Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **498,** 1726–1749 (2020) doi:10.1093/mnras/staa2438
Advance Access publication 2020 August 17

A search for transiting planets around FGKM dwarfs and subgiants in the *TESS* full frame images of the Southern ecliptic hemisphere

M. Montalto [1,2]★ L. Borsato [2] V. Granata,[1,2] G. Lacedelli,[1,2] L. Malavolta [1] E. E. Manthopoulou,[1,2] D. Nardiello [2,3] V. Nascimbeni [1,2] and G. Piotto[1,2]

[1]*Dipartimento di Fisica e Astronomia 'Galileo Galilei', Universitá di Padova, Vicolo dell'Osservatorio 3, I-35122 Padova, Italy*
[2]*Istituto Nazionale di Astrofisica – Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, I-35122 Padova, Italy*
[3]*Aix-Marseille Univ, CNRS, CNES, LAM, F-13013 Marseille, France*

Accepted 2020 August 5. Received 2020 July 27; in original form 2020 June 22

ABSTRACT

In this work, we present the analysis of 976 814 FGKM dwarf and subgiant stars in the *Transiting Exoplanet Survey Telescope* (*TESS*) full frame images (FFIs) of the Southern ecliptic hemisphere. We present a new pipeline, DIAmante, developed to extract optimized, multisector photometry from *TESS* FFIs and a classifier, based on the Random Forest technique, trained to discriminate plausible transiting planetary candidates from common false positives. A new statistical model was developed to provide the probability of correct identification of the source of variability. We restricted the planet search to the stars located in the least crowded regions of the sky and identified 396 transiting planetary candidates among which 252 are new detections. The candidates' radius distribution ranges between 1 $R_{\oplus}$ and 2.6 $R_{\rm J}$ with median value of 1 $R_{\rm J}$ and the period distribution ranges between 0.25 and 105 d with median value of 3.8 d. The sample contains four long period candidates ($P > 50$ d), one of which is new, and 64 candidates with periods between 10 and 50 d (42 new ones). In the small planet radius domain ($4R < R_{\oplus}$), we found 39 candidates among which 15 are new detections. Additionally, we present 15 single transit events (14 new ones), a new candidate multiplanetary system, and a novel candidate around a known TOI. By using *Gaia* dynamical constraints, we found that 70 objects show evidence of binarity. We release a catalogue of the objects we analysed and the corresponding light curves and diagnostic figures through the MAST and ExoFOP portals.

Key words: techniques: photometric – techniques: spectroscopic – catalogues – astrometry – planets and satellites: terrestrial planets – binaries: eclipsing.

1 INTRODUCTION

The *Transiting Exoplanet Survey Telescope* (*TESS*; Ricker et al. 2014) was launched on 2018 April 18 aboard a SpaceX Falcon 9 rocket out of Cape Canaveral. With its array of four cameras, covering in total 24 deg by 96 deg of the sky during each pointing, the satellite is delivering both short cadence (2 min) imagettes on pre-selected targets and full frame images (FFIs) with a cadence of 30 min. Each pointing corresponds to a sector of the sky and it is observed for a period of about 27 d, after which the satellite moves to the next one.

This is the first space-based transit search mission monitoring nearly the entire celestial sphere and focusing in particular on bright stars. Its eminent precursors were *CoRoT* (Baglin 2003) and *Kepler* (Basri, Borucki & Koch 2005) that monitored comparatively smaller, selected regions of the sky, targeting typically fainter stars than *TESS* and *K2* (Howell et al. 2014), which employed the same *Kepler* satellite to analyse 21 ecliptic fields. Focusing on bright stars is crucial to both facilitate the confirmation of transiting candidates and enable a galore of follow-up, characterization studies on confirmed planets. Very recently, the CHaracterising ExOPlanet Satellite (*CHEOPS*; Broeg et al. 2013) became operational. Planets around bright stars are primary targets for *CHEOPS* to measure their radii with unprecedented precision and probe their internal structure due to concomitant, ground-based, high-precision spectroscopic observations that permit to derive their masses. These objects are also of great importance for atmospheric characterization studies, for example, through transmission or emission spectroscopy and will be in particular prime targets for *JWST* (Gardner et al. 2006) and *ARIEL* (Tinetti et al. 2018).

The importance of *TESS* goes well beyond its primary mission goals. It will be also relevant for a wealth of different astrophysical studies. In this work, we introduce our project that consists to derive accurate, space-based photometry for all the dwarf and subgiant stars of the sky satisfying our constraints (Section 3). Such an effort will be important for the characterization of the variability properties of these objects and will be relevant also for the prioritization of targets of the future planet-hunting mission *PLATO* (Rauer et al. 2014). In particular, due to *TESS* FFIs it is possible to derive the photometry of unbiased lists of stars and extract information on: multiplicity (from the detection of eclipsing binaries down to small transiting planets), activity, pulsations, rotational periods are all physical quantities that can be obtained from *TESS* light curves. This can be done for millions of stars spread across the entire sky and with unprecedented photometric precision. In this work, we explore the potential of *TESS* FFIs to detect transiting planets, analysing a carefully selected sample of FGKM dwarf and subgiant stars in the Southern ecliptic hemisphere.

★ E-mail: marco.montalto@unipd.it

© 2020 The Author(s)
Published by Oxford University Press on behalf of the Royal Astronomical Society

Downloaded from https://academic.oup.com/mnras/article/498/2/1726/5893317 by CIS Maldura user on 24 January 2025

At the time of writing, *TESS* has nearly completed its core mission, monitoring both the Southern and the Northern ecliptic hemispheres and it is moving towards its already announced extended mission, where both the number of short cadence targets and FFIs will be increased. So far, 51 confirmed planets and 2040 transiting candidates have been announced (e.g. Gandolfi et al. 2018; Huang et al. 2018; Bouma et al. 2019a; King et al. 2019; Shporer et al. 2019; Trifonov, Rybizki & Kürster 2019; Vanderspek et al. 2019; Wang et al. 2019).[1]

In Section 2, we describe the observations and in Section 3 the stellar sample we analysed. We present the new pipeline we used to extract the photometry from *TESS* FFIs in Section 4. In Section 5–7, we discuss our post-reduction analysis and in Section 8 the photometric precision we achieved. In Section 9, we present the transit search approach and in Section 10 we discuss the classification algorithm we adopted to discriminate plausible transiting planets from several common false positives. The centroid motion algorithm is presented Section 11 and the selection of transiting candidates in Section 12. The results of the analysis are illustrated in Section 13 and the discussion of the results in Section 14. We finally draw our conclusions in Section 15.

2 OBSERVATIONS

In this work, we analysed the *TESS* FFIs delivered by the satellite during its first year of operation. The images cover the Southern ecliptic hemisphere. Each sector is imaged by four *TESS* cameras that are composed by four *TESS* CCDs. Therefore, each cadence image delivered by *TESS* is made up of 16 FFIs. The data set we analysed contains 15 347 epochs that correspond to 245 552 FFIs. The first observation started on 2018 July 25 at UT = 19:29:42.708 and the last observation on 2019 July 17 at UT = 19:59:29.974. We started our analysis from the calibrated FFIs. *TESS* cameras are read out every 2 s and the resulting images are stored into 30 min exposures. Cosmic rays are mitigated by an onboard algorithm (Vanderspek et al. 2019) and sent to the ground where they are processed by the Science Processing Operations Center (SPOC; Jenkins et al. 2016). The SPOC performs traditional CCD data reduction steps (e.g. correction for bias, dark current, and flat-field), as well as *TESS*-specific corrections (removing smear signals resulting from the lack of a shutter on the cameras). The resulting science data products are described by Tenenbaum & Jenkins (2018).

3 STELLAR SAMPLE

The stellar sample we analysed was built from the *Gaia* DR2 all-sky catalogue. It is a sample of FGKM (limited to F5) dwarf and subgiant stars. The selection was done in the absolute, intrinsic colour magnitude diagram by imposing conditions on the absolute magnitude and colours. Stellar distances were derived from Bailer-Jones et al. (2018) and reddening from the interpolation of Lallement et al. (2018) 3D maps. For FGK dwarfs and subgiants, the selection was limited to $V < 13$. For dwarfs, we imposed: $0.42 < (B-V)_0 < 1.38$, $M_{V,0} \geq 5\,(B-V)_0 + 0.4$, and $M_{V,0} < 5(B-V)_0 + 3.5$, while for subgiants: $M_{V,0} < 5(B-V)_0 + 0.4$ and $M_{V,0} > 5(B-V)_0 - 2$ in the colour range between $0.42 < (B-V)_0 \leq 0.8$ and $M_{V,0} < 4.5$ and $M_{V,0} > 5(B-V)_0 - 2$ in between $0.8 < (B-V)_0 \leq 1.0$. Here, $M_{V,0}$ is the absolute, intrinsic magnitude in the V band and $(B-V)_0$ is the absolute $(B-V)$ colour. Optical B- and V-band photometry was derived from *Gaia* colours using calibration relations in Evans et al. (2018). For M-dwarfs, the selection was performed down to $V \leq 16$ and distance < 600 pc, imposing $(G_{BP}\text{-}G_{RP})_0 \geq 1.84$ and $M_{G,0} > 2.334(G_{BP}\text{-}G_{RP})_0 + 2.259$, where $M_{G,0}$ is the absolute, intrinsic magnitude in the G band and $(G_{BP}\text{-}G_{RP})_0$ is the absolute intrinsic $(G_{BP}\text{-}G_{RP})$ colour. The optical V-band photometry was derived in this case as a function of the *Gaia* colour using a custom calibrated relation. Further details on the catalogue construction will be provided in Montalto et. al. (in preparation). This catalogue was restricted to stars falling in the footprint of *TESS* CCDs of the first 13 sectors and matched with the *TESS* Input Catalog (v8, Stassun et al. 2019) restricted to stars with $T_{\rm eff} < 7000$ K, $\log g > 3$, and TIC $V < 13$ for FGK dwarfs and subgiants, and to stars with $T_{\rm eff} < 3870$ K, $\log g > 3$, and TIC $V < 16$ for M-dwarfs. From the TIC catalogue, we extracted the stellar parameters used in this work. The number of stars and the number of sectors in which they are observed are reported in Table 1. In Fig. 1, we show the distributions of *TESS* magnitudes, stellar distances, contamination ratios, effective temperatures, stellar radii, and masses of the selected stars. Fig. 2 (left-hand panel) shows a pie chart representing the per centual number of targets that are observed in one, two, and more than two sectors and the targets' distribution across the celestial sphere (right-hand panel). The sample contains 889 411 FGK dwarfs and 87 403 M-dwarfs for a grand total of 976 814 stars.

Table 1. Number of stars (N. stars), per centual number of stars, and number of sectors (N. sectors) where the stars are imaged.

N. stars	Per cent	N. sectors
711 981	72.9	1
182 383	18.7	2
34 368	3.5	3
6464	0.7	4
3379	0.3	5
4084	0.4	6
2835	0.3	7
1503	0.2	8
1796	0.2	9
2082	0.2	10
4580	0.5	11
10 338	1.0	12
11 021	1.1	13

4 DATA REDUCTION

Several approaches have been presented so far for the analysis of *TESS* FFIs (e.g. Oelkers & Stassun 2018; Bouma et al. 2019b; Feinstein et al. 2019; Handberg & Lund 2019; Nardiello et al. 2019, 2020). Here, we employ a method based on the difference image analysis (DIA; Alard & Lupton 1998; Alard 2000; Bramich 2008; Miller, Pennypacker & White 2008). Difference imaging permits a very efficient subtraction of all constant sources in the field and therefore reduces the impact of contaminants on the target's photometry and permits a more accurate estimate of the sky background with respect to simple aperture photometry. Moreover, the subtracted images can also be exploited during the centroid motion analysis, as explained in Section 11. The technique is based on the subtraction of a high S/N reference image (R) of a stellar field to a target image (I) after convolving the reference with an optimal Kernel (K) to match the intensity, the background (B), and the point spread function (PSF) of the target image. As a kernel, we chose the delta function basis kernel (Miller et al. 2008) and

[1] A useful list of *TESS* publications can be found at https://heasarc.gsfc.nasa.gov/docs/tess/tpub.html.

Downloaded from https://academic.oup.com/mnras/article/498/2/1726/5893317 by CIS Maldura user on 24 January 2025

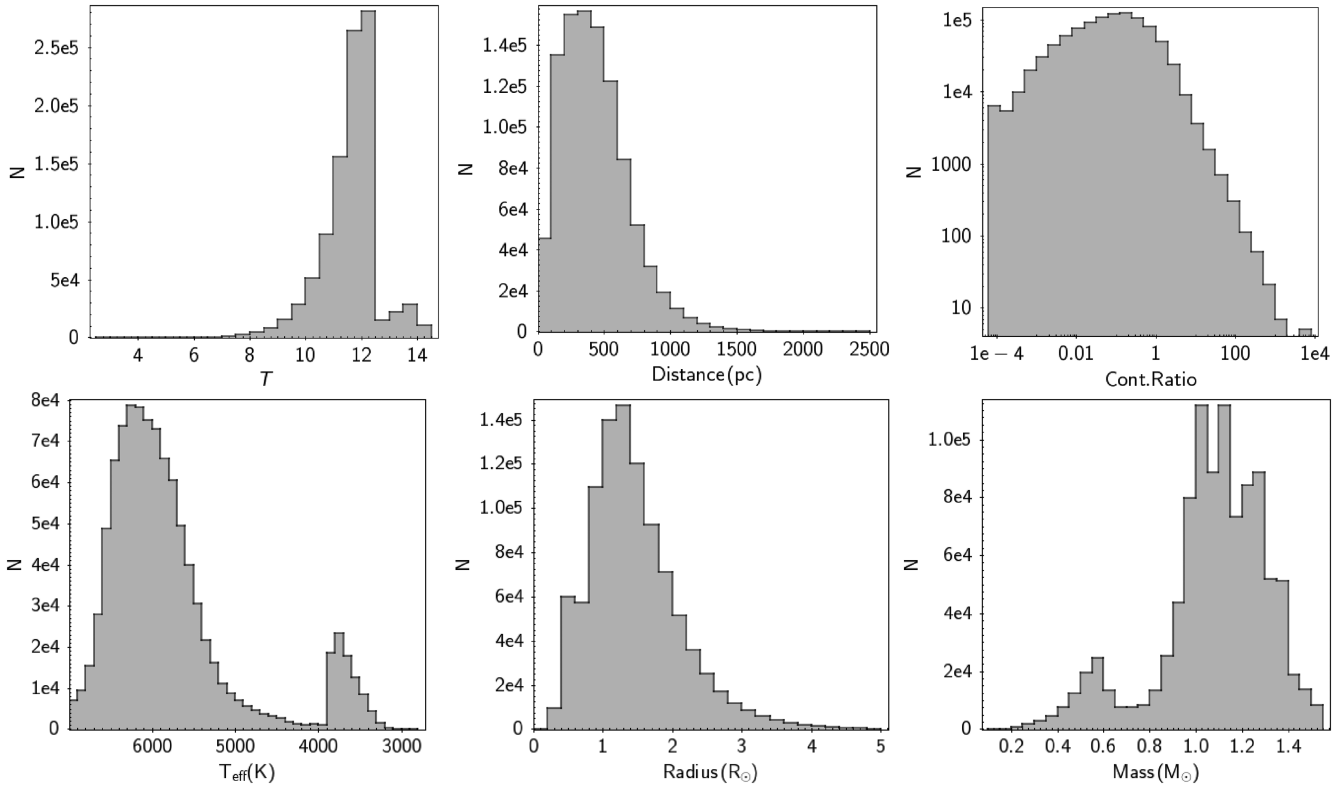

Figure 1. *Top, from the left:* Distribution of *TESS* magnitudes, distances, and contamination ratios for the stars analysed in this work. *Bottom, from the left:* Distribution of stellar effective temperatures, radii, and masses. Values of stellar parameters are taken from Stassun et al. (2019).

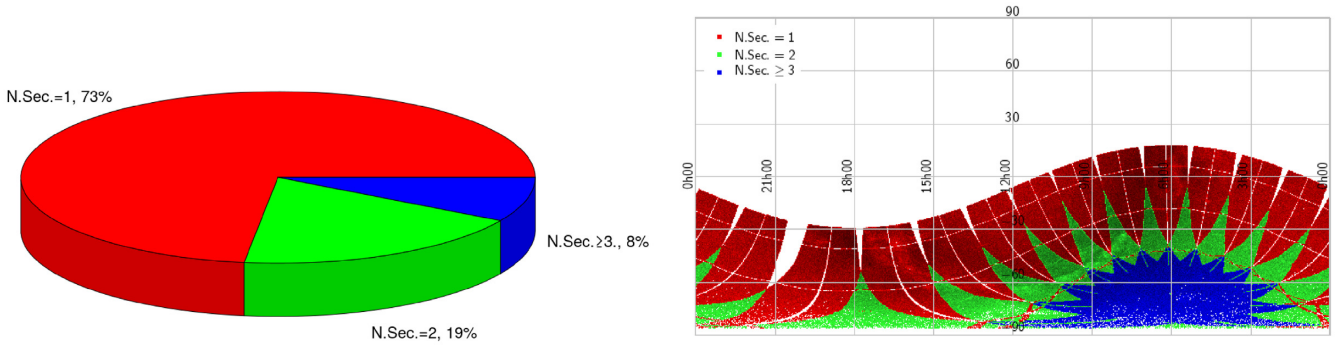

Figure 2. *Left:* Pie chart representing the per centual number of targets observed in one, two, or more than two sectors. *Right:* Spatial distribution of our selected targets across the celestial sphere in an equatorial coordinate reference system. Colours denote the number of sectors where a star was measured, as specified in the legend.

experimented with different kernel dimensions between 3 and 5 pix and constant, first- and second-order expansions. We found that in most cases, a constant kernel of dimension 3 pix × 3 pix provided the best compromise between accuracy of the subtraction and efficiency. Moreover, as demonstrated by Miller et al. (2008), such a kernel has the ability to compensate for small drifts of the target and reference image directly during the kernel solution step. This is an attractive feature because it permits to avoid the usual step of image registration and ensures a perfect flux conservation. In our approach, we therefore convolve the reference image with the optimal kernel and extract the photometry on the reference system of the target image by using the World Coordinate System (WCS) solution embedded in each FFI. Moreover, we also solve for the differential background between the reference and the target image simultaneously with the kernel, as explained below. The best kernel solution is then found by solving the equation

$$\mathbf{R} \otimes \mathbf{K} + \mathbf{B} = \mathbf{I}, \tag{1}$$

where the symbol ⊗ denotes the convolution operation and B is the differential background model that we initially assumed constant.

