Vetting and characterization of 24 new *K2* planetary candidates

Author: Marc del Alcázar i Julià
*Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain**

Advisor: Jorge Núñez de Murga

Abstract: For the analysis done in this work, we have used a sample of 220 sub-Neptune candidates from the `TFAW`-survey. Those candidates are the result of the application of `EVEREST 2.0` pipeline, the detrending and denoising `TFAW` algorithm, and `TLS` to light curves from the *K2* mission. They also passed through a set of vetting processes including visual inspection of the light curves, revision of high-resolution images and FPP< 2% from `VESPA` algorithm. From this sample, we choose the 38 candidates with more chances of being detected by the *TESS* mission or ground-based telescopes attending to their *TESS* magnitudes and transit depths. We apply to them two extra vetting procedures to better pick out between false positives and candidates: analysis of their host stars *Gaia* eDR3 astrometric parameters and a PSF centroid test. We finally obtain a set of 34 candidates, 24 of which have not been previously published. We present nine new candidates in the Radius gap and eight new multiplanetary systems, one of them with four candidates.

I. INTRODUCTION

The NASA *Kepler* mission [1] was launched in March 2009 with the aim of detecting new Earth-sized planets in the Habitable Zone (HZ) of Sun-like stars using the transit method, focusing its efforts in a single field located in the constellations of Cygnus and Lyra. In May 2013, the spacecraft had lost two of its four reaction wheels, making it impossible to continue with the original mission. In order to keep using the potential of the instruments installed on the spacecraft–a 0.95 m Schmidt telescope with a resolution of 4" per pixel and a FoV of 110 square degrees–, an alternative mission called *Kepler* 2 (shortened to *K2*) was proposed [2].
Taking advantage of the orbit of the spacecraft around the Sun, the *K2* mission was designed to make continuous observations of a group of fields along the ecliptic. The combination of the orbital velocity of the spacecraft, its mechanical adjustments, and the radiation pressure of the Sun, let the telescope observe each field for 75 uninterrupted days with a 30-minute cadence. With a typical photometric precision of 400 ppm at V=12 or 80 ppm integrating 6 hours of cadence, the *K2* mission became a powerful tool to detect and characterize exoplanets, among many other scientific goals.
The results of the *Kepler* and *K2* missions are outstanding, with more than 3200 confirmed planets discovered of a total amount of 5000 approximately (taking into account all missions and methods) [1] and with an enormous potential to exploit yet. Moreover, the ongoing Transiting Exoplanet Survey Satellite (*TESS*) [3] opens up a whole new world of possibilities to detect new planetary candidates and confirm or validate candidates from *Kepler* and *K2* missions.

This work is presented in the frame of the `TFAW`-survey, the application to the *K2* light curves of a new wavelet-based detrending and denoising method that improves the detection of small planets in weak stars and enhance their characterization [4]. Its optimization for `EVEREST 2.0` light curves (Section II) and the subsequent vetting procedure (Section III) let us to present a new set of exoplanet candidates with good chances of being detected by the *TESS* mission (Section IV), their characterization (Section V) and future work (Section VI).

II. DATA & OBSERVATIONS

The light curves of the targets as well as the Long and Short Cadence Target Pixel Files (TPL and TPS, respectively) from the 19 (+1 technical) *K2* campaigns, with more than 10000 targets per campaign, are available for the scientific community. In this work, we analyzed data from campaigns C1-C8 and C12-C18.
In order to increase the detection probability, it is necessary to filter and process the raw light curves. Using `EVEREST 2.0` [5], one of the most advanced pipelines for *K2* light curves which includes pixel decorrelation and detrending algorithms, it is possible to recover the original *Kepler* photometric precision for stars with a *Kepler* magnitude below 15^{th} mag. The application of this pipeline allowed us to rediscover an important fraction of the candidate and validated planets found in [6], [7] and [8].
Another high-frequency pass filter commonly used is a 2-day quadratic Savitsky-Golay filter–also called median filter–, which is applied with the purpose of smoothing the light curves. This filter consists of the subtraction of the median computed every 2-day range of points from the original light curve with the aim of removing features with typical lengths larger than two days coming from stellar activity.

