UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Master Degree in Astrophysics and Cosmology

Final Dissertation

Assessing the performance of filtering algorithms for Earth-analogues observed with PLATO

Thesis supervisor

Prof. Luca Malavolta

Candidate

Konrad Maliszewski

Academic Year 2022/23

Abstract

Context: Detecting Earth-like planets presents a formidable challenge, primarily due to the influence of stellar activity, granulation and systematic errors on transit signals observed through photometric techniques. Stellar phenomena, including flares, spots, and convection, as well as the instrumental errors introduce complexities that obscure the identification of genuine planetary transits.

Goal: This thesis aims to assess and compare the effectiveness of various filtering algorithms, with a specific focus on detecting Earth-like planets within the Habitable Zone, with orbital periods extending up to 2 years. The studied algorithms include the one developed by the research groups specifically to detrend the light curves from the stellar activity: Young Stars Detrending (YSD), as well as the all-purpose algorithms: biweight and Huber spline methods employed with 3 different window lengths (0.7, 1.4 and 2.0 days). The results hold particular relevance for the upcoming PLATO mission, as synthetic light curves were generated using the PLATO Solar-like Light-curve Simulator (PSLS). A series of injection-retrieval tests were conducted on these synthetic light curves to evaluate the performance of the selected filtering algorithms.

Results: The biweight method and YSD Lowess regression emerge as the most effective algorithms for conducting a blind search for Earth-analogous planets. However, the precise retrieval of planetary parameters from the recovered transit signals remains challenging, as filtering algorithms distort the original signal. For the P1 sample representing the target stars during the initial two years of the mission, these algorithms fail to recover the planetary signal when applied to F5 spectral type stars. This is primarily due to the larger radii of such stars, which complicates detection by extending the period duration and reducing the planet-to-star radius ratio.

Contents

1 Introduction

1.1 Exoplanet study

The exoplanetary science is a relatively young astrophysical field with the first discovery of a planet beyond our solar system being done by Wolszczan (1992), where the author provided the evidence for the companion presence by timing the arrival of radio signals from the millisecond pulsar PSR1257+12. This technique of the exoplanet detection is only possible around radio pulsars which by their violent nature can constitute a planetary systems which are unlikely to have a Habitable Zone (HZ) making the life formation (or its existence) around them impossible. The Nobel prize "for the first discovery of a planet orbiting a solar-type star outside our solar system" was awarded to Michael Mayor and Didier Queloz who in their work (Mayor and Queloz, 1995) inferred the presence of a Jupiter-mass companion orbiting around the 51 Pegasi star by measuring the periodic variations in the star's radial velocity (RV). This method along with the transit method (Charbonneau et al., 2000), which uses the variations of the flux coming from the star in the light curve (LC), are by far the largest contributors to the exoplanet detections. Other methods that are presently used include (but are not limited to): direct imaging (Marois et al., 2008), astrometry (Benedict et al., 1999) and microlensing (Gould and Loeb, 1992).

Since the first discovery using the transit method more than 5000 exoplanets have been confirmed up to the present day[1]. The exoplanet surveys were conducted using both ground-based searches, including: HATNet (Bakos et al., 2004), KELT (Pepper et al., 2007) and CHESPA (Zhang et al., 2019), as well as the space campaigns: *Kepler* (Borucki et al., 2010), K2 (Howell et al., 2014) and TESS (Ricker et al., 2014). Emphasising the importance of the *Kepler* mission, it discovered thousands of exoplanet candidates and led to the confirmation of the existence of over 2,600 exoplanets including numerous exoplanets located within the HZ, where the liquid water could potentially exist, increasing the possibility of finding habitable environments outside the Solar System. None of these planets are located in the HZ of Solar-type stars, however, could be definitively confirmed to be rocky through mass measurements, as the faintness of the host stars made the spectroscopic follow-up unfeasible. The mission, although extended beyond its primary scope, ended in 2018 due to the lack of fuel to power up its instruments. The majority of the exoplanets that it discovered are categorized as super-Earths and mini- or sub-Neptunes, which means that they possess radii and/or masses that fall between those of Earth and Neptune. Interestingly, no such equivalents for these types of planets exist in our Solar System. In terms of host stars, almost all known exoplanets orbit main-sequence stars that fall under the spectral categories of F, G, K, or M. The current detection techniques pose inherent biases that limit the types of exoplanets that we can study. The masses of detected objects must be quite high because they exert larger effect on the RVs or LC variations making the detection possible, based on the current precision of the instruments. Radial velocity method favours the massive planets, while the detection of the transiting planet is easier for ones that are orbiting closer to the host star, as the transits will happen more frequently. Most notably, Earth-sized planets are difficult to detect due to their small radius in comparison to their host stars resulting in relatively small dimming of the flux coming from the star. Earth-like planets are especially challenging to identify, as they require extended periods of telescope pointing and extremely high-precision measurements Petigura et al. (2013).

1.2 The PLATO mission

The search for Earth-analogues, planets that share similar characteristics with our own planet, has captivated the scientific community and the public alike. The discovery of exoplanets has opened up new avenues for understanding the potential prevalence of habitable environments and life beyond Earth. Among the missions dedicated to this pursuit, the PLATO (Planetary Transits and Oscillations of Stars) mission stands at the forefront as the natural successor to the *Kepler* mission, promising to revolutionize our understanding of exoplanetary systems (Rauer et al., 2014).

The PLATO mission, which is led by the European Space Agency (ESA) is scheduled for launch in 2026, aims to detect and then characterize exoplanets by observing the transit of a planet in front of the host star. By studying these transit signals, important planetary properties such as size, orbital period can be inferred. Differently form *Kepler*, PLATO will target bright stars amenable for the spectroscopic follow-up from the ground that measure the RV variations providing

---

[1]http://www.exoplanet.eu/

the mass estimates of the observed planets and their inclination angle such as ESPRESSO (Yant et al., 2018) or ANDES@E-ELT (Marconi et al., 2021) that can be later used to derive the bulk densities. Asteroseismic analyses of the photometric LCs can be also used to establish the age of the planetary system providing more information towards our understanding of how such systems are formed.

The instruments on board will include 24 "standard" cameras (N-CAM) grouped into four sets of 6, each with a 25-second readout cadence and 21-second integration time. Additionally, there will be 2 "fast" cameras (F-CAM) with a 2.5-second readout cadence. This multi-telescope mission will cover a total field of view (FoV) of about 2132 deg^2 (Nascimbeni et al., 2022). Each camera is a refracting telescope with an aperture diameter of around 120 mm. The total setup comprises of 104 CCDs with a pixel size of 18 m. Long pointings made with N-CAMs will search for Earth-like planets in the HZ of solar-like stars and will last about 2 yrs divided into quarters (every 3 months the line of sight pointing will rotate by 90°to adapt to the changing Earth position with respect to the Sun – the resulting pixel position of observed object might be different in subsequent quarters). Long pointings will be complemented by the short ones (made with F-CAMs) that are devoted to the search for planets around very bright stars, resulting in coverage of up to 40% of the sky. Such a large coverage will allow to observe over 1 mln stars in the missions planet searching range of $4 \leq m_v \leq 11$. Within this magnitude scope, asteroseismology becomes feasible, enabling the precise determination of planetary parameters. This includes obtaining accurate planet masses through follow-up spectroscopy using RV measurements as mentioned before.

1.3   Filtering algorithms

The detection and characterization of Earth-like planets in the HZ of a Solar-type star, or simply Earth-analogues, present unique challenges due to their small size and the presence of intrinsic noise in the data. Stellar noise is unique to each star and exhibits different features across a variety of timescales. The process of identifying potential transiting exoplanet candidates from photometric time series typically involves a two-step approach. Initially, long-term flux variations originating from either the star or the instrument are removed (also called the detrending of the LC). Following this, periodic transit-like signals are identified by folding the LC at various orbital periods, transit times, and duration. These signals are referred to as Threshold Crossing Events (TCEs), which can potentially be exoplanet candidates, but they can also include false positives. It is important to consider that several phenomena can mimic a transit signal, including eclipsing binaries, instrumental artifacts, noise, and stellar variability. The stellar activity cycles for example, like the 11 yrs one present in our Sun (Schwabe, 1844), are present in other stars and happen on different timescales (Rempel, 2008). The stellar spots and granulation also constitute an important factor that needs to be detrended from the original LC and its timescale depends on many factors like stellar activity, evolutionary stage and rotation period of the star. The significance of short-term variability has driven the adoption of a wide array of detrending methods.

Filtering algorithms employ a variety of techniques, such as signal processing and statistical analysis, to enhance the sensitivity and accuracy of planet detection. The performance of these filtering algorithms is paramount to the success of the PLATO mission and the advancement of our understanding of Earth-like planets. Therefore, the focus of this master's degree is to assess and evaluate the performance of these algorithms specifically designed for Earth-analogues observed with PLATO with duration period greater than 40 days up to 2 years. The upper value is the observational limit, and comes from the planned observation length. As this thesis is the continuation of the study carried out by Canocchi et al. (2023) in which the orbital period up to 40 days was studied, I will investigate the period range which was beyond the scope of that work.

There are many different filtering techniques which are summed up in Hippke et al. (2019) each having different advantages and throwbacks. These methods include sliding medians, sliding means, splines, polynomial filters, LOWESS regressions, Cosine Filtering with Autocorrelation Minimization (exo-moon optimized detrending method), Gaussian processes (that require the knowledge of the planetary period and rotational period of the star, and thus are not the best choice for the blind search), the Savitzky and Golay (1964a) filter, frequency filtering through Fourier decomposition and wavelets. Even neural networks (Morvan et al., 2020) or machine learning techniques (Kim et al., 2009; Kim, 2016) are being used but were not considered in this work because of their poor performance based on the recovery rate.

This research will involve a combination of theoretical analysis, computer simulations, and

statistical modeling. The evaluation metrics will be based on sensitivity, specificity, and reliability of the filtering algorithms, considering factors such as false positive and false negative rates and detection thresholds. The ultimate goal is to develop optimized strategy for the upcoming PLATO mission that can serve to identify and characterize Earth-analogues, enabling us to gain deeper insights into the prevalence and properties of habitable planets in our galaxy. The anaysis and results presented in this thesis will be implemented in the Exoplanetary Analysis System for the detection and characterization of planets in PLATO data through the connection with the PLATO Working Package (WP) 111 000 "Coordination of Tools for Lightcurve Filtering".

2 Simulations

In this section the process of generating the LCs that are subsequently the subject of analysis is presented. Because of the fact that the simulations were carried out using different codes, the in-depth explanation of how the light curves are obtained is needed. The section describes all the physical contributors to the LC variations and steps that needed to be taken while building the custom pipeline to handle the data generation.

2.1 Simulating the light curves

Simulations of the light curves were done using custom pipeline that used two codes: PSLS [2] (the PLATO Solar-like Light-curve Simulator Samadi et al., 2019) and the extracted python pipeline used to simulate the stellar activity that was provided to me by Suzanne Aigrain and is implemented in PlatoSim's [3] software toolkit called PLATOnium (Jannsen et al., submitted). The decision to run the simulations using PSLS and then inject the stellar activity was due to the heavy computational load of the PlatoSim, which would be an overkill for the purpose of this study. The raw light curves were the output of running both codes, which were then used to inject the planetary signals using the BATMAN [4] package (Kreidberg, 2015b). It is beneficial to keep the raw light curves and the planetary signal variations separate so that different planets could be injected into single light curve later on without the need to re-run the whole simulation again.
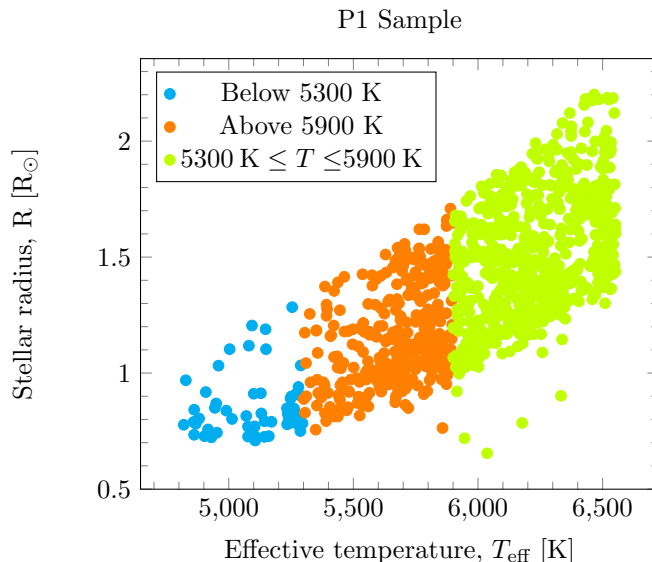

Figure 1: A stellar effective temperature vs. stellar radius plot for the P1 sample. It consists of the main sequence dwarf stars, the different color marks represent K5 (light blue), G5 (orange), and F5 (blue/light green) spectral type stars with spectral type boundaries from Pecaut and Mamajek (2013).

The sample contains 1000 stars which are the representants of the P1 sample (Montalto et al., 2021) from the Plato Input Catalogue (PIC). The P1 sample aims to include a minimum of 15,000 dwarf and subgiant stars ranging from spectral types F5 to K7, encompassing all sky fields. The stellar sample's dynamic range should have a V magnitude of no more than 11 with the emphasis within the sample to maximize the presence of brighter targets ($V \leq 10.5$). Additionally, the random noise within this sample must remain under 50 ppm in 1 hour. Observations of stellar sample 1 are scheduled to take place during a LOP (Long Observation Period). The target stars used in this work are taken from the PLATO Consortium technical note PLATO-PL-KUL-PL-TN-0020_i.1.2 named *Multi-camera and multi-quarter simulations I: Input for light curve stitching and detrending*. In Figure 1 their stellar parameters (radius vs stellar effective temperature) with

---

[2]`https://sites.lesia.obspm.fr/psls/`
[3]`https://ivs-kuleuven.github.io/PlatoSim3/`
[4]`https://lkreidberg.github.io/batman/docs/html/index.html`

the distinction based on the spectral type are presented. The PIC catalogue contains de-reddened V magnitude in the Johnson-Cousins passband, that was transformed from Gaia DR2 photometry. The reliance on color for flux estimations at the detector level was eliminated by converting the measurements to the synthetic photometric PLATO passband, denoted as $\mathcal{P}$. The transformation between the two is done with the help of the formula established by Marchiori et al. (2019):

$$V - \mathcal{P} = -1.184 \times 10^{-12} \, T_{\text{eff}}^3 + 4.526 \times 10^{-8} \, T_{\text{eff}}^2 - 5.805 \times 10^{-4} \, T_{\text{eff}} + 2.449$$

where T_{eff} is the stellar effective temperature.

2.1.1 Stellar oscillations and granulation

Stellar oscillations manifest as rhythmic variations in a star's brightness due to internal pressure waves propagating through its interior. Similarly, granulation, driven by convective currents near a star's surface, creates small-scale brightness fluctuations that reflect the convective cells' dynamic nature. They both contribute to the total flux variations, and they were addressed with the usage of the PSLS code.

PSLS models stellar oscillations starting from the oscillation spectrum, that is derived from the relevant oscillation profiles and must be supplemented with the mode frequencies, heights, and line-widths. Two separate cases are considered because for the sub-giant and main sequence stars, the simplified approach to treat the modes frequencies may be adopted, which requires less accurate models and thus improving the computation time. Theoretical mode frequencies are computed with the help of the ADIPLS adiabatic pulsation code developed by Christensen-Dalsgaard (2008). All sample stars used in this study are assumed to be the main sequence stars, so that the near surface effects may be ignored, as they will not influence the light curves in a substantial way. Stellar granulation is a stochastic process, that can be modelled in the Fourier space and then transferred back to the time domain (which essentially is a light-curve variation) through the inverse Fourier transform. For the detailed description of the treatment for both processes, one can refer to the Samadi et al. (2019).

2.1.2 Instrumental errors

The PSLS code has the utility to incorporate the simulated effects of instrumental errors. These errors come from a multitude of sources and can be divided into systematic errors of the instrument and the random noise. In order to simulate the effect of the instrumental errors the imagettes representing the PLATO CCDs are being generated using the PIS (Plato Image Simulator). Realistic simulations extending periods of 3 months, and targeting different magnitude stars were carried out by the authors of the simulator, in order to quantify the effect of the instrumental errors accounting for different observation conditions: the beginning of life (BOL) and the end of life (EOL). The difference between the two is that for the EOL, the charge transfer inefficiency (CTI) is being accounted for at the level expected 6 years after the telescope launch, whereas the EOL does not include it in the systematics at all. As this works main focus is the P1 sample and the observation period extending the first two years of the mission, the BOL conditions are set. For stars in the P5 sample, which LC is measured on board through aperture photometry, mask updates within a quarter may be needed to mitigate the effect of the long-term drift of the star, which causes the noise-to-signal ratio (NSR) to increase during the observation sequence, and they represent an additional source of systematics noise.

2.1.3 Stellar activity

To simulate the effect of the spots and plages on the photometric data the custom python code called PySpot, provided by Suzanne Aigrain and mostly based on the work by Meunier et al. (2019). As an input it takes the effective temperature of the star and its inclination angle. It then calculates the rotation period based on the $B - V$ value, that is a linear interpolation of the values contained in the previously mentioned paper, and depends on the effective temperature of the star. The spots and plages are first randomly generated, and then they are evolved with the period of the cycle based on the earlier estimated rotation period. The example of the flux variations of the star solely due to the stellar activity is presented in the Figure 2.
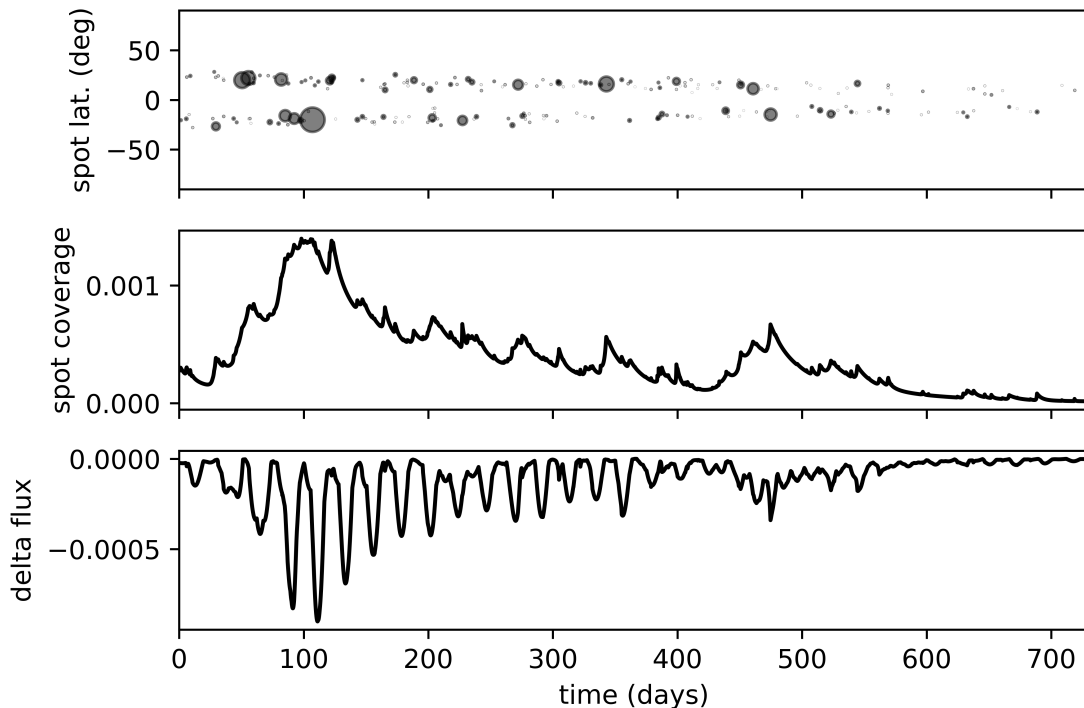
Figure 2: The example output of the PySpot code with the plots that represent: the change of spots latitude with time (spots are drawn to represent the real sizes), the spot coverage as the ratio between the area covered by the spots and the surface of the star, finally the relative flux variations due to the stellar activity.

2.1.4 Running the simulations

Due to the isolating of stellar activity from the rest of the flux variation contributors, as well as the incorporation of gaps between each quarter, a custom pipeline was build to run the simulations. This section provides the necessary steps that needed to be taken in order to produce full LCs for the P1 sample, also providing some of the assumed values for the configuration files. After obtaining the data of target stars of the P1 sample they were joined into a single table and the missing data was supplemented. Not all of the target stars contained the values of the rotation period or inclination angle, because not all of them were subject to the prior studies on their activity. These are steps that were taken to produce the light-curves:

1. Running PySpot: The input values for this pipeline were the star's effective temperature and its inclination angle. Before running the code, the missing inclination angle values were supplemented with the values drawn at random from the uniform distribution between 85 and 90°, as specified in the previously mentioned PLATO Technical Note. The simulation overall duration was set to 730 days and the cadence was 27×25 s. The choice of the cadence duration (time step of the simulation) was picked keeping a balance between the precision of the simulation and its computational cost. Having it represent the integer multiple of the time step in the PSLS simulator, aligned with PLATO's 25-second readout cadence, was equally advantageous.

2. Interpolation of PySpot data: The misalignment in the cadences required the stellar activity output data to be interpolated. The data points were linearly interpolated to produce a 25 s cadence set.