The kernel **K** was expanded in a set of $\mathbf{M}^2$ delta basis kernels $\mathrm{K}_{\mathrm{p,q}}$ which are **M** × **M** matrixes whose elements are

$$\mathbf{K}_{\mathrm{p,q}}(\mathrm{i},\mathrm{j}) = \begin{cases} 1 \text{ if } (\mathrm{i} = \mathrm{p} \wedge \mathrm{j} = \mathrm{q}) \\ 0 \text{ if } (\mathrm{i} \neq \mathrm{p} \vee \mathrm{j} \neq \mathrm{q}) \end{cases},$$

Downloaded from https://academic.oup.com/mnras/article/498/2/1726/5893317 by CIS Maldura user on 24 January 2025

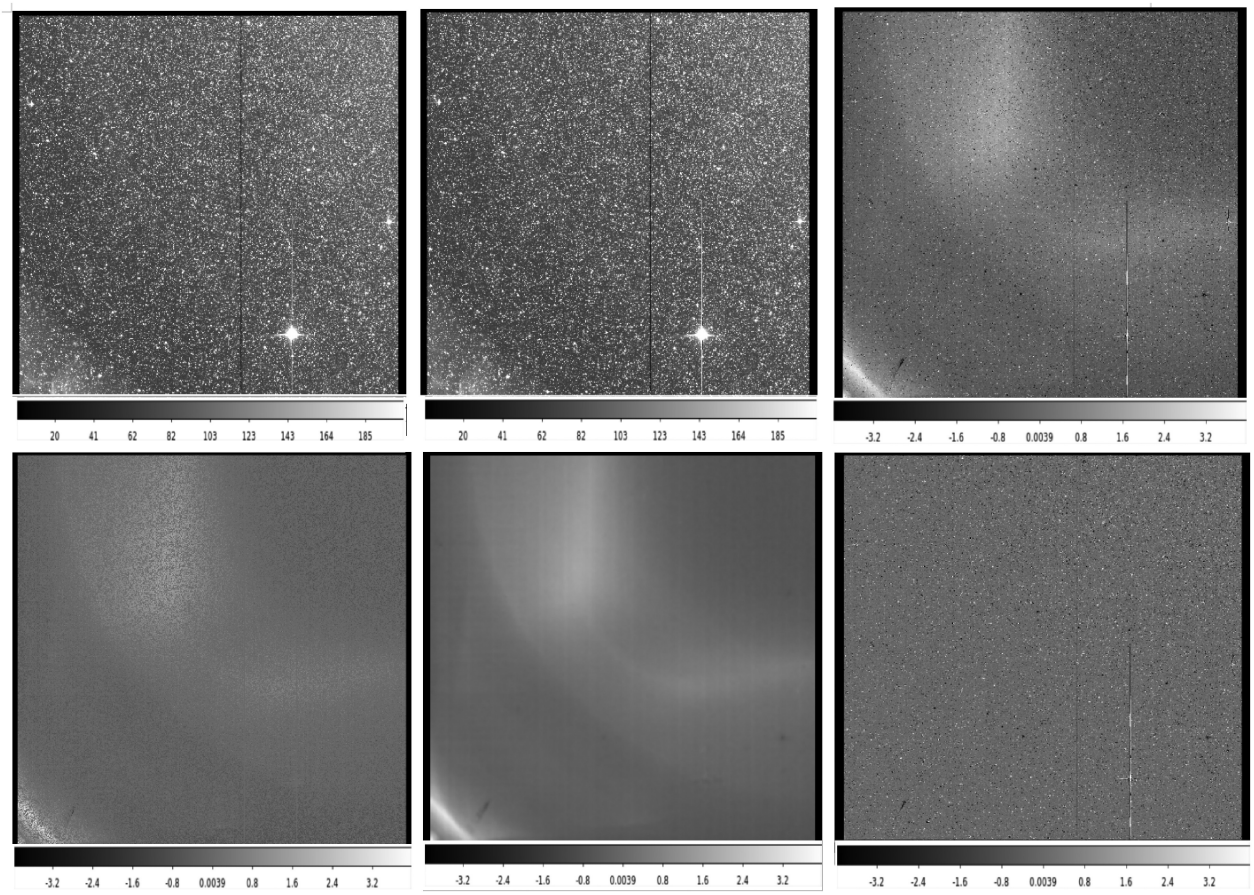

Figure 3. Overview of the reduction process. In the top row from the left side: original image, reference image, and first iteration subtracted image. On the bottom row, from the left side: filtered subtracted image, filtered and smoothed differential background model, and final subtracted image. The image represented is tess2019022075936-s0007-3-3-0131-s_ffic.fits of Sector 7, camera 3, CCD 3.

then equation (1) is rewritten as

$$\sum_{p=1}^{M}\sum_{q=1}^{M} \mathbf{A}_{p,q}\,(\mathbf{R} \otimes \mathbf{K}_{p,q}) + \mathbf{B} = \mathbf{I}, \tag{2}$$

where $\mathbf{A}_{p,q}$ are the Kernel coefficients. One of the peculiar characteristics of *TESS* FFIs is that they present rather erratic background variations depending on the boresight angle between each camera and the Sun and the Moon directions. Because of this, we then constructed a more accurate differential background model. We considered the first iteration subtracted image and filtered it to remove flux variations' high spacial frequencies that are typically associated with the residuals of stellar sources. To do this, we first determined the flux's dispersion (σ) of the entire subtracted image. We then calculated, for each pixel, the absolute flux differences ($|df|$) between that pixel and all surrounding pixels. If the condition $|df| > \sigma$ was met for at least one of these flux differences, the pixel was masked. We then set each image pixel's value equal to the average flux calculated in a square region centred on the pixel (considering only the unfiltered pixels in the averaging process). After some tests, we adopted a value of 20 pix for the dimension of the box smoothing region. Such radius allows to model both small- and large-scale *TESS* background variability. The filtered and smoothed differential background model ($\mathbf{B}_{im}$) was then simultaneously fit with the Kernel by replacing the constant term B in equation (1) with a first-order expansion

$$\mathbf{B}' = \mathbf{B}_1 \times \mathbf{B}_{im} + \mathbf{B}_2.$$

Once the optimal Kernel solution was found, the convolved reference image and the background terms were subtracted from the target image to form the final subtracted image. Fig. 3 shows an overview of the reduction process for the image tess2019022075936-s0007-3-3-0131-s_ffic.fits of Sector 7, camera 3, CCD 3. In the video at https://youtu.be/EpuIPCcgTo8, we also show the full sequence of 1087 subtracted images for Sector 7, Camera 3, CCD 3.[2] The frame rate is equal to 12 images per s. At the beginning of the video (0:00–0:01) and at (0:46–0:47) vertical 'straps' are visible. Towards the end of the sequence (1:09), a moving object entered the camera FOV from the top right corner.

4.1 Reference image

From the WCS solutions of each image, we deduced the zero-order shift of each image with respect to a reference astrometric image. This image was usually chosen as the first frame for which a WCS solution was reported and that did not have any obvious defect. For each image, we then considered the relative offset in the X and Y coordinates between that image and all other images in the set and determined the minimum relative offset. If such minimum offset was larger than 0.003 pix, we excluded the image from the set of images used to build the reference. This procedure served to eliminate images

[2]We excluded six images affected by momentum dumps.

Downloaded from https://academic.oup.com/mnras/article/498/2/1726/5893317 by CIS Maldura user on 24 January 2025

with relatively big isolated offsets with respect to the others. From the remaining set, we then produced a median stack image, which we used as reference frame for our analysis. We also calculated the median offset between all the selected frames used to build the reference and the astrometric reference frame, which we then used to project our master list on the reference frame to extract the reference frame's flux.

4.2 Photometry

We extracted the photometry on a set of two concentric circular apertures with radii of 1 and 2 pix. The aperture was centred on the *Gaia* catalogue predicted positions of our targets, corrected by proper motions at the beginning of each *TESS* sector campaign whenever the relative proper motion error on each coordinate was smaller than 10 per cent. The conversion between the sky coordinates and the image coordinates was performed using the WCS solution embedded in each image using the task `sky2xy` of the WCSTOOLS package (Mink 1997). We checked the quality of the astrometric solution for a sample of representative images. We considered the centroid positions derived from the WCS solution as initial guesses for a PSF fitting algorithm, by means of which we further refined the centroid coordinates. We used in particular the empirical PSFs provided by the *TESS* team.[3] We then compared the PSF centroid coordinates with the WCS solution coordinates and obtained the following median differences: $|\boldsymbol{dr}| = 0.03$ pix with standard deviation $\sigma_{|\mathbf{dr}|} = 0.18$ pix. Given that our aperture radii are either 1 or 2 pix, we considered this difference acceptable and adopted the WCS solution coordinates.

Denoting with $f_{i,j}(S)$ the flux of the subtracted image in the (i,j) image pixel and with f(I) and f(R) the integrated flux of an object in the target and reference image, we have

$$f(I) = C_s \left(f(R) + \frac{1}{||\mathbf{K}||} \sum_{(i,j)\in\Gamma} w(i,j) f_{i,j}(S) \right), \quad (3)$$

where the summatory was extended to all pixels included or intersecting a given aperture Γ. The w(i,j) function is a weighting function that gives the fractional area of pixel (i,j) included in the aperture and it was calculated using the Green's theorem. The term $||\mathbf{K}||$ is the Kernel norm that represents the flux scaling factor between the reference and the target image. The flux f(R) on the reference image was calculated from the *TESS* magnitude (*T*) reported in the TIC. For each CCD, we calculated the median magnitude difference (Δ m) between the instrumental magnitudes of a sample of 250 calibration stars (measured on the reference image adopting the same aperture used for the target image), and the corresponding *TESS* magnitudes. These stars were selected among the most bright, non-saturated, and isolated stars in each CCD. Then the flux f(R) was obtained from

$$f(R) = 10^{-0.4(T+\Delta m)}. \quad (4)$$

The constant C_s in equation (3) accounts for sector, camera, CCD systematic zero points. Denoting with $R_s = 10^{-0.4(\Delta m)}$ the median flux ratio between the comparison stars' instrumental fluxes of the s-th sector, camera, CCD, and the *TESS* catalogue fluxes, and with $\widehat{R}$ the median of all sectors, camera, CCDs median flux ratios, the constant C_s can be expressed as

$$C_s = \frac{\widehat{R}}{R_s}. \quad (5)$$

[3] https://archive.stsci.edu/missions/tess/models/prf_fitsfiles/

The flux f(i) in equation (3) is background subtracted and the background was evaluated from an annular region surrounding each target with inner radius equal to 5 pix and outer radius equal to 15 pix. The light curves obtained with the procedure described in this section are referred as LC0 light curves.

5 MASKING

Bad measurements were masked to avoid the most prominent systematics. We flagged all images that had *TESS* quality flag different from zero. In addition, we checked the kernel norm to identify possible problems with the subtracted images. We typically flagged images having kernel norm values that differ by more than a few per cent the median value of a given sector, camera, CCD data set. Such pipeline dependent flags are merged with the *TESS* pipeline flags and incorporated in the bitmask we released with the light curves.

6 COTRENDING

Light curves were then cotrended to correct for systematic effects. The log rms versus *T* magnitude diagram was interpolated with a third-order polynomial, downweighting positive residuals to allow the fit to converge towards the lower bound limit of the distribution. Stars within 1.5σ from the interpolated curve were then selected. We calculated the Pearson correlation coefficient of each light curve with respect to all other light curves in the sample, and then the median correlation coefficient. The stars were then sorted out in increasing order of the median correlation coefficient and the 50 per cent most correlated light curves were selected. We then sorted out the light curves in increasing order of magnitude and the first 1200 stars were used to perform a principal component analysis (PCA). The resulting eigenvectors were sorted in decreasing order of their eigenvalues (eigval). We chose then the first N eigenvectors in the sorted list satisfying the condition $\frac{\sum_1^N \text{eigval}}{\sum_1^{1200} \text{eigval}} > 0.9$, and selected a maximum of 10 eigenvectors. Each light curve was then linearly decorrelated against this set of eigenvectors. The criterium described above was adopted after testing it on several light curves and it was found appropriate to remove the most important systematics without affecting transit detection. It is also capable to preserve short term, intrasector variability, but it is less appropriate to preserve long-term variability. In multisector observations, each portion of the light curve belonging to a given sector, camera, CCD is decorrelated against the corresponding set of eigenvectors appropriate for the same sector, camera, CCD, but completely independently from measurements acquired on other sectors. Therefore, it is difficult to preserve variability extending on time-scales longer than one sector ($\sim$27 d). In any case, this fact is largely unimportant in the context of this work, which is focused on the detection of transits occurring on much shorter time-scales. The light curves obtained after cotrending are referred as LC1 light curves.

7 LIGHT-CURVE NORMALIZATION

Before searching for transits, we applied a high-pass filter to enhance transit detectability. First, we proceeded by averaging each light curve over 8 h time intervals and then B-spline interpolated the resulting averages. Any gap in the data for which it was not possible to calculate the time interval average was skipped in the splining process. Then, we identified possible outliers in the splined light curve searching for measurements satisying the condition $f_i < Q1 - 1.5 \times IQR$ or $f_i > Q3 + 1.5 \times IQR$, where f_i is the i-th

Downloaded from https://academic.oup.com/mnras/article/498/2/1726/5893317 by CIS Maldura user on 24 January 2025

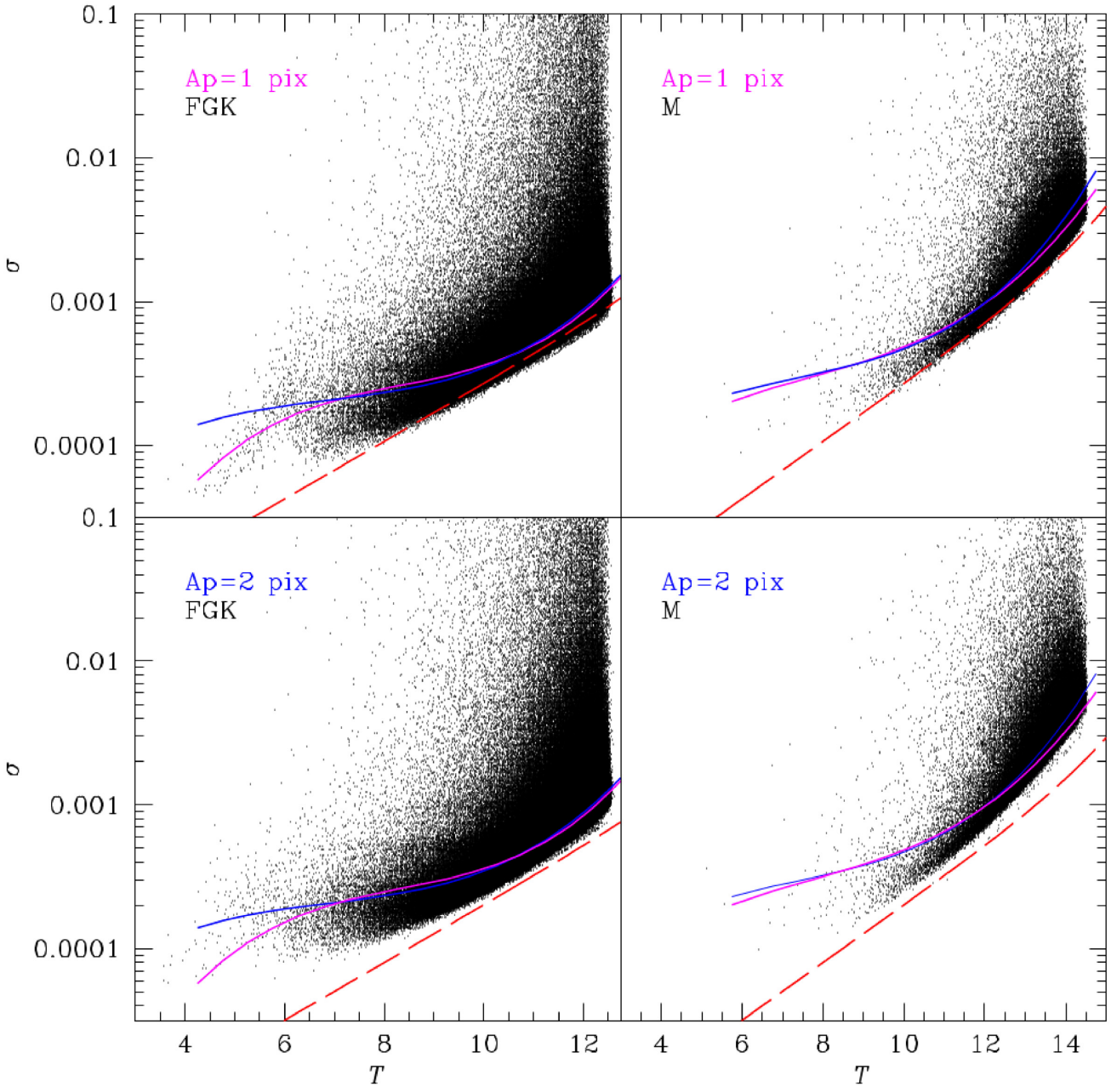

Figure 4. Photometric precision of the light curves for different samples and photometric apertures. The red dashed lines correspond to theoretical models, while the continuous coloured lines correspond to the best-fitting interpolation models of observed precisions. The magenta colour is related to aperture equal to 1 pix, while the blue colour is related to aperture equal to 2 pix.

flux measurement and Q1, Q3, and IQR are the first quartile, the third quartile, and the interquartile range of the flux measurements' distribution. We then repeated the splining procedure starting from the original light curve but excluding the identified outliers. This procedure was repeated 10 times by shifting the averaging intervals of 0.8 h each time. The resulting interpolating splines were then evaluated at the instant of each observation, and these final estimates were averaged together. The above procedure was repeated also after the transit search by excluding from the splining procedure also those measurements falling within the transit window. The light curves obtained after normalization are referred as LC2 light curves.

