procedures developed in Caceres et al. (2019b) and applied to the Kepler data set by Caceres et al. (2019a). We institute a variety of improvements to their methods in our application to TESS Year 1 light curves: injection-based training sets for both planetary transits and eclipsing binaries with sophisticated data augmentation procedures (Section 4.1); optimized Random Forest algorithm for imbalanced training set (Section 5); extensive engineering of feature selection including Gaia information (Section 5); multiple metrics for classification performance (Section 5.2); classification training and validation using both injections and Confirmed Planet samples (Sections 4, 9); and completeness heat maps (Section 9). Based on the results presented here for application to TESS, a number of further improvement can be envisioned:

Stellar variation removal: The linear ARIMA model seems effective in removing autocorrelated trend for $\gtrsim$90% of the DIAmante preprocessed TESS light curves considered here (Figure 7). However, ARIMA was less effective over the 4 yr Kepler data, where only 47% of the ARIMA residuals were consistent with white noise (Caceres et al. 2019a). More elaborate nonlinear autoregressive models might be better for TESS light curves near the ecliptic poles or multiyear light curves. However, detailed examination is needed in order to reduce overfitting of ARIMA and ARIMAX models for blended eclipsing binaries. Overfitting may disguise photometric behaviors arising from tidally distorted and mutually illuminating close binaries and erroneously lead to classification as transiting planets.

Transit depth estimation: The ARIMAX modeling often gave biased estimates based on the ARIMA residuals because some of the planetary signal is incorporated into the ARIMA model (Section 2.2.3). This may not have affected the Random Forest classifier greatly, as we use the ARIMAX S/N, rather than the ARIMAX depth, as a feature. However, it does affect the astronomical interpretation, and corrections to the estimated planet radius are therefore instituted in Paper II. Also, the trapezoidal-shaped model used as the exogenous variable is probably too simple. A more accurate exoplanet transit with curved ingress and egress is likely to be more effective, as in the Transit Least Squares procedure for identifying exoplanet transits (Hippke & Heller 2019), at the expense of adding astrophysical parameters to the statistical model.

TCF sensitivity: For Kepler 4 yr light curves, the Transit Comb Filter periodogram appears to be more sensitive to smaller planets than other periodicity search methods such as the Box Least Squares periodogram (Figures 9–10 in Caceres et al. 2019b; Figure 10 in Caceres et al. 2019a; Gondhalekar et al. 2023). However, in our TESS FFI application, the TCF periodogram had a low recall rate for injected planets with radii $\lesssim 4\ R_\oplus$. The reason for this difference needs to be elucidated. Is it a product of the number of points in the light curve available or is there another factor? Further investigation of TCF, BLS, and other periodograms is needed. This can lead to discovery of smaller planets in a given data set, which is a driving goal of the TESS and upcoming PLATO missions (Rauer et al. 2014).

Multiple planet systems: Currently, the ARPS procedure only treats the TCF periodogram peak with the strongest S/N and does not search for nor consider multiple transiting planets. Multi-planet systems could be searched for by an iterative "pre-whitening" procedure: the strongest planetary signal can be subtracted from the light curve, and ARIMA and TCF can be reapplied. The procedure would be repeated until the TCF peak effective signal-to-noise ratio falls below the threshold indicated in Figure 18.

Classifier features; New features can be added to the Random Forest classifier in order to better identify true exoplanets or astrophysical False Positives. A feature that quantifies the difference between the TCF periodogram peak and the spurious spike of non-exoplanet transits with a period between 13.5 and 15 days due to the TESS satellite orbit (Figure 13) might mitigate the bias against longer-period exoplanets seen in the current DTARPS-S classifier. Features that characterize an EB secondary eclipse or tidal distortions may help reduce EB contamination and reduce the human vetting effort.

Classifier training set: The training set planet properties were derived from injections based on confirmed planets in the Kepler sample. But with 4 yr light curves, rather than the 1 month typical of TESS FFIs, most of the injected planets are too small. The large number of undetectable planets in the positive training set may have distorted the classifier. A better match to TESS sensitivity might improve classifier performance. In addition, a larger number of planet injections may be helpful in regions of the period–radius diagram where the recall rate is transitioning between low and high (2–5 $R_\oplus$), where true hot Jupiters compete with False Positive EBs, and near the edges of the heat map (Figure 17). Finally, the distribution of injected EB signals might be adjusted to approximate the expected dilution in blended systems.

Specialized classifiers: Random Forest or other classifiers might be trained for particular subpopulations of TESS FFI stars, such as Sun-like stars, lower-mass K and M stars, subgiants, or stars in the continuous observing zone near the ecliptic poles. This would require new training sets of injections on light curves of just these stellar host subpopulations.