3. Running the PSLS: To mock the observations of the PLATO telescope, there is a need to introduce the quarters and gaps that are essential to redirect the telescope due to the orbital motion of the Earth around the Sun. The full observation period covers 2 years divided into 8 quarters each lasting 88.33 days, with the gaps in-between lasting for 2.9825 days. Four of the quarters were simulated separately using different random number seed because of the

6

fact that after the full year, the telescope will go back to its original position and the quarter cycle will repeat. Most of the psls configuration remained the same for all the input stars, differing only the part regarding the stellar parameters. The example configuration file is presented in Table 1.

4. Joining the simulations: The next step is to join the outputs of the two codes. The time step was already prepared to be the same for both simulations, the only thing that is needed to be done is to add the flux variations for each time stamp. The results of both simulations and the joined output were plotted for the example stars and presented in Figures 4, 5 and 6.

5. Re-doing the simulations for some LCs: As it is clearly seen on the example plot in Figure 6, the systematic errors for some quarters can become really high compared to the others. The root of the problem lays in the manner in which the quarters were simulated – the random number seed had to be changed to generate each quarter. This imposes that for each of the first 4 quarters (the following 4 are obtained by repeating them), the systematic errors can take independent values. In Figure 3 the systematic errors are presented, which were generated for 24 individual LCs by the authors of the psls (Samadi et al., 2019, Figure 9). In case of simulations that were carried out for the purpose of this thesis, the systematic errors in some cases exceeded 6%. This sudden jump in the flux variation would later lead to the misbehaviour of the transit search algorithms which detects it as strong signal. The replacement of such LCs was needed, which led to the decision of building a custom pipeline which would compensate for high systematic errors, by removing the quarters, for which the standard deviation was greater than 0.5% and running the psls for them again. Points 1-5 would be then repeated until we get all 1000 LCs with standard deviation for each quarter below this arbitrary threshold. It is safe to assume, that although the systematic errors cannot be known until the mission is launched, the data reducing pipelines will be correcting for these and other kind of errors, as in the case of missions like Kepler (Stumpe et al., 2012) or CHEOPS (Silva et al., 2020).
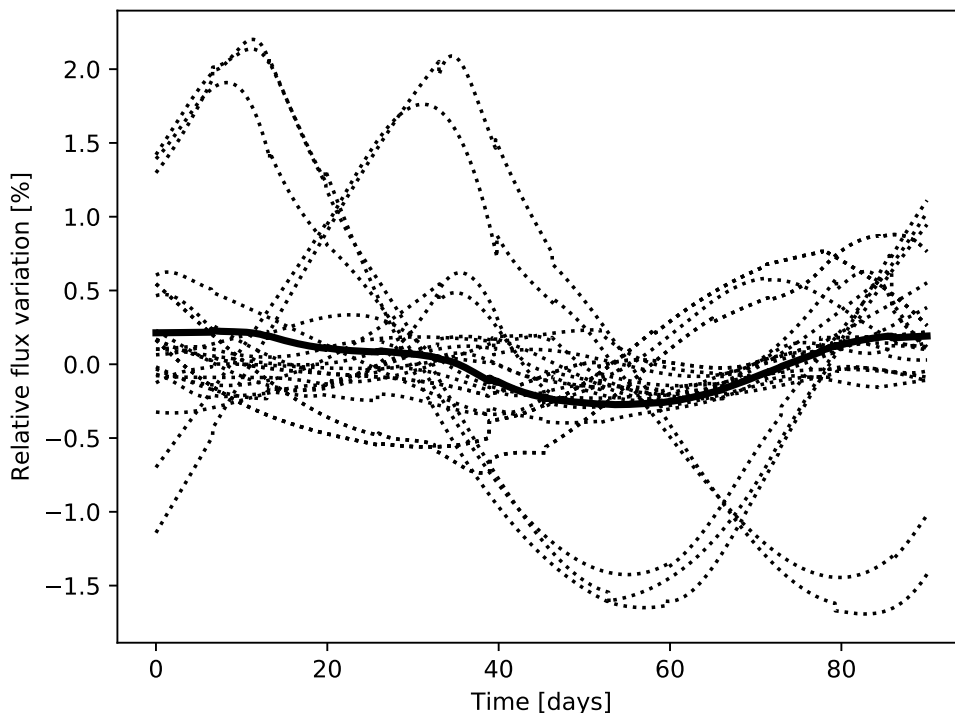

Figure 3: Plot taken from Figure 9 of the work by Samadi et al. (2019). It presents the systematic errors for 24 individual simulations that were carried out for a star of magnitude $V = 11$ and taking the EOL conditions. The dashed lines represent the LC of each simulation while the thick line is the averaged LC.

Table 1: Example configuration file of the PLATO Solar-like Light-curve Simulator (PSLS). Categories like activity or transit were not included in the table as they were disabled for all simulations. Parameters that were not used by the code because of the fixed model type (grid) are also removed. Values that were altered between simulations are marked with *. Each of the three levels ('low', 'medium', and 'high') of the drift level was assigned to one-third of the simulations, ensuring an equitable distribution of conditions across the study.

Observation conditions		
Duration [days]	88.33	Observation duration
Master seed	1704040900*	Master seed of PRNG
Instrument parameters		
Sampling [s]	25.0	Sampling cadence
Integration time [s]	21.0	
Number of camera groups	4	
Number of cameras per group	6	
Time shift [s]	6.25	Time shift between camera groups
Random noise		
Enable	1	Random noise enabled
Type	PLATO_SIMU	NSR taken from realistic simulated LCs
NSR [ppm in one hour]	73.0	Noise to Signal Ratio
Systematics		
Enable	1	Systematics enabled
Table	BOL_P1_V2.npy	BOL for the P1 sample
Version	2	
Drift level	medium*	low, medium or high
Stellar parameters		
Magnitude ($\mathcal{P}$ passband)	10.0*	
ID	12069449*	
Model directory	psls/models/m+0y27l	
Model type	grid	
Model name	0012069449*	
Evolutionary status	ms	main-sequence, red-giant or sub-giant
Effective temperature [K]	5750.0*	
$\log g$	4.353*	logarithm of gravity at the surface
Surface rotational period [days]	0.0*	
Inclination [deg.]	0.0*	
Oscillations parameters		
Enable	1	Oscillations enabled
Surface effects	0	Surface effects disabled
Granulation		
Enable	1	Granulation enabled

Figure 4: Produced LCs of the codes for stellar activity (PySpot) and of the PSLS simulator supplemented with the joined output for example star with ID: 10037521. On the *y*-axis is the normalized flux (1.0 value represents no change in the flux). The gaps between each quarter are also visible.
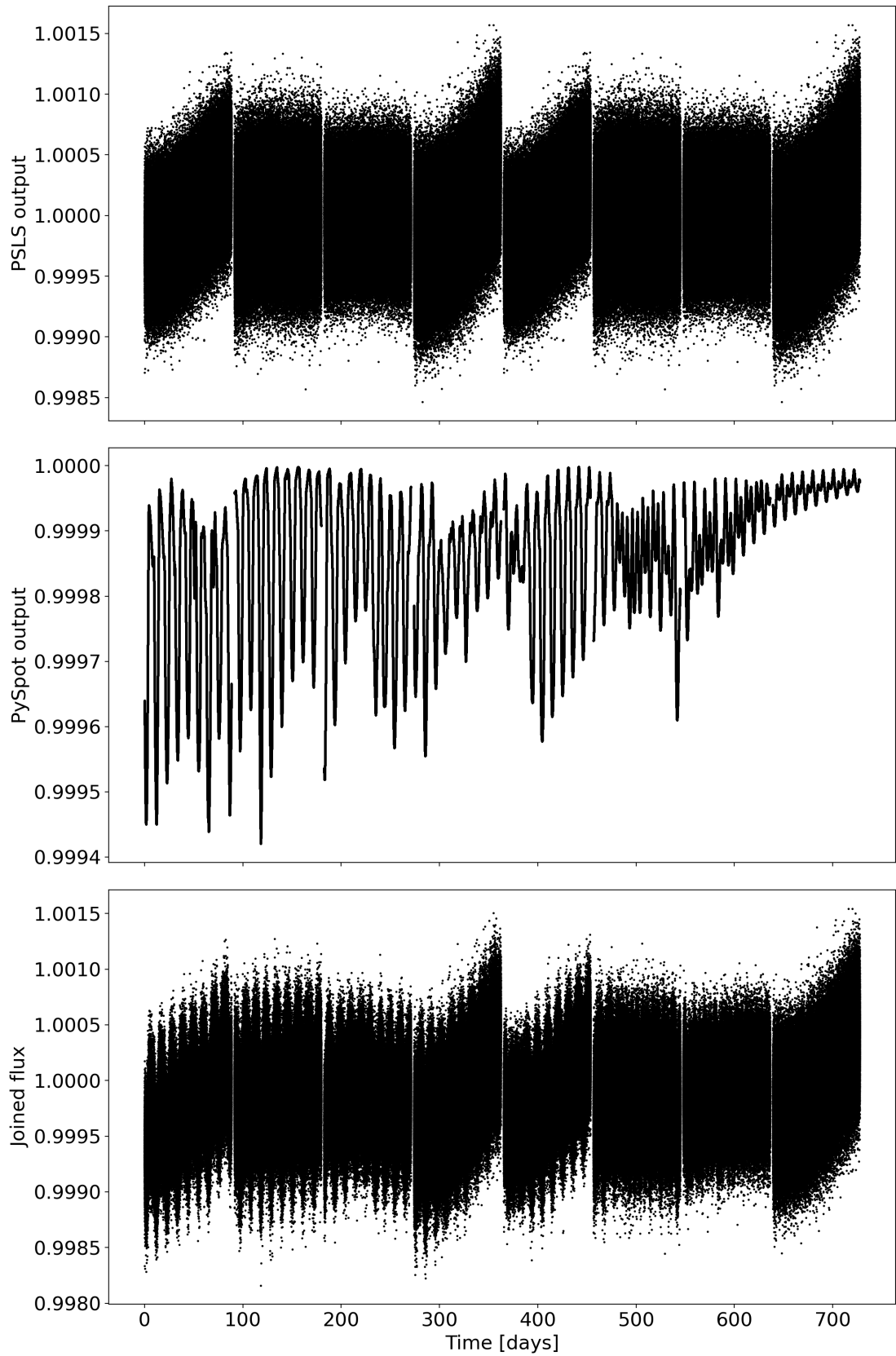
Figure 5: Produced LCs of the codes for stellar activity (PySpot) and of the PSLS simulator supplemented with the joined output for example star with ID: 10223302. On the y-axis is the normalized flux (1.0 value represents no change in the flux). The gaps between each quarter are also visible.
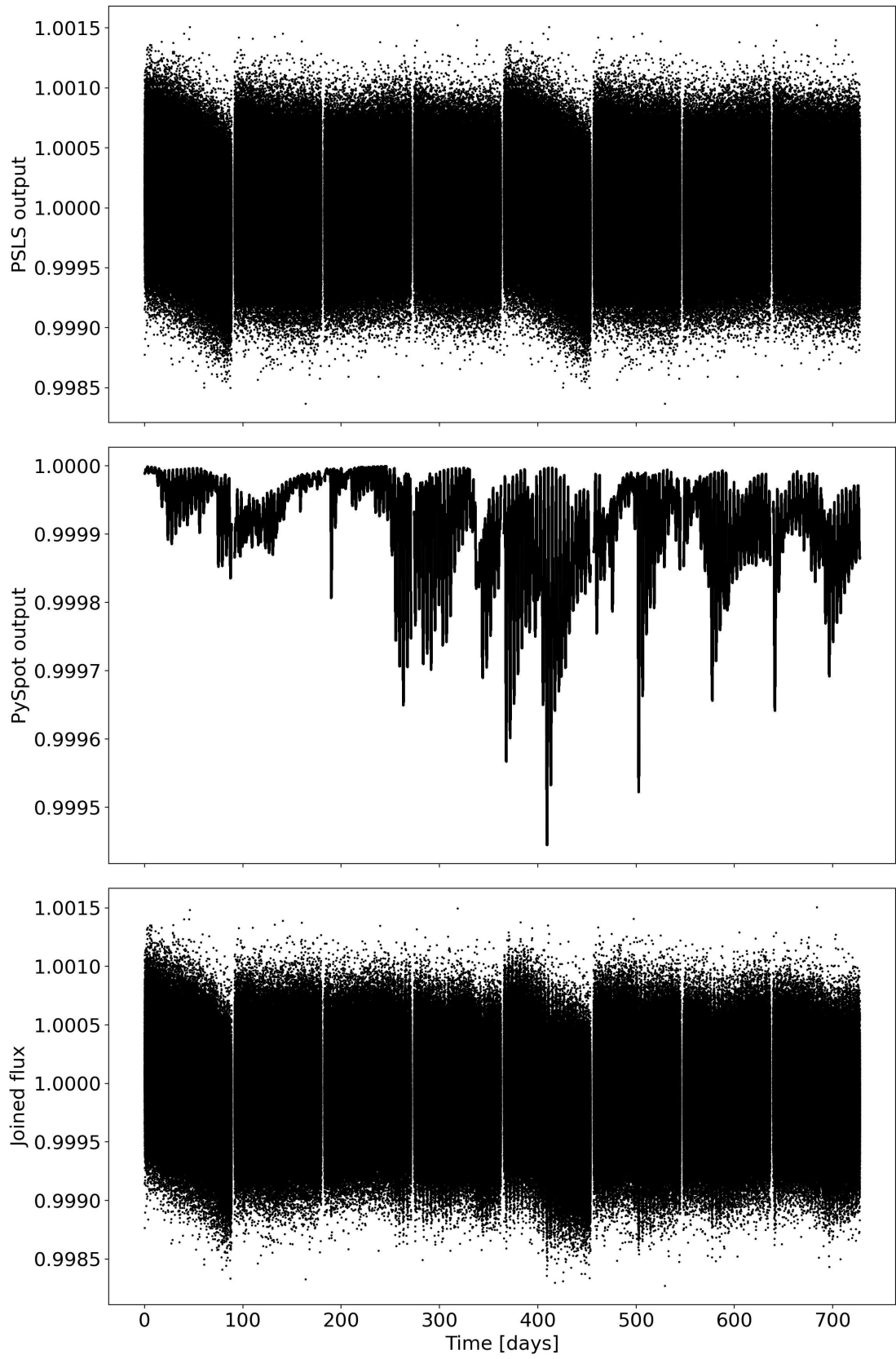
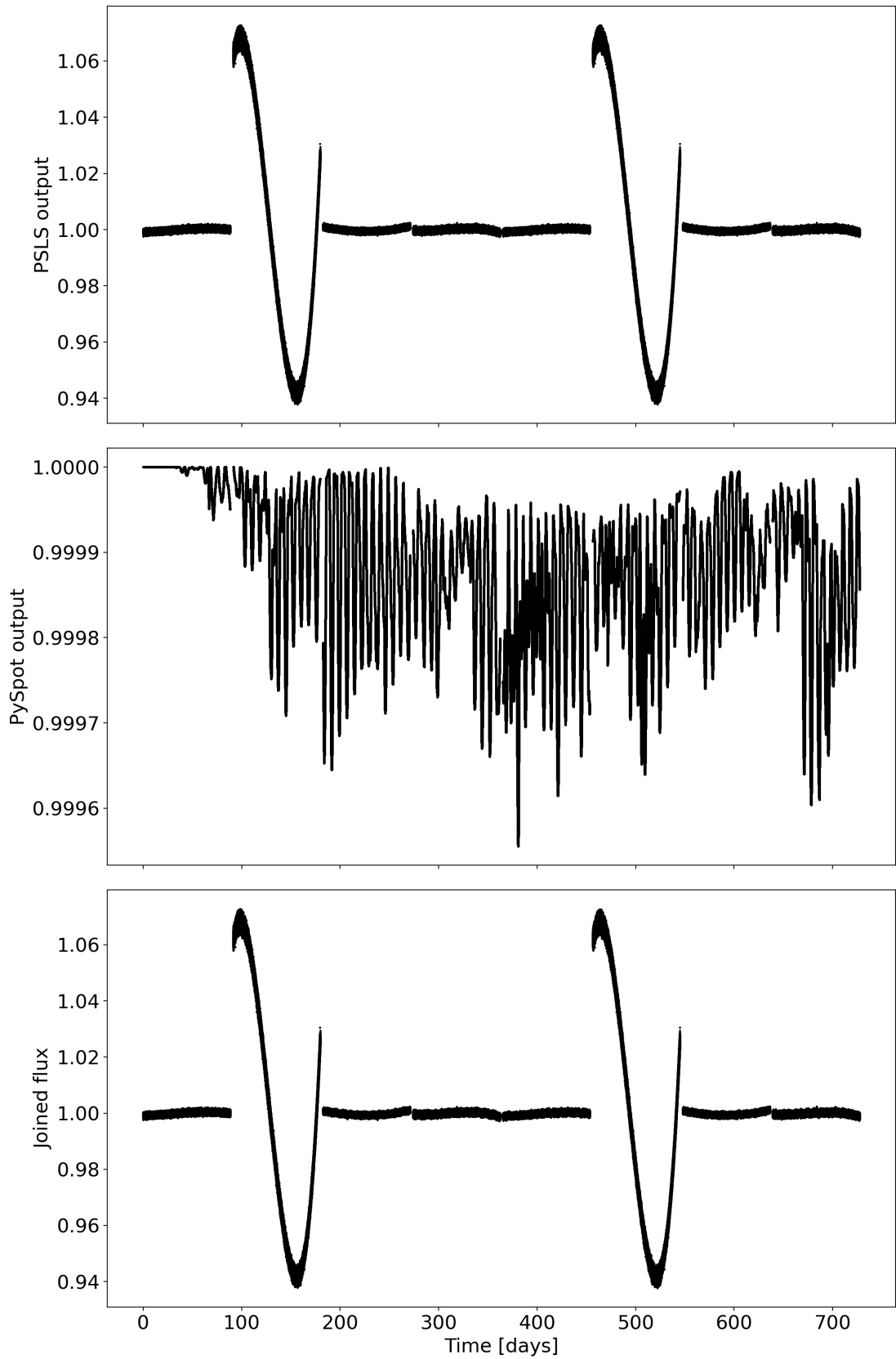Figure 6: Produced LCs of the codes for stellar activity (PySpot) and of the PSLS simulator supplemented with the joined output for example star with ID: 10279275. On the *y*-axis is the normalized flux (1.0 value represents no change in the flux). The gaps between each quarter are also visible. It is clear that the systematic errors for the second and sixth quarter are much larger than for others.

3 Filtering algortihms

In order to discover and characterize exoplanets, we must first search through vast data sets, which contain the noisy measurements of stars. These will include the errors that contaminate the LCs, and were explained in detail in the previous section. For the purpose of separating the planetary signal of a transiting planet from the noise, we must first detrend the light curve, and this is where the filtering algorithms come into play. These algorithms are essential tools, as they enable the removal of instrumental artifacts, systematic errors, and outliers from the observed data, allowing the further exoplanetary signals detections that would otherwise be obscured by the complexities of real-world observations. By mathematically modeling and reducing these sources of noise, filtering algorithms pave the way for more accurate and reliable exoplanet detection and characterization. As mentioned earlier this work is a continuation of Canocchi et al. (2023) paper which already assessed the performance of many algorithms. This is why the main focus are the filtering algorithms that had the best recovery rate in the injection-retrieval tests: biweight method, Huber spline and Lowess-YSD regression.

In Figure 7 the comparison of different algorithms is presented with the filter model (red line) in the left panel and the filtered data on the right. Filtered data is a raw flux divided by the filter model which will be later used as an input for the transit search algorithms. The input data does not contain the planetary signal and there is no significant difference between filtering algorithms in this specific example.

3.1 Biweight method

Tukey's biweight, also known as bisquare, is a well-established robust location estimator used in the context of detrending light curves (Mosteller and Tukey, 1977). Unlike ordinary least-squares, which assigns constant weights, the biweight assigns weights based on the distance from the central location. The weight function for the biweight is defined as:

$$
\begin{cases}
\left(-(a/c)^2\right)^2 & \text{if } |a/c| < 1 \\
0 & \text{otherwise}
\end{cases}
\tag{1}
$$

where:

- a represents the residuals, which are the differences between the observed data and the central location.

- c is a tuning parameter, set to a value of approximately 5 in default settings (as in Wōtan).

The choice of c affects both the efficiency and robustness of the estimate. For instance, setting $c = 4.685$ results in an estimate with 95% efficiency compared to the least-squares method for normally distributed data. In general, higher values of c result in a less robust estimate, but enhance its efficiency. For more details on the c tuning parameter values refer to the Wōtan paper Hippke et al. (2019).

Tukey's biweight method can be employed as a one-step estimate, where the initial guess for the central location is the median. As an alternative, one could approximate it using iterative algorithms like the Newton-Raphson method. In practical terms, the disparity between the single-step estimation and a fully converged solution is generally modest but not insignificant, typically falling within the range of a few to a few tens of parts per million (ppm). This supplementary detrending noise can be mitigated by employing a few iterations of the Newton-Raphson method, achieving convergence to levels lower than 10^{-6} or less than 1 ppm. Tools such as Wōtan provide the flexibility to customize the convergence threshold according to specific requirements.

Summing up, Tukey's biweight method is used with a specific value of c for robust location estimation in the presence of noise and outliers. The weights assigned to data points depend on their distance from the midpoint, allowing for effective noise reduction in the LC data.