*Electronic address: marcdelalcazar@gmail.com
[1] See more details of the statistics in `https://exoplanetarchive.ipac.caltech.edu/docs/counts_detail.html`.

III. METHODS

A. TFAW

The `TFAW` algorithm [9] is a combination of the original `Trend Filtering Algorithm`–labeled `TFA`–[10] and the stationary wavelet transform (SWT) filter [11].
The `TFA`, here customized for wide-field variability surveys, is based on the idea that similar systematic effects are shared by stars of a given field observed by the same instrument. These effects will not be intrinsic to the stars but they will mostly be caused by instrumental or environmental effects. The possibility of detecting the systematic features lets us generate a filter to be applied to other stars not part of the sample, correcting these undesired effects. The `TFA` conforms the first step of the `TFAW`, removing trends and systematics.
The second step is based on the SWT, which consists of the decomposition of a given series into its wavelets, that is, highly localized impulses obtained from scaling and shifting the mother wavelet function[2] [9]. This process lets us obtain the wavelet coefficients that characterize the correspondence between the mother wavelet and the time series. Attending to these coefficients it is possible to separate the original time series into its frequency levels without losing information. Taking into account that noise is mostly localized in high frequency levels, we can subtract them (denoising) without almost affecting the astrophysical signal of the star. It is also possible to clean the light curve from outliers.
In the third step of `TFAW`, the `Transit Least Squares` (`TLS`) algorithm [12] implements a frequency analysis which makes use of stellar information from [13] or [14] (when information from [13] is not available) to fit transit-like signals in order to search for significant periods. It is important to say that `TLS` is optimized to detect small Earth-sized planets. We considered a period to be significant if the Signal Detection Efficiency (SDE) given by `TLS` is greater than 9.0, which implies a false-positive rate $< 10^{-4}$ [13]. If a significant periodicity is found, the light curve is phase folded and the SWT is iteratively applied again to remove trends and noise and reconstruct the original signal.
As shown in [4], `TFAW` achieves a much better photometric precision and cleaner light curves than any previous processing of the light curves (see Figs. 7, 9, 10 and 11 in [4]), which increases the probability of detecting small planets in fainter stars and improves their characterization.

[2] The difference between the SWT and the Fourier transform is that the first one is irregular in shape and compactly supported, which let us to conserve the temporal information as well as signals with a non-stationary nature or discontinuities.

B. Vetting procedure

Once we processed the light curves and selected only those with SDE> 9.0, we visually inspected them keeping the ones that showed transit-like features and dismissing light curves probably coming from other sources like eclipsing binaries (EB) or strange shapes. We also checked the odd/even transit depths for all transit candidates in order to discard EB.
After that, we compared our database with cataloged planets in order to discard already found candidates, looking at the period (P) and the T_0 (date and time in Julian days of the first period detected) to make sure not to eliminate possible multiple systems. We also inspected the cutouts around the stars with candidates to reject stellar systems with contamination from other sources. The same inspection was done with almost *K2*-contemporaneous high-resolution images from *Pan-STARRS1* [15] and *Gaia* DR2 [16].
Furthermore, with the purpose of deleting any residual period not coming from the stellar system's behavior but from the instrument or the background, we searched for transits with similar P and T_0 in other light curves of the same CCD module. In addition, we compared the `TLS` periodograms (histogram with the most significant periods and their SDE) before and after the application of `TFAW` to check if the latter introduced some artificial signal.
Finally, we applied the Validation of Exoplanet Signals using a Probabilistic Algorithm (`VESPA`) [17] to our database, a false positive probability (FPP) calculator that makes use of stellar models to discard detections coming from six different astrophysical processes such as non-associated blended eclipsing binaries or EB. It is also considered a validation method for exoplanets candidates when FPP< 1% is reached. We only accepted candidates with FPP< 2%, which gave us a first sample of potential exoplanets waiting for new vetting procedures, crossmatch with the light curves from other missions like *TESS*, ground-based follow-up with speckle interferometry or adaptive optics.