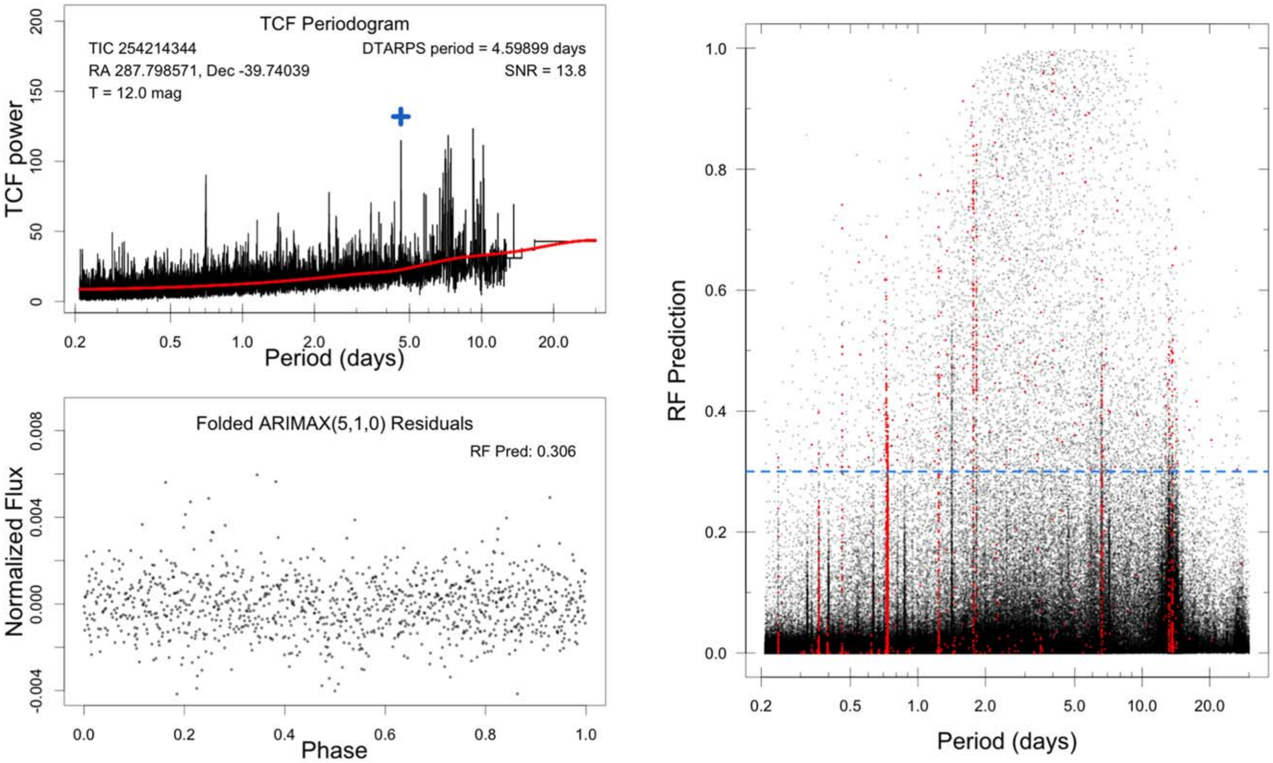


Figure 24. Two sources of contamination that affect the DTARPS-S Analysis List of potential planetary transits. (Left) TCF Periodogram and folded ARIMAX residual light curve of TIC 254214344 that shows an unconvincing peak in the TCF periodogram with no clear transit in the folded light curve. (Right) Random Forest prediction score for the full set of DIAmante-extracted light curves showing clusters of stars subject to ephemeris matching and periodic satellite operations as red points.

Classifier type: The Random Forest classifier developed in Section 5 is highly effective, but the 0.43% False Positive rate is high, given the huge class imbalance. This necessitates a complex vetting process (Paper II). Improved performance might be achieved with different machine learning classifiers such as XGBoost, LightGBM, GOSDT, or Explainable Boosting Machines. XGBoost (Chen & Guestrin 2016) is very similar to Random Forest, but rather than building decision trees independently, new trees are built iteratively to minimize classification error. LightGBM (Ke et al. 2017) is also similar to Random Forest and XGBoost but grows trees leaf-wise rather than level-wise. Explainable Boosting Machines (Lou et al. 2012) use small forests of decision trees for each feature in a linear regression ensemble. GOSDT (Lin et al. 2020) is a modern approach to fast decision tree optimization.