8 PHOTOMETRIC PRECISION

In Fig. 4, we show the photometric precision (σ) achieved in the final normalized light curves. The photometric precision is calculated as half the difference between the 84th and the 16th percentiles of the cumulative distribution of the flux measurements. The diagrams in Fig. 4 display the photometric precision for the two separate samples we analysed in this work, that is the FGK sample (left-hand panels) and the M-dwarfs sample (right-hand panels). They also represent the precision for the two different apertures we used and in particular for aperture = 1 pix (upper panels, hereafter aperture 1) and for aperture = 2 pix (lower panels, hereafter aperture 2).

Downloaded from https://academic.oup.com/mnras/article/498/2/1726/5893317 by CIS Maldura user on 24 January 2025

The observed precisions were fit with a third-order polynomial function and the results are represented by the continuous magenta (aperture 1) and blue (aperture 2) lines. On the diagrams relative to a given aperture (and for the same stellar sample), we represent the best-fitting models of both apertures, to facilitate the comparison. The dashed red lines are theoretical expectations and represent the photometric precision achievable for a given aperture in half-an-hour integration time, accounting for photon noise, assuming a sky equal to 150 $e^- s^{-1}$ and a RON = 8.5 e^-. The two apertures have in general comparable performances, but aperture 2 photometry appears to have a more stable behaviour, in particular at the bright end of the FGK magnitude sample. Aperture 1 photometry has slightly better performances for *TESS* magnitudes $T > 11$. Considering these results, we decided to perform the search for transiting planets in the full sample for all magnitudes for aperture 1 and up to $T < 11$ for aperture 2.

9 TRANSIT DETECTION

To search for planetary transits, we used the box-fitting algorithm (BLS) of Kovács, Zucker & Mazeh (2002). We searched for signals with periods in between $P_{\rm min} = 0.25$ d and ΔT, where ΔT was the total time spanned from the first to the last measurement of each light curve. The period step ΔP was determined by imposing $\Delta P = \frac{P}{\Delta \rm T}\epsilon$ (Ofir 2014). We set $\epsilon = 5$ min to ensure a precision equal to 5 min in the folding process between the first and the last transit in the observable window, for any tested period P. The fractional transit length q was adjusted for any trial period P and target star radius $R_\star$ and mass $M_\star$. In particular, we estimated the maximum transit duration $\tau_{\rm max}$ at period P considering a circular orbit and a 90° transiting planetary orbital inclination which gives $\tau_{\rm max} \simeq \frac{P}{\pi} \arcsin \frac{R_\star}{a}$, where a is the orbital semimajor axis. We then considered fractional transit lengths q in between $q_{\rm min} = 0.1\frac{\tau_{\rm max}}{P}$ and $q_{\rm max} = 1.1\frac{\tau_{\rm max}}{P}$. The number of bins ($n_{\rm bins}$) in which each folded light curve was subdivided to evaluate the BLS metric was varied as well for any tested period and set equal to $n_{\rm bins} = \frac{2}{q_{\rm min}}$.

10 CLASSIFICATION

After the search for transits, we applied a classification algorithm in order to identify the most promising candidates. It is known that the BLS can be sensitive to different kind of variables and/or events that mimic the shape of transiting planet signals. Some of these false positives can be usually recognized by a morphological inspection of the light curves. For example, it is customary to identify eclipsing binaries by analysing odd/even transits, checking for variations of transit depths between them. Also the presence of secondary eclipses is generally an indicator of false positive signals. Ultra short period variables may be difficult to correctly cotrend/detrend and the residual signals may as well trigger the BLS. In general, the detection of these classes of objects interferes with the identification of plausible transiting planets, decreasing the efficiency of the search. It is desirable therefore to build a filter to decrease the rate of false positives. This is especially important when dealing with massive searches for planets. To tackle this problem, we used a Random Forest Classifier (RFC; Breiman 2001). The Random Forest is an ensemble method of machine learning that uses several individual decision trees to assign a class to a given input. In order to apply it, it is necessary to define a set of *features*, which can be considered as quantities related to some qualifying aspects of the signal we want to detect and of the false positive signals we want to reject. Each tree is trained on a random subset of the input training sample, and a random subset of the input features is used at each branch of the trees to split them into sub-branches. Splits are done maximizing a given metric that optimizes correct classification. At the end of this process, each tree provides the likelihood that a given input belongs to a certain class. Bootstrap aggregation (known as *bagging*; e.g. Breiman 2001) is used then to summarize the results of the single trees into a unique probabilistic estimate. The RFC is a popular classification algorithm, known for being very robust and easy to implement. It also permit a straightforward analysis of the classification results by judging, for example, the importance that each variable has on the classification process. RFCs have been already described in detail and applied and tested in the context of exoplanet searches in recent literature works (e.g. Caceres et al. 2019a, b; Schanche et al. 2019).

In our implementation of the RFC, we focused the attention on the morphological aspects of the problem related to exoplanet transit identification. By morphological aspects, we intend the fact that usually many of the features used by humans to isolate planetary transit candidates are purely morphological, in the sense that they relate one property of the light curve to other properties extracted from the same light curve. The presence of secondary eclipses, the light-curve modulations, the morphological appearance of the primary eclipse, the odd/even transit depths, the significance of periodic signals are all qualifying morphological aspects of folded light curves (or of periodograms). Therefore, the set of features we defined is nearly all unitless and involve S/N estimates or similar normalized quantities evaluated at particular critical points of the light curves (or periodograms). Facing the problem of planetary transit classification in a purely morphological sense, it is a useful way to simplify the procedure to identify good candidates. There is evidently other important information that should be considered for a correct classification (such as dynamical information, centroid motion information for detectors like *TESS*), but at the first stage in the analysis, when essentially only the light curves are scrutinized, a morphological analysis can already eliminate a large fraction of contaminant signals. Nevertheless, several false positives are expected to pass this test. For example, it would be evidently too pretentious to require a morphological classifier to discriminate between planetary signals and low-mass eclipsing binary signals since it is known that these objects produce essentially indistinguishable transit photometry. In fact these objects can be identified only by subsequent follow-up radial velocity (or timing) analysis. For this reason, when evaluating the performances of the classifier, these kind of false positives should not be considered. Moreover, feature design should also take into account the need to define quantities that can be easily and robustly calculated for any light curve. Feature definition is described below in Section 10.1.

One of the problems related to the construction of a classifier of this kind is which sample of planets to consider and which classes of false positives to include in the analysis. By considering that the physics of transiting planets and of eclipsing binaries is relatively well understood (at least for what concerns the modellization of their light curves), in the following we resort to use simulations to produce traning sets to feed the classifier. This has the undoubtful vantage to overcome issues related to class imbalance, where planetary signals are usually strongly underrepresented in samples drawn from the real world. However, in order to reproduce as closely as possible the conditions to which the classifier will be applied, we selected a random, representative sample of stars from the total sample we analysed in this work, and injected the artificial signals of planets or false positives into their light curves, as described in Section 10.2.

Downloaded from https://academic.oup.com/mnras/article/498/2/1726/5893317 by CIS Maldura user on 24 January 2025

10.1 Features' definition

10.1.1 Effective S/N of the primary eclipse

One of the most important parameters to define in the contex of transit searches is the effective S/N of the primary eclipse, as already recognized by Kovács et al. (2002)

$$\mathrm{SN_I} = \frac{\delta_1}{\sigma}\left(\frac{1}{\sqrt{N_{\mathrm{in}}}} + \frac{1}{\sqrt{N_{\mathrm{out}}}}\right), \tag{6}$$

where δ_1 is the transit depth of the primary eclipse estimated by the difference between the average of the in-transit measurements and the average of the out-of-transit measurements. Such quantity is divided by the error (of the average's difference of the in- and out-of-transit measurements), where we assumed uniform noise across the light curve, represented by the standard deviation of the out-of-transit measurements (σ).

The effective signal to noise defined above is calculated (as most of the quantities we defined) on the light curve folded with the period corresponding to the highest peak in the BLS power spectrum. It is a measure of how well transit signals occurring at regular intervals of time identical to the chosen period and phase add up constructively.

10.1.2 Effective S/N of the secondary eclipse

To check for the presence of secondary eclipses, the effective S/N of the secondary eclipse can be used

$$\mathrm{SN_{II}} = \frac{|\delta_2|}{\sigma}\left(\frac{1}{\sqrt{N_{\mathrm{in}}}} + \frac{1}{\sqrt{N_{\mathrm{out}}}}\right), \tag{7}$$

where δ_2 is the estimated depth of the secondary eclipse. Such quantity is the difference between the average of the measurements within the secondary eclipse phase interval (located exactly at mid-phase between primary eclipses) and the out-of-secondary eclipse measurements. The time interval corresponding to the measurements within the secondary eclipse was considered identical to the transit duration reported by the BLS. Since we are uninterested to descriminate between positive and negative flux variations at the secondary eclipse phase, we took the absolute value of the eclipse depth in equation (7).

10.1.3 Effective S/N of the tertiary eclipse

Similarly to the previous two cases, the S/N of the tertiary eclipse can be define as follows:

$$\mathrm{SN_{III}} = \left\{\frac{|\delta_3|}{\sigma}\left(\frac{1}{\sqrt{N_{\mathrm{in}}}} + \frac{1}{\sqrt{N_{\mathrm{out}}}}\right)\right\}_{\mathrm{MAX}(0.2<\phi<0.8)}. \tag{8}$$

By tertiary eclipses, we intend here any eclipse occurring at phases possibly different from the primary and secondary phase. Eccentric eclipsing binaries can produce secondary eclipses that are not found exactly at mid-phase between primary eclipses, so it is useful to check for these signals. The quantity δ_3 is the depth of the tertiary eclipse and it is calculated as the maximum absolute difference of the flux measurements within and outside a sliding window (the same length of the transit duration given by the BLS) centred at orbital phases comprised between ϕ =0.2 and ϕ =0.8 (where the primary transit occurs at ϕ =0).

10.1.4 Signal detection efficiency (SDE)

The SDE was defined in Kovács et al. (2002) and it is equal to

$$\mathrm{SDE} = \frac{SR_{\mathrm{peak}} - < SR >}{sd(SR)}. \tag{9}$$

On the contrary of the previous quantities, this feature is extracted from the BLS power spectrum, where SR_{peak} is the power of the BLS peak, $<SR>$ the average power, and *sd(SR)* the standard deviation of the BLS power spectrum.

10.1.5 Average signal detection efficiency of aliasing peaks

Periodic signals usually originate a cascade of aliases of the primary peak in the BLS power spectrum. To check for their presence, we defined the average SDE over the nine peaks closest to the primary on both the low and high frequency intervals of the spectral window (N_{aliases})

$$\mathrm{SDE_{AL}} = \frac{\sum_{\mathrm{i}=1}^{N_{\mathrm{aliases}}} \frac{SR_{\mathrm{i}} - <SR>}{sd(SR)}}{N_{\mathrm{aliases}}}. \tag{10}$$

This feature (as well as the SDE) is extracted from the BLS power spectrum, where SR_{i} is the power of the i-th BLS peak considered, $<SR>$ the average power, and *sd(SR)* the standard deviation of the BLS power spectrum.

10.1.6 Effective S/N of odd/even transits

As reported above, eclipsing binary signals may be revealed by calculating the odd (δ_{odd}) and even (δ_{even}) transit depths. A simple metric can be introduced to detect odd/even transit depth variations

$$\mathrm{SN_{OE}} = \frac{|\delta_{\mathrm{odd}} - \delta_{\mathrm{even}}|}{\sigma}\left(\frac{1}{\sqrt{N_{\mathrm{in,even}}}} + \frac{1}{\sqrt{N_{\mathrm{in,odd}}}}\right), \tag{11}$$

where $N_{\mathrm{in,\ even}}$ and $N_{\mathrm{in,\ odd}}$ are the number of in-transit measurements in the even and in the odd transits, respectively.

10.1.7 Out-of-transit (OOT) variability

Another interesting morphological feature that can permit to identify likely eclipsing binary stars is the presence of phase-locked flux modulations with the same periodicity of the transiting body. Usually these flux variations can be produced by proximity effects (e.g. reflections, ellipsoidal modulations) between two close orbiting bodies and are expected to be much stronger for stellar than for planetary companions. To identify these flux modulations, we performed an harmonic fit of the light curve by assuming the following model:

$$f = A\cos(\omega t) + B\sin(\omega t) + C, \tag{12}$$

where f is the flux, $\omega = \frac{2\pi}{P}$, and P is the period corresponding to the highest peak in the BLS power spectrum. We then calculated

$$R^2_{\mathrm{OOT}} = 1 - \frac{\sum_{\mathrm{i}=1}^{N_{\mathrm{out}}}(f_{\mathrm{i}} - [A\cos\omega t_{\mathrm{i}} + B\sin\omega t_{\mathrm{i}} + C])^2}{\sum(f_{\mathrm{i}} - C)^2}, \tag{13}$$

where N_{out} is the number of out-of-transit measurements where the summatory is performed. R^2_{OOT} is a measure of how much variance is explained by the complete harmonic model with respect to the sole constant component of the model. We have that $R^2_{\mathrm{OOT}} < 1$ and, the better the harmonic model fits the data, the closer R^2_{OOT} is to unity. This parameter was calculated both using the LC1 light curves and using the LC2 light curves and the maximum value was taken.

Downloaded from https://academic.oup.com/mnras/article/498/2/1726/5893317 by CIS Maldura user on 24 January 2025

10.1.8 Fractional transit duration

The fractional transit duration (q) is equal to

$$q = \frac{\tau}{P}, \tag{14}$$

where τ is the transit duration and P is the orbital period.

10.1.9 Point-to-point statistic inside and outside the transit

A point-to-point noise estimate within and outside the transit window can help to identify potentially spurious transits. It is defined as

$$P2P_{\rm IO} = \sqrt{\frac{\sum_{\rm in}(f_i - f_{i-1})^2}{\sum_{\rm out}(f_i - f_{i-1})^2}}, \tag{15}$$

where the numerator is evaluated on the in-transit measurements and the denominator on the out-of-transit measurements (in the phase folded light curve).

10.1.10 Symmetry of folded light curve

A measure of the symmetry of the folded light curve is also a useful quantity to consider and it can be obtained with

$$P2P_{\rm S} = \sqrt{\frac{\sum_{i=1}^{N}(f_i - f_{i-1})^2_{|\phi|}}{\sum_{i=1}^{N}(f_i - f_{i-1})^2_{\phi}}}, \tag{16}$$

where the numerator is evaluated after ordering the measurements as a function of the absolute value of the phase ($|\phi|$)[4] and the denominator is evaluated on the nominal phase-folded light curve.

10.1.11 Estimated transiting body radius

The radius of the transiting body can be estimated from

$$r = \sqrt{\delta_1} R_\star, \tag{17}$$

where δ_1 is the transit depth of the primary eclipse and $R_\star$ is the stellar radius.

As it can be noted, the transiting object radius is the only feature that has a physical dimension in our set. The reason to include it stems from the fact that the stellar radius is now well constrained by *Gaia* and thus the transit depth can be effectively converted to transiting object radius for all the targets. The gain is huge, since without such information many transiting candidates could be erroneously considered plausible or erroneously considered implausible due to the larger ambiguity inherent to the use of the sole transit depth.

10.2 Simulations

To train the model to recognize transiting planets, we used a set of simulations. From the pool of stars we used to construct eigenvectors described in Section 6, we randomly selected a sample of 20 000 stars taking care of the fact that the sample selected reproduced the global properties of the whole sample of stars we analysed. By assuming the Mandel & Agol (2002) model, we first injected a set of transiting planets (one per star) with radii uniformly distributed between one Earth radius and 2.5 Jupiter radii. The orbital period was uniformly chosen between 0.25 d and the total time spanned from the first to the last measurement of each light curve. The transits were injected in the raw light curves (Section 4.2) and then all the post-correction procedure described in the previous sections was applied. The light curves were searched for transits using the BLS (Section 9). We then isolated the light curves for which the injected planets were recovered. This was done by looking at the period corresponding to the highest peak in the BLS periodogram and selecting those light curves for which the relative difference between the recovered period and the injected period was smaller than 1 per cent.

A similar procedure was applied to generate false positives. We first considered detached eclipsing binaries with equal mass ratio, radius of the secondary uniformly chosen between 2.5 $R_{\rm J}$ and the primary radius, secondary to primary eclipse ratios comprised between 0.1 and 0.9, orbital periods spanning the same temporal range as in the planet case and circular orbits. The eclipsing binary signal was injected into the same pool of constant stars used for the planet sample, and the same procedure explained above was applied to isolate the sample of recovered eclipsing binaries. In this case, we considered both the light curves for which the relative difference between the recovered period and the injected period was less than 1 per cent and those for which the relative difference between twice the recovered period and the simulated period was less than 1 per cent. Once primary and secondary eclipse depths are very similar, the BLS tends to recover half of the correct period. These stars are useful to train the odd/even metric, while those for which the recovered period is consistent with the simulated one are useful to train the secondary eclipse metric.