C. Extra vetting

1. *Gaia eDR3 parameters*

One of the specific tasks that have been done for this work is the application of two extra vetting tests to a subsample of our exoplanet candidates (see Section IV). We inspected *Gaia* eDR3 [18] to search for neighboring stars close to our targets–as we did with *Gaia* DR2–taking into account that the spatial resolution and the relation δmag vs angular separation in *Gaia* eDR3 is better than the one in *Gaia* DR2. Besides, we analyzed three more parameters given by *Gaia* eDR3 as indicators of non-resolved stars. These parameters are the Astrometric Goodness of Fit in the Along-Scan direction (`GOF_AL`),

the Significance of the Astrometric Excess Noise (D), and the Re-normalised Unit Weight Error (RUWE).
The first one, GOF_AL, is the χ^2 applied to a Gaussian model with mean zero and standard deviation one. Theoretically, values over 3 are considered an indicator of a bad correspondence between the single-star model and data [19]. For its part, D is an additional uncertainty related to the possible modeling errors. In [19], the thresholds of these two parameters are set at GOF_AL<20 and D<5 (taking into account the best match of the boundary between confirmed binaries and confirmed singles), considering non-resolved stars with greater parameters.
The Unit Weight Error (UWE) is an indicator similar to the Astrometric Excess Noise but referred not to the noise but to the quality of the five-parameters solution of the model. It is also useful when D=0, a very common situation that affects half of the *Gaia* eDR3 targets. The RUWE is a re-normalization of the UWE taking into account the magnitude and the color of the stars [20]. We took a threshold value of RUWE<1.4 as an indicator of a good solution of the single-star model. Targets with larger values of RUWE are considered as possible binary or contaminated systems. This value comes from our analysis of a supposed single TFAW-survey target with D=15.1, GOF_AL=7.19 and RUWE=1.41 which finally turned out to be contaminated after considering SOAR Speckle imaging data observations.

2. PSF centroid testing

The second extra vetting step we applied to our subsample was the point spread function (PSF) centroid test. It consists in measuring the changes in the position of the centroid of a target star during the candidate transit. Taking into account that K2 was a redesigned mission, it was possible that some kind of mismatches in the pointing induced noise with a duration similar to typical exoplanet transits.
In order to eliminate possible false positives coming from background eclipsing sources, we used a Python-based implementation of the centroid test called vetting [21] that considers the *K2* motion. Making use of the TPL and TPS, P, T_0, and the duration and depth of the transit, it returns two distributions of centroids (during the transit and out of it) and a p-value corresponding to the likelihood of both distributions. It also gives the maximum angular distance to which another star should be to be the cause of the signal with one sigma error. We consider false positives those candidates with p<0.05, the same threshold used in [22]. We also checked the *Gaia* eDR3 catalog to make sure no other stars were inside the radius of maximum distance, though high-angular resolution images are needed to fully discard the presence of very close stars.

IV. TARGET SELECTION

The application of the basic vetting procedure to the TFAW-corrected light curves in the frame of the TFAW-survey resulted in the transition from an initial sample of approximately 40000 possible candidates before the visual inspection to a preliminary set of 220 new planetary candidates presented in [23].
In this work, we generated and analyzed a sub-sample of those candidates that have more chances to be followed up with the *TESS* mission or ground-based telescopes. In fact, because of the difficulty of obtaining time from large-aperture ground-based telescopes (4m-class or more for the photometric precision we need), we focused our search on candidates with host stars fitting the *TESS* requirements: they are visible in one or more of its observing sectors and they have a combination between the *TESS* magnitude ($T_{\rm mag}$) and the transit depth (δ) that make them potentially detectable. We considered a $T_{\rm mag} \leftrightarrow \delta$ relation a little bit more flexible than the one presented in Fig. 2 in [24] to contemplate both possibilities of applying TFAW to *TESS* light curves (which probably would increase the photometric precision), and make use of large-aperture ground-based telescopes.
We applied the following criteria to select the sub-sample:

1. Overlapping with *TESS* sectors 1-55 (until September 2022).
2. $\delta > 0.1$ mmag for stars with $T_{\rm mag} \leq 11$.
3. $\delta > 0.2$ mmag for stars with $11 < T_{\rm mag} \leq 12$.
4. $\delta > 0.5$ mmag for stars with $12 < T_{\rm mag} \leq 13$.
5. $\delta > 0.7$ mmag for stars with $13 < T_{\rm mag} \leq 14$.
6. $\delta > 0.9$ mmag for stars with $14 < T_{\rm mag} \leq 15$.
7. Stars with a $T_{\rm mag}$ >15 have not been considered according to the *TESS* web[3].