11.3. The Need for Vetting

A False Positive Rate of 0.43% multiplied by a test sample of 823,099 DIAmante light curves predicts that ∼3500 False Positives, about half of 7377, will be present in the DTARPS-S Analysis List of stars passing the classifier threshold. This includes possible False Alarms (i.e., cases where no significant periodicity is present despite the high RF classification probability) or False Positive signals from unmodeled stellar or instrumental variability. Despite our careful statistical efforts, the DTARPS-S Analysis List is dominated by non-planetary signals. The full list thus cannot be accepted for reliable calculation of exoplanet populations, such as converting the completeness heat maps into planetary occurrence rates, and it is an inefficient list for follow-up observations and spectroscopy with valuable telescope resources.

Fortunately, a suite of "vetting" operations can be conducted to cull many of the False Positives and False Alarms. Two examples of the effects to be removed are shown in Figure 24: A TCF periodogram and folded light curve without convincing signal that nonetheless passed the RF threshold; and groups of stars with similar periods due to satellite operations associated with the 13.7 day orbit or light contamination from a bright star (this is known as the "ephemeris matching" problem; Coughlin et al. 2014). Other effects to be removed include centroid wobbling in the FFI image, possible photometric binaries in Gaia photometry, and deep secondary eclipses inconsistent with planetary radii. These and other vetting operations are applied in Paper II to give a more reliable, though less complete, catalog of DTARPS-S Candidate Planets for the Year 1 TESS FFI DIAmante light curves.

Acknowledgments

E.J.M. and E.D.F. benefit from the vibrant community of Penn State's Center for Exoplanets and Habitable Worlds that is supported by the Pennsylvania State University and the Eberly College of Science. This study is also a product of the Center for Astrostatistics supported by the Eberly College of Science. We benefited from comments on the manuscript by members of these centers: Ian Czekela, Rebekah Dawson, Hyungsuk Tak, Jason Wright, as well as Joel Hartman (Princeton). We also appreciate the comments provided by a thoughtful anonymous referee.

This paper includes data collected by the TESS mission. Funding for the TESS mission is provided by NASA's Science Mission directorate. We acknowledge the use of public TOI Release data from pipelines at the TESS Science Office and at the TESS Science Processing Operations Center. This research also uses the NASA Exoplanet Archive (NASA Exoplanet Science Institute 2022) operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program. Data from the European Space Agency mission Gaia, processed by the Gaia Data Processing and Analysis Consortium (DPAC). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the Gaia Multilateral Agreement.

We acknowledge use of *Lightkurve* software, a Python package for Kepler and TESS data analysis (Lightkurve Collaboration et al. 2018), as well as Astropy, a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013, 2018).

Facilities: TESS, Gaia.

Software: R core (R Core Team 2020): *forecast* (Hyndman & Athanasopoulos 2021), *goftest* (Faraway et al. 2019), *moments* (Komsta & Novomestky 2015), *nortest* (Gross et al. 2015), *randomForestSRC* (Ishwaran 2022), *ROCR* (Sing et al. 2005), *smotefamily* (Siriseriwan 2019), *tseries* (Trapletti & Hornik 2019), *KernSmooth* (Wand 2020); Python: `Jupyter` (Kluyver et al. 2016), `LightKurve` (Lightkurve Collaboration et al. 2018), `Astropy` (Astropy Collaboration et al. 2013, 2018), `Astroquery` (Ginsburg et al. 2019), `tesscut` (Brasseur et al. 2019), `NumPy` (van der Walt et al. 2011; Harris et al. 2020), `Matplotlib` (Hunter 2007; Droettboom et al. 2016).

Appendix
Other Planet Searches

External lists and surveys that are compared to DIAmante samples and DTARPS-S Analysis List are briefly described here. The list of confirmed planets was taken from the list of confirmed planets on the NASA Exoplanet Archive (NEA; accessed 2022 March 15), some of which are TESS Objects of Interest[1] (TOIs). Previously identified exoplanet candidates were found in the list of community TOIs[2] (cTOIs) at the TESS EXOFOP website (accessed 2022 March 15) with contributions from Mayo et al. (2018), Dressing et al. (2019), Feinstein et al. (2019), Kostov et al. (2019), Kruse et al. (2019), Yu et al. (2019), M20, Dong et al. (2021), Eisner et al. (2021), and Olmschenk et al. (2021). Previously identified astrophysical False Positives were found in Affer et al. (2012), Collins et al. (2018), Schanche et al. (2019), von Boetticher et al. (2019), and Tu et al. (2020).