This work uses the Wōtan implementation of the Tukey's biweight method, that in general have the one main parameter which value needs to be considered before using – the window length. In their paper, authors suggest the use of a window size 2.5 times the transit duration. The latter for a central transit of the planet on a circular orbit is:

$$T_{14,\text{max}} = \frac{2(R_s + R_p)}{v_p} = \frac{2(R_s + R_p)}{2\pi a/P} \tag{2}$$

which due to the Kepler's third law (assuming that the star is much more massive than the planet) becomes:

$$T_{14,\text{max}} = (R_\text{s} + R_\text{p}) \left(\frac{4P}{\pi G M_\text{s}} \right)^{1/3} \tag{3}$$

where physical quantities denoted with 'p' refer to the planet and the ones denoted with 's' – to the star. Note that for the transits that are not central, the transit times will be smaller, which means that, calculated in this way, $T_{14,\text{max}}$ will, in fact, be the maximum duration time of the planetary transit. If we substitute $R_\text{s} = R_\odot$ and $R_\text{p} = R_\oplus$, as well as $M_\text{s} = M_\odot$ and $P = 1$ yr. We then get the maximum transit time for the Earth which is approximately 13 hours.

The Wōtan recommendation is to use the window length (which is the length of the filter window) 2.5 times larger than the transit duration. For the Earth-analogues this will be around $w = 1.35$ days, consequently the three window sizes around this value will be tested in order to better understand their effect on the detection efficiency of Earth-analogues. The selected windows size are: 0.7 days, 1.4 days, and 2.0 days.

3.2 Huber spline

In statistics and estimation techniques, it is beneficial to search for robust estimators that are less sensitive to the outliers. While conventional methods like ordinary least-squares and maximum likelihood estimation are widely used, more robust alternatives have been proposed, with one of them being the concept of M-estimators, which was initially introduced by Huber (1964).

M-estimators are providing generalization of maximum likelihood estimation and include various methods for estimating parameters. The common methods like nonlinear least-squares and maximum likelihood estimation are considered to be a special cases of M-estimators. The main idea behind it is to adjust the estimator's properties to the specific biasvariance trade off that is desired, assuming an underlying distribution that characterizes the data. These estimators are considered optimal when the observed data closely follow the assumed distribution.

What distinguishes the estimators between one another are their loss functions. The least-squares method employs the loss function in form: $L(a) = a^2$, with a representing the differences between observed values and the midpoint (so called residuals). The alternative developed by Huber (1964) that was furtherly developed by Huber (2011), incorporates the Huber loss function, which behaves quadratically for small residual values and linearly for larger ones. The transition between the two different behaviors is controlled by a tuning parameter c. When it approaches infinity, the Huber loss function converges to the typical least-squares estimator.

Mathematically, the Huber loss function can be written as:

$$L(a) = \begin{cases} \frac{1}{2}a^2 & \text{for } |a| \leq c \\ c\left(|a| - \frac{c}{2}\right) & \text{for } |a| > c \end{cases} \tag{4}$$

In this context, c refers to the tuning parameter that modifies the sharpness of the transition from quadratic to linear loss, while a represents the residuals.

The Huber spline is just an application of this Huber loss function with a specific parameter c. For example the Huber spline with window length of $w = 0.3$ days can be described using the following loss function:

$$L(a) = c^2 \left(\sqrt{1 + \left(\frac{a}{c}\right)^2} - 1 \right) \tag{5}$$

The Huber spline is a versatile tool for robust estimation in various statistical applications, providing a balance between quadratic and linear loss functions to accommodate different data characteristics and requirements. This is why it can be commonly used given the aim of detrening the LCs.

In this study the tuning parameter c is set at 1.5, which serves as the default value in Wōtan implementation. For the same reasons as explained in the section devoted to the biweight function method, the window lengths that are being tested are as well: 0.7, 1.4 and 2.0 days.

3.3 YSD Lowess Regression

The YSD Lowess regression method, was developed by Battley et al. (2020), being a Python pipeline employed to model stellar variability using the Lowess smoothing regression. The conventional Lowess regression method applies a low-order polynomial to data segments determined by a designated window size at every position on the x-axis. The window can take the form of any function along the x-axis (such as a boxcar or a Gaussian). Regression on the y-axis, on the other hand, can be carried out using the common least-squares methods or other robust estimators, such as the biweight.

The weighting function which is used in YSD Lowess smoothing is the standard tricube function:

$$w(x) = \left(1 - \left|\frac{d}{3}\right|^3\right)^3 \tag{6}$$

In this context, the weight w assigned to each data point x is determined by its distance d from the point on the fitted curve, normalized to fall between $[0, 1]$.

The detrending pipeline consists of the following steps:

1. Identifying and eliminating outliers, specifically peaks and troughs, is carried out with a prominence threshold of 0.001 and a span of 20 data points. This task is accomplished using the `find_peaks` function from the `scipy.signal` package

2. The estimation of variability trends is achieved using the `Lowess` function from the `statsmodels` module, with the `frac` parameter representing the scaled weight. For the purpose of filtering the LCs in this work, `frac` is set to 0.02.

The choice of parameters, such as the `frac` parameter, may depend on the specific data set and is crucial for achieving optimal performance. Different values of `frac` parameters were tested on the sample of 10 arbitrarily chosen LCs. The best recovery rate outcome proved to be when setting the parameter value to 0.02.

The YSD Lowess regression method is versatile approach while modeling the stellar variability. It combines the strengths of Lowess smoothing, which is based on locally weighted polynomial regression, with the ability to handle swift evolution in young star light-curves. This weighting of data points based on their proximity to the point being estimated sets it apart from more common methods like the Savitzky and Golay (1964b) filter.
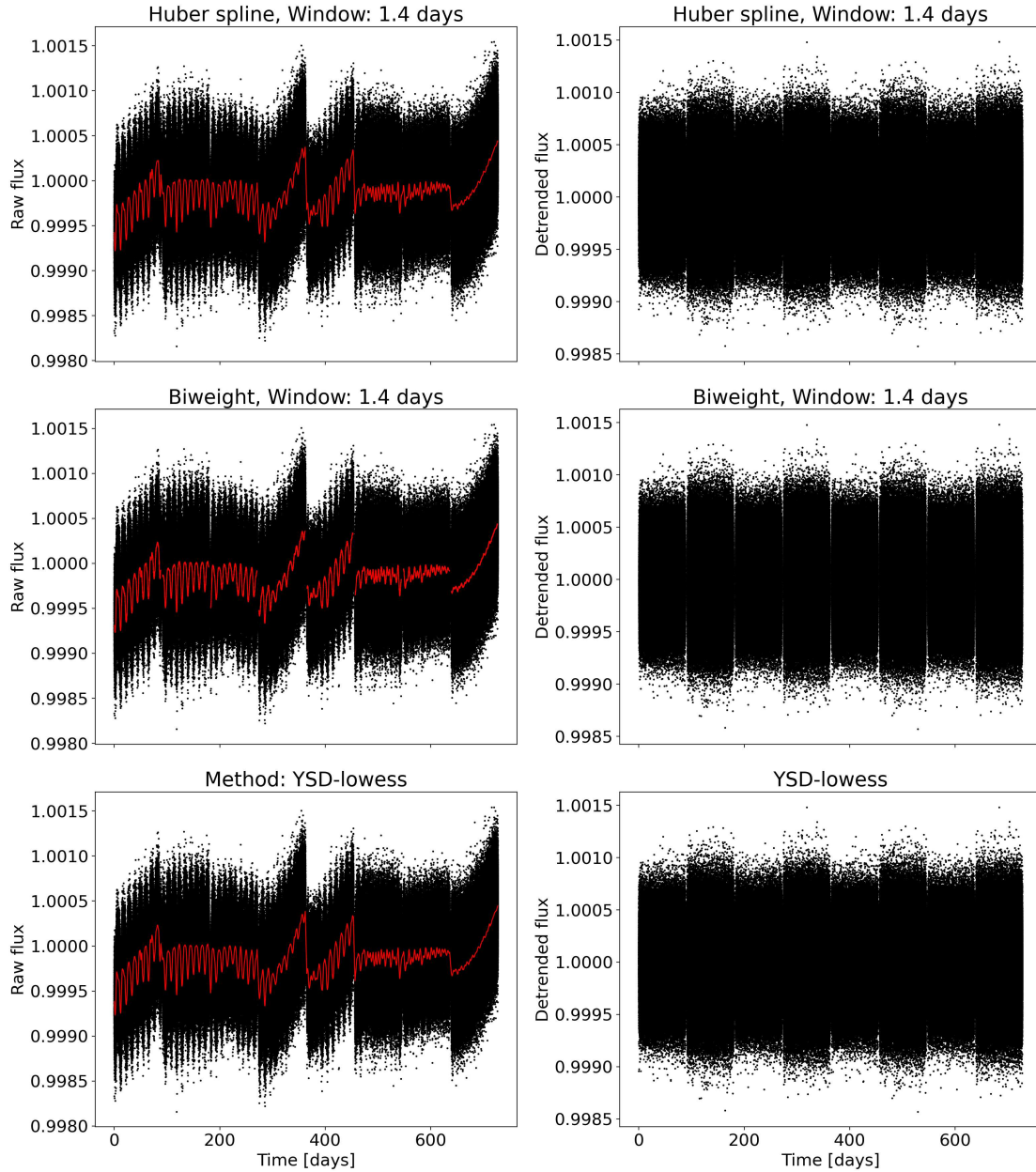
Figure 7: The comparison of filtering algorithms output for the sample star with id 10037521. Each row corresponds to different filtering algorithm. On the top there is the Huber spline method with window length of 1.4 days. Middle one shows the Biweight method with the same window length of 1.4 days. Finally, the bottom one is done by YSD-Lowess algorithm. In the left panel, the raw flux is marked with black dots, while the filter model is drawn with red line. The right panel presents the detrended flux which is the outcome of dividing the flux by the filter model. The filtering was done on empty LC without any planetary signal.

4 Planet injection and retrieval

The primary objective of this study is to compare the effectiveness and performance of various filtering techniques in identifying Earth-analogue planets during the initial two-year mission phase of PLATO. To accomplish this, the simulated light curves, discussed in Section 2, must be populated with the planetary data for analysis.

4.1 Planetary parameters selection

The selection of planetary parameters should be randomized, while also ensuring that the planet remains within the HZ. Understanding the boundaries of the HZ, specifically the Inner Edge (IHZ) and Outer Edge (OHZ), is crucial and is largely influenced by the star's effective temperature. The assumptions regarding the dependence of IHZ and OHZ on stellar parameters are grounded in the work by Kopparapu et al. (2013). The planet's parameter selection begins with the orbital period choice that is picked at random from the truncated normal centered on the Earth-like HZ. The standard deviation will be equal to the difference between the Earth HZ and the IHZ. To ensure that the periods are not too large, the upper limit must be restricted to the full cycle of the LC, which is in this case ~ 727 days (full cycle is 730.5 days but the last 3 days do not contain any data due to the observational gap). This choice is justified by the objective to test the efficiency of the filtering algorithms, rather than replicating a realistic distribution of planets, as seen in the computation of the mission's planetary yield. The dependence of the HZ period range and the temperature resulting from previously mentioned paper is presented in Figure 8.
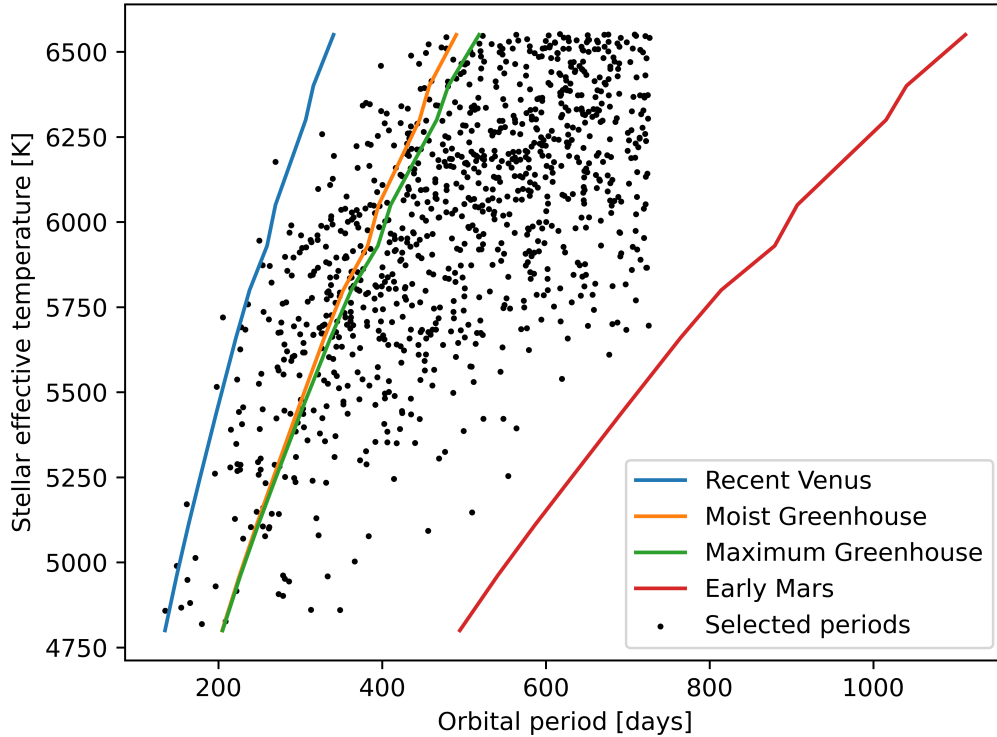

Figure 8: Plot presenting the selected orbital periods (in days) and the host star's effective temperature (in Kelvins). The recent Venus, moist greenhouse, maximum greenhouse and early Mars edges are also present in the plot and are referred from the paper by Kopparapu et al. (2013). Some of the planets orbital periods are below the rather optimistic inner edge of the HZ restricted by the recent Venus because of the fact that the upper limit had to be moved by ~ 3 days to avoid falling into the observational gap. This will have no significant meaning in the assessment of the filtering algorithms.

The Recent Venus term is based on the empirical studies Solomon and Head (1991), and fully relies on the fact that Venus does not have liquid water on its surface since 1 Gyrs. It constitutes rather optimistic limit on the IHZ, but can be used when considering the space missions that search for exoplanets, not to miss any of the potential candidates for the planet within HZ. On the opposite edge there is an Early Mars outer limit that is based on the similar considerations regarding Mars not having the liquid water since 3.8 Gyrs.

The moist green house effect is a phenomenon that occurs when the stratosphere becomes saturated with water, resulting in accelerated hydrogen escape into space. The latter happens because the water vapor molecules in the upper part of the atmosphere can be broken down by ultraviolet (UV) radiation, which will result in the significant water loss, potentially desiccating the planet. On the contrary there is a maximum greenhouse limit that refers to the point at which having more CO_2 (or another greenhouse gas) in the planet's atmosphere would not cause any additional warming due to the fact that it is already saturated with the gas, and effectively is trapping as much heat as it possibly can.

All the planets are assumed to have their eccentricities fixed at $e = 0$, which means that they are on circular orbits. Having the planets orbit the host star on eccentric orbits would not change much in the assessment of the filtering algorithms which is the main goal of this thesis. It only simplifies the problem, and this way the the argument of pericenter is undefined, as in a circular orbit all the points are equidistant to the center of mass. Out of the convenience, to avoid any computational problems, the argument of the pericenter for all planets is fixed at $o = 90°$. The planet's radii (relative to the host stars radii) were drew at random from the uniform distribution between 0.01 and 0.03. The T_0 of the transit (the time of the first transit counted from the beginning of the observations) were picked from the uniform distribution between 0 and the orbital period. For a reliable detection of the planetary signal to be possible and to establish the orbital period, at least 2 transit are needed to be present in the LC. Because the data is limited to the first 2 years of observations, and in most cases only 2 transit would be observable due to the long orbital periods of the planets, the condition on the T0 value of not falling into one of the eight gaps was imposed.

The two last parameters that are required later to inject the planet onto the LC are: semi-major axis and the inclination angle. The semi-major axis can be inferred from the stellar density using the Kepler's Thrid Law:

$$P^2 = \frac{4\pi^2}{GM_\mathrm{s}}a^3 = \frac{3\pi}{G\rho_\mathrm{s}R_\mathrm{s}^3}a^3 \tag{7}$$

where the term M_s/a^3 have been substituted with $4\pi\rho_\mathrm{s}/3$. Solving the equation for a (scaled to stellar radius) one gets:

$$a/R_\mathrm{s} = \left(\frac{GP^2\rho_\mathrm{s}}{3\pi}\right)^{1/3} \tag{8}$$

Having calculated the semi-major axis, which in this case is just a circular's orbit radius, one can move on to the final parameter – the inclination angle. The inclination angle must be such to allow for the planetary transit observations. The impact parameter b, that can range between 0 (central transit – planets movement from the observers point of view happens grazing the centre of the star) and 1 (the planet grazes through the edge of the star), is directly related to the inclination angle. For an eccentric orbit this relation is:

$$b = \frac{a\cos i}{R_\mathrm{s}} \cdot \frac{\sqrt{1 - e^2}}{1 + e\sin o} \tag{9}$$

so in general, the inclination angle will depend on b – the impact parameter, a/R_s – the semi-major scaled to the stellar radii, e – the eccentricity of the orbit and o – the argument of the pericenter. For circular orbits it can be simplified to:

$$b = \frac{a\cos i}{R_\mathrm{s}} \quad \rightarrow \quad i = \arccos\left(\frac{b}{a/R_\mathrm{s}}\right) \tag{10}$$

17

The inclination angle is thus calculated by first drawing the impact parameter from the uniform distribution between 0 and 1.

4.2 Limb darkening coefficients

Finally, to inject the planetary signal into the LC, one needs to know the limb darkening coefficients of the star, which basically reflect the shape of the transit. The phenomenon has its origin in observation that the center of the star appears brighter than its edge (limb). The underlying reason for this behaviour is the nature of the stellar atmosphere's opacity. The light that is emitted from deeper, hotter layers of a star is more intense than that from the cooler, outer layers. When we observe a star directly head-on, we are seeing the light coming from these deeper layers. As the observer's line of sight moves towards the limb or edge of the star, they are looking through an increasingly thicker part of the star's cooler, outer atmosphere. This causes the observed light to be dimmer, hence the star appears to be darker at the edges.

These coefficients are used to quantitatively describe this effect in models, when analyzing LCs from the transiting exoplanets. They are employed directly in the limb darkening laws, which serve to model the intensity variation across the stellar disk. In this work, the quadratic model is used (Kopal, 1950), which means that the two parameters are required to describe the previously mentioned effect on the intensity. The equation to describe the quadratic law is:

$$\frac{I(\cos\theta)}{I(1)} = 1 - u_1(1 - \cos\theta) - u_2(1 - \cos\theta)^2 \tag{11}$$

where $I(1)$ – is the radiation intensity measured at the center of the star, $\cos\theta$ – is the cosine value of the angle formed between the line of sight and the outgoing intensity, u_1, u_2 – are the limb darkening coefficients that need to be established.

In order to calculate the values of the limb darkening coefficients a prior knowledge of the effective temperature of the star (T_{eff}), the logarithm of gravitational acceleration at the surface ($\log g$), and the metallicity of the star (Z) is required. The evaluation was carried out using the PyLDTK (Python Limb Darkening ToolKit) that have been developed by Parviainen and Aigrain (2015) and is based on the PHEONIX library (Husser et al., 2013). The crucial element of running the code is defining the filter that will be used to observe the star. The PLATO filter profile is not yet publicly available but the shape is clearly seen in Figure 9, which resembles the one of the CHEOPS mission (Benz et al., 2021) but with a steeper separation at lower wavelengths (strong cut-off around 500 nm).

To imitate the real profile, the Boxcar filter was used with the transmission level at 1.0 between 500 and 1000 nm. The CHEOPS and Boxcar limb darkening oefficients were compared with the values reported in the PLATO-PL-KULPL-TN-0020 i.1.2 (already mentioned PLATO Consortium Technical Note), which provided the limb darkening coefficients for around 1/3 of the input stars. The Boxcar proved to be a better representation of the PLATO passband rather than the CHEOPS filter, at least in the computation of the quadratic limb darkening coefficients. This is reasonable, if we think to the higher impact of limb darkening in transit observations at shorter wavelength (e.g., Knutson et al., 2007). The discrepancy between the empirically obtained values of the coefficients and the ones obtained using PHOENIX models can be up to 0.2 (Patel and Espinoza, 2022). It is imperative that the difference in coefficient values remains within this range to ensure the reliability of the derived stellar parameters. Secondly, the same limb darkening coefficients that are used to model the transit signal will be later used to retrieve the planet from the LC, so theoretically the recovery rate should not be influenced by the accuracy of the estimated coefficients.

4.3 Planetary signal injection

The capability to inject synthetic planetary transits into stellar LCs provides invaluable insights into the robustness of detection algorithms and the assessment of potential systematic errors. Those are the main reasons for which the precise modelling of planetary signal becomes valuable. With the help of the `batman` (BAsic Transit Model cAlculatioN) python package (Kreidberg, 2015a), this process becomes streamlined, at the same time offering a high degree of customization to simulate different transit scenarios.