38 candidates in 26 planetary systems passed these criteria and form the sub-sample to which we applied the extra vetting methods explained in Section III C.

V. RESULTS

A. Extra vetting results

After the start of the TFAW-survey, many other teams have been analyzing *K2* light curves in order to find new exoplanets. From our sub-sample of 38 candidates, 13 of them appear in the literature: two confirmed planets from [25] (K2-356b) and [26] (K2-149b); nine candidates from [27] (five), [28] (one), [29] (one) and [30] (two); and two candidates catalogued as eclipsing binaries in [31].
With regard to the extra vetting results, 25 of our candidates passed all the vetting steps, that is the reason

[3] https://tess.mit.edu/

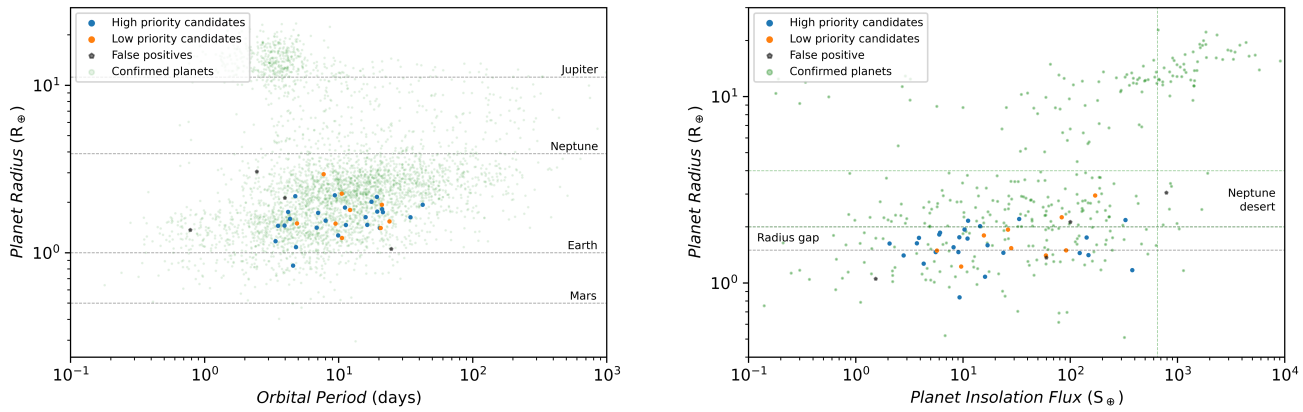

FIG. 1: **Left:** Planet radius as a function of the orbital period, and **Right:** Planet radius as a function of the planet insolation flux for `TFAW`-survey sub-sample high priority candidates (blue points), low priority candidates (orange points), false positives (black penthagons) and confirmed planets from *NASA* Exoplanet Archive (green points).

why we label them high-priority candidates. Nine other candidates are labeled as low priority candidates for different reasons: two of them do not pass the `D` threshold, four of them have close stars, and three of them failed the centroid test with periods $T_1 = 10.59$ d, $T_2 = 10.54$ d and $T_3 = 21.02$ d, which can be harmonics of a systematic error around 10.5-10.6 d that is eclipsing real transits. In addition, one of them is a confirmed planet in [25] (K2-356b) and the other one is a candidate in [27]. These nine low-priority candidates could benefit from follow-up observations to clarify their nature.

On the other hand, we consider four candidates as false positives. One of them does not pass the centroid test with a period of $T = 3.98$ d (not related to the possible systematic error previously mentioned) and has close stars. It is important to note we classify this signal as a false positive while it appears as a strong candidate in [30]. Two other candidates do not pass the *Gaia* eDR3 vetting steps, `GOF_AL`, `D`, and `RUWE`, which indicates a high probability for these candidates to be EB. Also, one of them has already been classified as an EB in [31]. Finally, a candidate has been refused after visually inspecting its light curve and classifying its host as a variable star.