A.1. Confirmed Planets from NEA

The Confirmed Planets used here include planets with published refereed planet confirmation papers. The NEA Confirmed Planets in the DIAmante data set were identified with the TIC ID number. In the case of multi-planet systems, we used the planet whose reported period best matched the TCF peak period. For planets with multiple entries, an average of reported orbital parameters were used. In total, 184 Confirmed Planet hosts on the NASA Exoplanet Archive were matched with objects in the DIAmante data set.

A.2. TOI List

The TOI list[1] reports dispositions of known planet (KP), confirmed planet (CP), planetary candidate (PC), ambiguous planetary candidate (APC), or False Alarm or False Positive (FP). We combined CPs and KPs to be confirmed planets. Objects with recent confirmation in unpublished papers on arXiv are considered to be Confirmed Planets. We combined the APCs and PCs when considering planet candidates, and combined FAs and FPs in when considering False Positives.

Of the 1,036 objects in the TOI catalog that overlap with the DIAmante data set, 185 were labeled confirmed planets, 670 were labeled planetary candidates, and 181 were labeled False Positives.

A.3. cTOI List

The cTOI list[2] has a wide range in the quality of planet candidates: follow-up EXOfop examination shows some are False Positives while others are promoted to planetary candidates on the TOI list. The DIAmante sample has 566 cTOIs; 364 are planetary candidates from M20, discussed below.

A.4. Affer et al. (2012)

Affer et al. (2012) measured rotation and binarity of field stars from the Convection Rotation and planetary Transits (CoRoT) satellite for stars in the solar neighborhood. Forty objects in Table 2 of Affer et al. (2012) were matched with objects in the DIAmante data set, one of which is in our DTARPS-S Analysis List. Affer et al. report a rotation period of 72 days for TIC 234091431, and we report a TCF period of 2.76592 days. Of the other 39 objects, only three have rotational or pulsational periods or pulsation periods that matched the TCF peak period. We label these as False Positives in the DIAmante data set.

A.5. Collins et al. (2018)

Collins et al. (2018) identified and classified False Positives in Kilodegree Extremely Little Telescope (KELT) light curves. They classified over 1000 transit-like signals in KELT as False Positives through photometric and spectroscopic observations in several classes: single-line spectroscopic binaries, multi-line spectroscopic binaries, spectroscopic giant stars, eclipsing binaries, blended eclipsing binaries, variable stars, nearby eclipsing binaries (blended in the KELT aperture), and stars with no significant radial velocity detected. The DIAmante samples has 156 objects matched in Collins et al. (2018), 19 of which are in the DTARPS-S Analysis List. We consider these to be previously identified False Positives.

A.6. Mayo et al. (2018)

Mayo et al. (2018) identified 275 planet candidates in the NASA's K2 mission, Campaigns 0-10, and estimated False Positive probability with the *vespa* package. The DIAmante sample has 21 objects examined by Mayo et al. (2018), one of which, TIC 21184505, is in the DTARPS-S Analysis List. Another object, TIC 68694240, was a probable eclipsing binary that we label as a False Positive.

A.7. Dressing et al. (2019)

Dressing et al. (2019) performed spectroscopic and photometric characterization for 172 K2 target stars identified as candidate hosts of transiting planets. They identified giants, likely eclipsing binaries, and cool dwarf stars. The DIAmante sample matches eight of these stars with one, TIC 438338723, in the DTARPS-S Analysis List. It is a probable eclipsing binary that we label as a False Positive.

A.8. Feinstein et al. (2019)

Feinstein et al. (2019) developed `elenor`, an open-source tool for extracting light curves from TESS FFIs. They applied the method to TESS Sector 1 Year 1 data and vetted by visual examination. The DIAmante sample matches 16 of their objects, three of which are in the DTARPS-S Analysis List: TIC 159835004 and TIC 299780329 previously identified as planetary candidates, and TIC 38813184 is identified as an eclipsing binary. The reported EB period for TIC 38813184 matches the TCF peak period. We include TIC 38813184 in the previously identified False Positive list.

A.9. Kostov et al. (2019)

Kostov et al. (2019) created an open source automatic vetting pipeline for K2 data called Discovery and Vetting of Exoplanets (DAVE). They applied DAVE to 772 planet candidates from K2 and vetted the candidates either as planet candidates or False Positives. Of the 30 objects that match the DIAmante stars, TIC 21184505, TIC 294301883, and TIC 366443576 in the DTARPS-S Analysis List were labeled as planetary candidates by DAVE. All three objects had a TCF peak period that matched the reported DAVE period.