A different set of eclipsing binary simulations was performed, but this time the secondary eclipse was arbitrarily shifted of at most a quarter of the orbital period before or after the nominal secondary eclipse instant. This procedure was adopted to account for eccentric binaries that can be photometrically recognized precisely from the shift of the secondary relative to the primary eclipses.

A final set was prepared to simulate purely rotationally modulated variables (with perfectly sinusoidal shape) with very short orbital periods (comprised between 0.25 and 1 d) and amplitudes between 0.05 and 0.3 mag.

In total, we therefore constructed six different samples: constant stars, planets, eclipsing binaries (nominal secondary eclipse timing), eclipsing binaries (with recovered period equal to half the injected one, odd/even variables), eclipsing binaries (with shifted secondary eclipses), and sinusoidal ultrashort variables. For simplicity, in the rest of this discussion the eclipsing binaries and the ultrashort variables are referred globally as variables.

To build a final global training set of simulated objects, we regrouped in different proportions the various categories reported above. In sample 1, we considered 50 per cent of constant stars plus all variables (with equal proportions of the constant and all variables' subcategories) plus a 50 per cent of planets. In sample 2, we considered a 33 per cent proportion of constant stars, 33 per cent of all kinds of variables, 33 per cent of planets. In sample 3, we considered 50 per cent of planets and 25 per cent of constant stars and the remaining 25 per cent equally subdivided among all variables' categories. These training data set were built separately for the light curves derived with aperture 1 and for the light curves derived with aperture 2 for which the *TESS* magnitude of the simulated stars was limited to $T < 11$. The total dimension of the three samples was equal to 16 181, 16 175, and 16 344 for aperture 1 and it was equal to 15 893, 15 981, and 15 951 for aperture 2. We also build six corresponding testing data sets (one for each combination of samples and photometries). The composition of these test samples was equal to 50 per cent planets

[4]Assuming $\phi = 0$ at the transit time and the phase running from $\phi = -0.5$ and $\phi = 0.5$.

Downloaded from https://academic.oup.com/mnras/article/498/2/1726/5893317 by CIS Maldura user on 24 January 2025

Table 2. Performance measures of the RFC for the three samples described in the text and photometric aperture equal to 1 pix. FPR = False Positive Rate; TPR = True Positive Rate, FNR = False Negative Rate, TNR = True Negative Rate, AUROC = Area Under the ROC curve.

Sample	FPR (per cent)	TPR (per cent)	FNR (per cent)	TNR (per cent)	AUROC
1	1.0	89.6	10.4	99.0	0.9956
2	1.0	90.4	9.6	99.0	0.9937
3	1.0	92.0	8.0	99.0	0.9936

Table 3. Performance measures of the RFC for the three samples described in the text and photometric aperture equal to 2 pix. FPR = False Positive Rate; TPR = True Positive Rate, FNR = False Negative Rate, TNR = True Negative Rate, AUROC = Area Under the ROC curve.

Sample	FPR (per cent)	TPR (per cent)	FNR (per cent)	TNR (per cent)	AUROC
1	1.0	94.3	5.7	99.0	0.9968
2	1.0	94.8	5.2	99.0	0.9968
3	1.0	95.4	4.6	99.0	0.9962

and 50 per cent all the remaining categories (in equal proportion). The dimension of the testing samples was equal to 4000 stars.

In all cases, the problem was treated as a binary classification problem, that is we did not distinguished between variables and constant stars attributing to them the negative class, while we attributed the positive class to the simulated planets. Below we compared the performances of the RF algorithm when trained on these different samples.

10.3 Training

We then train the Random Forest model using the CARET package *train* function in R, using a 10-fold cross validation method with five repeats. This approach first randomly shuffles the data, then creates 10 partitions, 9 of which are used to train the model and 1 is used for testing. Each partition is hold out one time for testing and the others used for training the model and a score is attributed to each one of these combinations of testing/training subsets. The entire procedure is repeated five times with a different random shuffling and partitioning of the data. Scores are then averaged to evaluate the model performances. We repeated the above procedure by tuning the *mtry* parameter of the RF on a grid of integer values comprised between 1 and 6, while the number of trees (*ntree*) was hold fixed at its default value of 500 trees. The *mtry* parameter controls the number of predictors randomly chosen to split each node in a tree. The area under the receiving operating characteristics curve (AUROC) was chosen as the metric to evaluate the model performances for each value of *mtry*. In all cases, the result was that the best value for *mtry* was either *mtry* = 3 or *mtry* = 4. The AUROC values corresponding to the best models of each sample and aperture are reported in Tables 2 (aperture 1) and 3 (aperture 2). To choose the best model to adopt for each aperture, we decided to fix the false positive rate (FPR) at a value of 1 per cent, and selected the model providing the largest value of the true positive rate (TPR). Choosing a low false positive threshold is crucial in this kind of experiment, in order to avoid an overwhelmingly large number of false positive candidates with respect to the true positive candidates. As a comparison, Barclay, Pepper & Quintana (2018) estimated an hit rate (the ratio of planets detected to observed stars) < 0.75 per cent, for stars in the FFIs. By looking at the results in Tables 2 and 3, we decided to adopt the model relative to sample 3 for both apertures. The corresponding probability thresholds of the RFC are equal to $P_{\mathrm{RCF}_1} = 0.7401$ and $P_{\mathrm{RCF}_2} = 0.6421$ for aperture 1 and aperture 2, respectively. The ROC curves we obtained for the best models of each aperture are shown in Fig. 5. The continuous lines represent the TPR versus FPR for the best models, while the dotted lines along the diagonals represent the perfectly random classifier. The more a classifier is able to discriminate between the two classes, the more the curve should be closer to the top left corner of the diagram (which indicates the perfect classifier). The black dots in Fig. 5, visualize on the ROC curves the TPR and FPR corresponding to our adopted detection thresholds.

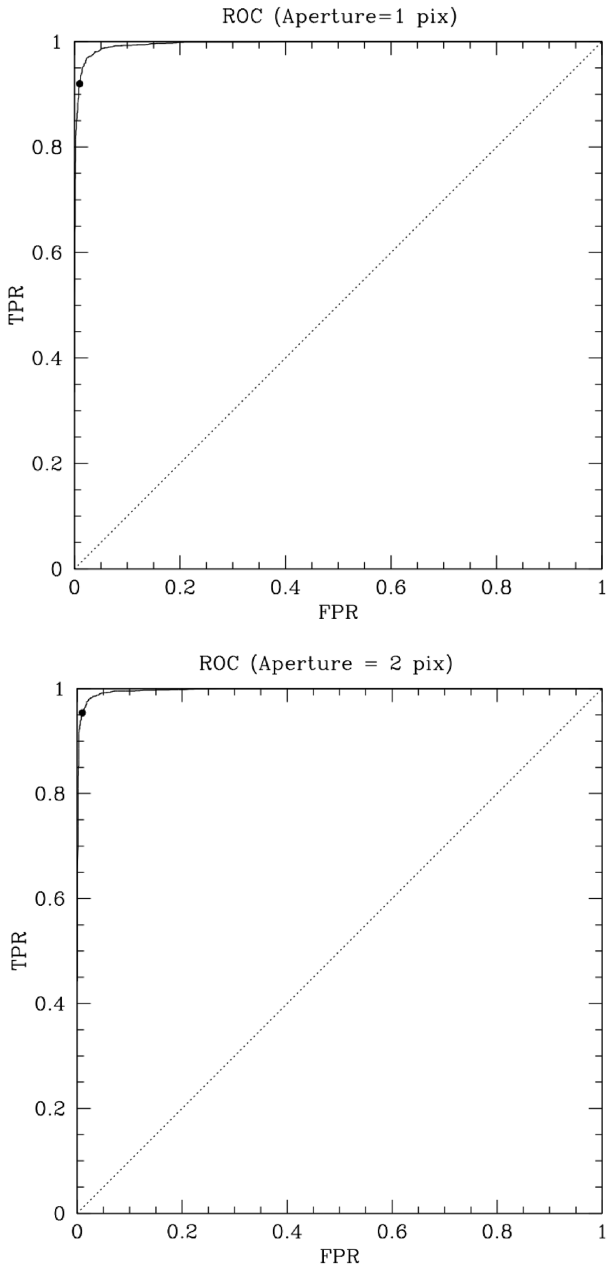

Figure 5. The solid curves denote the ROC curves for the best models of aperture 1 (top) and aperture 2 (bottom). The black points indicate the FPR and TPR corresponding to the adopted detection thresholds, while the dotted lines represent the perfectly random classifier.

Downloaded from https://academic.oup.com/mnras/article/498/2/1726/5893317 by CIS Maldura user on 24 January 2025

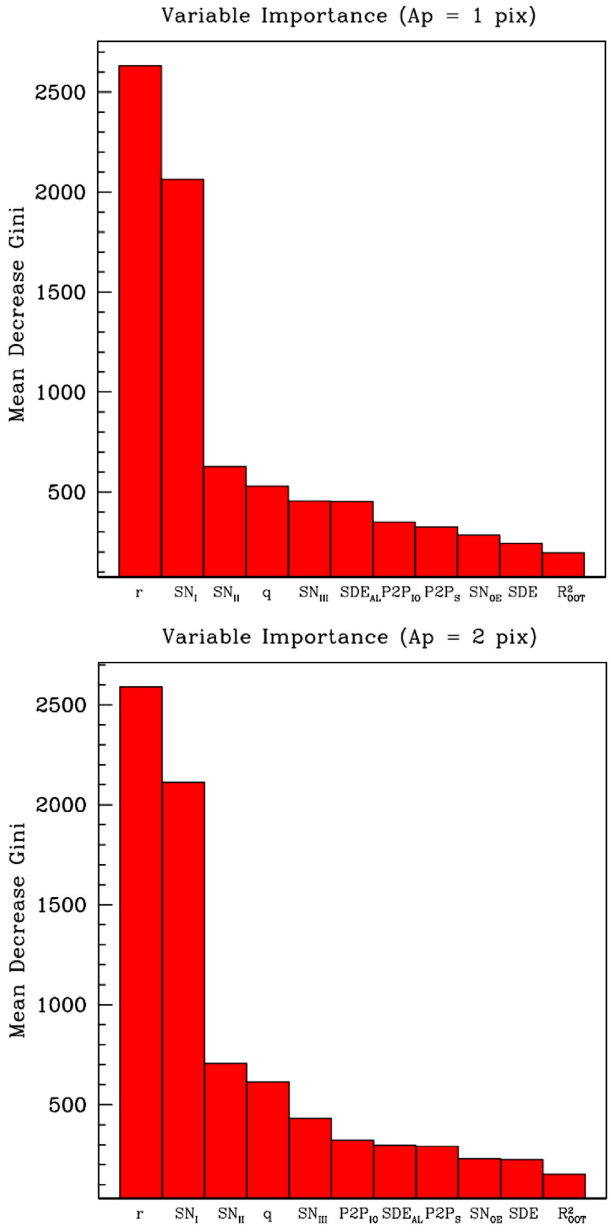

Figure 6. Mean decrease in the Gini impurity index for each variable involved in the classification process. The higher the decrease in the Gini Impurity, the more the variable is important.

10.4 Variable importance

In Fig. 6, we represent the importance that each variable has in the classification process in terms of the Mean Decrease Gini it produces. Such metric quantifies the total decrease in node impurity weighted by the proportion of samples reaching that node, averaged over all trees in the forest. The higher the Mean Decrease Gini, the more important is the variable. From Fig. 6 it results that for both apertures, the radius of the transiting object (r) and the signal to noise of the primary eclipse (SN_I) stand out in their importance with respect to the other variables. Also interesting to note is the fact that the signal detection efficiency metric (SDE) appears in general less important in the classification process than the SDE_{AL} which quantifies the power in the periodogram of aliases of the primary peak. Evaluating the power split on the moltitude of alias peaks appears more important for the classification process than evaluating the power of the sole primary peak.

10.5 Performances

The procedure described in the previous section permits to determine the performances of the RFC algorithm to disentangle plausible transiting planetary candidates from other kind of false positive events. It is important, however, to establish also the overall performance of the transit search algorithm that should also account for the performance of the BLS algorithm to correctly recover transit events. By using the same sample of stars we previously employed, we simulated a sample of 35 000 transiting planets with radii randomly selected between 1 and 25 $R_\oplus$ and periods randomly spanning the range between 0.25 d and the interval of time between the first and last observation of each light curve. Transits were injected in the raw light curves and we then repeated the full post-correction analysis, the BLS search, and applied the RFC algorithm. If the absolute relative difference between the BLS recovered period and the original injected period was smaller than 1 per cent and the probability returned by the RFC was larger than the adopted detection threshold, we considered the injected planet as recovered. The transit detection efficiency was then defined as the ratio of the number of recovered planets to the number of simulated planets. We analysed the results in a bidimensional grid presenting planetary radii against periods subdivided in steps sizes of 1 $R_\oplus$ and 1 d, respectively.

The results are presented in Fig. 7 for the sample of stars analysed with aperture equal to 1 pix (top) and equal to 2 pix (bottom). In general, the detection efficiency appears similar for both samples. It quickly drops below 50 per cent for planets with radii below 3 $R_\oplus$ in the short period planets domain and for orbital periods $P > 20$ d.

11 CENTROID MOTION

As it is well known, one of the issues of transit search programs is that they usually employ detectors with large pixel scales in order to monitor large field of views and measure the brightness of many stellar sources. In this way, it is expected that several stars may lie on the same photometric aperture especially in the most crowded fields, and consequently the signal of background eclipsing binaries can be diluted with the light of a brighter target star mimicking shallower planetary transit signals. One of the methods used to recognize these false positive signals is to monitor the centre of light motion of a given target and judge if any correlated motion occurs during a transit event pointing away from the target. Here, we used the difference images created during data reduction (Section 4) to study centroid shifts. For each one of these images, we extracted the flux weighted centroids on four concentric circular apertures centred on the targets. The circular apertures' radii were set equal to 1, 2, 3, and 4 pix, respectively. The contribution of each pixel to the weighted mean centroid of a given aperture was set equal to the absolute pixel flux multiplied by the area of the pixel contained within the aperture. The centroid measurements were registered on the same astrometric reference system of the reference astrometric image of a given sector, camera, and CCD (Section 4) and converted to RA and Dec. using the corresponding WCS information reported on the image header. Multisector observations were merged together after conversion to sky coordinates. Since the fluxes of both the target and contaminant stars change over time even outside transit windows, the centroid position based on the flux weighted measurements typically varies over time as well. This low frequency variations can be modelled out and

Downloaded from https://academic.oup.com/mnras/article/498/2/1726/5893317 by CIS Maldura user on 24 January 2025

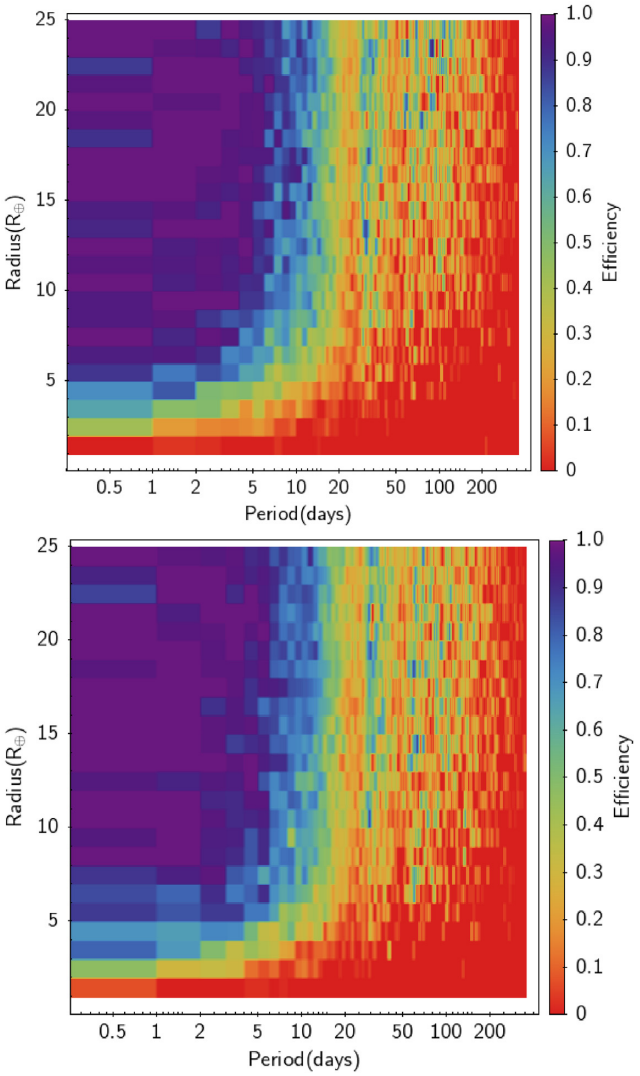

Figure 7. Transiting planets' detection efficiency as a function of injected planet radius and orbital period for the sample of stars analysed with photometric aperture equal to 1 pix (top) and equal to 2 pix (bottom).

corrected for. By using only out-of-transit centroid measurements, we repeated the splining procedure reported in Section 7 (applying it to centroid measurements). We therefore calculated the shift both in right ascension and in declination between the spline model and the median centroid and corrected all measurements to the median centroid estimate. After that, we considered only in-transit measurements and performed a PCA decomposition of the centroid shifts measurements calculating the two principal components and the standard deviations along these components. This procedure was repeated for each aperture separately. The centroid measurements, the orientation of the principal components' axes, and the associated standard deviations can be used to define four bivariate Gaussian distributions (one for each aperture). We then considered all sources within a radius of 3 arcmin from each target drawn from the *Gaia* catalogue and sorted them as a function of their Mahalanobis distance (Mahalanobis 1936) from the centroid of each aperture distribution. The source with the smallest distance was elected as the most likely source associated with the observed distribution of each aperture. If the elected source was the target star, we attributed a rank = 1 otherwise a rank = 0. Therefore, a criterium that can be adopted

Table 4. Best-fitting parameters of the logistic regression model related to the centroid metric $\log_{10}\,\eta$.