B. Sub-sample characterization

Adopting the parameters obtained from `TLS` and using a planet radius distribution similar to the one from [32], our sub-sample of final candidates is comprised of one sub-Earth planet ($R_p < 0.8R_\oplus$), three Earth-sized planets ($0.8R_\oplus < R_p < 1.25R_\oplus$), 25 super-Earth planets ($1.25R_\oplus < R_p < 2R_\oplus$) and five sub-Neptune planets ($R_p < 4R_\oplus$). Besides, five of our super-Earth planet candidates are hot desert worlds, planets with an insolation between $1.5S_\oplus < S < 5S_\oplus$ [33].

We also find one Ultra Short Period (USP) candidate (finally considered a false positive after reanalyzing its light curve), that is, a planet with an orbital period of less than one day. USPs, also called hot Earths or lava worlds as the temperature of their surface on the dayside is higher than the melting point of most rock-forming minerals [34], are important in planet formation and evolution theories. The proximity of these planets to their host stars presents challenges to current models [35].

It is also important to mention the results in [36], where a bimodal distribution of small-sized planets splits sub-Neptunes and super-Earths into two different classes. This bimodality is separated by a region between $1.5R_\oplus < R_p < 2R_\oplus$ called the Radius gap, where we can find a deficit of factor ≥ 2 in the planet occurrence rate distribution. Planets in the Radius gap can be useful to improve our knowledge of planet formation, atmosphere loss, and evolution theories. As can be seen in Fig. 1, 14 of our candidates lie in this region.

The photoevaporation desert or Neptunian desert is a lack of planets between $2R_\oplus - 4R_\oplus$ at very high insolations ($S \geq 650S_\oplus$). This absence of planets is explained by the evaporation of volatile elements, which tears planet envelopes, diminishing their radius [37]. We find one candidate in the Neptunian desert which turns out to be a false positive taking into account the *Gaia* eDR3 astrometric parameters.

Finally, we present eight new multiplanetary systems, mostly with sub-Neptune planets: four systems with two planets (we detect a new candidate in a system where [27] already found one candidate not detected by us; in the other three systems we discover or rediscover all the candidates or confirmed in the system until now), three systems with three planets and one system with four planets. These numbers represent an important input to the *K2* list of confirmed and candidate planets, which has 114 systems with two planets, 28 systems with three planets, and only 11 systems with four planets up to date. It is important to study multiplanetary systems in order to characterize the interaction between planets (orbital resonances), know more about formation and evolution

theories (especially attending to the difference between rock and gas giant planets), and detect non-transiting planets.

VI. CONCLUSIONS & FUTURE WORK

The results presented in this work, including 24 new planet candidates, show that the data from the *K2* mission still hides an important amount of planetary candidates, specially sub-Neptunes. To pull ahead with new discoveries will be necessary to improve filtering and processing algorithms and vetting procedures. To do the former, we made use of a sub-sample of 38 candidates from the `TFAW`-survey set of 220 candidates elected attending to the possibility of making follow up of them with *TESS* mission or ground-based telescopes. Concerning the latter, apart from the basic vetting procedures implemented to the whole `TFAW`-survey candidates, we applied extra vetting methods to the sub-sample, including high-resolution imaging with *Gaia* eDR3, the PSF centroid test, and the analysis of the *Gaia* eDR3 astrometric parameters. Finally, we characterized them using the `TLS` values for the planet features, resulting in 14 new candidates in the Radius gap and eight new multiplanetary systems, one of them with four planets. This last outcome can be important in the study of the orbital resonances and the differentiation between rocky and gas giant planet formation processes, among others.

Even though, it will be necessary to implement additional vetting procedures to the new candidates in order to validate and confirm them. Some future work remains undone for the currently sub-sample: follow-up with *TESS*, *CHEOPS* or ground-based telescopes that make use of techniques such as speckle, multi-filter or adaptive optics, analysis of the forthcoming *Gaia* DR3, or the application of other extra vetting methods like Linearized Field Deblending Photometry [38] (which implements a deblending technique in order to obtain better light curves in crowded fields) or Pixel Response Function Photometry [39] (an approximation to multi-band photometry from monochromatic data). Finally, we will have to run Markov Chain Monte Carlo to better determine planet features, their uncertainties and validate them. Furthermore, all the mentioned work can also be done for the whole `TFAW`-survey sample.