A.10. Kruse et al. (2019)

Kruse et al. (2019) identified 818 planetary candidates and 1060 eclipsing binary systems in Campaigns 0-8 of the K2 mission using the EVEREST pipeline. The DIAmante samples matches 44 objects, two of which are in the DTARPS-S Analysis List: planetary candidate TIC 294301883 and eclipsing binary TIC 438338723. The reported periods from Kruse et al. match the TCF peak period.

A.11. Schanche et al. (2019)

Schanche et al. (2019) presented a catalog of 1,041 False Positives from the SuperWASP survey of the northern hemisphere that had previously been identified as potential planetary candidates but were rejected after follow-up observations. The False Positives were classified as eclipsing binaries, blended eclipsing binaries, and low-mass eclipsing binaries. The DIAmante sample matches 47 objects, 12 of which lie in the DTARPS-S Analysis List with the following classifications: TIC 16490297 was labeled as an eclipsing binary system; TIC 61069470, TIC 117549305, TIC 13675776, TIC 271269442, and TIC 271374913 as blended eclipsing binaries; and TIC 9433212, TIC 12529950, TIC 264537668, TIC 277712294, TIC 443618156, and TIC 449050248 as low-mass eclipsing binary systems. Most, but not all, have TCF periods matching the SuperWASP periods.

A.12. von Boetticher et al. (2019)

von Boetticher et al. (2019) characterized 10 low-mass stars part of low-mass eclipsing binary systems as part of the EBLM project. The DIAmante sample has six of these systems with TCF periods matching the reported period. Four lie in the DTARPS-S Analysis List: TIC 101395259, TIC 277712294, TIC 350480660, and TIC 734505581.

A.13. Yu et al. (2019)

Yu et al. (2019) modified an neural network classifier, developed by Shallue & Vanderburg (2018) for identifying Kepler planet candidates, for TESS data. Applying the classifier to Year 1 Sector 6 TESS data, and accompanied by visual vetting, 288 new planetary candidates were identified. The DIAmante sample matches 140 objects, of which 65 are in the DTARPS-S Analysis List. In all cases, the TCF peak period agreed with the period reported in Yu et al. We label these as previously identified planetary candidates.

A.14. Montalto et al. (2020)

Of the 394 candidates identified by M20 in the DIAmante study, 364 were in the set of light curves classified by the DTARPS-S Random Forest. These are identified by a flag in the DTARPS-S Analysis List; see Section 10.1 for details. These include 221 in the NEA Confirmed Planet list, the TOI list on the NEA, or in other external surveys. Altogether, 82 are identified as Confirmed Planets. The M20 objects were placed on the cTOI list: 82 are labeled as planet candidates, 18 as ambiguous planet candidates, and 26 as False Positives. These include 13 DIAmante candidates independently listed as planetary candidates by Yu et al. (2019), and two identified as low-mass EBs by von Boetticher et al. (2019).

A.15. Tu et al. (2020)

Tu et al. (2020) studied superflares and other properties of 400 solar-type stars in TESS Year 1 data. Of the 277 stars in the DIAmante data set that had flares identified by Tu et al., only 57 had TCF peak periods that matched their stellar rotational periods. Six flare stars are in the DTARPS-S Analysis List. In the two cases where stellar rotational period matched the TCF peak period, TIC 121048789 and TIC 373844472, DTARPS-S is likely identifying the rotational period rather than a transiting planetary period.

A.16. Dong et al. (2021)

Dong et al. (2021) identified and characterized 55 Warm Jupiters in TESS Year 1 FFIs. The DIAmante sample has 40 of these systems, of which 21 lie in the DTARPS-S Analysis List. Of these, 20 had TCF periods matching those reported by Dong et al.; the exception is TIC 73038411.

A.17. Eisner et al. (2021)

Eisner et al. (2021) presented results from the Planet Hunters TESS citizen science project for the first two years of the TESS survey. They identified 90 new planetary candidates, of which 18 lie in the DIAmante sample. However, none of the overlapped objects have a TCF peak period that matches the reported period from Eisner et al.. This is partly due to their single transit events where the period of the planet was estimated from the transit duration. Two of their objects are in the DTARPS-S Analysis List. TIC 142087638 and TIC 404518509 were identified as single-transit events by Eisner et al. (2021); the TCF periodogram gives accurate periods within the error bars of their single-transit event estimate.

A.18. Olmschenk et al. (2021)

Olmschenk et al. (2021) applied a convolutional neural network to TESS FFI light curves to identify planetary candidates, followed by visual vetting. Of their 185 planet candidates, 25 overlap with the DIAmante sample, all of which have TCF peak periods that match their reported periods. The DTARPS-S Analysis List recovers 15 of their planet candidates.