To initiate the process, one must first define a set of transit parameters:
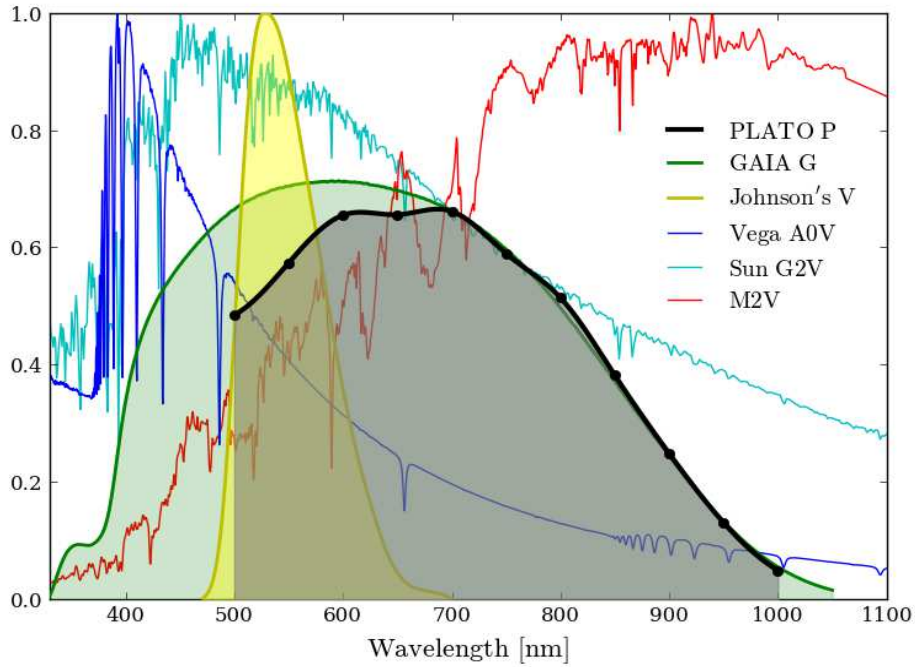
(a)
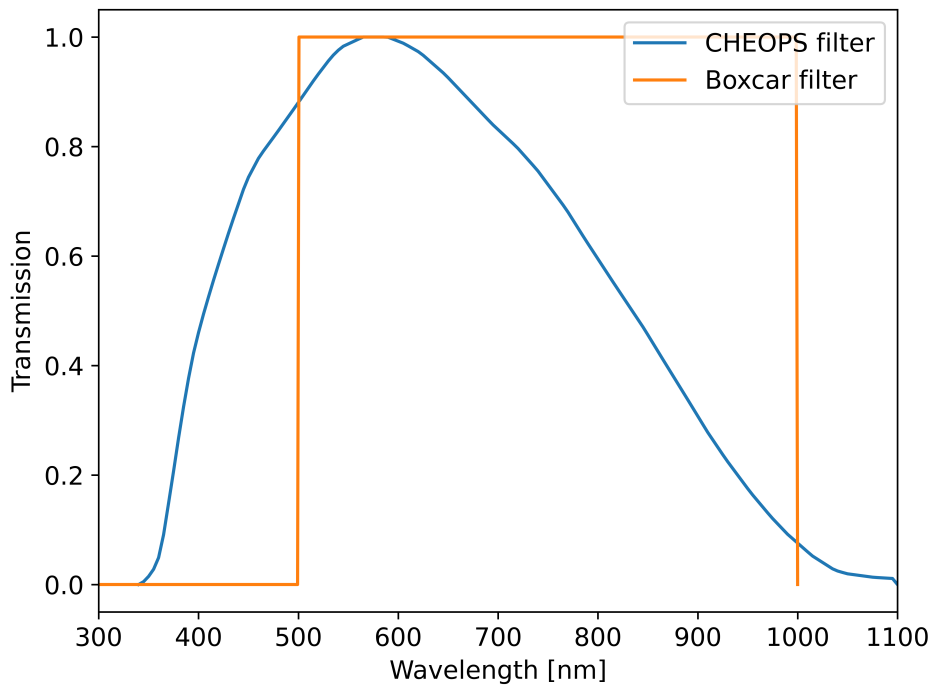
(b)

Figure 9: (a) A comparison of the preliminary spectral response of PLATO N-CAM at BOL conditions with other known bands, filters and spectra. GAIA G refers to the Gaia mission G band, Johnson's V is the Johnson's V filter. Labels starting with the name of the star are pointing to the normalized spectra of the given stars. The figure is taken from the work by Marchiori et al. (2019) (b) Filter profiles plot for the CHEOPS mission taken from the SVO Profile Service (Rodrigo and Solano, 2020) and the one which will be used in this work (Boxcar filter) to imitate the PLATO passband filter.

- T_0: time of inferior conjunction in days

- P: orbital period in days

- R_p: planet radius in the units of stellar radii

- a: semi-major axis in the units of stellar radii

- i: orbital inclination angle in degrees

- e: eccentricity of the orbit

- o: argument (longitude) of periastron in degrees

- limb_dark: limb darkening model

- u: limb darkening coefficients

for which the selection process has been already described in subsections 4.1 and 4.2.

The output of the code is a transit model which can be later injected into the LC. The model is normalized, meaning that the out-of-transit flux is simply 1.0. The examples of the resulting planetary signals produced by the `batman` package are presented in Figures 10 and 11.
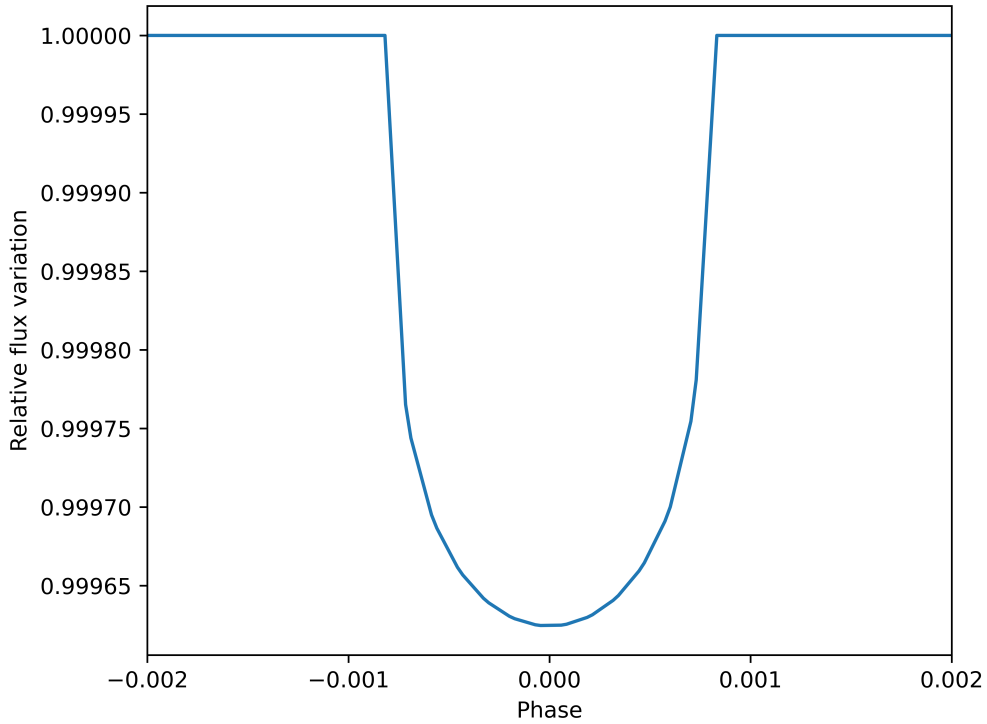

Figure 10: The modelled planetary transit for the target star with id 17451622. The transit parameters are: $T_0 = 57.28$ days, $P = 565.94$ days, $R_\mathrm{p} = 0.0176$, $a = 210.40$, $i = 89.98°$, $e = 0$, $o = 90°$, $u_1 = 0.458$, $u_2 = 0.135$.

Injection of the signal is rather straightforward and does not require in-depth explanation. Using the `batman package`, the model for a given time series, planetary and stellar parameters can be created. It is then incorporated into the previously produced LC (for details see section 2). Figures 12 and 13 present the generated LCs for the same example stars. In these stacked plots, the top LC is missing any planetary signal, the middle one shows the curve with a planetary signal, while the bottom one is the binned LC with the injected planet.

Binning is a method in data processing used to mitigate the impacts of minor observational discrepancies. Original data values that are falling within a specific small range, also known as a bin, are substituted by a value that characterizes given interval, typically the central value.
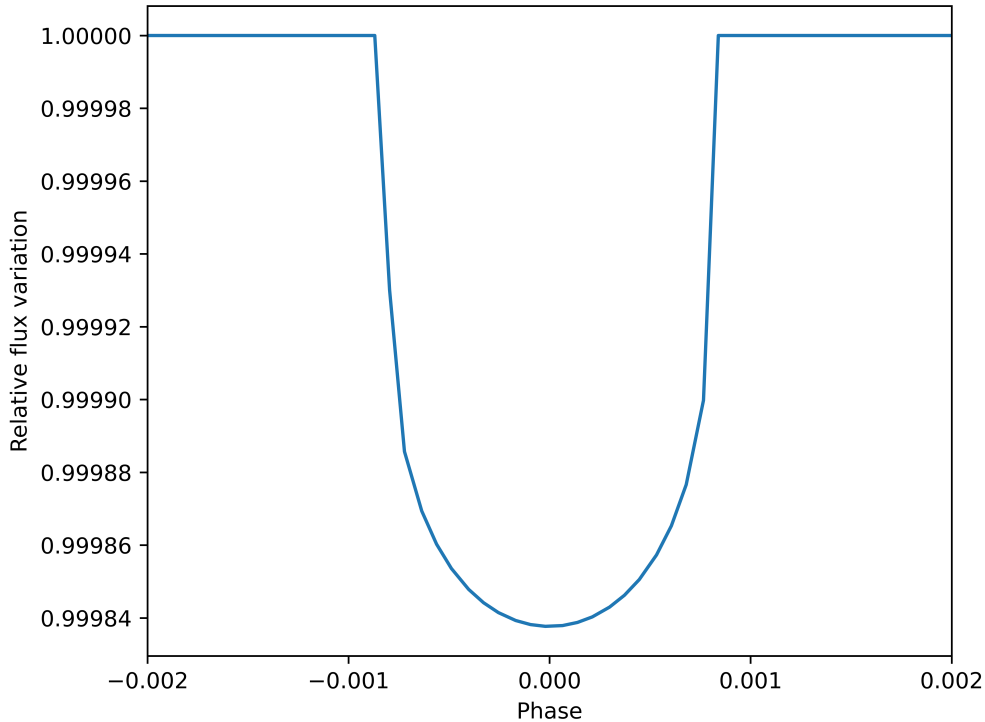
Figure 11: The modelled planetary transit for the target star with id 15958871. The transit parameters are: $T_0 = 102.41$ days, $P = 312.52$ days, $R_{\mathrm{p}} = 0.0117$, $a = 186.35$, $i = 89.89°$, $e = 0$, $o = 90°$, $u_1 = 0.442$, $u_2 = 0.155$.

Among the binning methods the most popular ones include smoothing with mean or median value. In this work, binning is done with the help of the `scipy.stats` package and in particular the `binned_statistics` function, which apart from the statistics method choice also allows the easy implementation of the minimum number of counts per bin. The bin size was chosen to be 10 min, which is a balanced compromise between the further computational load of executing the transit search method, at the same time not altering the transit signal in substantial way. There must be at least 5 data points inside a bin and the chosen statistics is the mean value.

From Figures 12 and 13 one can clearly see that for some LCs the planetary transit can be detected 'by eye' (Figure 12) while for others it might not be an easy task (Figure 13). There are many factors , which contribute to that but the main one is the position on the LC of the transit occurrence. The amplitude and the width of the signal itself are also crucial but in our case similar planets (Earth-analogues) are considered and the planetary parameters do not differ substantially. The binned LCs containing the injected planetary signal are then subject (input) to the filtering algorithms which should help in the transit recovery.

4.4 Planetary signal retrieval

The identification of exoplanets using the transit method is based on the detection of periodic reductions in the brightness of a star, that indicates an orbiting planet occulting a portion of the starlight. In order to enhance the accuracy of this process, specialized algorithms are employed. Amongst them, the Box-fitting Least Squares (BLS, Kovács et al., 2002) and Transit Least Squares (TLS, Hippke and Heller, 2019) algorithms stand out due to their widespread use and efficiency.

The BLS method seeks to identify potential transits by fitting a box-shaped function to the observed LCs. Essentially, it scans the data for consistent, periodic reductions in brightness that align with the profile of a box. While BLS is effective in scenarios where the approximate depth and duration of the transit are anticipated, its restrictive shape can sometimes result in overlooked transits, especially if they are of shorter duration or deviate from the anticipated box-like profile.

The actual profile of a transit diverges from the box shaped one making the use of accurate shape fitting method more beneficial. This is where more refined alternative, the TLS is useful.

Figure 12: The stacked plot of the flux without planetary signal (top), flux with planetary signal (middle) and the binned flux with planetary signal (bottom) for the target star with id 17451622. The transit happens with $T_0 = 57.28$ days and $P = 565.94$ days. The bin width is 10 min and the chosen statistics is the mean value.
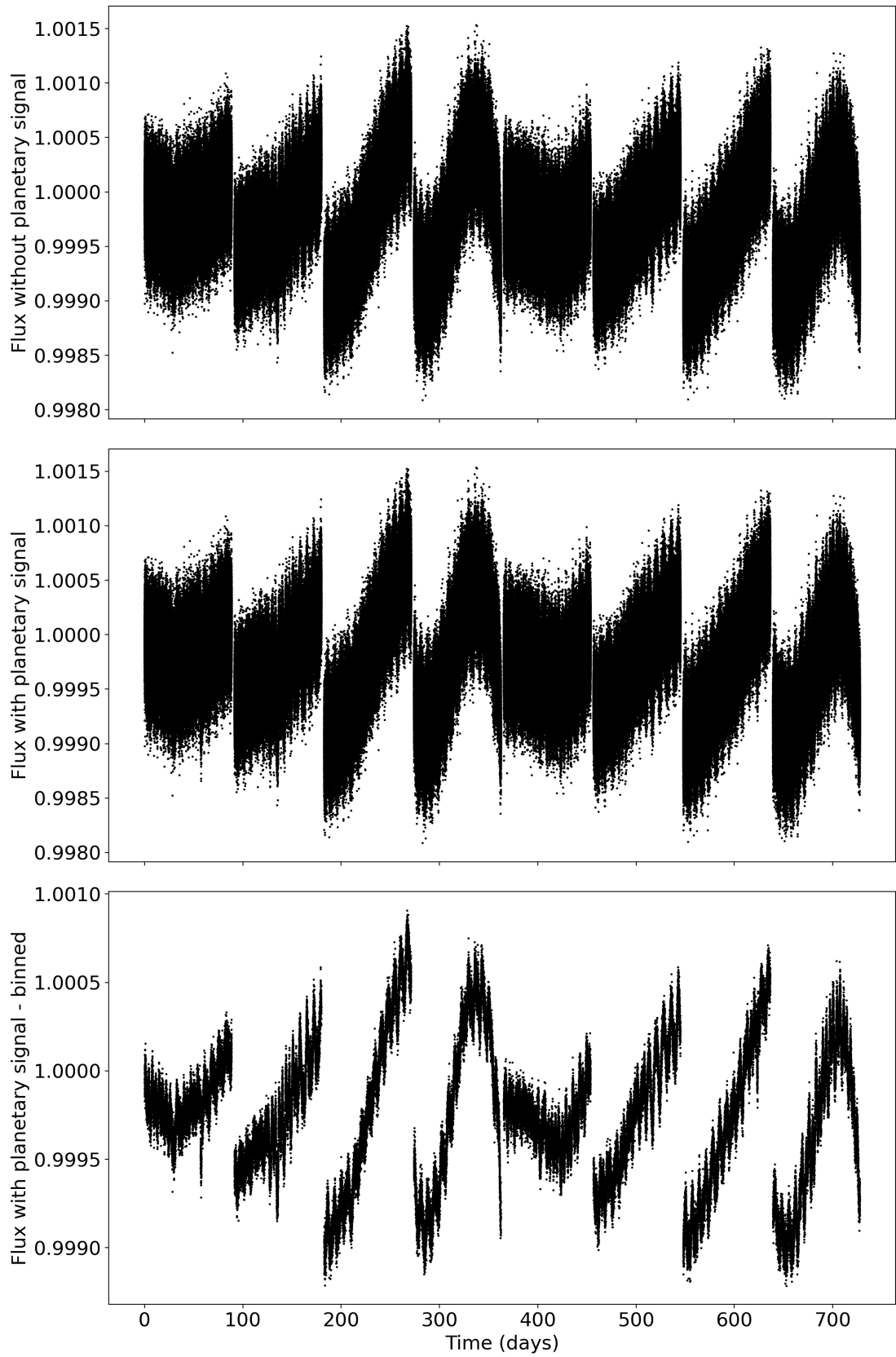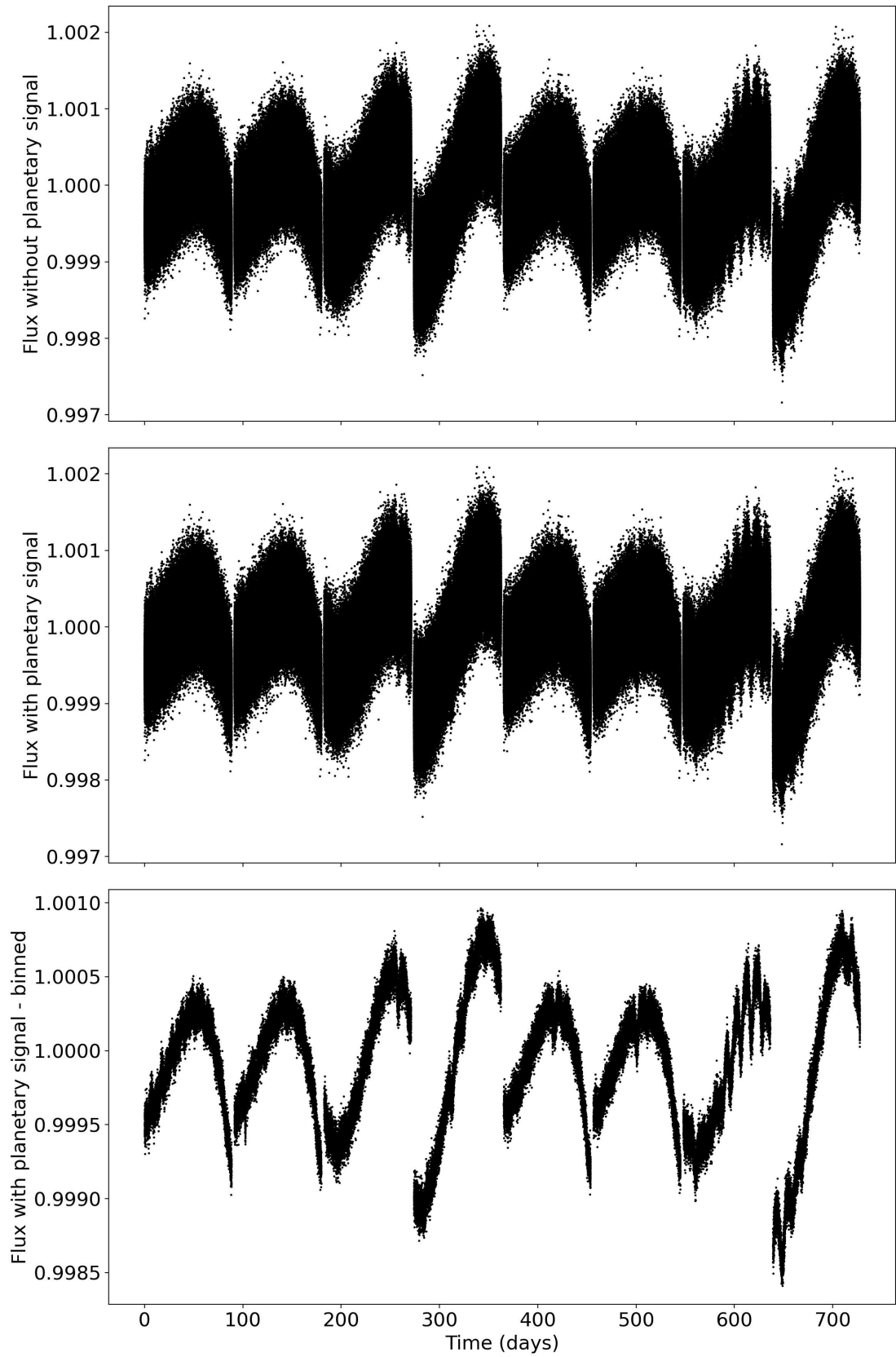
Figure 13: The stacked plot of the flux without planetary signal (top), flux with planetary signal (middle) and the binned flux with planetary signal (bottom) for the target star with id 15958871. The transit happens with $T_0 = 102.41$ days and $P = 312.52$ days. The bin width is 10 min and the chosen statistics is the mean value.

It is tailored to identify transits by modeling the light curve with shapes that more accurately reflect the profile of real planetary signals. Instead of a simplistic box, TLS employs a composite of potential transit shapes, enhancing its sensitivity. This makes it a great choice for detecting smaller planets or transits that might not perfectly conform to the box shape assumed by BLS. In fact, the TLS uses the previously described package `batman` that allows for the customization of limb darkening profile. In the realm of exoplanet detection, determining the authenticity of a signal is crucial and this is why the Signal Detection Efficiency (SDE) definition needs to be introduced.

Recovery Criterion: Signal Detection Efficiency (SDE)

The Signal Detection Efficiency (SDE) serves as a yardstick to quantify the significance of a detected signal. It computes the contrast between the strength of a potential transit signal and the prevailing background noise:

$$\text{SDE} = \frac{\text{Transit Signal Strength}}{\text{Mean of Background Noise}} \tag{12}$$

The TLS calculates the SDE as a metric to pinpoint the most noticeable signal in the periodogram (period decomposition) that aligns with a potential exoplanetary transit. The procedure for determining the SDE distribution with respect to the orbital period is adapted from the methodology by Kovács et al. (2002):

$$\text{SDE}(P_{\text{orb}}) = \frac{SR_{\text{peak}} - \langle SR(P_{\text{orb}}) \rangle}{\sigma(SR(P_{\text{orb}}))} \tag{13}$$

In this relation:

- $SR(P_{\text{orb}})$ signifies the signal residual, which is the disparity between the theoretical model and the actual observational data. This is inferred from the distribution of the minimum χ^2.