Acknowledgments

I acknowledge Dr. Jorge Núñez for the opportunity of working with him and his scientific team for two years in the frame of the UB-RACAB collaboration grant 2020.2.RACAB.1 and this *Treball de Fi de Grau*. Also, I acknowledge Dr. Daniel del Ser and Dr. Octavi Fors their support during the development of this work, including corrections, recommendations, and the possibility of using the `TFAW`-survey exoplanet candidates data set.

[1] W. J. Borucki et al., Science **327**, 977 (2010).
[2] S. B. Howell et al., PASP **126**, 398 (2014).
[3] G. R. Ricker et al., JATIS **1**, 014003 (2014).
[4] D. del Ser and O. Fors, MNRAS **498**, 2778 (2020).
[5] R. Luger et al., AJ **156**, 99 (2018).
[6] I. J. M. Crossfield et al., ApJS **226**, 7 (2016).
[7] A. Vanderburg et al., ApJS **222**, 14 (2016).
[8] A. W. Mayo et al., AJ **155**, 136 (2018).
[9] D. del Ser, O. Fors, and J. Núñez, A&A **619**, A86 (2018).
[10] G. Kovács, G. Bakos, and R. W. Noyes, MNRAS **356**, 557 (2005).
[11] M. Holschneider et al., in *Wavelets*, edited by J.-M. Combes et al. (Springer Berlin Heidelberg, Berlin, Heidelberg, 1990), pp. 286–297.
[12] M. Hippke and R. Heller, A&A **623**, A39 (2019).
[13] K. K. Hardegree-Ullman et al., ApJS **247**, 28 (2020).
[14] D. Huber, S. T. Bryson, and et al., VizieR Online Data Catalog IV/34 (2017).
[15] N. Kaiser et al., in *Ground-based and Airborne Telescopes III*, edited by L. M. Stepp, R. Gilmozzi, and H. J. Hall (2010), vol. 7733 of *SPIE Conference Series*, p. 77330E.
[16] A. G. A. Brown et al., A&A **616**, A1 (2018).
[17] T. D. Morton, ApJ **761**, 6 (2012).
[18] A. G. A. Brown et al., A&A **650**, C3 (2021).
[19] D. F. Evans, RNAAS **2**, 20 (2018).
[20] L. Lindegren, in *Re-normalising the astrometric chi-square in Gaia DR2* (2018).
[21] C. Hedges, RNAAS **5**, 262 (2021).
[22] D.-C. Chen et al., AJ **163**, 249 (2022).
[23] M. del Alcázar, D. del Ser, and O. Fors, in *CS20.5* (2021), Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun, p. 316.
[24] M. Kunimoto et al., arXiv e-prints (2022).
[25] A. Castro-González et al., MNRAS **509**, 1075 (2022).
[26] T. Hirano et al., AJ **155**, 127 (2018).
[27] J. K. Zink et al., AJ **162**, 259 (2021).
[28] G. Kovacs, A&A **643**, A169 (2020).
[29] A. Dattilo et al., AJ **157**, 169 (2019).
[30] H. N. H. Weishaupt, *Implementing a pipeline to search for transiting exoplanets: application to the K2 survey data* (Linnaeus University, 2018).
[31] D. Canal, *Estrellas variables en la campaña 18 del telescopio espacial Kepler en su misión extendida K2* (Universidad de Oviedo, 2021).
[32] W. J. Borucki et al., ApJ **736**, 19 (2011).
[33] A. Zsom et al., AJ **778**, 109 (2013).
[34] J. N. Winn, R. Sanchis-Ojeda, and S. Rappaport, New Astronomy Reviews **83**, 37 (2018), ISSN 1387-6473.
[35] E. R. Adams, B. Jackson, and M. Endl, AJ **152**, 47 (2016).
[36] B. J. Fulton et al., AJ **154**, 109 (2017).
[37] M. S. Lundkvist et al., Nature Communications **7**, 11201 (2016).
[38] C. Hedges et al., AJ **162**, 107 (2021).
[39] C. Hedges et al., AJ **161**, 95 (2021).