ORCID iDs

Eric D. Feigelson https://orcid.org/0000-0002-5077-6734
Marco Montalto https://orcid.org/0000-0002-7618-8308

References

Affer, L., Micela, G., Favata, F., & Flaccomio, E. 2012, MNRAS, 424, 11

Akosa, J. S. 2017, Proc. SAS Global Forum 942-2017, https://support.sas.com/resources/papers/proceedings17/0942-2017.pdf

Alard, C., & Lupton, R. H. 1998, ApJ, 503, 325

Angus, R., Morton, T., Aigrain, S., Foreman-Mackey, D., & Rajpaul, V. 2018, MNRAS, 474, 2094

Ansdell, M., Ioannou, Y., Osborn, H. P., et al. 2018, ApJL, 869, L7

Armstrong, D. J., Günther, M. N., McCormac, J., et al. 2018, MNRAS, 478, 4225

Aschwanden, M. J., & McTiernan, J. M. 2010, ApJ, 717, 683

Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B., et al. 2018, AJ, 156, 123

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33

Barclay, T., Pepper, J., & Quintana, E. V. 2018, ApJS, 239, 2

Boisse, I., Bouchy, F., Hébrard, G., et al. 2011, A&A, 528, A4

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. 2015, Time Series Analysis: Forecasting and Control (New York: Wiley), https://books.google.com/books?id=rNt5CgAAQBAJ

Brasseur, C. E., Phillip, C., Fleming, S. W., Mullally, S. E., & White, R. L. 2019, Astrocut: Tools for creating cutouts of TESS images, Astrophysics Source Code Library, ascl:1905.007

Breiman, L. 2001, ML, 45, 5

Breiman, L., Bouchy, F., Hébrard, G., et al. 1984, Classification and Regression Trees (New York: Chapman and Hall)

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

Caceres, G. A., & Feigelson, E. D. 2022, TCF: Transit Comb Filter periodogram, Astrophysics Source Code Library, ascl:2206.002

Caceres, G. A., Feigelson, E. D., Jogesh Babu, G., et al. 2019b, AJ, 158, 58

Caceres, G. A., Feigelson, E. D., Jogesh Babu, G., et al. 2019a, AJ, 158, 57

Cañas, C. I., Mahadevan, S., Cochran, W. D., et al. 2022, AJ, 163, 3

Chakraborty, J., Wheeler, A., & Kipping, D. 2020, MNRAS, 499, 4011

Chatfield, C., & Xing, H. 2019, The Analysis of Time Series: An Introduction with R (Boca Raton, FL: CRC Press)

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. 2002, JAIR, 16, 321

Chen, C., Liaw, A., & Breiman, L. 2004, Using Random Forest to Learn Imbalanced Data 666, Berkeley, https://statistics.berkeley.edu/tech-reports/666

Chen, T., & Guestrin, C. 2016, Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, KDD '16 (New York: ACM), 785

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. 2016, ApJ, 828, 99

Christiansen, J. L., Clarke, B. D., Burke, C. J., et al. 2020, AJ, 160, 159

Cleveland, W. S., & Devlin, S. J. 1988, JASA, 83, 596

Collins, K. A., Collins, K. I., Pepper, J., et al. 2018, AJ, 156, 234

Coughlin, J. L., Mullally, F., Thompson, S. E., et al. 2016, ApJS, 224, 12

Coughlin, J. L., Thompson, S. E., Bryson, S. T., et al. 2014, AJ, 147, 119

Davenport, J. R. A. 2016, ApJ, 829, 23

Delisle, J. B., Hara, N., & Ségransan, D. 2020, A&A, 635, A83

Dong, J., Huang, C. X., Dawson, R. I., et al. 2021, ApJS, 255, 6

Dressing, C. D., Hardegree-Ullman, K., Schlieder, J. E., et al. 2019, AJ, 158, 87

Droettboom, M., Hunter, J., Caswell, T. A., et al. 2016, matplotlib: matplotlib, v1.5.1, Zenodo, doi:10.5281/zenodo.44579

Eisner, N. L., Barragán, O., Lintott, C., et al. 2021, MNRAS, 501, 4669

Faraway, J., Marsaglia, G., Marsaglia, J., Baddeley, A., et al. 2019, goftest: Classical Goodness-of-Fit Tests for Univariate Distributions, https://CRAN.R-project.org/package=goftest

Feigelson, E. D., Babu, G. J., & Caceres, G. A. 2018, FrP, 6, 80

Feinstein, A. D., Montet, B. T., Foreman-Mackey, D., et al. 2019, PASP, 131, 094502