- $\langle SR(P_{\text{orb}}) \rangle$ represents the arithmetic mean of the signal residual.

- $\sigma(SR(P_{\text{orb}}))$ denotes its standard deviation.

- SR_{peak}, the peak value

In TLS the SR is calculated from the distribution of the minimum χ^2 as a function of P:

$$SR(P) = \frac{\chi^2_{\text{min,glob}}}{\chi^2_{\text{min}}(P)} \tag{14}$$

This calculation inherently confines the $SR(P)$ distribution to fall within the interval [0, 1]. The χ^2 statistic is assessed on the phase-folded LC, comparing the data points from the transit model to the observed values. After examining the SDE distribution, the period with the highest SDE value is identified as the orbital period for the detected transit.

In Figure 14 the discrepancy between the SDE values for TLS is presented and is taken directly from the paper by Hippke and Heller (2019) (Figure 7). It can be clearly seen that the TLS outperforms the BLS, even after the median-smoothing of the LC, before employing the BLS method on it. The difference is in the SDE values is essential as for TLS, the signal with the maximum SDE was found at 66.7 while for BLS it reached only 16.9 without processing the LC and 24.5 after smoothing.

The planetary signal is considered to be recovered when it fulfills these two criteria:

1. The recovered period matches the injected one with the maximum discrepancy of 1%

Figure 14: The plot taken directly from Figure 7 of Hippke and Heller (2019), which compares the performance of TLS with respect to the BLS. The left plots are the phase folded LC with the maximum SDE transit fit marked with red line. Plots on the right are the periodograms, which decompose the SDE values for each period searched. The top figure are the results for TLS, middle one is the BLS with no prior data processing and the bottom one is the BLS with median smoothing.

2. SDE value is above certain threshold value which is calibrated using the data containing white noise only

Both conditions must be satisfied concurrently. If the identified signal aligns with the period of the introduced planet yet the SDE falls beneath the established threshold, the signal is treated as unretrieved.

4.5   Estimating the SDE threshold

In this work the same reasoning as in Hippke and Heller (2019) will be adapted to determine the SDE threshold, above which we consider the detection of the planetary signal to be qualified for the further period check. The approach for estimating this threshold involves executing the TLS on data without any injected transit, aiming to compute the SDE values corresponding to the white noise only. Consequently, we could set the false positive rates at 1%, implying that in actual observations, only 10 out of 1000 empty light curves would be flagged as potential transit candidates.

Previously, in their work Canocchi et al. (2023) assumed the SDE threshold value to be 7.0 and was directly taken from the Hippke and Heller (2019), who performed the calculations on 10 000 realizations of the synthetic LCs. The results of their study are presented in Figure 15 where they compare it with BLS performance as well. For both TLS and BLS, the false positives rates at the 1% level were obtained for 7.0 SDE threshold value, but the TLS outperformed in terms of the planetary signal recovery efficiency.

In their research, various empirical thresholds for SDE were highlighted and compared. Specifically, they list the following thresholds that have been proposed by different authors:

- SDE > 6.0 as suggested by Dressing and Charbonneau (2015).

- SDE > 6.5 according to Livingston et al. (2018).

- SDE > 7.0 as presented by Siverd et al. (2012).

- SDE $\in [6, 8]$ (varying based on the orbital period) as discussed by Pope et al. (2016); Aigrain et al. (2016).

- SDE > 9.0 adopted by Heller et al. (2022)

- An even higher threshold of SDE > 10 is proposed by Wells et al. (2018).

This assortment of thresholds offers a diverse perspective on potential SDE values for the transit detection. It is, however, worth noting that higher SDE threshold values will result in reduced recovery rates, while lower ones carry the risk of having a high number of false positives. For this reason, the in-depth study of the SDE threshold within the particular data set is beneficial to maintain the perfect balance between the detection efficiency and the false positive rate.

To provide the SDE threshold values for different levels of the false positives rates, the TLS method was employed on the P1 sample LCs before the injection of the planetary signal. Most of the parameters were chosen to be the default ones, but the period range have been restricted for the lower limit. The injected planets will have orbital periods larger than 100 days, this is why running TLS for low values would be only an unnecessary computational expense. The lower limit of period range was set to 40 days and the minimal number of transits was set to 2. The detrending was carried out before running TLS with arbitrarily chosen biweight method, for which the window length was set to be 0.5 days.

The resulting SDE threshold value estimate should be 10.5 to be casting the 1% rate of false positives. However, there are several factor that should be considered as well. First of all, the sample was quite small compared to the one that Hippke and Heller (2019) have used to generate their results. Secondly, the generated LCs representative of the P1 sample, are divided into the quarters which differ between one another by the systematic errors relative amplitude. In general, they will not have standard variation above 0.5% meaning that systematic errors variations cannot exceed this value as well. Nevertheless, the highest SDE values on the white noise data are achieved for the LCs that have the largest variations of the systematic between adjacent quarters, which is clearly seen in Figure 16.
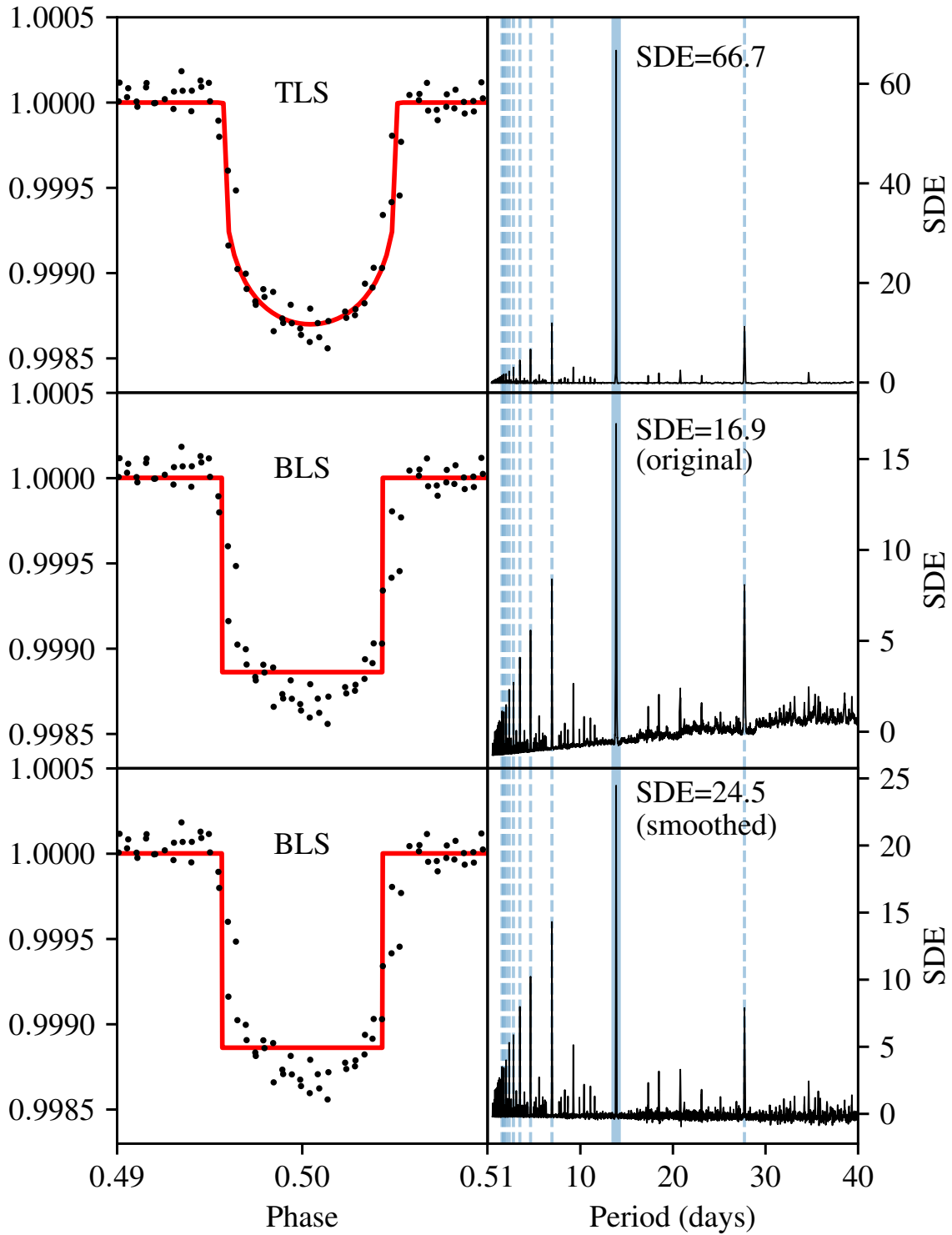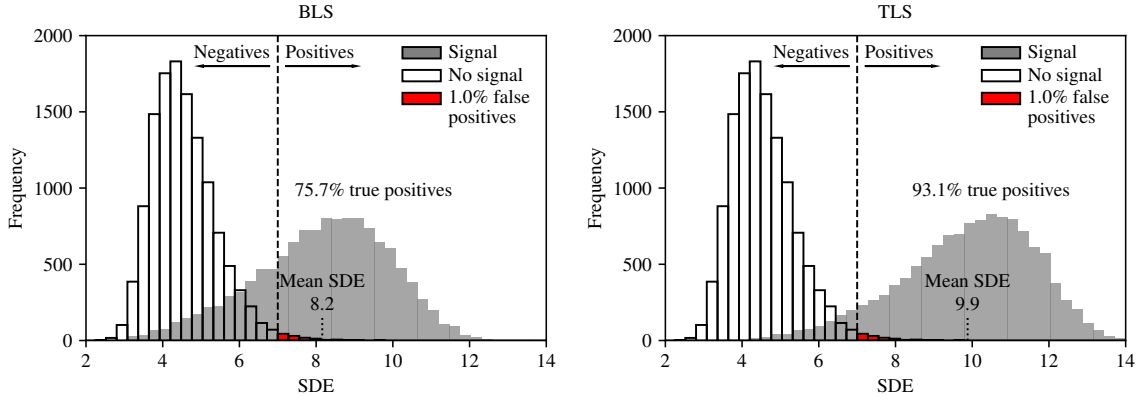
Figure 15: The plot taken directly from Figure 6 of Hippke and Heller (2019), which compares the performance of TLS with respect to the BLS in terms of the transit recovery (true positives). For both methods the SDE threshold is at 7.0 and is set from the condition of 1% of false positives for the white noise data. The data with signal has the mean SDE of 9.9 for TLS, which means better performance of this method compared to the BLS, that has the mean value at 8.2.

False Positive Rate	SDE Threshold
1%	10.54
2%	9.93
3%	9.49
4%	9.01
5%	8.82

Table 2: The estimated SDE thresholds for different levels of false positive rates. The threshold is provided for the rate of up to 5% and was calculated for the 1000 LC containing only white noise data (no planetary signal). The curves were first detrended with the biweight method using the window length of 0.5 days.

The 1% false positive rate is generally accepted within the astronomical community, primarily because of the precious telescope time that is at stake. Allocating the telescope time for the follow-up of objects based on a higher false positive rate may lead to a considerable amount of wasted effort on celestial bodies that are irrelevant or non-beneficial to the core objectives of the study. Such a misallocation also affects the rate at which significant discoveries are made, as the observation time is finite and highly desired by many groups.

Hence, for a more comprehensive analysis in this study, we will be exploring three distinct values for the SDE threshold. These are:

1. > 7.0, which aligns with the threshold set by Canocchi et al. (2023) in their work

2. > 9.0, following the value chosen by Heller et al. (2022) in their studies

3. > 10.5, which is the value deduced from current analysis

Examining these thresholds will allow to shed light on their respective influence on both the recovery rate and false positive rates. This comprehensive approach will allow to make more informed decisions on the threshold value, that keeps the right balance between discovery potential and resource allocation.
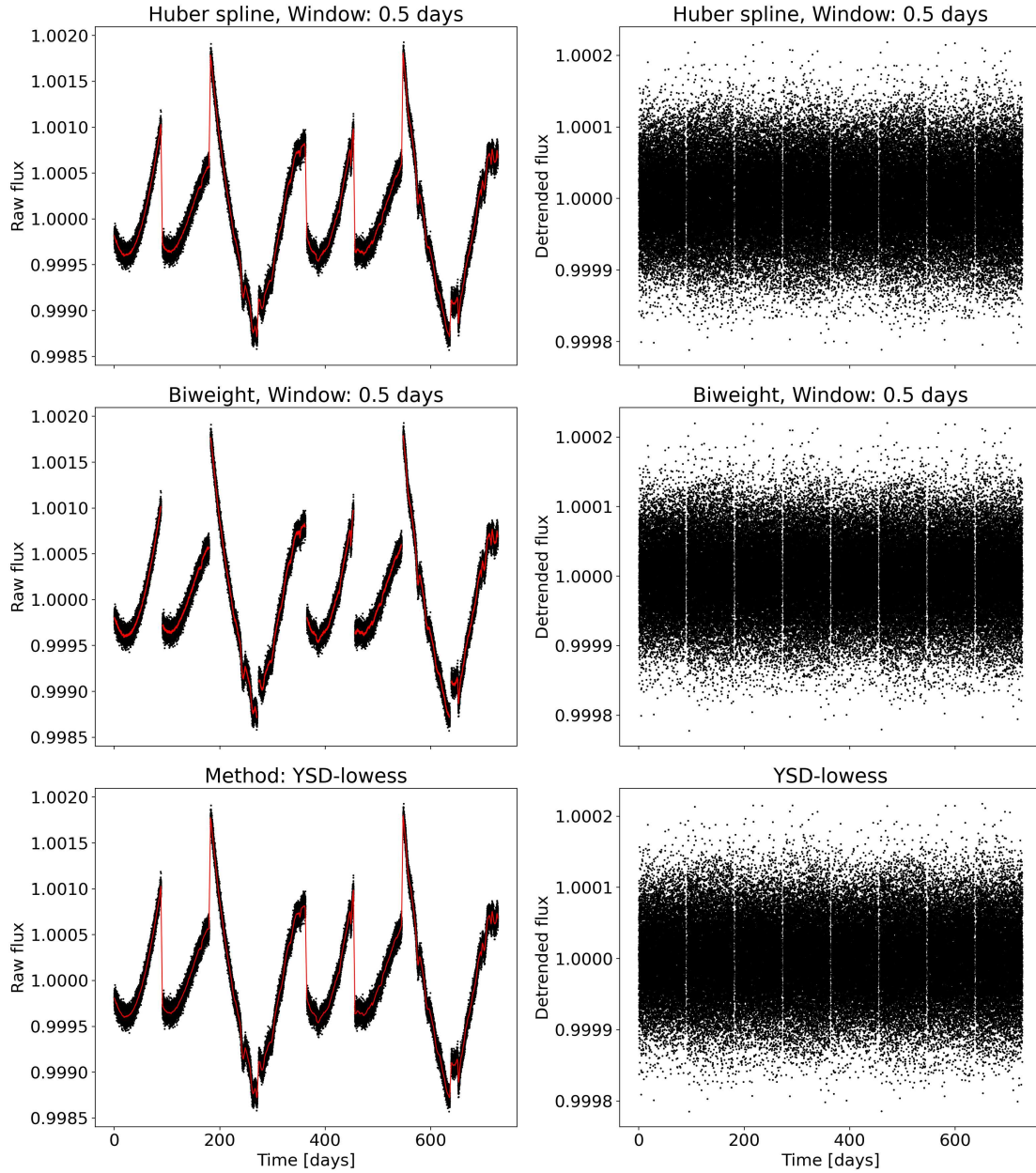
Figure 16: The detrended light curves for the target star with id 17266742 for which the highest value of SDE on empty data is reached (13.54). Such high value is reached due to the large systematic errors variation between adjacent observational quarters. The raw flux is a binned data to the 600 seconds cadence, with the minimum number of data points constituting the bin set to 5.

5   Results and discussion

The core objective of this master's thesis revolved around the evaluation and efficiency comparison of various filtering algorithms tailored for the detection of exoplanets, with a specific focus on the Earth-analogues. The algorithms subjected to the testing were: the Huber spline method (with window lengths of 0.7, 1.4, and 2.0), as well as the biweight method (employing identical window lengths), and the YSD-Lowess technique. Each of these algorithms was carefully examined to uncover their unique traits and computational complexities, in the context of exoplanet detection.

In this section the culmination of this extensive analysis will be presented, providing a detailed exposition of the outcomes, specifically the comparison of recovery rates of the injected planets and its dependence on the planetary and stellar parameters.

5.1   Recovery rate

As already explained in subsection 4.4 there are 2 conditions that need to be fulfilled simultaneously to consider a successful recovery of the planetary signal: the recovered period needs to be within 1% agreement with the injected period, and the SDE value needs to be larger than the actual threshold value. However, there are edge cases that go beyond these general criteria and incorporating them to the total efficiency will provide more accurate outlook on the real recovery rates. All cases that will be distinguished in this study include:

1. **Detection**: In this scenario, all transits are successfully recovered, and the planet is identified with the correct period, demonstrating a high level of precision ($\Delta P < 1\%$).

2. **Aliased Detection**: This case occurs when most of the transits are recovered, but the retrieved period is half of the injected period. It implies that some aliasing is present, leading to a period misinterpretation. This occurs when an observational gap is present midway between the two detected transits, causing the transit search algorithm to falsely identify a transit period precisely half of the injected one.

3. **Partial Detection**: In contrast, the partial detection scenario arises when only one transit is recovered, and the period estimate is "wrong" due to the presence of noise that is mistaken as another transit. When the first transit is not recovered correctly both T_0 and P will have wrong values. However, their combination matches the injected one with the 1% precision. The second case happens, when the first transit is recovered correctly but the second one is missed. Then the recovered value of T_0 will match the injected one given the same precision.

4. **Wrongly recovered**: Case in which the SDE crossed the established threshold value, but the recovered period does not overlap with the injected one, in any of the ways mentioned above.

It is worth noting that, given the mission's objectives, borderline cases '2' and '3' hold particular significance. Performing photometric and ground-based follow-up observations are likely to shed light on these intriguing cases. Recovering the correct period of the planet might be a hard task and goes far beyond the scope of this work. Numerous instances exist where planets have been validated or confirmed despite the lack of knowledge regarding their orbital periods. Those include (but are not limited to): TOI-2076 (Hedges et al., 2021), V1298 Tau (David et al., 2019) and HIP41378 (Vanderburg et al., 2016). To provide the maximum efficiency of the tested filtering algorithms the relevant sum of '1 + 2' and '1 + 2 + 3' will be presented as well. Previous work comparing the detection efficiency of filtering algorithms (Canocchi et al., 2023) did not need to include these additional criteria as their injection tests always included at least three transits, thus avoiding scenarios '2' and '3'.

In Figures 17, 18 and 19 the recovery rates are presented for all the algorithms that were being tested, with the distinction based on the data categorization, that was introduced earlier.
In Figure 20, the SDE value histograms for various methods are presented. Notably, the SDE distributions for YSD-Lowess, biweight (1.4-day window length), and Huber spline (1.4-day window length) exhibit remarkable similarities. These distributions have a nearly bimodal pattern, characterized by a prominent peak in the range of 25-30 SDE values, accompanied by a secondary, albeit less significant peak, located between 5-10 SDE. Additionally, it is clearly seen why the
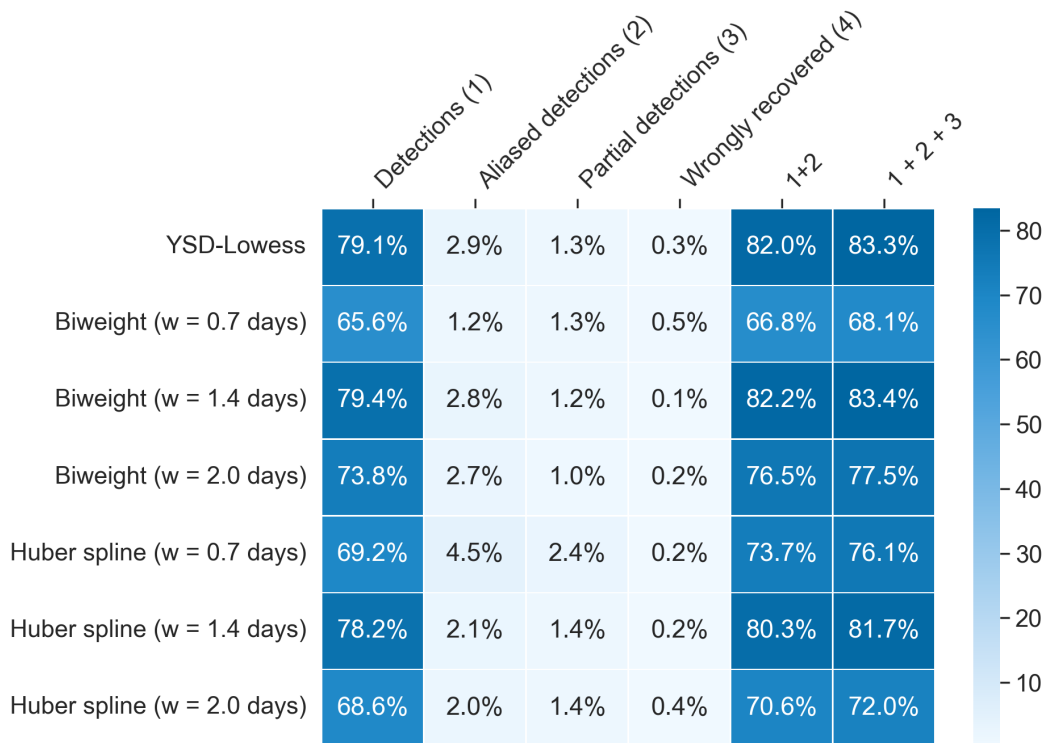
	Detections (1)	Aliased detections (2)	Partial detections (3)	Wrongly recovered (4)	1+2	1+2+3
YSD-Lowess	79.1%	2.9%	1.3%	0.3%	82.0%	83.3%
Biweight (w = 0.7 days)	65.6%	1.2%	1.3%	0.5%	66.8%	68.1%
Biweight (w = 1.4 days)	79.4%	2.8%	1.2%	0.1%	82.2%	83.4%
Biweight (w = 2.0 days)	73.8%	2.7%	1.0%	0.2%	76.5%	77.5%
Huber spline (w = 0.7 days)	69.2%	4.5%	2.4%	0.2%	73.7%	76.1%
Huber spline (w = 1.4 days)	78.2%	2.1%	1.4%	0.2%	80.3%	81.7%
Huber spline (w = 2.0 days)	68.6%	2.0%	1.4%	0.4%	70.6%	72.0%

Figure 17: Heatmap representation of the filtering techniques efficiency outcome for SDE threshold 10.5. The percentages are derived from a sample of 1000 LCs. Additionally, cumulative sums for categories "1+2" and "1+2+3" are provided.