α_1	β_1
-0.2 ± 0.1	-2.8 ± 0.3

to consider a given target as the source of variability is to require all ranks associated with the four different apertures to be equal to 1. Hereafter, we will refer to these calculated ranks as the Mahalonobis ranks.

In practice, the effectiveness of the above procedure depends also on different factors. For example, we can expect that the larger the local number density of stars, the more likely is that a given source will fall by chance close enough to the centroid distributions that it could be erroneously identified as the source of variability. To study this problem more in detail, we analysed a random sample of known eclipsing binary sources and known transiting planets retrieved from the International Variable Star Index data base (VSX). We derived the centroid distributions for each one of them and analysed the cases where the selected variable was considered as the source of variability accordingly with the procedure described above (that is all Mahalanobis ranks were equal to 1), and the cases in which it was not. In our final list of 316 stars, half of them belonged to the first category and the other half to the second category. For each variable, we also registered the local average number density of stars (η) obtained as the average of the number of stars per *TESS* pixel measured in each one of the four apertures we considered. All the stars we considered in the calculation were retrieved from the *Gaia* DR2 catalogue. The density of the considered objects was ranging from approximately 0.01 to 100 stars px^{-2}. The result is shown in Fig. 8, where the cumulative distribution of the correctly identified sources is represented in blue and the cumulative distribution of the misidentified sources is represented in red. Both distributions are presented as a function of the $\log_{10}\,\eta$. Objects in the first group were attributed to Class = 1 and objects in the second group to Class = 0. The Class variable is then a binary dependent variable that can be statistically modelled to provide the probability that an object belongs to one class as a function of the predictor variable $\log_{10}\,\eta$. To this purpose we can use a logistic function, for which the log-odds l of a star to belong to Class = 1 is a linear combination of the predictor variable

$$l = ln\frac{1}{1-\mathrm{P}_\eta} = \alpha_1 + \beta_2\,\log_{10}\,\eta, \tag{18}$$

where P_η is the probability that Class = 1

$$\mathrm{P}_\eta = \frac{1}{1+e^{-(\alpha_1+\beta_1\,\log_{10}\,\eta)}} \tag{19}$$

the coefficients α_1 and β_1 are obtained by iterative non-linear optimization techniques. The best-fitting coefficients we found are reported in Table 4 and the best-fitting logistic function is represented by the continuous black line in Fig. 8. The result we obtained indicates that for a local average number density equal to 0.84 stars pix^{-2}, the probability of correct source identification is 50 per cent. Using this model, a threshold can be conveniently chosen for the local number stellar density in order to increase the chances of correctly identify a source of variability.

To further strengthen variable identification, other information can be exploited. For example, we should expect that if a variable source corresponds with the target (which is by definition at the centre of the apertures), then the centroid drift from the centre should be small and the centroid measurements should be clustered

Downloaded from https://academic.oup.com/mnras/article/498/2/1726/5893317 by CIS Maldura user on 24 January 2025

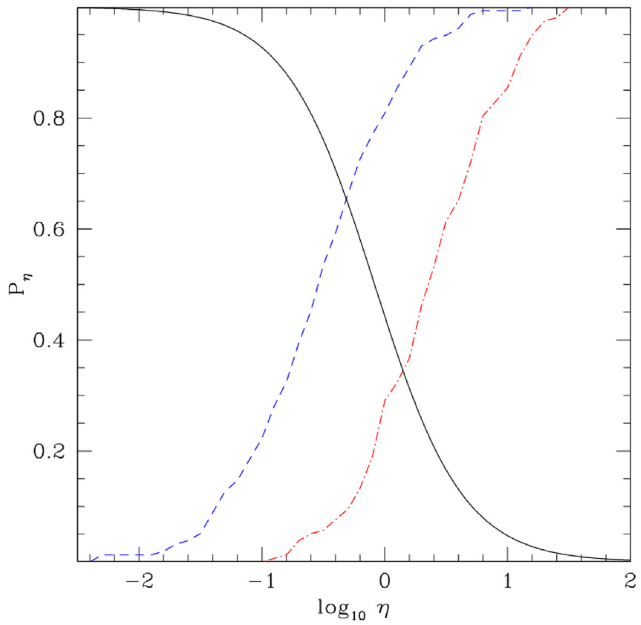

Figure 8. Cumulative distributions of the average local density metric ($\log_{10}\eta$) for targets identified as the correct source of variability (blue dashed line) and for misidentified sources of variability (red dash–dotted line) along with the best-fitting logistic model (black line) representing the probability of correct source identification (P_η) as a function of $\log_{10}\eta$ (as defined in equation 19).

around the target with possibly small dispersions. If instead the source of variability lies apart from the target, we could expect the centroid mesurement to be shifted in the direction of that source and the principal components' axes oriented correspondingly. Fig. 9 presents an example. On the left-hand panel, the apertures (denoted by the four concentric circles) are centred on a known eclipsing binary star (star 6782396682063717888 according to *Gaia* DR2). The coloured ellipses represent the centroid motion distributions for the four different apertures. In this case, all centroid measurements are clustered around the target which is effectively the source of variability. The Mahalanobis ranks are as well reported on the bottom of the figure and they are all equal to 1. On the right-hand panel, the apertures are centred on a close-by star (star 67823967164234506) on the South-East of the eclipsing binary. In this case, the centroid distribution measurements progressively depart from the target, precisely in the direction of the eclipsing binary and the outermost Mahalonobis rank is equal to 0.

It is possible to exploit this behaviour of the centroid motion measurements to construct a metric that will provide the probability of association with a given target. In this case, we analysed a sample of known variable sources and surrounding stars. For each one of them, we calculated the following quantity:

$$\log_{10} \mathrm{D} = \frac{1}{2}\log_{10}\left(\overline{r}^2 + \overline{\sigma}^2\right), \tag{20}$$

where

$$\overline{r} = \frac{\sum_{i=1}^{i=4} w_i\, r_i}{\sum_{i=1}^{i=4} w_i}$$

$$\overline{\sigma} = \frac{1}{\sqrt{\sum_{i=1}^{i=4} w_i}}$$

$$w_i = \frac{1}{\sigma_{1,i}^2 + \sigma_{2,i}^2}$$

with $\overline{r}$ the weighted average of the apertures' flux weighted centroids (r_i), $\overline{\sigma}$ the corresponding weighted average error, and $\sigma_{1,i}$, $\sigma_{2,i}$

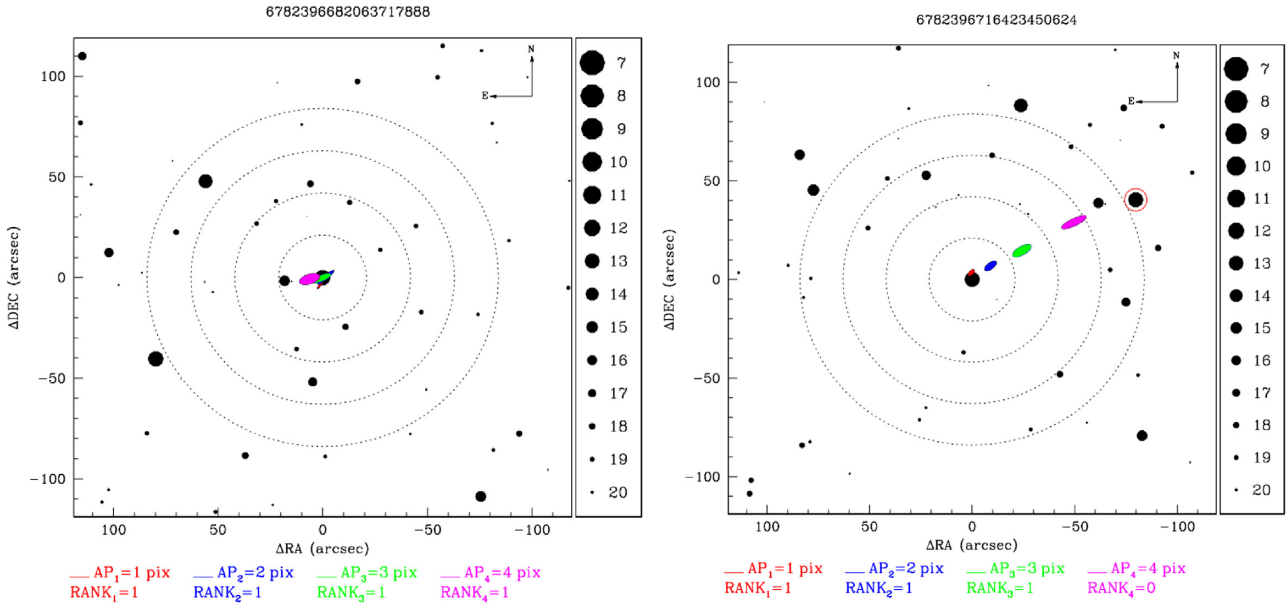

Figure 9. Two examples of application of the centroid algorithm. The plot on the left is centred on the star 6782396682063717888 (*Gaia* DR2 ID), a known eclipsing binary. The coloured ellipses represent the position and dispersion of the centroid metric measurements relative to this target for the four concentric apertures discussed in the text (and represented by the four dotted circles). The probability of source association P_D calculated from equation (22) is equal to 91 per cent. On the bottom, the Mahalanobis distance ranks are also reported. On the right, the same measurements are repeated for the close-by star 6782396716423450624, South-East of the eclipsing binary (highlighted in this panel with the red circle). The probability of source association P_D is equal to 20 per cent, in this case. The legend on the right of each plot is relative to the stars' magnitudes in the *Gaia G* band. All stars represented in the figure are taken from the *Gaia* DR2 archive.

Downloaded from https://academic.oup.com/mnras/article/498/2/1726/5893317 by CIS Maldura user on 24 January 2025

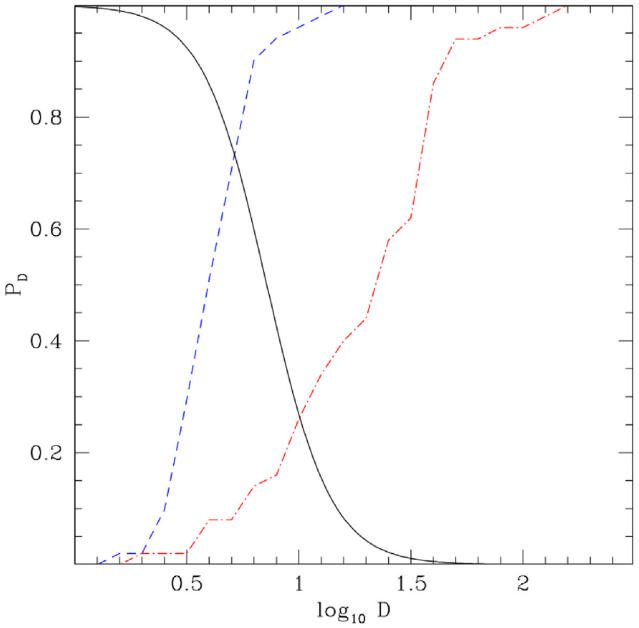

Figure 10. Cumulative distributions of the centroid metric ($\log_{10}$ D) for transit events associated with the target source (blue dashed line) and associated with surrounding sources (red dash–dotted line) along with the best-fitting logistic model (black line) representing the probability of correct source identification (P_D) as a function of $\log_{10}$ D (as defined in equation 22).

the principal components' standard deviations of each aperture probability density distribution.

In this case, our sample consisted of 50 stars known to be eclipsing binaries or planets and 50 close-by surrounding stars randomly chosen in a region between 3 arcsec and about 3 arcmin from the binaries. The distributions of the $\log_{10}$ D metric for these two objects' categories are represented by the histograms shown in Fig. 10. Also in this case we can model the probability that an object is correctly associated with the known variables with a logistic function using this time as predictor variable $\log_{10}$ D

$$l = ln\frac{1}{1-P_D} = \alpha_2 + \beta_2 \log_{10} D, \quad (21)$$

where P_D is the probability of correct association

$$P_D = \frac{1}{1+e^{-(\alpha_2+\beta_2 \log_{10} D)}} \quad (22)$$

the best-fitting coefficients α_2 and β_2 are reported in Table 5 and the best-fitting logistic function is represented by the continuous black line in Fig. 10.

The $\log_{10}$ D metric combines both the centroid disposition and its errors into a single quantity, averaged over all apertures. The result of the analysis can be interpreted as the fact that when the combined effect of the centroid disposition and its uncertainty amount to about 7.2 arcsec (that is ∼34 per cent of a TESS pixel dimension), the probability that the target source is the source of variability is 50 per cent. This model can be used to quantify the probability (P_D) that a given target is associated with the observed centroid motion distributions (and therefore with the observed transit events). For the situation represented in Fig. 9, the eclipsing binary on the left-hand panel has a probability of association equal to 91 per cent, whereas the close-by star on the right-hand panel has a probability equal to 20 per cent.

Table 5. Best-fitting parameters of the logistic regression model related to the centroid metric $\log_{10}$ D.

α_2	β_2
$6. \pm 1$	-7 ± 1

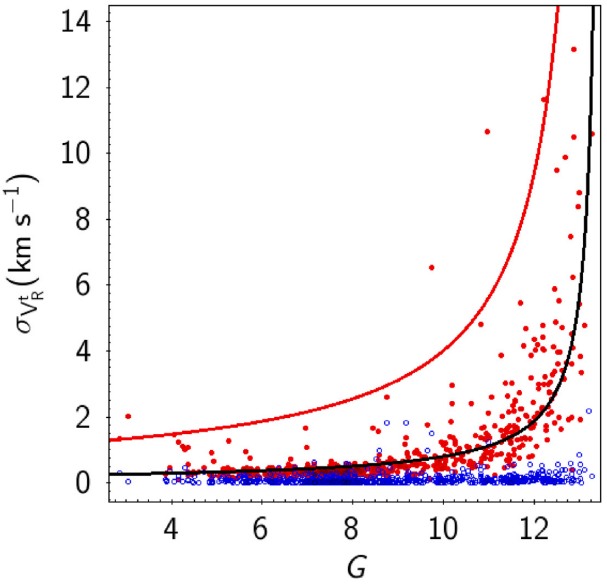

Figure 11. Standard deviation of the *Gaia* DR2 radial velocity measurements ($\sigma_{V_R^t}$, red dots) versus the *Gaia* *G*-band magnitude for a sample of 746 known planet hosts stars. Blue open circles show the radial velocity semi-amplitudes measured with high-precision spectrographs. The black line denotes the best fit of the standard deviation measurements and the red line the 5σ limit adopted in this work to flag suspected radial velocity variables.

11.1 Dynamical constraints

Additional information on the candidate planet's host stars can be obtained directly from the *Gaia* DR2 catalogue. In particular both radial velocities and their errors, as well as information on the quality of the astrometric solution, can be exploited to flag stars that potentially have stellar companions.

We retrieved from the exoplanet orbit data base (Han et al. 2014), a sample of known planet host stars for which the radial velocity semi-amplitude was determined and cross-matched it with the *Gaia* DR2 catalogue, requiring *Gaia* radial velocities and their uncertainties to be defined. The sample consists of 746 planet hosts.

The standard deviation ($\sigma_{V_R^t}$) of the *Gaia* radial velocity measurements can be obtained inverting equation (1) of Katz et al. (2019) which gives

$$\sigma_{V_R^t} = \sqrt{\frac{2N}{\pi}\left(\epsilon_{V_R}^2 - 0.11^2\right)}, \quad (23)$$

where ϵ_{V_R} is the radial velocity uncertainty and N the number of eligible transits[5] used to derive the median radial velocity, both of them reported in the *Gaia* DR2 catalogue.

In Fig. 11, we show the *Gaia* radial velocity standard deviation versus the *G*-band magnitude of the planet host stars (red points). Open blue circles show the radial velocity semi-amplitudes associated with

[5]The term *transit* used in the Gaia documentation is synonymous of visit.

Downloaded from https://academic.oup.com/mnras/article/498/2/1726/5893317 by CIS Maldura user on 24 January 2025

the planets known to orbit around these stars. It is apparent that the *Gaia* radial velocities are not precise enough to permit the detection of such planets. The radial velocity standard deviations shown in Fig. 11 simply reflect the limiting precision of *Gaia* measurements (Soubiran et al. 2018; Katz et al. 2019). For most of the stars, the radial velocity standard deviation is smaller than about 1 km s^{-1} down to $G \simeq 10$. After that, the *Gaia* radial velocity precision quickly deteriorates.

We fit the $\sigma_{V_R^t}$ as a function of the G-band magnitude adopting an hyperbolic relation of the form

$$f(G) = \frac{A}{13.5 - G} \tag{24}$$

valid for $G < 13.5$ where the best-fitting coefficient A is equal to $A = 2.808$. The black line in Fig. 11 represents this equation, which is also our best-fitting model for the *Gaia* radial velocity precision. It is important to recall that the *Gaia* radial velocity precision depends on several factors beyond the apparent magnitude (Katz et al. 2019), such as the effective temperature, for example, which implies that such precision cannot be in general described by a unique relationship as we did above. However, we are here only interested to set a conservative upper limit to flag potential radial velocity variability and therefore we adopted a 5σ threshold that corresponds to the following equation:

$$5f(G) = \frac{14.04}{13.5 - G} \tag{25}$$

and it is represented by the red line in Fig. 11. Stars for which the *Gaia* $\sigma_{V_R^t} > 5f(g)$ were flagged as potential binaries.