Feliz, D. L., Plavchan, P., Bianco, S. N., et al. 2021, AJ, 161, 247

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1

Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1

Genuer, R., Poggi, J.-M., & Tuleau-Malot, C. 2010, PaReL, 31, 2225

Gilliland, R. L., Chaplin, W. J., Dunham, E. W., et al. 2011, ApJS, 197, 6

Ginsburg, A., Sipőcz, B. M., Brasseur, C. E., et al. 2019, AJ, 157, 98

Gondhalekar, Y. P., Feigelson, E. D., Montalto, M., & Saha, S. 2023, ApJL, 959, L16

Greco, G., Kondrashov, D., Kobayashi, S., et al. 2016, in ASSP 42, The Universe of Digital Sky Surveys, ed. N. R. Napolitano et al. (Cham: Springer), 105

Gross, J., & Ligges, U. 2015, nortest: Tests for Normality, https://CRAN.R-project.org/package=nortest

Guerrero, N. M., Seager, S., Huang, C. X., et al. 2021, ApJS, 254, 39

Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Natur, 585, 357

Hippke, M., & Heller, R. 2019, A&A, 623, A39

Howard, A. W., Marcy, G. W., Bryson, S. T., et al. 2012, ApJS, 201, 15

Huang, C. X., Vanderburg, A., Pál, A., et al. 2020, RNAAS, 4, 204

Huijse, P., Estevez, P. A., Protopapas, P., Zegers, P., & Principe, J. C. 2012, ITSP, 60, 5135

Hunter, J. D. 2007, CSE, 9, 90

Hyndman, R., & Athanasopoulos, G. 2021, Forecasting: Principles and Practice (Melbourne: OTexts), https://otexts.com/fpp3/

Ishwaran, H., & Kogalur, U. 2022, Fast Unified Random Forests for Survival, Regression, and Classification RF-SRC, v3.2.3, https://cran.r-project.org/package=randomForestSRC

Jara-Maldonado, M., Alarcon-Aquino, V., Rosas-Romero, R., Oleg, S., & Ramirez-Cortes, J. 2020, Earth Sci. Inform., 13, 573

Jenkins, J. M., Seader, S., & Burke, C. J. 2017a, Kepler Data Processing Handbook: A Statistical Bootstrap Test KSCI-19081-002, https://archive.stsci.edu/kepler/manuals/KSCI-19081-002-KDPH.pdf

Jenkins, J. M., Tenenbaum, P., Seader, S., et al. 2017b, Kepler Data Processing Handbook KSCI-19081-003, https://archive.stsci.edu/kepler/manuals/KSCI-19081-002-KDPH.pdf

Jenkins, J. M., Twicken, J. D., McCauliff, S., et al. 2016, Proc. SPIE, 9913, 99133E

Ke, G., Meng, Q., Finely, T., et al. 2017, in Advances in Neural Information Processing Systems 30 NIPS 2017, ed. I. Guyon et al. (New Orleans, LA: NIPS), https://papers.nips.cc/paper_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html

Kluyver, T., Ragan-Kelley, B., Pérez, F., et al. 2016, in Positioning and Power in Academic Publishing: Players, Agents and Agendas, ed. F. Loizides & B. Scmidt (Amsterdam: IOS Press), 87

Koen, C. 2021, AJ, 161, 281

Komsta, L., & Novomestky, F. 2015, Moments: Moments, Cumulants, Skewness, Kurtosis and Related Tests, v.0.14.1, https://CRAN.R-project.org/package=moments

Kostov, V. B., Mullally, S. E., Quintana, E. V., et al. 2019, AJ, 157, 124

Kovács, G., Zucker, S., & Mazeh, T. 2002, A&A, 391, 369

Kruse, E., Agol, E., Luger, R., & Foreman-Mackey, D. 2019, ApJS, 244, 11

Kunimoto, M., Winn, J., Ricker, G. R., & Vanderspek, R. K. 2022, AJ, 163, 290

Lightkurve Collaboration, Cardoso, J. V. d. M., Hedges, C., et al. 2018, Lightkurve: Kepler and TESS Time Series Analysis in Python, Astrophysics Source Code Library, ascl:1812.013

Lin, J., Zhong, C., Hu, D., Rudin, C., & Seltzer, M. 2020, PMLR, 119, 6150, https://proceedings.mlr.press/v119/lin20g.html