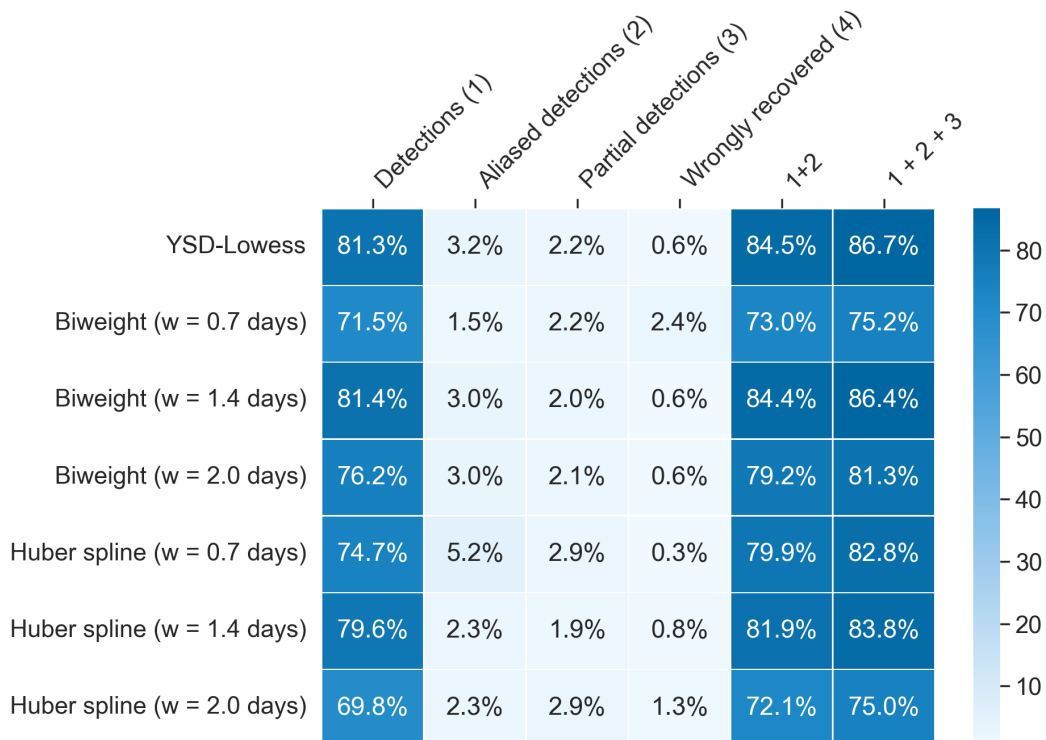|	Detections (1)	Aliased detections (2)	Partial detections (3)	Wrongly recovered (4)	1+2	1+2+3
YSD-Lowess	81.3%	3.2%	2.2%	0.6%	84.5%	86.7%
Biweight (w = 0.7 days)	71.5%	1.5%	2.2%	2.4%	73.0%	75.2%
Biweight (w = 1.4 days)	81.4%	3.0%	2.0%	0.6%	84.4%	86.4%
Biweight (w = 2.0 days)	76.2%	3.0%	2.1%	0.6%	79.2%	81.3%
Huber spline (w = 0.7 days)	74.7%	5.2%	2.9%	0.3%	79.9%	82.8%
Huber spline (w = 1.4 days)	79.6%	2.3%	1.9%	0.8%	81.9%	83.8%
Huber spline (w = 2.0 days)	69.8%	2.3%	2.9%	1.3%	72.1%	75.0%

Figure 18: Heatmap representation of the filtering techniques efficiency outcome for SDE threshold 9.0. The percentages are derived from a sample of 1000 LCs. Additionally, cumulative sums for categories "1+2" and "1+2+3" are provided.
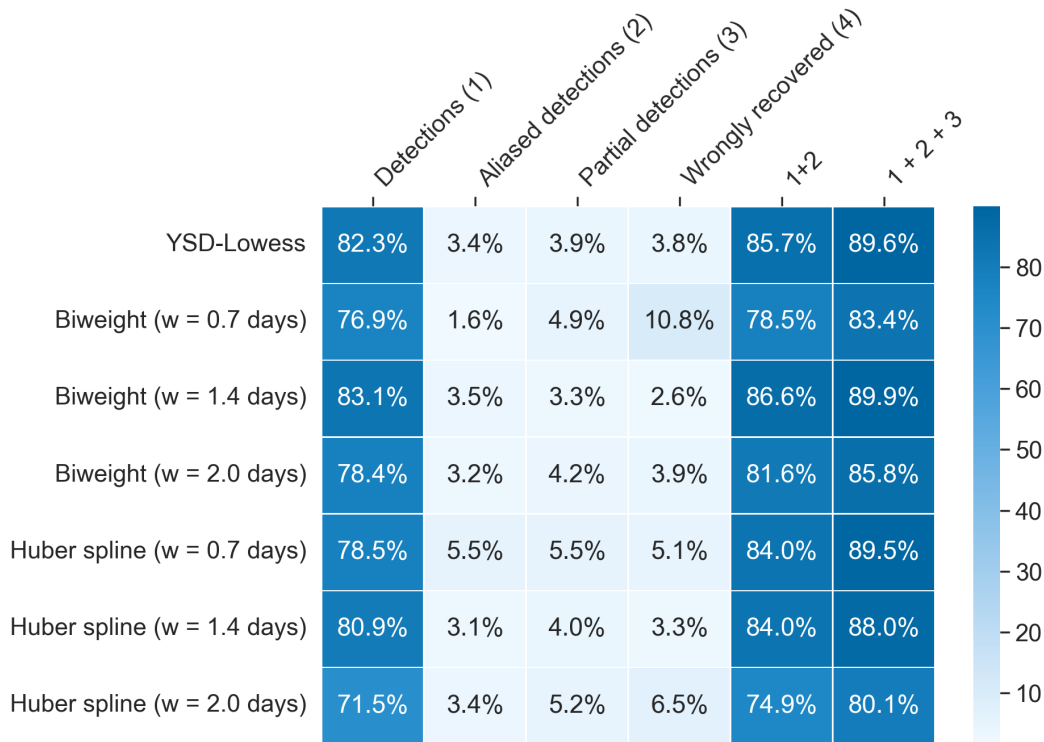
Figure 19: Heatmap representation of the filtering techniques efficiency outcome for SDE threshold 7.0. The percentages are derived from a sample of 1000 LCs. Additionally, cumulative sums for categories "1+2" and "1+2+3" are provided.

Huber spline method with a 2.0 window length yields the poorest performance. The lower SDE value peak becomes more pronounced, resulting in an increase in the number of missed transits. Given the SDE threshold of 10.5, anticipated to achieve a false positive rate of approximately 1%, the analysis shows that the biweight method with a window length of 1.4 days stands out as the most efficient, delivering an efficacy of 83.4%. Concurrently, it manifests a mere 0.1% rate of wrongly recovered transits. Closely following is the YSD-Lowess technique, exhibiting an efficiency of 83.3% and a slightly elevated 0.3% rate of incorrect transit recovery. Another noteworthy method is the Huber spline with a 1.4 days window, having an efficiency of 81.7% and also low (0.2%) percentage of wrongly recovered cases. What can be seen is that the number of partial detections is nearly the same for all these methods. Lowering the SDE threshold generally enhances overall detection efficiency; however, this comes at the expense of an increased number of incorrectly recovered cases. Intriguingly, at diminished SDE threshold values, the Huber spline method with a 0.7 days window length emerges as a competitive technique. At SDE threshold value of 7.0, it even surpasses the Huber spline with a 1.4-day window in terms of efficiency. The surge can be primarily attributed to the escalated count of aliased detections, which reach a noteworthy 5.5%. This higher performance, nevertheless, is accompanied by an increase in the rate of false recoveries.

Among the various methods evaluated, the biweight approach with a 1.4-day window length stands out as the preeminent technique, showcasing overall the best efficiency as well as the lowest number of incorrect recoveries. It is closely rivaled by YSD-Lowess, which also achieves commendable, nearly indistinguishable from the biweight method results. Determining an optimal SDE threshold proves to be challenging. While this study implies a specific limit, the data reduction pipelines could potentially decrease the false positive rates derived from the white noise data. This might cause the lower SDE values to be more appealing. In this study, however, transitioning from a 10.5 to a 7.0 SDE threshold boosted the efficiency of the biweight method with a 1.4 days window length by 6.5%. Nevertheless, this improvement came at the expense of a rise in wrongly recovered transits, which raised from 0.1% to 2.6%. Given that elevated false positive rates would necessitate
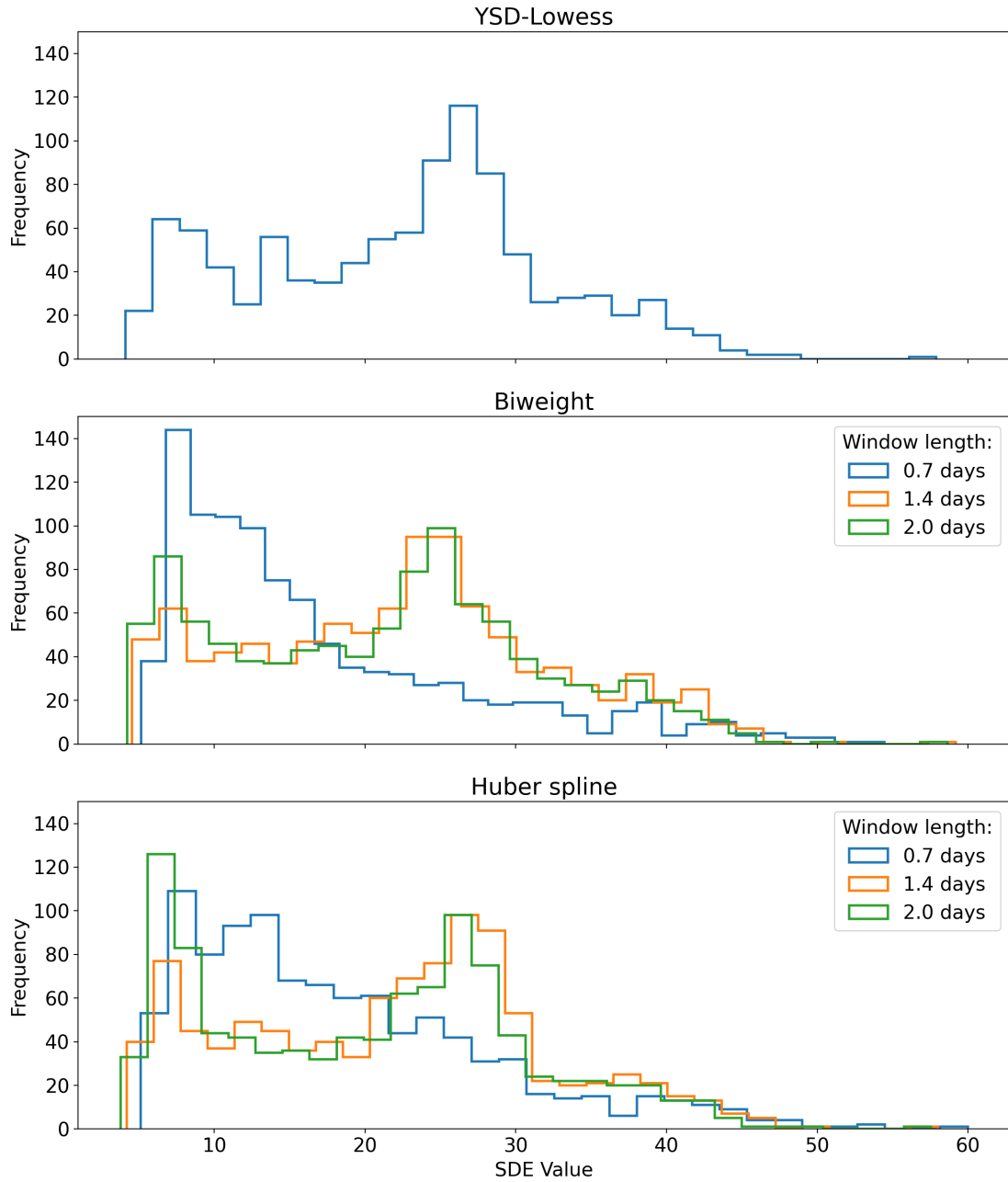
Figure 20: Histogram representation of the SDE values distribution for all studied methods. The sample contains 1000 P1 stars with the injected planets. The top panel illustrates the YSD-Lowess method, the middle panel displays the biweight method, with three distinct window lengths color-coded, and the bottom panel showcases the Huber spline, categorized in the same manner.

allocating more telescope time for the follow-up observations, it seems beneficial to lean towards the higher SDE threshold value to limit the occurrence of such cases.

The PLATO pipeline foresees the use of more than one filtering algorithm. Thus, joining the best performing methods: YSD-Lowess and Biweight with 1.4-day window length can provide insights into the cumulative detection efficiency. A planet might be overlooked by one filter but identified by another, and such instances would still be considered detections. Upon evaluating the combined efficiency of these two proficient methods, it is clear that the difference compared to using only one filter is not substantial. With an SDE threshold of 7.0, the combined efficiency reaches 89.9% (same as in case of using only biweight method), whereas with an SDE threshold of 10.5, efficiency is at 84.3% level (0.9% higher when using only the biweight method).

5.2 Recovering the planetary parameters

The transit curve shape encodes a wealth of information about an exoplanet and its orbit. By analyzing the features and characteristics of this curve, various planetary parameters can be extracted.

Planet-to-star radii ratio: The depth of the transit, representing the fraction by which the star's light dims during a planetary passage, is directly proportional to the square of the ratio of the planet's radius R_p to the stellar radius R_s. A deeper transit implies a larger planet in relation to its host star:

$$\frac{R_p}{R_s} = \sqrt{\delta}$$

where δ is the transit depth. This relation is obtained by assuming the spherical shape of the planet and star, as well as neglecting the effect of the limb darkening. In reality, the latter should be included as it can cause quite large overshoot of the recovered planet-to-star radii ratio. The discrepancy is approximately 15% for spectral type A, main sequence stars and can raise up to 20% for solar-like stars (Heller, 2019). The TLS package uses the analytical solution provided by the Heller (2019) accounting for the limb darkening effect:

$$\frac{R_p}{R_s} = \sqrt{\delta \frac{\langle I \rangle_A}{I_p}}$$

where $\langle I \rangle_A$ is the average intensity across the entire stellar disk area and I_p is the intensity covered by the planet in the middle of the transit.

As observed by Canocchi et al. (2023) the recovered ratio is underestimated and is caused by the deformation of the transit shape by the filtering algorithms. It can be clearly seen in Figure 21 where the injected and recovered transit shapes for one of the sample LC are drawn (with id 17451622). The difference is substantial as in this case the recovered ratio $R_p/R_s = 0.0104$, while the injected one was $R_p/R_s = 0.0176$. It also implies that all the other parameters that are directly inferred from the planet-to-star ratio will be prone to the estimation errors. This is not something we should be worried about, as the detected transits would go through a more refined modelling, for example by employing Markov Chain Monte Carlo analysis, in order to retrieve accurate planetary parameters.

Stellar density: Stellar density is a parameter that can be precisely extracted from the planetary transit because it is estimated using the Kepler's third law as shown in Equation 7. We can solve the equation for ρ_s and expressing it in units of solar density we get:

$$\rho_{s,\odot} = \frac{1}{(P(\text{days})/365.25)^2}$$

The period P is known with at least 1% precision and the only assumption that was made in this approximation was that planetary mass $M_p << M_s$ and it was neglected. For our example star the recovered period was 565.95 days so $\rho_{s,\odot} = 0.42$ which is really close to the real value of
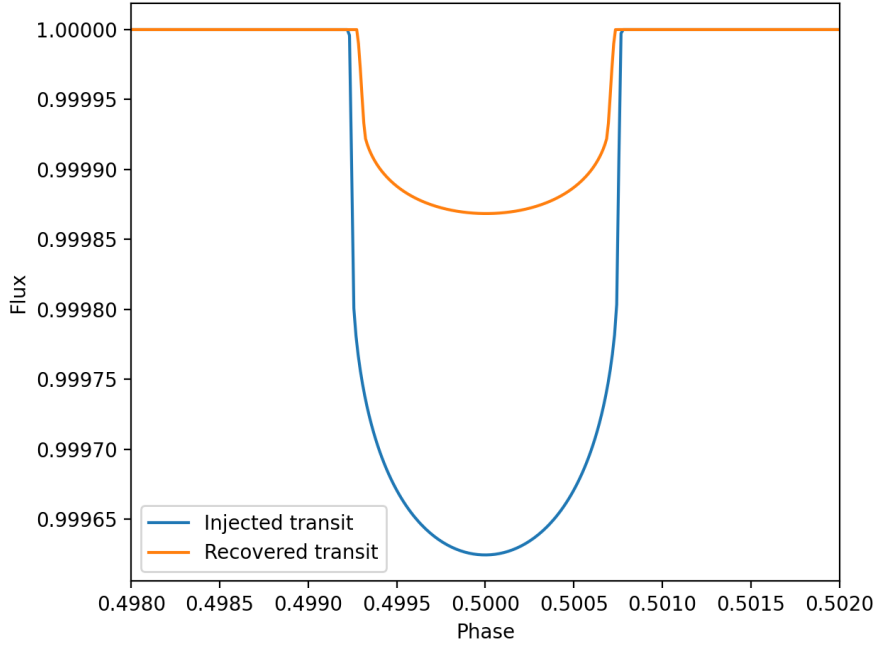
Figure 21: The comparison of the injected transit signal (blue color) and the recovered planetary signal (orange color) for the star with id 17451622. The LC was first detrended with the biweight method using the 1.4 window length. The deformation is affecting both the transit duration T_{dur} and more significantly – the transit depth δ.

$\rho_{s,\odot} = 0.39$. Also in this case, a detailed analysis of the detected transits will improve the accuracy of the stellar density, for a direct comparison with asteroseismic determinations.

Scaled semi-major axis: We can also determine the scaled semi-major axis in terms of stellar radii with Equation 8. For the same light curve example, we obtain:

$$\frac{a}{R_s} = \left(\frac{GP^2 \rho_{s,\odot}\rho_\odot}{3\pi}\right)^{1/3} = 210.23,$$

while the actual injected value is 210.39. With the stellar mass obtained from spectroscopic observations, the stellar radius could also be calculated, providing the information about the semi-major axis value.

Impact parameter: The impact parameter is the projected sky distance between the center of the stellar disc and the planetary disc's center during conjunction. For a circular orbit, it is described with Equation 10. The distance l which the planet needs to traverse across the star's disk is directly related to the impact parameter and the transit duration. If the impact parameter is 0 then this distance is the maximum possible and the transit duration is the longest as well. Increasing it will result in the transit duration (and the distance – l) gradual decrease. The distance l can be expressed as:

$$l = \sqrt{(R_s + R_p)^2 - (bR_s)^2}$$

The distance l is directly related to the transit duration, by the relation:

$$T_{dur} = P\frac{\alpha}{2\pi} = \frac{P}{\pi}\arcsin\left(\frac{l}{a}\right) = \frac{P}{\pi}\arcsin\left(\frac{\sqrt{(R_s + R_p)^2 - (bR_s)^2}}{a}\right)$$
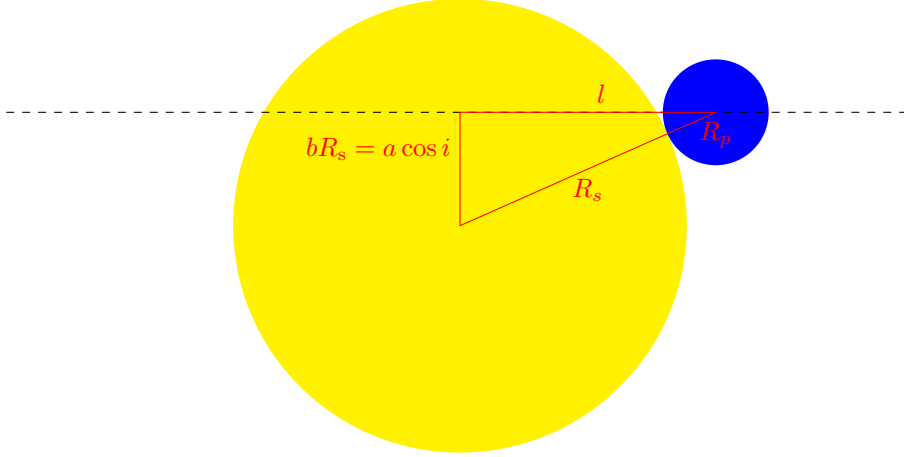
34

Figure 22: Auxiliary plot depicting the geometry at the transit ingress (T_1). The yellow circle represents the star, while the blue circle represents the planet. Throughout the transit, the planet must traverse a distance equal to $2l$. The path of the planet is drawn with the dashed line.