Similar considerations can be done also for what concerns the astrometric signal. Sources that are unresolved or barely resolved in *Gaia* DR2 may have poor astrometric solutions. Here, we follow the approach described in Evans et al. (2018) who used the Astrometric Goodness of Fit in the Along-Scan direction (gofAl) and the Significance of the Astrometric Excess Noise (astroExcessSig) as indicators of poorly resolved binaries. Accordingly to Evans et al. (2018), confirmed binaries have astroExcessSig > 5 and gofAL > 20 and we adopted the same thresholds in this work.

Both radial velocities and astrometric information should not be considered necessarily as conclusive indications that a given planetary candidate is a false positive. That is because the astrometric or radial velocity signals may be not associated with the observed transit events. They could also indicate the presence of additional (likely stellar) companions in the system. In our analysis, we decided to report candidates for which either the radial velocity or the astrometric conditions defined above are satisfied and to flag them as suspected binaries.

12 SELECTION OF PLANETARY CANDIDATES

To select planetary candidates, we started calculating their RFC probability (P_{RFC}). We then retrieved, from the *Gaia* archive, the list of contaminant stars of each target out of a distance of 3 arcmin from each target and calculated the probability P_η described in Section 11, related to the stellar field density. Fig. 12 (top panels) shows the P_{RFC} versus P_η diagrams for all the target stars of aperture 1 (left) and aperture 2 (right). For aperture 1, the entire target list (976 814 stars) was used and for aperture 2 we restricted the analysis to the brightest stars (with *TESS* magnitude $T < 11$, 201 510 stars). The continuous red lines in the same figures denote our adopted detection thresholds. The RFC thresholds have been described in Section 10, while for P_η we adopted a 30 per cent probability threshold.

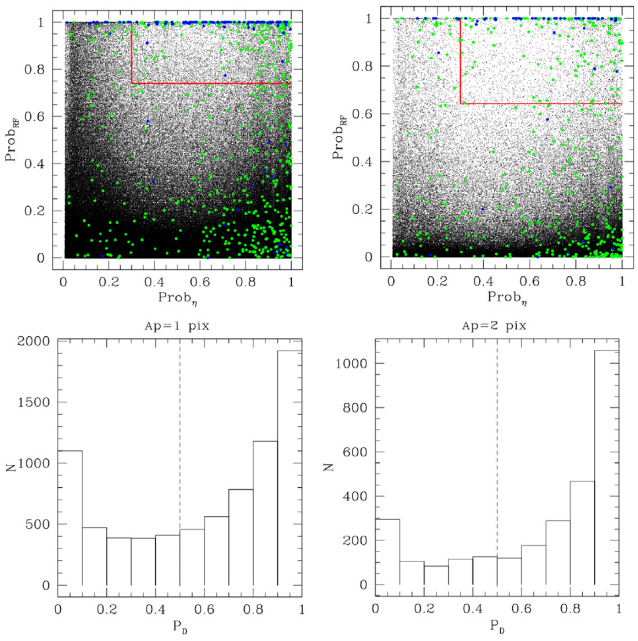

Figure 12. Selection of planetary candidates. The top figures show the P_{RF} versus P_η diagrams for light curves extracted with aperture = 1 pix (left) and aperture = 2 pix (right). Targets selected for the subsequent analysis are located within the top right rectangles delimited by the red lines. The bottom histograms show the distribution of P_D probabilities for targets passing the selection in P_{RF} versus P_η. Green dots denote known TOIs, while blue dots known transiting planets from the TEPcat compilation of Southworth (2011).

The number of objects that remained after applying these selection criteria was equal to 7658 stars for aperture 1 and 2836 stars for aperture 2. Green dots in Fig. 12 denote known TOIs, while blue dots known transiting planets from the TEPcat compilation of Southworth (2011) and will be discussed in Section 14. For each one of these stars, we then calculated the probability P_D related to the centroid motion, as described in Section 11. Fig. 12 (bottom panels) shows the distribution of candidates that passed the P_{RFC}, P_η thresholds as a function of the probability P_D. We applied a 50 per cent detection threshold for P_D as indicated by the vertical dashed lines in the same figure. The number of candidates that passed this criterium was equal to 4904 stars for aperture 1 and to 2111 stars for aperture 2.

The whole set of criteria applied up to this point served to identify the candidates with the highest probability of being a transiting planet accordingly to the morphological analysis performed by the RFC, located in the most isolated stellar fields, and having the smallest centroid motion. We then proceeded with a more detailed scrutiny of each individual candidate. For convenience, we merged the lists of candidates of the two apertures into a single unique list. We then analysed separately the photometry of each object first in aperture 1 and then in aperture 2. As a first step, we fit a transit model to each detected event, as described in the next section.

12.1 Transit analysis

All light curves that passed the selection criteria previously described were analysed using the Mandel & Agol (2002) transit model algorithm. We assumed a circular orbit, a rectified path across the

Downloaded from https://academic.oup.com/mnras/article/498/2/1726/5893317 by CIS Maldura user on 24 January 2025

transit window and that the mass of the transiting body was much smaller than the mass of the primary. We fit for the instant of transit minimum, the transit duration, the planet-to-star radius ratio, the stellar density, the linear and quadratic limb darkening coefficients, and a constant multiplicative factor of the form (1 + const). The solutions were obtained using a Levemberg–Marquardt algorithm (LM). The initial guess parameters for the LM algorithm were obtained from the BLS results and the stellar parameters from the TIC catalogue. Limb darkening coefficients were guessed using the Espinoza & Jordán (2015) software, plugging in the *TESS* response function. In some cases, we decided to fix some of the fitting parameters (usually the limb darkening coefficients). After obtaining the results of the analysis, we split the transits into odd and even and performed again the fitting analysis on these two groups of transits separately. In this case, we fixed all parameters to their best-fitting values, with the exception of the radius ratio and the time of minimum. Due to that a more refined estimate, the odd/even transit depth variability was calculated.

We then proceeded by fitting individually each single transit event. In this case, we fixed all parameters to the values of the best-fitting model but the instant of transit minimum. The times of the individual transits (T) we obtained were then interpolated with a linear model in order to obtain the best-fitting linear ephemerides

$$\mathrm{T} = \widehat{T}_0 + \widehat{P} \times \mathrm{E}, \tag{26}$$

where $\widehat{T}_0$ is the time of the reference epoch and $\widehat{P}$ is the orbital period and E is the epoch. The null epoch is set by definition as the epoch before the first observed transit. By assuming normally distributed errors, we also calculated the uncertainties on $\widehat{P}$ and $\widehat{T}_0$

$$\Delta\widehat{P} = \sqrt{\frac{\frac{1}{\mathrm{N}-2}\sum_{i=1}^{\mathrm{N}}\widehat{\varepsilon}_i^2}{\sum_{i=1}^{\mathrm{N}}(\mathrm{E_i}-\bar{\mathrm{E}})}} \tag{27}$$

$$\Delta\widehat{T}_0 = \Delta\widehat{P}\sqrt{\frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}\mathrm{E}_i^2}, \tag{28}$$

where $\widehat{\varepsilon}_i$ is the residuals of the fit, $\mathrm{E_i}$ and $\bar{\mathrm{E}}$ is the i-th epoch and the average epoch, and N is the number of transits. We also estimated the confidence bands around the regression line as

$$(T_0 + P\,\mathrm{E}) \in \left[\widehat{T}_0 + \widehat{P}\,\mathrm{E} \pm \{\alpha + \beta \times \mathrm{E}\}\right], \tag{29}$$

where

$$\alpha = t^*_{\mathrm{N}-2}\sqrt{\left(\frac{1}{\mathrm{N}-2}\sum\widehat{\varepsilon}_i^2\right)\cdot\left(\frac{1}{\mathrm{N}}\right)} - t^*_{\mathrm{N}-2}\bar{E}\sqrt{\left(\frac{1}{\mathrm{N}-2}\right)\cdot\left(\frac{\sum\widehat{\varepsilon}_i^2}{\sum(\mathrm{E_i}-\bar{\mathrm{E}})^2}\right)} \tag{30}$$

$$\beta = t^*_{\mathrm{N}-2}\sqrt{\left(\frac{1}{\mathrm{N}-2}\right)\cdot\left(\frac{\sum\widehat{\varepsilon}_i^2}{\sum(\mathrm{E_i}-\bar{\mathrm{E}})^2}\right)} \tag{31}$$

and $t^*_{\mathrm{N}-2}$ is the $(1\text{-}\frac{\gamma}{2})$-th quantile of the Student's t-distribution with N-2 degrees of freedom. We adopted $\gamma = 0.05$ that corresponds to a 95 per cent confidence level. Equation (29) permits a straightforward estimate of the uncertainty of transit times at any arbitrary epoch ($E \geq 0$). An example is shown in Fig. 13, where the 95 per cent confidence interval is represented by the dotted lines in the bottom panel.

12.2 Individual object analysis

After the results of the fit analysis were obtained, we scrutinized each object individually. For each candidate, we produced two diagnostic figures where we summarized all the information known up to this point of the analysis. The first plot is illustrated in Fig. 14 for the case of HAT-P-30/WASP-50. It presents several useful diagrams including the phase folded light curve with the best-fitting model and the residuals of the fit, the folded light curve at secondary phase, out-of-transit diagrams to determine the presence of out-of-transit variability, a stellar radius versus effective temperature diagram, the BLS periodogram, the odd/even diagrams with the corresponding fits in addition to several quantitative measurements related to the candidate, the star, the classification algorithm, and the dynamical constraints. The second plot is entirely similar to the ones presented in Fig. 9 and was used to analyse the centroid motion distributions. The meaning of all diagrams and quantities reported in these figures is explained in the documentation released with the data products. Both aperture photometries were inspected together with the centroid motion diagram. We searched in particular for differential transit depths in the two apertures. In some cases, we accepted some ambiguous situations (e.g. some transits detected only in one aperture, or Mahalanobis rank failures when the probability thresholds are met, etc.) and highlighted them with a bitmask flag in our catalogue (see Table 7). For all objects that were not obviously spurious, we then further refined the transit fit analysis reported in Section 12.1. This time an LM algorithm with bootstrap analysis with 1000 iterations was performed to obtain the best-fitting parameters and their uncertainties.

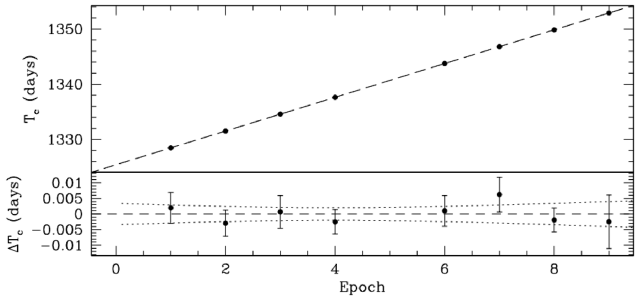

Figure 13. The top diagram shows the measured central transit times (dots) versus the epoch for the candidate 6612853122929259264 (*Gaia* DR2 ID). The bottom diagram shows the residuals after subtraction of the best-fitting linear ephemerides (denoted by the dashed line in the upper diagram). The upper and lower confidence intervals of the central transit times defined in equation (29) are represented by the dotted lines in the bottom diagram.

12.3 Dynamical constraints

By using *Gaia* DR2, we determined the possible presence of binaries in our sample, as described in Section 11.1. In Fig. 15, we show the candidates' significance of the astrometric excess noise versus the astrometric goodness of fit diagram (left) and the standard deviation of the radial velocity measurements versus the *G*-band diagram (right). Among the candidates, 51 show exceedingly large radial velocity standard deviations, 27 significant astrometric excess, and 8 of them both. The total list of suspected binaries, based on our adopted thresholds, is therefore equal to 70 objects. In Fig. 15 (left), we represented with green open circles stars that are either known TOIs (*TESS* objects of interest) or known CTOIs (Community TOIs). Eight of these stars are in the list of suspected binaries, based on their RV standard deviations. They are: star 260130483 (TOI 933.01), 207081058 (TOI 948.01), 9033144 (TOI 367.01), 49899799 (TOI 416.01), 219345200 (TOI 706.01), 382068562 (TOI 924.01),

Downloaded from https://academic.oup.com/mnras/article/498/2/1726/5893317 by CIS Maldura user on 24 January 2025

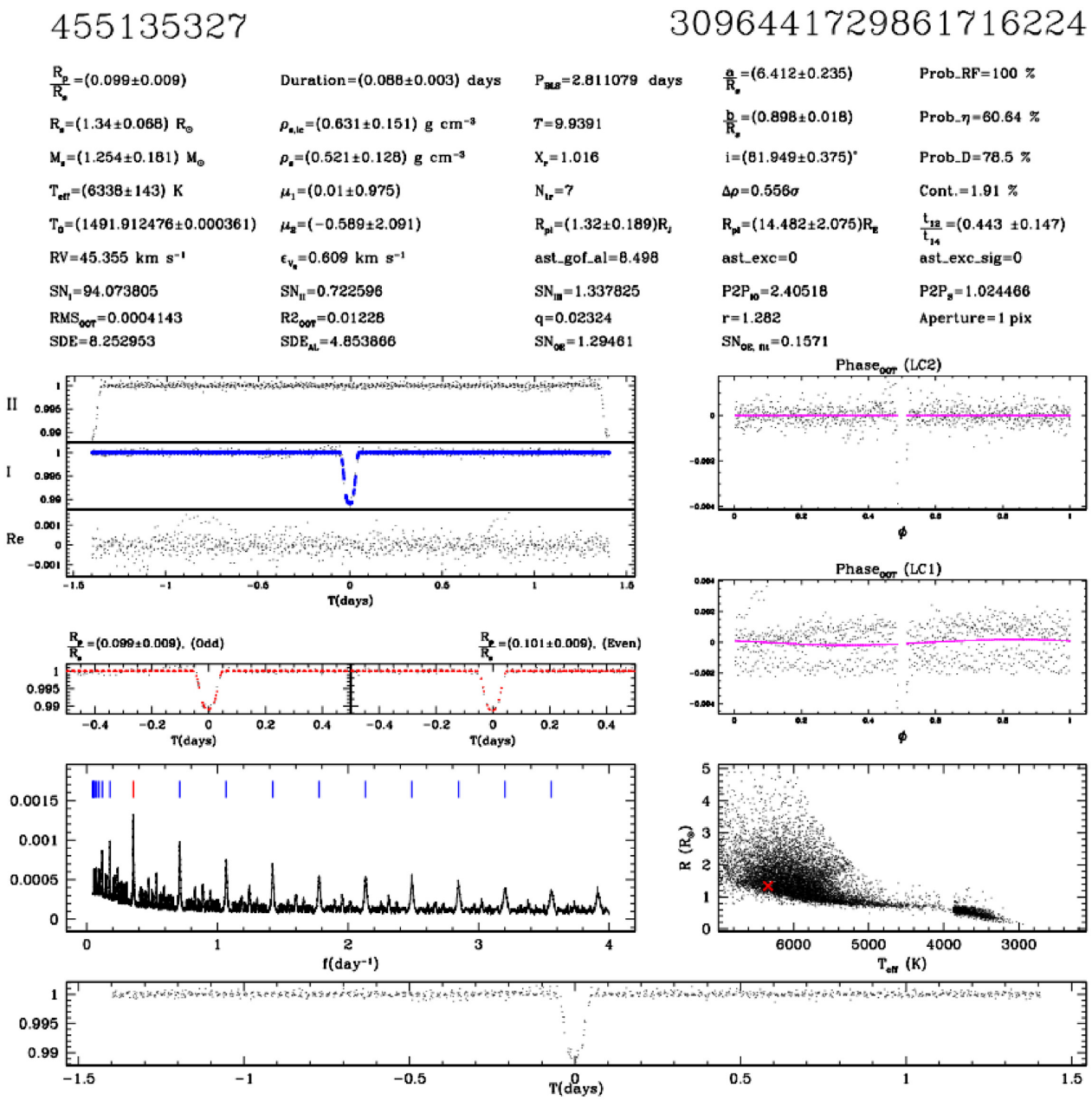

Figure 14. Example of diagnostic plot used during the screening process of the candidates. The object represented is HAT-P-30/WASP-50.

423670610 (TOI 850.01), and 146438872 (TOI 948.01). Seven TOIs/CTOIs are instead flagged on the basis of the astrometric indicators (Fig. 15, right): star 122612091 (TOI264.01, which is a known planet, WASP-72) 257567854 (TOI403.01/WASP-22), 40083958 (TOI851.01), 183593642 (TOI355.01), 261261490, 429302040 (TOI1905.01/WASP-107b), and 219345200 (TOI706.01). This last object is common to both lists.