Ljung, G. M., & Box, G. E. P. 1978, Biometrika, 65, 297

Lou, Y., Caruana, R., & Gehrke, J. 2012, in KDD'12 (New York: ACM), 150

Luger, R., Agol, E., Kruse, E., et al. 2016, AJ, 152, 100

Mayo, A. W., Vanderburg, A., Latham, D. W., et al. 2018, AJ, 155, 136

McCauliff, S. D., Jenkins, J. M., Catanzarite, J., et al. 2015, ApJ, 806, 6

Mellor, A., Boukir, S., Haywood, A., & Jones, S. 2015, JPRS, 105, 155

Melton, E. J., Feigelson, E. D., Montalto, M., et al. 2024a, AJ, 167, 203

Melton, E. J., Feigelson, E. D., Montalto, M., et al. 2024b, AJ, submitted

Montalto, M. 2020, DIAmante, STScI/MAST, doi:10.17909/t9-p7k6-4b32
Montalto, M. 2023, MNRAS, 518, L31
Montalto, M., Borsato, L., Granata, V., et al. 2020, MNRAS, 498, 1726
Nardiello, D., Piotto, G., Deleuil, M., et al. 2020, MNRAS, 495, 4924
NASA Exoplanet Archive 2022, Exoplanet Follow-up Observing Program - TESS, Version: 2022-03-15, NExScI-Caltech/IPAC, doi:10.26134/ExoFOP3
NASA Exoplanet Science Institute 2022, Planetary Systems Table, 2022-03-15, IPAC, doi:10.26133/NEA12
Oelkers, R. J., & Stassun, K. G. 2018, AJ, 156, 132
Ofir, A. 2014, A&A, 561, A138
Olmschenk, G., Ishitani Silva, S., Rau, G., et al. 2021, AJ, 161, 273
Osborn, H. P., Ansdell, M., Ioannou, Y., et al. 2020, A&A, 633, A53
Pont, F., Zucker, S., & Queloz, D. 2006, MNRAS, 373, 231
Powers, D. 2011, J. Mach. Learn. Technol, 2, 2229
Rao, S., Mahabal, A., Rao, N., & Raghavendra, C. 2021, MNRAS, 502, 2845
Rauer, H., Catala, C., Aerts, C., et al. 2014, ExA, 38, 249
R Core Team 2020, R: A Language and Environment for Statistical Computing, https://www.R-project.org/
Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2015, JATIS, 1, 014003
Roberts, S., McQuillan, A., Reece, S., & Aigrain, S. 2013, MNRAS, 435, 3639
Sakamoto, Y., Ishiguro, M., & Kitagawa, G. 1986, Akaike Information Criterion Statistics (Dordrecht: D. Reidel), 26853
Schanche, N., Collier Cameron, A., Almenara, J. M., et al. 2019, MNRAS, 488, 4905
Shahaf, S., Zackay, B., Guterman, P., et al. 2021, fBLS — a fast folding algorithm to produce BLS periodograms in search for transiting planets, Zenodo, doi:10.5281/zenodo.5559886
Shallue, C. J., & Vanderburg, A. 2018, AJ, 155, 94
Sing, T., Sander, O., Beerenwinkel, N., & Lengauer, T. 2005, Bioinform., 21, 7881
Siriseriwan, W. 2019, smotefamily: A Collection of Oversampling Techniques for Class Imbalance Problem Based on SMOTE, v1.4.0, https://CRAN.R-project.org/package=smotefamily
Siriseriwan, W., & Sinapiromsaran, K. 2017, Songklanakarin J. Sci. Technol, 39, 565
Stassun, K. G., Oelkers, R. J., Paegert, M., et al. 2019, AJ, 158, 138
STScI 2022, TESS Calibrated Full Frame Images: All Sectors, STScI/MAST, doi:10.17909/0CP4-2J79
Tenenbaum, P., & Jenkins, J. M. 2018, TESS Science Data Productions Description Document EXP-TESS-ARC-ICD-0014 RevD, NASA, https://ntrs.nasa.gov/citations/20180007935
Trapletti, A., & Hornik, K. 2019, tseries: Time Series Analysis and Computational Finance, v0.10-55, https://CRAN.R-project.org/package=tseries
Tu, Z.-L., Yang, M., Zhang, Z. J., & Wang, F. Y. 2020, ApJ, 890, 46
van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, CSE, 13, 22
VanderPlas, J. T. 2018, ApJS, 236, 16
von Boetticher, A., Triaud, A. H. M. J., Queloz, D., et al. 2019, A&A, 625, A150
Waldmann, I. P. 2012, ApJ, 747, 12
Wand, M. 2020, KernSmooth: Functions for Kernel Smoothing Supporting Wand & Jones 1995, v2.23-22, https://CRAN.R-project.org/package=KernSmooth
Wheatland, M. S. 2000, ApJL, 536, L109
Yu, L., Vanderburg, A., Huang, C., et al. 2019, AJ, 158, 25