Where α is the angle between the ingress (T_1) and egress (T_4) of the transit drawn from the center of the star.

The relation can be transformed to express the impact parameter in terms of T_{dur}:

$$b^2 = \frac{(R_{\mathrm{s}} + R_{\mathrm{p}})^2 - a^2 \sin^2\left(\frac{\pi T_{\mathrm{dur}}}{P}\right)}{R_{\mathrm{s}}^2} = \left(1 + \frac{R_{\mathrm{p}}}{R_{\mathrm{s}}}\right)^2 - \left(\frac{a}{R_{\mathrm{s}}}\right)^2 \sin^2\left(\frac{\pi T_{\mathrm{dur}}}{P}\right)$$

Transit duration found by the TLS is in this case 0.82 days. The real value, for the injected planet calculated from the expression above would be 0.87 days. Both discrepancies in the T_{dur} and $R_{\mathrm{p}}/R_{\mathrm{s}}$ accumulate, making the accurate estimation of the impact parameter impossible. Again, this issue will be dealt by a detailed analysis of the LC.

5.3 Planetary and stellar parameters

In this subsection, the aim is to examine the overlap between planets detected using various algorithms. Specifically, we want to identify stellar and planetary parameters common to those detected by a majority of the algorithms. It is expected that larger planets, which induce more significant dips in the light curve, would be more easily detected. While shorter orbital periods might typically be favored for detection due to more frequent transits within a given observation time, this dependency may not be reflected in this analysis. Given the studied sample, most planets will likely transit across the stellar disk no more than two times during the two-year observational cadence. This is attributed to the fact that their orbital periods, in most cases, exceed 300 days.

In Figure 23, a scatter plot illustrating the relationship between the injected planets radius (expressed in units of stellar radius) and orbital period is presented. As anticipated, a visible correlation emerges between the number of algorithms detecting the signal and the planet radius (relative to stellar radius). Specifically, a larger relative planet radius corresponds to a higher detection rate, due to the greater impact of the planet on the LC. There is also a modest correlation with orbital period; notably, for orbital periods up to ~ 300 days, even when the planet-to-star radius ratio is small, most algorithms successfully recovered the signal. This is attributed to the presence of more than two transits in the LC.

In Figure 24, a similar as before scatter plot is presented, this time focusing on stellar parameters: effective temperature versus stellar radius. The data points are color-coded to represent the normalized count of algorithms that detected the planetary signal. Notably, a noteworthy reduction for the algorithms performance arises in the case of F5 spectral type stars. This difficulty is primarily due to their larger stellar radii, resulting in a relatively smaller ratio of planetary to stellar radius and longer transit durations. Both of these factors combine to make the detection of planets around F5 stars more challenging.
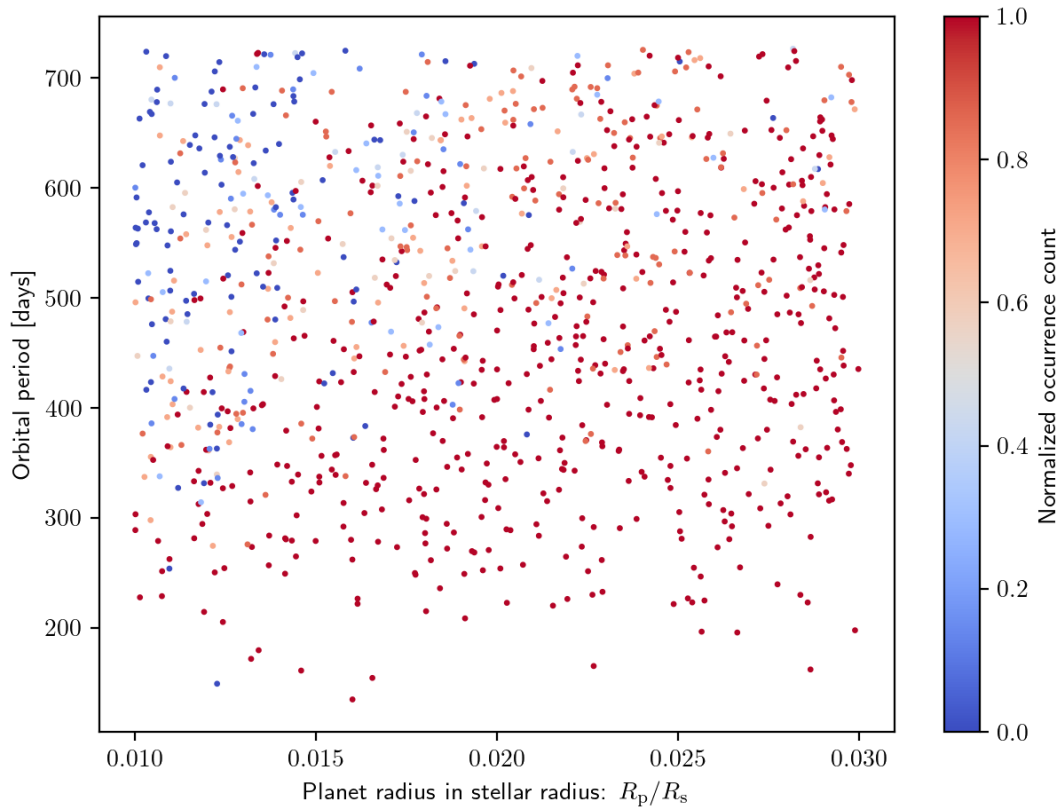
Figure 23: Plot attributing the number of recoveries by tested filtering algorithms (normalized to the maximum number of occurrence count) to the planetary parameters: injected planets radius in units of stellar radius and orbital period in days. The successful recovery is treated as the '1', '2' or '3' from the previous subsection. The plot is made for SDE threshold value of 10.5.
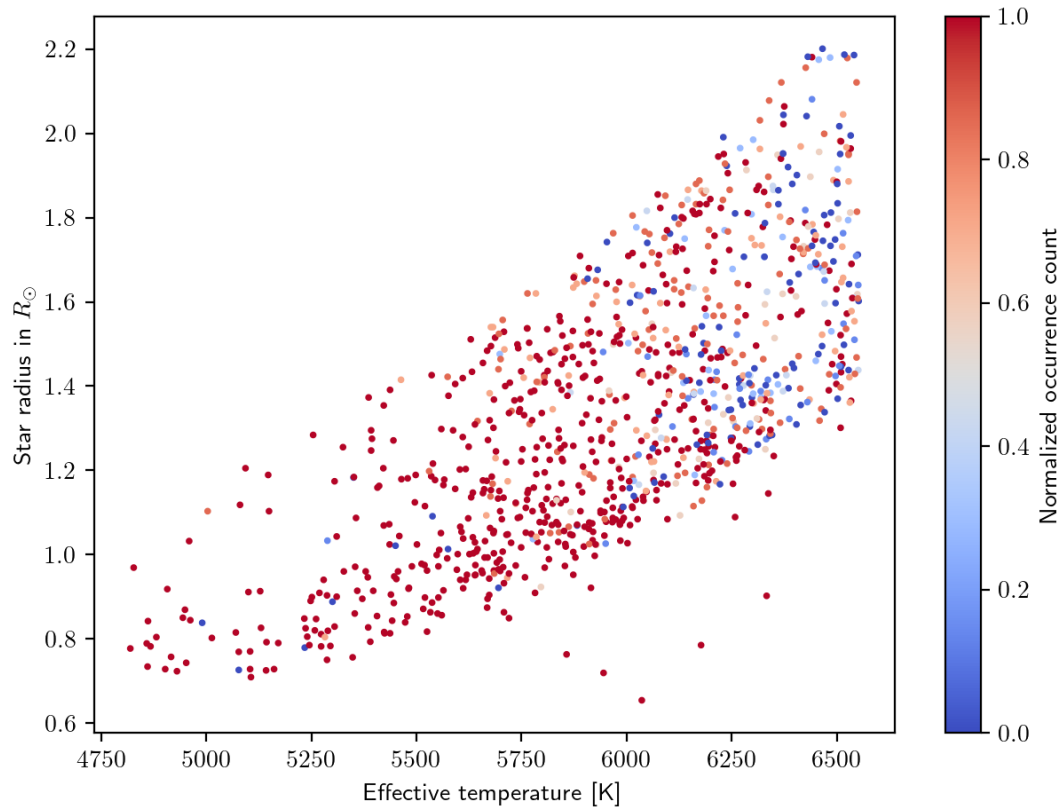
Figure 24: Plot attributing the number of recoveries by tested filtering algorithms (normalized to the maximum number of occurrence count) to the stellar parameters: effective temperature in Kelvins and stellar radius in $R_\odot$. The successful recovery is treated as the '1', '2' or '3' as defined in the previous subsection. The plot is made for SDE threshold value of 10.5.

5.4 Conclusions

The evaluation of the selected filtering algorithms has revealed that the most effective choices are the biweight method with a 1.4-day window length and YSD-Lowess, both achieving an overall efficiency of 89.9% and 89.6%, respectively, given the SDE threshold of 7.0. Combining the two methods together would not increase the overall efficiency, as all transits that are missed by the biweight method are missed by the YSD-Lowess as well. It is worth noting that the Huber spline method with a 0.7-day window length demonstrates comparable efficiency at the 7.0 SDE threshold level. Nonetheless, its use comes with an elevated risk of incorrectly recovered transits and a higher likelihood of false positives. This heightened risk is primarily due to the SDE distribution, which peaks around 10 SDE for LCs with injected signals. The retrieved planetary parameters should be interpreted with caution, as substantial discrepancies exist between the shape of injected and recovered planetary signals. This divergence underscores the imprecise nature of planetary parameter retrieval in this context as the usage of filtering algorithms alters the original signal.

It is worth acknowledging that not all algorithms from the Canocchi et al. (2023) paper were considered in this work, and there may be additional algorithms worth exploring. The *K2SC* algorithm that bases on the Gaussian Process regression Aigrain et al. (2016) was excluded due to its relatively poor performance resulting from the mentioned work. Conversely, the Notch and LOCoR (N&L) algorithm Rizzuto et al. (2017) was not tested due to its high computational cost. Indeed, this is the reason why at the moment there is no plan to include this algorithm in the official Exoplanet Analysis System for the analysis of PLATO LCs. Nevertheless, it was found to exhibit the highest performance for Earth-sized planets in both active and quiet samples in the referenced paper and might be the algorithm worth testing in the future, if the algorithm can be implemented in a computationally more efficient way, for example by employing Machine Learning techniques.

In the search for Earth-analogues, it is crucial to conduct a thorough study when a signal is detected. The limited number of transits during the 2-year observational period may lead to partial detections (where one out of two transits is not correctly recovered) or aliased recoveries (when an observational gap falls midway between two transits, resulting in a recovered period that is half of the actual one). These scenarios highlight the importance of careful analysis and follow-up observations to validate the potential planetary detections. Indeed, this works demonstrates that with PLATO mission we will be able to detect Earth-analogues with very high efficiency (close to 90%), if at least two transits are observed.

References

S. Aigrain, H. Parviainen, and B. J. S. Pope. K2SC: flexible systematics correction and detrending of K2 light curves using Gaussian process regression. *Monthly Notices of the Royal Astronomical Society*, 459(3):2408–2419, July 2016. doi: 10.1093/mnras/stw706.

G. Bakos, R. W. Noyes, G. Kovács, K. Z. Stanek, D. D. Sasselov, and I. Domsa. Wide-Field Millimagnitude Photometry with the HAT: A Tool for Extrasolar Planet Detection. *PASP*, 116 (817):266–277, March 2004. doi: 10.1086/382735.

Matthew P. Battley, Don Pollacco, and David J. Armstrong. A search for young exoplanets in Sectors 1-5 of the TESS full-frame images. *Monthly Notices of the Royal Astronomical Society*, 496(2):1197–1216, August 2020. doi: 10.1093/mnras/staa1626.

G. Fritz Benedict, Barbara McArthur, D. W. Chappell, E. Nelan, W. H. Jefferys, W. van Altena, J. Lee, D. Cornell, P. J. Shelus, P. D. Hemenway, Otto G. Franz, L. H. Wasserman, R. L. Duncombe, D. Story, A. L. Whipple, and L. W. Fredrick. Interferometric astrometry of proxima centauri and barnard's star using hubble space telescope fine guidance sensor 3: Detection limits for substellar companions*. *The Astronomical Journal*, 118(2):1086, aug 1999. doi: 10.1086/300975. URL https://dx.doi.org/10.1086/300975.

W. Benz, C. Broeg, A. Fortier, N. Rando, T. Beck, M. Beck, D. Queloz, D. Ehrenreich, P. F. L. Maxted, K. G. Isaak, N. Billot, Y. Alibert, R. Alonso, C. António, J. Asquier, T. Bandy, T. Bárczy, D. Barrado, S. C. C. Barros, W. Baumjohann, A. Bekkelien, M. Bergomi, F. Biondi, X. Bonfils, L. Borsato, A. Brandeker, M. D. Busch, J. Cabrera, V. Cessa, S. Charnoz, B. Chazelas, A. Collier Cameron, C. Corral Van Damme, D. Cortes, M. B. Davies, M. Deleuil, A. Deline, L. Delrez, O. Demangeon, B. O. Demory, A. Erikson, J. Farinato, L. Fossati, M. Fridlund, D. Futyan, D. Gandolfi, A. Garcia Munoz, M. Gillon, P. Guterman, A. Gutierrez, J. Hasiba, K. Heng, E. Hernandez, S. Hoyer, L. L. Kiss, Z. Kovacs, T. Kuntzer, J. Laskar, A. Lecavelier des Etangs, M. Lendl, A. López, I. Lora, C. Lovis, T. Lüftinger, D. Magrin, L. Malvasio, L. Marafatto, H. Michaelis, D. de Miguel, D. Modrego, M. Munari, V. Nascimbeni, G. Olofsson, H. Ottacher, R. Ottensamer, I. Pagano, R. Palacios, E. Pallé, G. Peter, D. Piazza, G. Piotto, A. Pizarro, D. Pollaco, R. Ragazzoni, F. Ratti, H. Rauer, I. Ribas, M. Rieder, R. Rohlfs, F. Safa, M. Salatti, N. C. Santos, G. Scandariato, D. Ségransan, A. E. Simon, A. M. S. Smith, M. Sordet, S. G. Sousa, M. Steller, G. M. Szabó, J. Szoke, N. Thomas, M. Tschentscher, S. Udry, V. Van Grootel, V. Viotto, I. Walter, N. A. Walton, F. Wildi, and D. Wolter. The CHEOPS mission. *Experimental Astronomy*, 51(1):109–151, February 2021. doi: 10.1007/s10686-020-09679-4.

William J. Borucki, David Koch, Gibor Basri, Natalie Batalha, Timothy Brown, Douglas Caldwell, John Caldwell, Jørgen Christensen-Dalsgaard, William D. Cochran, Edna DeVore, Edward W. Dunham, Andrea K. Dupree, Thomas N. Gautier, John C. Geary, Ronald Gilliland, Alan Gould, Steve B. Howell, Jon M. Jenkins, Yoji Kondo, David W. Latham, Geoffrey W. Marcy, Søren Meibom, Hans Kjeldsen, Jack J. Lissauer, David G. Monet, David Morrison, Dimitar Sasselov, Jill Tarter, Alan Boss, Don Brownlee, Toby Owen, Derek Buzasi, David Charbonneau, Laurance Doyle, Jonathan Fortney, Eric B. Ford, Matthew J. Holman, Sara Seager, Jason H. Steffen, William F. Welsh, Jason Rowe, Howard Anderson, Lars Buchhave, David Ciardi, Lucianne Walkowicz, William Sherry, Elliott Horch, Howard Isaacson, Mark E. Everett, Debra Fischer, Guillermo Torres, John Asher Johnson, Michael Endl, Phillip MacQueen, Stephen T. Bryson, Jessie Dotson, Michael Haas, Jeffrey Kolodziejczak, Jeffrey Van Cleve, Hema Chandrasekaran, Joseph D. Twicken, Elisa V. Quintana, Bruce D. Clarke, Christopher Allen, Jie Li, Haley Wu, Peter Tenenbaum, Ekaterina Verner, Frederick Bruhweiler, Jason Barnes, and Andrej Prsa. Kepler Planet-Detection Mission: Introduction and First Results. *Science*, 327(5968):977, February 2010. doi: 10.1126/science.1185402.

G. Canocchi, L. Malavolta, I. Pagano, O. Barragán, G. Piotto, S. Aigrain, S. Desidera, S. Grziwa, J. Cabrera, and H. Rauer. Discovering planets with PLATO: Comparison of algorithms for stellar activity filtering. *A&A*, 672:A144, April 2023. doi: 10.1051/0004-6361/202244067.

David Charbonneau, Timothy M. Brown, David W. Latham, and Michel Mayor. Detection of Planetary Transits Across a Sun-like Star. *ApJ*, 529(1):L45–L48, January 2000. doi: 10.1086/312457.

Jørgen Christensen-Dalsgaard. ADIPLS—the Aarhus adiabatic oscillation package. *Ap&SS*, 316 (1-4):113–120, August 2008. doi: 10.1007/s10509-007-9689-z.

Trevor J. David, Erik A. Petigura, Rodrigo Luger, Daniel Foreman-Mackey, John H. Livingston, Eric E. Mamajek, and Lynne A. Hillenbrand. Four Newborn Planets Transiting the Young Solar Analog V1298 Tau. *ApJ*, 885(1):L12, November 2019. doi: 10.3847/2041-8213/ab4c99.

Courtney D. Dressing and David Charbonneau. The Occurrence of Potentially Habitable Planets Orbiting M Dwarfs Estimated from the Full Kepler Dataset and an Empirical Measurement of the Detection Sensitivity. *The Astrophysical Journal*, 807(1):45, July 2015. doi: 10.1088/0004-637X/807/1/45.

Andrew Gould and Abraham Loeb. Discovering Planetary Systems through Gravitational Microlenses. *The Astrophysical Journal*, 396:104, September 1992. doi: 10.1086/171700.

Christina Hedges, Alex Hughes, George Zhou, Trevor J. David, Juliette Becker, Steven Giacalone, Andrew Vanderburg, Joseph E. Rodriguez, Allyson Bieryla, Christopher Wirth, Shaun Atherton, Tara Fetherolf, Karen A. Collins, Adrian M. Price-Whelan, Megan Bedell, Samuel N. Quinn, Tianjun Gan, George R. Ricker, David W. Latham, Roland K. Vanderspek, Sara Seager, Joshua N. Winn, Jon M. Jenkins, John F. Kielkopf, Richard P. Schwarz, Courtney D. Dressing, Erica J. Gonzales, Ian J. M. Crossfield, Elisabeth C. Matthews, Eric L. N. Jensen, Elise Furlan, Crystal L. Gnilka, Steve B. Howell, Kathryn V. Lester, Nicholas J. Scott, Dax L. Feliz, Michael B. Lund, Robert J. Siverd, Daniel J. Stevens, N. Narita, A. Fukui, F. Murgas, Enric Palle, Phil J. Sutton, Keivan G. Stassun, Luke G. Bouma, Michael Vezie, Jesus Noel Villaseñor, Elisa V. Quintana, and Jeffrey C. Smith. TOI-2076 and TOI-1807: Two Young, Comoving Planetary Systems within 50 pc Identified by TESS that are Ideal Candidates for Further Follow Up. *The Astronomical Journal*, 162(2):54, August 2021. doi: 10.3847/1538-3881/ac06cd.

René Heller. Analytic solutions to the maximum and average exoplanet transit depth for common stellar limb darkening laws. *A&A*, 623:A137, March 2019. doi: 10.1051/0004-6361/201834620.

René Heller, Jan-Vincent Harre, and Réza Samadi. Transit least-squares survey. IV. Earth-like transiting planets expected from the PLATO mission. *A&A*, 665:A11, September 2022. doi: 10.1051/0004-6361/202141640.

Michael Hippke and René Heller. Optimized transit detection algorithm to search for periodic transits of small planets. *A&A*, 623:A39, March 2019. doi: 10.1051/0004-6361/201834672.

Michael Hippke, Trevor J. David, Gijs D. Mulders, and René Heller. Wōtan: Comprehensive Time-series Detrending in Python. *The Astronomical Journal*, 158(4):143, October 2019. doi: 10.3847/1538-3881/ab3984.

Steve B. Howell, Charlie Sobeck, Michael Haas, Martin Still, Thomas Barclay, Fergal Mullally, John Troeltzsch, Suzanne Aigrain, Stephen T. Bryson, Doug Caldwell, William J. Chaplin, William D. Cochran, Daniel Huber, Geoffrey W. Marcy, Andrea Miglio, Joan R. Najita, Marcie Smith, J. D. Twicken, and Jonathan J. Fortney. The K2 Mission: Characterization and Early Results. *PASP*, 126(938):398, April 2014. doi: 10.1086/676406.

Peter J. Huber. Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35(1):73 – 101, 1964. doi: 10.1214/aoms/1177703732. URL https://doi.org/10.1214/aoms/1177703732.

Peter J. Huber. *Robust Statistics*, pages 1248–1251. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. ISBN 978-3-642-04898-2. doi: 10.1007/978-3-642-04898-2_594. URL https://doi.org/10.1007/978-3-642-04898-2_594.