12.4 Cross-match with external catalogues

We checked if the list of objects we found was included in known lists of variables or false positives. We cross-matched our catalogue with the International VSX catalogue searching for all variables within a distance of 3 arcmin from our targets and we found 60 stars. Almost all of them have also a reported period. By imposing a precision of 1 per cent between the period we found and the VSX period (or half or twice this period), the list is restricted to 30 objects. All of them are known planets from the WASP, HAT, and HATS surveys, with the exception of one object that corresponded to star 1SWASPJ055532.69−571726.0 in VSX and to star 734505581 in our catalogue. Such object is reported to be a detached eclipsing binary and it was therefore eliminated from the list. The matched objects without a reported period were SN 1995V, an SN II Supernova sitting 159 arcsec from 35857242, the variable ASASSN-V J071237.50−530912.6 at about 54 arcsec from 344087362, NSV 4303 at 0.2 arcsec from 13737885 which is

Downloaded from https://academic.oup.com/mnras/article/498/2/1726/5893317 by CIS Maldura user on 24 January 2025

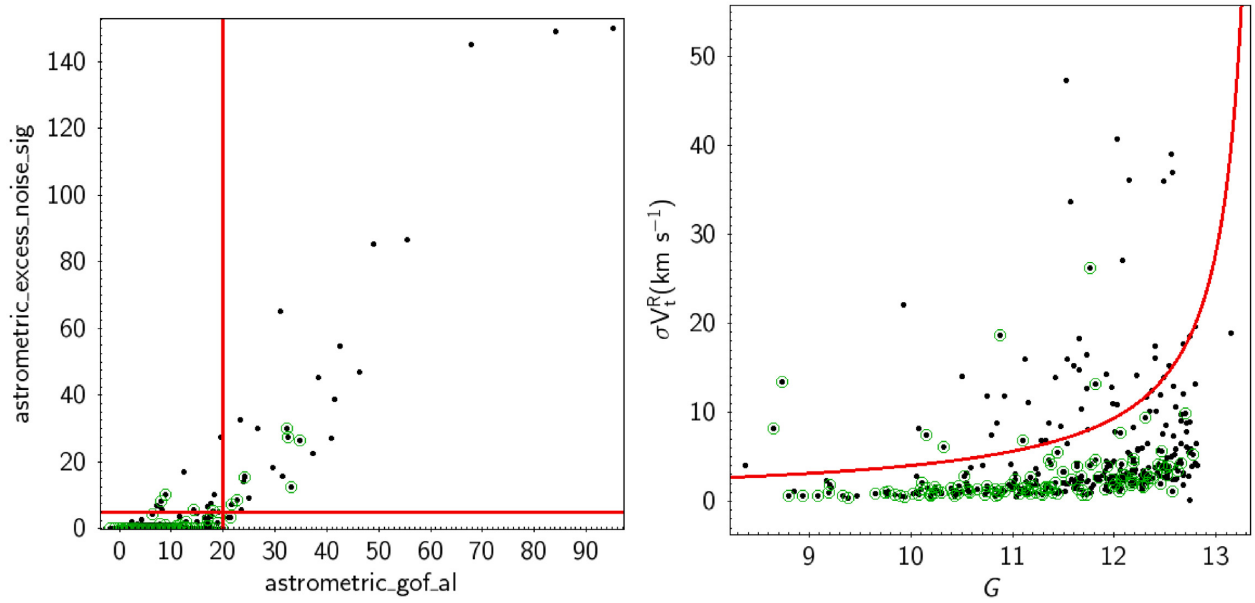

Figure 15. *Left:* Significance of the astrometric excess noise versus astrometric goodness of fit in the along-scan direction from the *Gaia* DR2 archive relative to the planetary candidates discussed in this work. *Right:* Planetary candidates' standard deviation of *Gaia* radial velocity measurements against *Gaia G*-band magnitude. The red lines denote selection criteria we used to flag potential binaries (see text), while green circles show known TOIs and CTOIs.

classified as CST (that is a retracted variable star), and CR Gru at 171 arcsec from 265612438 classified as LB, a slow irregular variable of late spectral type.

We then considered the KELT Follow-Up Network and Transit False Positive Catalog (KELT-FUN; Collins et al. 2018), an all-sky catalogue of 1128 bright stars ($6 < V < 13$) showing transit-like features in the KELT light curves that have been then ruled out as false positives by follow-up observations. By repeating the same procedure reported above, we found one match with our list, which corresponds to star 144426921 (TIC ID). This object is classified as an SB2 (multilined binary) and it is has been therefore eliminated from our list.

Another valuable compilation of known false positives is the SuperWASP dispositions and false positive catalogue (Schanche et al. 2019) which lists 1041 Northern hemisphere SuperWASP targets, rejected as false positives by follow-up observations. Part of our targets are present also in the equatorial region and northern sky. We found in this case three matches with our catalogue, which correspond to stars 443618156, 16490297, and 9727392 (TIC IDs). These objects are classified as EB or EBLM binaries and were therefore eliminated. We note that star 9727392 is also included in the list of known TOIs (TOI236.01). The public comment reports a 1700 ppm secondary detection and flagged it as a likely EB.

13 RESULTS

The search yielded 396 candidates among which 144 are known TOIs or community TOIs and 252 are new candidates.[6] We also compared our candidates' list with the SPOC multisector TCE list available from MAST[7] and found 22 matches among which 20 were known TOIs or CTOIs. The other two stars are stars 141770592 (TIC ID) and 177350401. Table 6 reports the minimum, the first quartile, the median, the mean, the third quartile, and the maximum values of the period and radius distributions. The median values of the distributions correspond to Jupiter planets in short orbital period (Hot Jupiters). The radii distribution is extended down to $1R_\oplus$ and the orbital period distribution up to ∼105 d. By considering the distribution of the candidates' impact parameters (b) obtained from the fitting analysis, we found that ∼50 per cent (197) of the candidates have $b \leq 0.8$. Because of the long temporal cadence of TESS FFIs, planetary candidates, especially around late-type dwarfs, may have preferentially V-shaped transits. Nevertheless, the statistical argument for the planetary nature of $b < 0.8$ objects is generally stronger than for $b > 0.8$ objects (e.g. Seager & Mallén-Ornelas 2003). To facilitate follow-up analysis and target prioritization, we subdivided our list of candidates in five tiers as a function of the impact parameter value, with candidates having $b \leq 0.2$ belonging to the first tier (see also Table 7).

We discovered a long-period candidate planet ($P \sim 92$ d) with a radius of about 0.7 R_J around a likely early K dwarf (TIC ID = 382200986). We recovered a Jupiter like candidate around star 309792357. We note that this object is included in the list of TOIs (TOI ID = 199.01). The public comments indicate the presence of a single transit, but we recovered three transits and inferred a period of ∼105 d. According to our analysis therefore this object is a long

Table 6. Minimum, first quartile (1stQ), median, mean, third quartile (3rdQ), and maximum values of candidates' orbital periods (P) and radii (R_p) distributions.

	Min	1stQ	Median	Mean	3rdQ	Max
$R_p(R_\oplus)$	1.0	7.0	11.5	11.8	15.1	28.9
P (d)	0.25	2.1	3.8	6.6	7.2	104.9

[6]We refer to the candidates' relase of 2020 June 19.
[7]http://archive.stsci.edu/tess/bulk_downloads/bulk_downloads_tce.html the relevant list is tess2018206190142-s0001-s0013_dvr-tcestats.csv.

Downloaded from https://academic.oup.com/mnras/article/498/2/1726/5893317 by CIS Maldura user on 24 January 2025

Table 7. Content of the catalogue released with this work.

Column number	Column name	units	Description
1	ticID	–	TIC catalogue ID number
2	gaiaID	–	*GAIA* DR2 source ID number
3	ra	deg	Right ascension from *GAIA* DR2
4	dec	deg	Declination from *GAIA* DR2
5	p	–	Radius ratio
6	pErr	–	Error on radius ratio
7	t0Fit	d	Central transit time from transit fit
8	t0FitErr	d	Error on central transit time from transit fit
9	periodBLS	d	Orbital period from BLS analysis
10	duration	d	Transit duration
11	durationErr	d	Error on transit duration
12	denFit	g cm^{-3}	Stellar density from transit fit
13	denFitErr	g cm^{-3}	Error on stellar density from transit fit
14	u1	–	Linear limb darkening coefficient
15	u1Err	–	Error on linear limb darkening coefficient
16	u2	–	Quadratic limb darkening coefficient
17	u2Err	–	Error on quadratic limb darkening coefficient
18	const	–	Constant term in transit fit
19	constErr	–	Error on constant term in transit fit
20	chir	–	Reduced chi squared of the transit fit
21	deltarho	–	Difference of stellar density from the TIC and from transit fit, normalized by the square sum of the errors
22	trdepth	ppm	Transit depth
23	rmsoot	–	Root mean square of out of transit measurements
24	ar	–	Ratio of semimajor axis to stellar radius
25	arErr	–	Error on ratio of semimajor axis to stellar radius
26	b	–	Impact parameter
27	bErr	–	Error on impact parameter
28	i	deg	Inclination
29	iErr	deg	Error on inclination
30	t12t14	–	Ratio of ingress time to total transit time
31	t12t14Err	–	Error on ratio of ingress time to total transit time
32	rpj	$R_{\rm J}$	Candidate radius
33	rpjErr	$R_{\rm J}$	Error on candidate radius
34	rpe	$R_{\oplus}$	Candidate radius
35	rpeErr	$R_{\oplus}$	Error on candidate radius
36	rank1	–	First Mahalanobis rank
37	rank2	–	Second Mahalanobis rank
38	rank3	–	Third Mahalanobis rank
39	rank4	–	Fourth Mahalanobis rank
40	prf	per cent	Random Forest probability
41	pd	per cent	Centroid motion probability (equation 22)
42	peta	per cent	Local average stellar number density probability (equation 19)
43	t0Lin	d	Central transit time from linear fit
44	t0LinErr	d	Error on central transit time from linear fit
45	periodLin	d	Orbital period from linear fit
46	periodLinErr	d	Error on orbital period from linear fit
47	alpha	d	α parameter from equation (30)
48	beta	d	β parameter from equation (31)
49	teff	K	Stellar effective temperature from TIC
50	teffErr	K	Error on stellar effective temperature from TIC
51	radius	$R_{\odot}$	Stellar radius from TIC
52	radiusErr	$R_{\odot}$	Error on stellar effective temperature from TIC
53	mass	$M_{\odot}$	Stellar mass from TIC
54	massErr	$M_{\odot}$	Error on stellar mass from TIC
55	rho	g cm^{-3}	Stellar density from TIC
56	rhoErr	g cm^{-3}	Error on stellar density from TIC
57	contratio	per cent	Contaminatio ratio from TIC
58	tmag	–	*TESS* magnitude from TIC
59	vmag	–	*V* magnitude from TIC
60	gmag	–	*G* magnitude from TIC
61	rv	km s^{-1}	Radial velocity from *Gaia* DR2
62	rvErr	km s^{-1}	Error on radial velocity ($\epsilon_{\rm RV}$) from *Gaia* DR2
63	nrv	–	Number of eligible transits used to derive the median radial velocity from *Gaia* DR2
64	srv	km s^{-1}	Standard deviation of radial velocity measurements

Downloaded from https://academic.oup.com/mnras/article/498/2/1726/5893317 by CIS Maldura user on 24 January 2025

Table 7 – *continued*

Column number	Column name	units	Description
65	gofAl	–	Goodness of fit in the along scan direction from *Gaia* DR2
66	astroExcess	mas	Excess of astrometric noise from *Gaia* DR2
67	astroExcessSig	–	Significance of Excess of astrometric noise from *Gaia* DR2
68	snI	–	S/N ratio of primary transit from equation (6)
69	snII	–	S/N ratio of secondary eclipse from equation (7)
70	snIII	–	S/N ratio of tertiary eclipse from equation (8)
71	snOE	–	Odd/even S/N ratio from equation (11)
72	snOEFit	–	Odd/even S/N ratio from transit fit
73	r2oot	–	R^2_{OOT} parameter from equation (13)
74	q	–	Fractional transit duration from equation (14)
75	sde	–	Signal detection efficiency from equation (9)
76	sdeAL	–	Signal detection efficiency of alias peaks from equation (10)
77	p2pio	–	In/Out of transit point-to -point noise from equation (15)
78	p2ps	–	$P2P_s$ parameter from equation (16)
79	r	R_J	Estimated candidate radius from equation (17)
80	ntr	–	Number of transits
81	apnum	pix	Aperture photometry for which the transit analysis was performed
82	btier	–	A level flag based on the impact parameter value btier = 1, $b \leq 0.2$ btier = 2, $0.2 < b \leq 0.4$ btier = 3, $0.4 < b \leq 0.6$ btier = 4, $0.6 < b \leq 0.8$ btier = 5, $b > 0.8$
83	bitmask	–	Bitmask flag bitmask = 1 – binary according to *Gaia* DR2 dynamical constraints bitmask = 2 – Single transit bitmask = 4 – At least one Mahalanobis rank failed bitmask = 8 – Present only in Aperture 1 bitmask = 16 – Present only in Aperture 2
84	mult	–	Candidate numeration within multiple systems

period transiting candidate. We also confirmed the presence of a long-period candidate ($P \sim 89$ d) around a late G sub-giant star (TIC = 260130483, TOI ID = 933.01) and a long period ($P \sim 53$ d) giant planet candidate around star with TIC = 350618622 (TOI ID = 201.01). Star 270341214 (TOI173.01) is another interesting object. The public comment of this TOI indicates the presence of a single transit. We detected, however, 2 transits, one in sector 1 and one in sector 13. The BLS period is ∼25 d likely because of a saturation effect on the periodogram due to the presence of a large gap in the data. Indeed, this star has been observed only in these two sectors and therefore this object could have potentially a very long period (∼327 d). Because of the large gap in the data, it is entirely possible that the period is a submultiple of this value. We note that the transit duration is also long (8.42 h) and the star radius is ∼1.4 $R_\odot$. This somewhat supports the fact that this object could be a long-period planetary candidate. Moreover, we additionally detected 15 single transit events (one of which is known, TOI 706.01, TIC = 219345200) that may potentially be long period candidates.

We detected a conspicuous population of candidates with orbital periods between 10 and 50 d and radii between 2 $R_\oplus$ and 2.6 R_J. In total this sample amounts to 64 objects, 42 of which are new candidates. There are 39 candidates with radius $R_p < 4$ $R_\oplus$, where 15 are new candidates.

13.1 Multiples

We searched for multiple planetary candidates after subtracting the best-fitting model of the primary candidate and repeating the transit search as described in the previous sections. We adopted the same detection tresholds adopted for the primary transits. In this way, we detected a new candidate around star 47601197 with radius $R_p = (3.4 \pm 0.6)$ $R_\oplus$ and period equal to ∼4.9 d, nearly half the one of the primary candidate (which is itself a new candidate with $P = 8.6$ d, $R_p = 9.0 \pm 1.0$ $R_\oplus$). We confirmed the super-Earth around star 259377017 (TOI ID = 270.02, Günther et al. 2019) with period $P \sim 11.4$ d and radius $R_p = (2.2 \pm 0.3)$ $R_\oplus$.[8] We found a new candidate around star 260417932 with period $P \sim 8$ d and $R_p = (1.8 \pm 0.3)$ $R_\oplus$.

Fig. 16 shows some representative examples of candidates we detected. Along with the folded light curves (on the left) and the BLS periodograms (on the right), we report the TIC ID number, the orbital period (P), the radius of the transiting object (R_p), the number of observed transits (N_{tr}), and the random forest probability (P_{RF}).

14 DISCUSSION

Fig. 17 shows the stellar radius versus effective temperature diagram of the transiting candidates' host stars. The dimensions of the points and their colours denote the planetary radii and periods, as reported in the legend. We plot only stars for which the relative errors on the planetary radius is smaller than 30 per cent and excluded single

[8] We detected also the signal of TOI 270.03 on aperture = 2 pix with period ∼3.4 d and radius ∼1.5 $R_\oplus$, but the random forest probability was 52.8 per cent, below our detection threshold.

Downloaded from https://academic.oup.com/mnras/article/498/2/1726/5893317 by CIS Maldura user on 24 January 2025

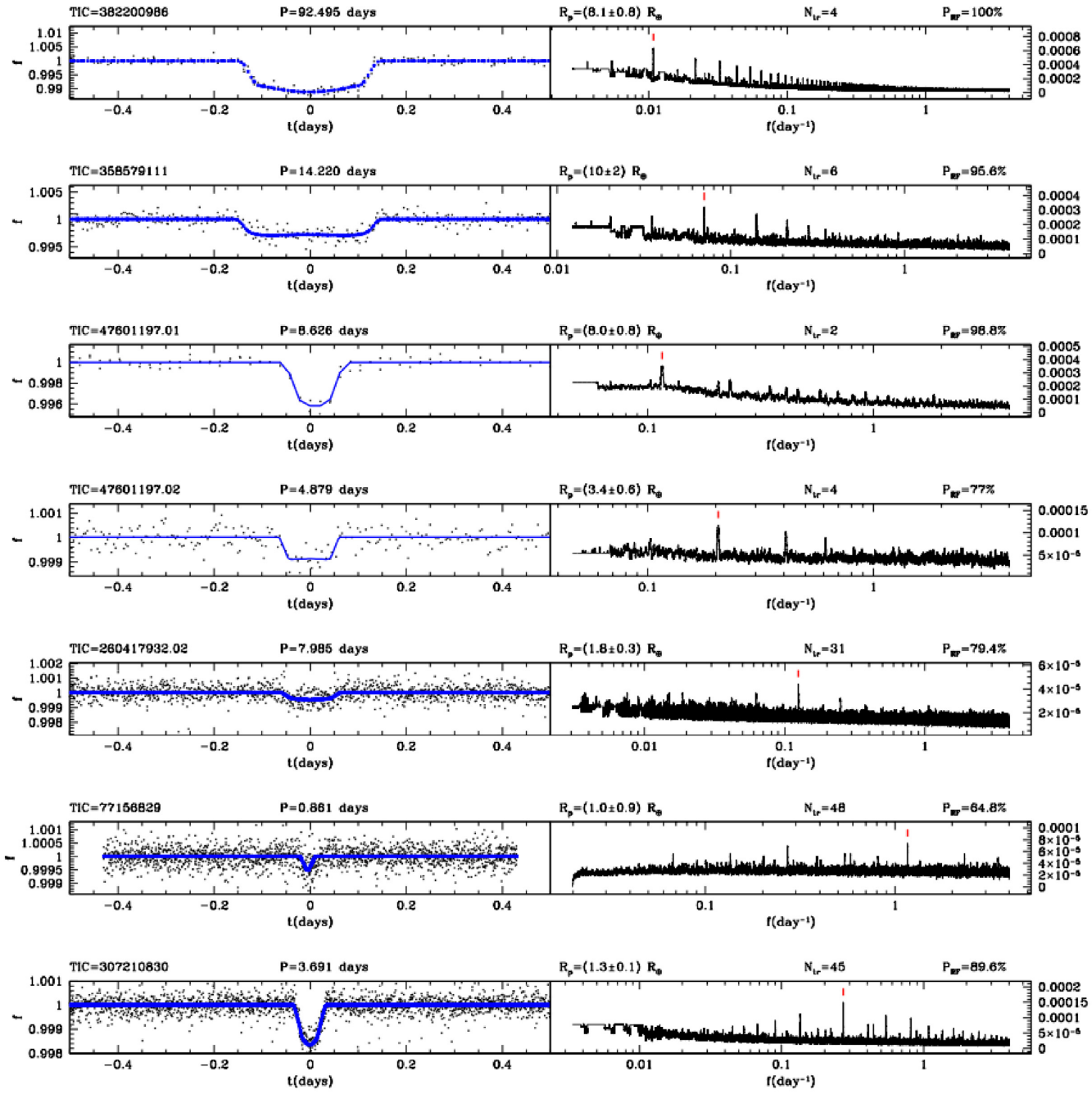

Figure 16. A few representative examples of transiting planetary candidates we detected. On the left, the folded light curves and on the right the corresponding BLS periodograms. The labels on the top report the TIC ID number, the orbital period (P), the transiting object's radius (R_p), the number of observed transits (N_{tr}), and the random forest probability (P_{RF}).

transit hosts.[9] In total, the diagram represents 304 candidates. From this figure, we can deduce that candidates are present in all the portions of the diagram, from cool late-type dwarfs to evolved subgiants stars.

It is important to keep in mind some caveats once interpreting the results discussed in this work. First, our analysis has been purposely limited to less crowded regions of the sky by imposing a cut-off threshold on the local average stellar density. This is a reasonable choice for a work like this one, since it increases the chances of correct source identification. Unavoidably, however, as demonstrated in Fig. 12, this likely reduced the efficiency of the search. We found 618 TOIs in our initial target list in aperture 1 (436 in aperture 2) and 139 were eliminated on the basis of the stellar density condition, which gives a fractional loss of 20 per cent, 30

[9]An error of 10 per cent was added in quadrature to the stellar radius error (and propagated to the planetary radius error) for stars with $T_{\rm eff} < 3950$ K because the error reported in the TIC catalogue was very small.

Downloaded from https://academic.oup.com/mnras/article/498/2/1726/5893317 by CIS Maldura user on 24 January 2025

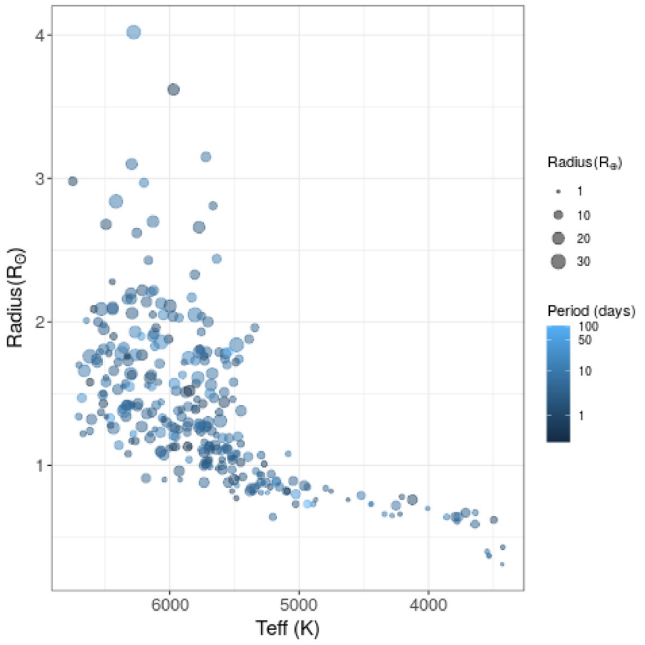

Figure 17. Stellar radius versus $T_{\rm eff}$ for planetary candidates' host stars. The different sizes and colours of the points in the diagram codify the planetary candidates' radii and orbital periods as explained in the legend.

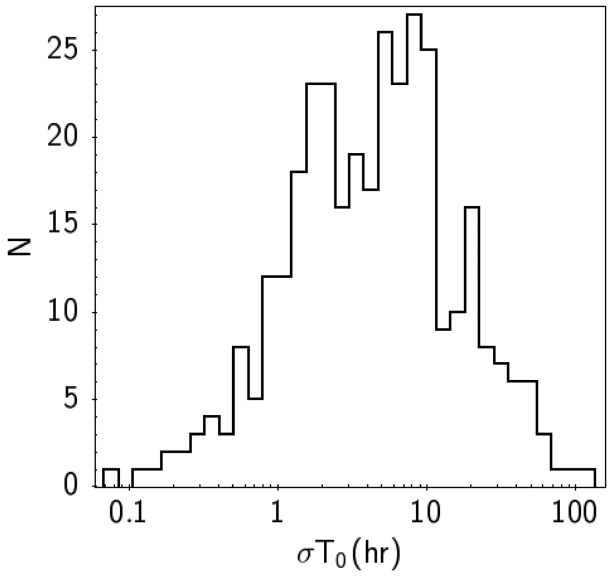

Figure 18. Uncertainty on the central transit time of planetary candidates calculated with equation (29), assuming all objects will be observed on 2020 August 1.

per cent for aperture 1 and 2, respectively. If we consider the TEPcat catalogue (Southworth 2011), we found 118 stars in common with aperture 1 sample (56 for aperture 2). Out of them 28 and 16 were eliminated by the condition on the stellar density which gives a fractional loss of 23 per cent and 29 per cent for aperture 1 and 2, respectively. Then, by restricting the analysis to only those targets that passed the stellar density condition, we obtain that the overall TOIs recovery rate is 29 per cent−42 per cent and that the TEPcat recovery rate is 57 per cent−83 per cent. The gap between these two samples' recovery rates can be explained by the fact that the TOI sample is strongly biased towards small radii objects. In the TOI sample in common with our target list, 48 per cent of the candidates have a radius < 5 $R_\oplus$, whereas in the TEPcat sample this percentage decreases to 10 per cent. As shown in Fig. 7, we expected a significant drop in our efficiency in this regime. It should be also recalled that our detection thresholds were set to relatively conservative values, as demonstrated for example by the ROC curves of Fig. 5. There would be certainly room to increase the detection efficiencies in both apertures by tweaking the detection thresholds, at the price, however, of tolerating higher false positive rates than the 1 per cent level we decided to adopt. Certainly other factors impact efficiencies beyond the ones we considered. A complete examination of them goes well beyond the purpose of this work, but it is sufficient to recall for example that our training algorithm was based on simulated data sets and that these data sets certainly are not exhaustively incorporating all the possible sources of false positives which may affect the detection of transiting planets. Despite these limitations, it is remarkable that we managed to discover and to recover a substantial sample of small size and of long period candidates, as demonstrated in Section 13. Because our detection efficiencies are clearly reduced in these two domains (e.g. Fig. 7), this clearly points towards the conclusion that the abundance of these objects is large.

As a last point, it is also worth to have a look at the ephemerides' uncertainties of the detected objects. We calculated the uncertainty in the central transit time as the difference between the maximum and minimum ephemerides obtained from equation (29), assuming to observe on 2020 August 1. Fig. 18 shows the distribution of the uncertainties that range from a minimum of 5 min to a maximum of 4.8 d with a median value of 4.7 h. This fact is in itself a very good reason for an extended mission to observe again these objects and refine their ephemerides. For example, Bouma et al. (2017) showed that if we detect an additional transit 1 yr after the final observed transit from the Primary Mission, the uncertainty on the mid-transit time decreases by an order of magnitude. This will permit to observe several candidates for many years to come.

15 CONCLUSIONS

In this work, we explored the potential of *TESS* FFIs to detect transiting planets around a well defined sample of FGKM dwarfs and subgiant stars in the Southern ecliptic hemisphere. We discussed our new reduction pipeline, DIAmante, based on the differential imaging analysis to optimally extract multisector photometry from FFIs. We then presented our post-correction analysis method and transit search approach. In particular, we discussed a morphological classification algorithm based on the Random Forest technique which permits to discriminate planetary transiting candidates from different categories of false positives. We discussed the ROC analysis and the overall algorithm performances. We then focused our attention on the centroid motion algorithm, introducing different quantitative metrics to reduce the chances of wrong source identification. First, we discussed the Mahalanobis distance classification, which is based on the distance of the target star from the centroid motion distributions derived in four concentric apertures centred on the target. We then developed a probabilistic model to account for the probability to misidentify a source of variability when the local average stellar density increases. We showed that when the local stellar average number density is equal to 0.84 stars pix^{-2}, the probability of correct

Downloaded from https://academic.oup.com/mnras/article/498/2/1726/5893317 by CIS Maldura user on 24 January 2025

source identification is 50 per cent. We then analysed the empirical behaviour of different quantities related to the centroid motion distributions, combining them into a unique metric. We showed that when the combined effect of the centroid motion disposition and its uncertainty amount to about 7.2 arcsec (that is 34 per cent of the *TESS* average pixel dimension), the probability that the target source is the source of variability is equal to 50 per cent. We then considered *Gaia* dynamical constraints and presented the method of selection of planetary candidates. Our search produced a list of 396 planetary candidates, out of which 252 are new candidates. By applying dynamical constraints from *Gaia*, we found that 70 (18 per cent) candidates show evidence of unresolved binarity. This list includes 14 TOIs/CTOIs. The planetary radii distribution ranges from $1R_\oplus$ to 2.6 R_J with a median value 1 R_J. The orbital period distribution ranges between ∼0.25 and ∼105 d with a median value of 3.8 d. We discovered a new long period candidate with a period of 92 d and R = 0.7 R_J and inferred a long period for TOI 199.01 (P = ∼105 d). Other two TOIs with periods larger than 50 d were detected. We also discussed the curious case of TOI 173.01, observed only in Sector 1 and Sector 13 and for which two transits separated by ∼327 d were detected. Additionally, 15 single transit events have been found and could potentially be long period candidates. We found 64 candidates with orbital periods between 10 and 50 d, 42 of which are newly discovered. We also detected 39 candidates with radii smaller than 4 $R_\oplus$, and 15 of them are new. Finally, we found a new multiple system around star 4701197 and found a new candidate planet around star 260417932.

During our work, we largely benefited from the public compilations of variable stars and known false positives which have been made publicly available. This is worth mentioning, given that these lists are extremely valuable resources that permit to avoid duplication of efforts and we certainly encourage a broad diffusion of similar lists available to other groups.

All of the objects discussed in this work have been detected by *TESS* about 1 or 2 yr ago. We showed that the median uncertainty of the candidates' central transit time is ∼4.7 h. This is a very good reason for *TESS* coming back during its extended mission refreshing their ephemerides.

As time passes, the value of *TESS* data increases. Adding novel observations to objects measured in the past reduces the impact of systematics, and increases the chances to detect smaller and longer period planets. The importance of *TESS* goes well beyond its primary mission goals. The data base of high-precision light curves that is created from *TESS* FFIs and from *TESS* imagettes is an important legacy also for the preparation of future planet hunting missions like *PLATO*. Nevertheless, a wealth of other different studies in nearly any field of astrophysics will benefit of it. What it is sure is that as *TESS* will continue now into its extended mission, *TESS* data miners will continue to follow its swings around the sky to unveil the hidden gems it will observe.

With this work, we release the catalogue of the 396 candidates we found together with their light curves and diagnostic plots. This material is submitted to the Mikulsky Archive for Space Telescopes (MAST) under the section High Level Science Products (HLSP) and the project's name DIAmante. The catalogue description is reported in Table 7. Newly discovered candidates are also reported in the ExoFOP portal as Community TOIs (CTOIs).

ACKNOWLEDGEMENTS

This paper includes data collected by the *TESS* mission. Funding for the *TESS* mission is provided by the NASA Explorer Program. This research has made use of the Exoplanet Follow-up Observation Program website, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program, and of the Mikulski Archive for Space Telescopes (MAST) which is a NASA funded project to support and provide to the astronomical community a variety of astronomical data archives, with the primary focus on scientifically related data sets in the optical, ultraviolet, and near-infrared parts of the spectrum. This research also made use of the International Variable Star Index (VSX) data base, operated at AAVSO, Cambridge, Massachusetts, USA. The authors are grateful to the anonymous referee for all the comments and suggestions that permitted to improve the manuscript and increased the visibility of our work.

DATA AVAILABILITY

The data underlying this article are available in MAST, at doi:10.17909/t9-p7k6-4b32 and https://archive.stsci.edu/hlsp/diamante and at the ExoFOP portal at https://exofop.ipac.caltech.edu/tess/.

REFERENCES

Alard C., 2000, A&AS, 144, 363
Alard C., Lupton R. H., 1998, ApJ, 503, 325
Baglin A., 2003, Adv. Space Res., 31, 345
Bailer-Jones C. A. L., Rybizki J., Fouesneau M., Mantelet G., Andrae R., 2018, AJ, 156, 58
Barclay T., Pepper J., Quintana E. V., 2018, ApJS, 239, 2
Basri G., Borucki W. J., Koch D., 2005, New Astron. Rev., 49, 478
Bouma L. G., Winn J. N., Kosiarek J., McCullough P. R., 2017, preprint (arXiv:1705.08891)
Bouma L. G. et al., 2019a, AJ, 157, 217
Bouma L. G., Hartman J. D., Bhatti W., Winn J. N., Bakos G. Á., 2019b, ApJS, 245, 13
Bramich D. M., 2008, MNRAS, 386, L77
Breiman L., 2001, Mach. Learn., 45, 5
Broeg C. et al., 2013, EPJ Web Conf., 47, 03005
Caceres G. A., Feigelson E. D., Jogesh Babu G., Bahamonde N., Christen A., Bertin K., Meza C., Curé M., 2019a, AJ, 158, 57
Caceres G. A., Feigelson E. D., Jogesh Babu G., Bahamonde N., Christen A., Bertin K., Meza C., Curé M., 2019b, AJ, 158, 58
Collins K. A. et al., 2018, AJ, 156, 234
Espinoza N., Jordán A., 2015, MNRAS, 450, 1879
Evans D. W. et al., 2018, A&A, 616, A4
Feinstein A. D. et al., 2019, PASP, 131, 094502
Gandolfi D. et al., 2018, A&A, 619, L10
Gardner J. P. et al., 2006, Space Sci. Rev., 123, 485
Günther M. N. et al., 2019, Nat. Astron., 3, 1157
Han E., Wang S. X., Wright J. T., Feng Y. K., Zhao M., Fakhouri O., Brown J. I., Hancock C., 2014, PASP, 126, 827
Handberg R., Lund M. N., 2019, T'DA Data Release Notes – Data Release 4 for TESS Sectors 1 + 2, Zenodo, Genève, Switzerland
Howell S. B. et al., 2014, PASP, 126, 398
Huang C. X. et al., 2018, ApJ, 868, L39
Jenkins J. M. et al., 2016, in Chiozzi G., Guzman J. C., eds, Proc. SPIE Conf. Ser. Vol. 9913, Software and Cyberinfrastructure for Astronomy IV. SPIE, Bellingham, p. 99133E
Katz D. et al., 2019, A&A, 622, A205
King G. W., Wheatley P. J., Bourrier V., Ehrenreich D., 2019, MNRAS, 484, L49
Kovács G., Zucker S., Mazeh T., 2002, A&A, 391, 369
Lallement R. et al., 2018, A&A, 616, A132
Mahalanobis P. C., 1936, Proc. Natl. Inst. Sci. India, 2, 49
Mandel K., Agol E., 2002, ApJ, 580, L171

Downloaded from https://academic.oup.com/mnras/article/498/2/1726/5893317 by CIS Maldura user on 24 January 2025

Miller J. P., Pennypacker C. R., White G. L., 2008, PASP, 120, 449
Mink D. J., 1997, in Hunt G., Payne H., eds, ASP Conf. Ser. Vol. 125, Astronomical Data Analysis Software and Systems VI. Astron. Soc. Pac., San Francisco, p. 249
Nardiello D. et al., 2019, MNRAS, 490, 3806
Nardiello D. et al., 2020, MNRAS, 495, 4924
Oelkers R. J., Stassun K. G., 2018, AJ, 156, 132
Ofir A., 2014, A&A, 561, A138
Rauer H. et al., 2014, Exp. Astron., 38, 249
Ricker G. R. et al., 2014, in Oschmann Jr. J. M., Clampin M., Fazio G. G., MacEwen H. A., eds, Proc. SPIE Conf. Ser. Vol. 9143, Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave. SPIE, Bellingham, p. 914320
Schanche N. et al., 2019, MNRAS, 488, 4905
Seager S., Mallén-Ornelas G., 2003, ApJ, 585, 1038
Shporer A. et al., 2019, AJ, 157, 178
Soubiran C. et al., 2018, A&A, 616, A7
Southworth J., 2011, MNRAS, 417, 2166
Stassun K. G. et al., 2019, AJ, 158, 138
Tenenbaum P., Jenkins J. M., 2018, TESS Science Data Products Description Document. Available at: https://archive.stsci.edu/missions/tess/doc/EXP-TESS-ARC-ICD-TM-0014.pdf
Tinetti G. et al., 2018, Exp. Astron., 46, 135
Trifonov T., Rybizki J., Kürster M., 2019, A&A, 622, L7
Vanderspek R. et al., 2019, ApJ, 871, L24
Wang S. et al., 2019, AJ, 157, 51

This paper has been typeset from a TEX/LATEX file prepared by the author.

Downloaded from https://academic.oup.com/mnras/article/498/2/1726/5893317 by CIS Maldura user on 24 January 2025