T.-O. Husser, S Wende-von Berg, S Dreizler, D Homeier, A Reiners, T. Barman, and Peter H Hauschildt. Astrophysics A new extensive library of PHOENIX stellar atmospheres. *A&A*, 553:A6, 2013. doi: 10.1051/0004-6361/201219058.

Dae-Won Kim. PDT: Photometric DeTrending Algorithm Using Machine Learning. Astrophysics Source Code Library, record ascl:1605.008, May 2016.

Dae-Won Kim, Pavlos Protopapas, Charles Alcock, Yong-Ik Byun, and Federica B. Bianco. Detrending time series for astronomical variability surveys. *Monthly Notices of the Royal Astronomical Society*, 397(2):558–568, July 2009. doi: 10.1111/j.1365-2966.2009.14967.x.

Heather A. Knutson, David Charbonneau, Robert W. Noyes, Timothy M. Brown, and Ronald L. Gilliland. Using Stellar Limb-Darkening to Refine the Properties of HD 209458b. *The Astrophysical Journal*, 655(1):564–575, January 2007. doi: 10.1086/510111.

Zdenek Kopal. Detailed effects of limb darkening upon light and velocity curves of close binary systems. *Harvard College Observatory Circular*, 454:1–12, January 1950.

Ravi Kumar Kopparapu, Ramses Ramirez, James F. Kasting, Vincent Eymet, Tyler D. Robinson, Suvrath Mahadevan, Ryan C. Terrien, Shawn Domagal-Goldman, Victoria Meadows, and Rohit Deshpande. Habitable Zones around Main-sequence Stars: New Estimates. *The Astrophysical Journal*, 765(2):131, March 2013. doi: 10.1088/0004-637X/765/2/131.

G. Kovács, S. Zucker, and T. Mazeh. A box-fitting algorithm in the search for periodic transits. *A&A*, 391:369–377, August 2002. doi: 10.1051/0004-6361:20020802.

Laura Kreidberg. batman: BAsic Transit Model cAlculatioN in Python. *PASP*, 127(957):1161, November 2015a. doi: 10.1086/683602.

Laura Kreidberg. batman: BAsic Transit Model cAlculatioN in Python. *PASP*, 127(957):1161, November 2015b. doi: 10.1086/683602.

John H. Livingston, Michael Endl, Fei Dai, William D. Cochran, Oscar Barragan, Davide Gandolfi, Teruyuki Hirano, Sascha Grziwa, Alexis M. S. Smith, Simon Albrecht, Juan Cabrera, Szilard Csizmadia, Jerome P. de Leon, Hans Deeg, Philipp Eigmüller, Anders Erikson, Mark Everett, Malcolm Fridlund, Akihiko Fukui, Eike W. Guenther, Artie P. Hatzes, Steve Howell, Judith Korth, Norio Narita, David Nespral, Grzegorz Nowak, Enric Palle, Martin Pätzold, Carina M. Persson, Jorge Prieto-Arranz, Heike Rauer, Motohide Tamura, Vincent Van Eylen, and Joshua N. Winn. 44 Validated Planets from K2 Campaign 10. *The Astronomical Journal*, 156 (2):78, August 2018. doi: 10.3847/1538-3881/aaccde.

V. Marchiori, R. Samadi, F. Fialho, C. Paproth, A. Santerne, M. Pertenais, A. Börner, J. Cabrera, A. Monsky, and N. Kutrowski. In-flight photometry extraction of PLATO targets. Optimal apertures for detecting extrasolar planets. *A&A*, 627:A71, July 2019. doi: 10.1051/0004-6361/201935269.

A. Marconi, M. Abreu, V. Adibekyan, M. Aliverti, C. Allende Prieto, P. Amado, M. Amate, E. Artigau, S. Augusto, S. Barros, S. Becerril, B. Benneke, E. Bergin, P. Berio, N. Bezawada, I. Boisse, X. Bonfils, F. Bouchy, C. Broeg, A. Cabral, R. Calvo-Ortega, B. L. Canto Martins, B. Chazelas, A. Chiavassa, L. Christensen, R. Cirami, I. Coretti, S. Covino, G. Cresci, S. Cristiani, V. Cunha Parro, G. Cupani, I. de Castro Leão, J. Renan de Medeiros, M. A. Furlande Souza, P. Di Marcantonio, I. Di Varano, V. D'Odorico, R. Doyon, H. Drass, P. Figueira, A. Belen Fragoso, J. P. Uldall Fynbo, E. Gallo, M. Genoni, J. González Hernández, M. Haehnelt, J. Hlavacek-Larrondo, I. Hughes, P. Huke, A. Humphrey, H. Kjeldsen, A. Korn, D. Kouach, M. Landoni, J. Liske, C. Lovis, D. Lunney, R. Maiolino, L. Malo, T. Marquart, C. Martins, E. Mason, P. Molaro, J. Monnier, M. Monteiro, C. Mordasini, T. Morris, A. Mucciarelli, G. Murray, A. Niedzielski, N. Nunes, E. Oliva, L. Origlia, E. Pallé, G. Pariani, P. Parr-Burman, J. Peñate, F. Pepe, E. Pinna, N. Piskunov, J. L. Rasilla Piñeiro, R. Rebolo, P. Rees, A. Reiners, M. Riva, D. Romano, S. Rousseau, N. Sanna, N. Santos, M. Sarajlic, T. C. Shen, F. Sortino, D. Sosnowska, S. Sousa, E. Stempels, K. Strassmeier, F. Tenegi, A. Tozzi, S. Udry, L. Valenziano, L. Vanzi, M. Weber, M. Woche, M. Xompero, E. Zackrisson, and M. R. Zapatero Osorio. HIRES, the High-resolution Spectrograph for the ELT. *The Messenger*, 182:27–32, March 2021. doi: 10.18727/0722-6691/5219.

Christian Marois, Bruce Macintosh, Travis Barman, B. Zuckerman, Inseok Song, Jennifer Patience, David Lafrenière, and René Doyon. Direct Imaging of Multiple Planets Orbiting the Star HR 8799. *Science*, 322(5906):1348, November 2008. doi: 10.1126/science.1166585.

Michel Mayor and Didier Queloz. A Jupiter-mass companion to a solar-type star. *Nature*, 378 (6555):355–359, November 1995. doi: 10.1038/378355a0.

N. Meunier, A. M. Lagrange, T. Boulet, and S. Borgniet. Activity time series of old stars from late F to early K. I. Simulating radial velocity, astrometry, photometry, and chromospheric emission. *A&A*, 627:A56, July 2019. doi: 10.1051/0004-6361/201834796.

M. Montalto, G. Piotto, P. M. Marrese, V. Nascimbeni, L. Prisinzano, V. Granata, S. Marinoni, S. Desidera, S. Ortolani, C. Aerts, E. Alei, G. Altavilla, S. Benatti, A. Börner, J. Cabrera, R. Claudi, M. Deleuil, M. Fabrizio, L. Gizon, M. J. Goupil, A. M. Heras, D. Magrin, L. Malavolta, J. M. Mas-Hesse, I. Pagano, C. Paproth, M. Pertenais, D. Pollacco, R. Ragazzoni, G. Ramsay, H. Rauer, and S. Udry. The all-sky PLATO input catalogue. *A&A*, 653:A98, September 2021. doi: 10.1051/0004-6361/202140717.

Mario Morvan, Nikolaos Nikolaou, Angelos Tsiaras, and Ingo P. Waldmann. Detrending Exoplanetary Transit Light Curves with Long Short-term Memory Networks. *The Astronomical Journal*, 159(3):109, March 2020. doi: 10.3847/1538-3881/ab6aa7.

Frederick Mosteller and John W. Tukey. *Data analysis and regression. A second course in statistics.* 1977.

V. Nascimbeni, G. Piotto, A. Börner, M. Montalto, P. M. Marrese, J. Cabrera, S. Marinoni, C. Aerts, G. Altavilla, S. Benatti, R. Claudi, M. Deleuil, S. Desidera, M. Fabrizio, L. Gizon, M. J. Goupil, V. Granata, A. M. Heras, D. Magrin, L. Malavolta, J. M. Mas-Hesse, S. Ortolani, I. Pagano, D. Pollacco, L. Prisinzano, R. Ragazzoni, G. Ramsay, H. Rauer, and S. Udry. The PLATO field selection process. I. Identification and content of the long-pointing fields. *A&A*, 658:A31, February 2022. doi: 10.1051/0004-6361/202142256.

Hannu Parviainen and Suzanne Aigrain. ldtk: Limb Darkening Toolkit. *MNRAS*, 453(4):3821–3826, November 2015. doi: 10.1093/mnras/stv1857. URL http://mnras.oxfordjournals.org/lookup/doi/10.1093/mnras/stv1857.

Jayshil A. Patel and Néstor Espinoza. Empirical Limb-darkening Coefficients and Transit Parameters of Known Exoplanets from TESS. *The Astronomical Journal*, 163(5):228, May 2022. doi: 10.3847/1538-3881/ac5f55.

Mark J. Pecaut and Eric E. Mamajek. Intrinsic Colors, Temperatures, and Bolometric Corrections of Pre-main-sequence Stars. , 208(1):9, September 2013. doi: 10.1088/0067-0049/208/1/9.

Joshua Pepper, Richard W. Pogge, D. L. DePoy, J. L. Marshall, K. Z. Stanek, Amelia M. Stutz, Shawn Poindexter, Robert Siverd, Thomas P. O'Brien, Mark Trueblood, and Patricia Trueblood. The Kilodegree Extremely Little Telescope (KELT): A Small Robotic Telescope for Large-Area Synoptic Surveys. *PASP*, 119(858):923–935, August 2007. doi: 10.1086/521836.

Erik A. Petigura, Andrew W. Howard, and Geoffrey W. Marcy. Prevalence of Earth-size planets orbiting Sun-like stars. *Proceedings of the National Academy of Science*, 110(48):19273–19278, November 2013. doi: 10.1073/pnas.1319909110.

Benjamin J. S. Pope, Hannu Parviainen, and Suzanne Aigrain. Transiting exoplanet candidates from K2 Campaigns 5 and 6. *Monthly Notices of the Royal Astronomical Society*, 461(4):3399–3409, October 2016. doi: 10.1093/mnras/stw1373.

H. Rauer, C. Catala, C. Aerts, T. Appourchaux, W. Benz, A. Brandeker, J. Christensen-Dalsgaard, M. Deleuil, L. Gizon, M. J. Goupil, M. Güdel, E. Janot-Pacheco, M. Mas-Hesse, I. Pagano, G. Piotto, D. Pollacco, Ċ. Santos, A. Smith, J. C. Suárez, R. Szabó, S. Udry, V. Adibekyan, Y. Alibert, J. M. Almenara, P. Amaro-Seoane, M. Ammler-von Eiff, M. Asplund, E. Antonello, S. Barnes, F. Baudin, K. Belkacem, M. Bergemann, G. Bihain, A. C. Birch, X. Bonfils, I. Boisse, A. S. Bonomo, F. Borsa, I. M. Brandão, E. Brocato, S. Brun, M. Burleigh, R. Burston, J. Cabrera, S. Cassisi, W. Chaplin, S. Charpinet, C. Chiappini, R. P. Church, Sz. Csizmadia, M. Cunha, M. Damasso, M. B. Davies, H. J. Deeg, R. F. Díaz, S. Dreizler, C. Dreyer, P. Eggenberger, D. Ehrenreich, P. Eigmüller, A. Erikson, R. Farmer, S. Feltzing, F. de Oliveira Fialho, P. Figueira, T. Forveille, M. Fridlund, R. A. García, P. Giommi, G. Giuffrida, M. Godolt, J. Gomes da Silva, T. Granzer, J. L. Grenfell, A. Grotsch-Noels, E. Günther, C. A. Haswell, A. P. Hatzes, G. Hébrard, S. Hekker, R. Helled, K. Heng, J. M. Jenkins, A. Johansen, M. L. Khodachenko, K. G. Kislyakova, W. Kley, U. Kolb, N. Krivova, F. Kupka, H. Lammer, A. F.

Lanza, Y. Lebreton, D. Magrin, P. Marcos-Arenal, P. M. Marrese, J. P. Marques, J. Martins, S. Mathis, S. Mathur, S. Messina, A. Miglio, J. Montalban, M. Montalto, M. J. P. F. G. Monteiro, H. Moradi, E. Moravveji, C. Mordasini, T. Morel, A. Mortier, V. Nascimbeni, R. P. Nelson, M. B. Nielsen, L. Noack, A. J. Norton, A. Ofir, M. Oshagh, R. M. Ouazzani, P. Pápics, V. C. Parro, P. Petit, B. Plez, E. Poretti, A. Quirrenbach, R. Ragazzoni, G. Raimondo, M. Rainer, D. R. Reese, R. Redmer, S. Reffert, B. Rojas-Ayala, I. W. Roxburgh, S. Salmon, A. Santerne, J. Schneider, J. Schou, S. Schuh, H. Schunker, A. Silva-Valio, R. Silvotti, I. Skillen, I. Snellen, F. Sohl, S. G. Sousa, A. Sozzetti, D. Stello, K. G. Strassmeier, M. Švanda, Gy. M. Szabó, A. Tkachenko, D. Valencia, V. Van Grootel, S. D. Vauclair, P. Ventura, F. W. Wagner, N. A. Walton, J. Weingrill, S. C. Werner, P. J. Wheatley, and K. Zwintz. The PLATO 2.0 mission. *Experimental Astronomy*, 38(1-2):249–330, November 2014. doi: 10.1007/s10686-014-9383-4.

Matthias Rempel. Solar and stellar activity cycles. In *Journal of Physics Conference Series*, volume 118 of *Journal of Physics Conference Series*, page 012032, October 2008. doi: 10.1088/1742-6596/118/1/012032.

George R. Ricker, Joshua N. Winn, Roland Vanderspek, David W. Latham, Gáspár. Á. Bakos, Jacob L. Bean, Zachory K. Berta-Thompson, Timothy M. Brown, Lars Buchhave, Nathaniel R. Butler, R. Paul Butler, William J. Chaplin, David Charbonneau, Jørgen Christensen-Dalsgaard, Mark Clampin, Drake Deming, John Doty, Nathan De Lee, Courtney Dressing, E. W. Dunham, Michael Endl, Francois Fressin, Jian Ge, Thomas Henning, Matthew J. Holman, Andrew W. Howard, Shigeru Ida, Jon Jenkins, Garrett Jernigan, John A. Johnson, Lisa Kaltenegger, Nobuyuki Kawai, Hans Kjeldsen, Gregory Laughlin, Alan M. Levine, Douglas Lin, Jack J. Lissauer, Phillip MacQueen, Geoffrey Marcy, P. R. McCullough, Timothy D. Morton, Norio Narita, Martin Paegert, Enric Palle, Francesco Pepe, Joshua Pepper, Andreas Quirrenbach, S. A. Rinehart, Dimitar Sasselov, Bun'ei Sato, Sara Seager, Alessandro Sozzetti, Keivan G. Stassun, Peter Sullivan, Andrew Szentgyorgyi, Guillermo Torres, Stephane Udry, and Joel Villasenor. Transiting Exoplanet Survey Satellite (TESS). In Jr. Oschmann, Jacobus M., Mark Clampin, Giovanni G. Fazio, and Howard A. MacEwen, editors, *Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave*, volume 9143 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 914320, August 2014. doi: 10.1117/12.2063489.

Aaron C. Rizzuto, Andrew W. Mann, Andrew Vanderburg, Adam L. Kraus, and Kevin R. Covey. Zodiacal exoplanets in time (zeit). v. a uniform search for transiting planets in young clusters observed by k2. *The Astronomical Journal*, 154(6):224, nov 2017. doi: 10.3847/1538-3881/aa9070. URL https://dx.doi.org/10.3847/1538-3881/aa9070.

C. Rodrigo and E. Solano. The SVO Filter Profile Service. In *XIV.0 Scientific Meeting (virtual) of the Spanish Astronomical Society*, page 182, July 2020.

R. Samadi, A. Deru, D. Reese, V. Marchiori, E. Grolleau, J. J. Green, M. Pertenais, Y. Lebreton, S. Deheuvels, B. Mosser, K. Belkacem, A. Börner, and A. M. S. Smith. The PLATO Solar-like Light-curve Simulator. A tool to generate realistic stellar light-curves with instrumental effects representative of the PLATO mission. *A&A*, 624:A117, April 2019. doi: 10.1051/0004-6361/201834822.

A. Savitzky and M. J. E. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36:1627–1639, January 1964a. doi: 10.1021/ac60214a047.

A. Savitzky and M. J. E. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36:1627–1639, January 1964b. doi: 10.1021/ac60214a047.

Heinrich Schwabe. Sonnenbeobachtungen im Jahre 1843. Von Herrn Hofrath Schwabe in Dessau. *Astronomische Nachrichten*, 21(15):233, February 1844. doi: 10.1002/asna.18440211505.

André M. Silva, Sérgio G. Sousa, Nuno Santos, Olivier D. S. Demangeon, Pedro Silva, S. Hoyer, P. Guterman, Magali Deleuil, and David Ehrenreich. ARCHI: pipeline for light curve extraction of CHEOPS background stars. *Monthly Notices of the Royal Astronomical Society*, 496(1):282–294, July 2020. doi: 10.1093/mnras/staa1443.

Robert J. Siverd, Thomas G. Beatty, Joshua Pepper, Jason D. Eastman, Karen Collins, Allyson Bieryla, David W. Latham, Lars A. Buchhave, Eric L. N. Jensen, Justin R. Crepp, Rachel Street, Keivan G. Stassun, B. Scott Gaudi, Perry Berlind, Michael L. Calkins, D. L. DePoy, Gilbert A. Esquerdo, Benjamin J. Fulton, Gábor Fűrész, John C. Geary, Andrew Gould, Leslie Hebb, John F. Kielkopf, Jennifer L. Marshall, Richard Pogge, K. Z. Stanek, Robert P. Stefanik, Andrew H. Szentgyorgyi, Mark Trueblood, Patricia Trueblood, Amelia M. Stutz, and Jennifer L. van Saders. KELT-1b: A Strongly Irradiated, Highly Inflated, Short Period, 27 Jupiter-mass Companion Transiting a Mid-F Star. *The Astrophysical Journal*, 761(2):123, December 2012. doi: 10.1088/0004-637X/761/2/123.

Sean C. Solomon and James W. Head. Fundamental issues in the geology and geophysics of venus. *Science*, 252(5003):252–260, 1991. doi: 10.1126/science.252.5003.252. URL `https://www.science.org/doi/abs/10.1126/science.252.5003.252`.

Martin C. Stumpe, Jeffrey C. Smith, Jeffrey E. Van Cleve, Joseph D. Twicken, Thomas S. Barclay, Michael N. Fanelli, Forrest R. Girouard, Jon M. Jenkins, Jeffery J. Kolodziejczak, Sean D. McCauliff, and Robert L. Morris. Kepler presearch data conditioning iarchitecture and algorithms for error correction in kepler light curves. *Publications of the Astronomical Society of the Pacific*, 124(919):985–999, 2012. ISSN 00046280, 15383873. URL `http://www.jstor.org/stable/10.1086/667698`.

Andrew Vanderburg, Juliette C. Becker, Martti H. Kristiansen, Allyson Bieryla, Dmitry A. Duev, Rebecca Jensen-Clem, Timothy D. Morton, David W. Latham, Fred C. Adams, Christoph Baranec, Perry Berlind, Michael L. Calkins, Gilbert A. Esquerdo, Shrinivas Kulkarni, Nicholas M. Law, Reed Riddle, Maïssa Salama, and Allan R. Schmitt. Five Planets Transiting a Ninth Magnitude Star. *ApJ*, 827(1):L10, August 2016. doi: 10.3847/2041-8205/827/1/L10.

R. Wells, K. Poppenhaeger, and C. A. Watson. Three small transiting planets around the M-dwarf host star LP 358-499. *Monthly Notices of the Royal Astronomical Society*, 473(1):L131–L135, January 2018. doi: 10.1093/mnrasl/slx171.

A. Wolszczan. Discovery of Two Planets Around a Millisecond Pulsar. In *23rd Lunar and Planetary Science Conference*, volume 781 of *LPI Contributions*, page 53, January 1992.

M. H. Yant, S. M. Hörst, A. H. Parker, S. Protopapa, K. Nowicki, C. A. Thomas, J. Hanley, W. M. Grundy, and Project Espresso Team. Project ESPRESSO: Optical Constants for Quantitative Spectral Analysis. In *49th Annual Lunar and Planetary Science Conference*, Lunar and Planetary Science Conference, page 2758, March 2018.

Hui Zhang, Zhouyi Yu, Ensi Liang, Ming Yang, Michael C. B. Ashley, Xiangqun Cui, Fujia Du, Jianning Fu, Xuefei Gong, Bozhong Gu, Yi Hu, Peng Jiang, Huigen Liu, Jon Lawrence, Qiang Liu, Xiaoyan Li, Zhengyang Li, Bin Ma, Jeremy Mould, Zhaohui Shang, Nicholas B. Suntzeff, Charling Tao, Qiguo Tian, C. G. Tinney, Syed A. Uddin, Lifan Wang, Songhu Wang, Xiaofeng Wang, Peng Wei, Duncan Wright, Xuefeng Wu, Robert A. Wittenmyer, Lingzhe Xu, Shi-hai Yang, Ce Yu, Xiangyan Yuan, Jessica Zheng, Hongyan Zhou, Ji-lin Zhou, and Zhenxi Zhu. Exoplanets in the Antarctic Sky. II. 116 Transiting Exoplanet Candidates Found by AST3-II (CHESPA) within the Southern CVZ of TESS. , 240(2):17, February 2019. doi: 10.3847/1538-4365/aaf583.